

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ
М.В.ЛОМОНОСОВА»

На правах рукописи

Алексеев Андрей Юрьевич

ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА:
КОНЦЕПТУАЛЬНЫЙ СТАТУС
КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА

Специальность 09.00.08 – Философия науки и техники

ДИССЕРТАЦИЯ
на соискание ученой степени
доктора философских наук

Научный консультант –
доктор философских наук, профессор
Д.И.Дубровский

Москва – 2015

ВВЕДЕНИЕ.....	4
ГЛАВА I. ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: МЕЖДИСЦИПЛИНАРНЫЙ ПОХОД.....	23
1.1. Анализ дискуссий по философско-методологическим проблемам искусственного интеллекта.....	23
1.1.1. Становление философии искусственного интеллекта	24
1.1.2. Анализ дискуссий на мероприятиях Научного совета РАН по методологии искусственного интеллекта (НСМИИ РАН).....	35
1.1.3. Итоги десятилетней работы НСМИИ РАН	73
1.2. Роль теста Тьюринга в современных исследованиях искусственного интеллекта.....	107
1.2.1. Становление проблематики теста Тьюринга в отечественных дискуссиях.....	107
1.2.2. Роль теста Тьюринга в философии искусственного интеллекта.....	116
1.2.3. Роль теста Тьюринга в исследованиях электронной культуры.....	134
Выводы	158
ГЛАВА II. ОТ ОРИГИНАЛЬНОГО ТЕСТА ТЬЮРИНГА К КОМПЛЕКСНОМУ ТЕСТУ ТЬЮРИНГА.....	160
2.1. Идея оригинального теста Тьюринга	162
2.1.1. От машины Тьюринга к тесту Тьюринга: истоки конструирующей функции комплексного теста Тьюринга	162
2.1.2. Игра в имитацию интеллектуального поведения: истоки дефинитной функции комплексного теста Тьюринга	188
2.1.3. Полемический стандарт теста Тьюринга: истоки критической функции комплексного теста Тьюринга	197
2.2. Частные тесты Тьюринга	207
2.2.1. Становление частных тестов Тьюринга.....	208
2.2.2. Крупные версии теста Тьюринга.....	213
2.2.3. Возможности прохождения тестов Тьюринга.....	220
2.3. Идея комплексного теста Тьюринга	229
2.3.1. Версии и интерпретации теста Тьюринга.....	229
2.3.2. Построение комплексного теста Тьюринга.....	234
2.3.3. Формулирование комплексного теста Тьюринга	238
Выводы	247

ГЛАВА 3. ФУНКЦИИ КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА ...	248
3.1. Дефинитная функция	248
3.1.1. Дефиниция социальных феноменов: «другой» в электронной культуре	249
3.1.2. Дефиниция персональных феноменов: «личность» в электронной культуре	256
3.1.3. Дефиниция ментальных феноменов: «сознание» в электронной культуре	274
3.2. Критическая функция	286
3.2.1. Критика технологии: комплексный тест Тьюринга против комплексной технологии динамических экспертных систем реального времени	286
3.2.2. Критика методологии: комплексный тест Тьюринга против комплексного проекта искусственной личности	302
3.2.3. Критика мировоззрения: комплексный тест Тьюринга против комплексного тестирования в вузовском образовании	319
3.3. Конструирующая функция	339
3.3.1. Философско-методологические аспекты: обоснование стратегической роли проблемы «творчество-компьютер»	339
3.3.2. Научно-методологические аспекты: нейрокомпьютинг – базовая технология искусственного интеллекта	355
3.3.3. Инженерно-методологические аспекты: комплексная машина Корсакова–Тьюринга как принципиальный проект интеллектуального компьютера	392
3.4. Конституирующая функция	425
3.4.1. Функционализм как доминирующая парадигма современной аналитической философии	425
3.4.2. Компьютеризм как высшая стадия функционализма	435
3.4.3. От машинного функционализма к тестовому компьютеризму	444
Выводы	449
ЗАКЛЮЧЕНИЕ	450
СПИСОК СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ	457
БИБЛИОГРАФИЯ	458

Введение

Актуальность темы исследования

Современный человек, неожиданно «заброшенный» в фундированный компьютерными технологиями мир – мир электронной культуры, сталкивается с многочисленными философскими проблемами: технологического конструирования искусственного бытия – В.Л. Макаров [Макаров, 2006; 2010; 2011], онтологического обживания его адекватными мировоззренческими универсалиями – В.С. Стёпин [Стёпин, 1991; 1992; 2007], В.В. Миронов [Миронов, 2006]), этического обустройства (А.А. Гусейнов, Б.М. Шалютин [Шалютин, 1976; 1985; 2006], В.О. Лобовиков [Лобовиков, 2005], А.А. Скворцов [Скворцов, 2005; 2006] и эстетического выражения в нём собственно человеческого – В.И. Самохвалова [Самохвалова, 2005; 2006]). Возникает необходимость пересмотра фундаментальных отношений к смешанной, естественно-искусственной, реальности: её эпистемологического удостоверения (В.А. Лекторский [Лекторский, 2006-1, 2; 2011]), манипулирования её субъективной и альтерсубъективной формами (Д.И. Дубровский [Дубровский, 1971-2012]), конституирования интерсубъективных параметров (Н.М. Смирнова [Смирнова, 2005-2011]). Бурное развитие электронной технологии искусственного интеллекта выпукло обозначает проблемы гносеологической идентичности естественного человеческого мышления (И.Ю. Алексеева [Алексеева, 1992-2006], А.Ф. Зотов [Зотов, 2005; 2006], А.Н. Кочергин [Кочергин, 2005-2006], А.Н. Портнов [Портнов, 1994]), философско-методологической компетентности его исследования (Б.В. Бирюков [Бирюков, 1989; 2002; 2004], Т.Б. Кудряшова [Кудряшова, 2005-2013], А.П. Огурцов [Огурцов, 2005; 2006]), экстраполяции искусственных форм из биологического наследия (И.П. Меркулов [Меркулов, 2005; 2006]) в контексте эмерджентного эволюционного мышления (Е.Н. Князева [Князева, 2012; 2013]).

Развитие электронной культуры актуализирует логические проблемы её рационализации (В.К. Финн [Финн, 1999-2012]), семиотические проблемы формирования дискурсов новых форм культуры в её онтологическом, эпистемологическом и системологическом аспектах (В.М. Розин [Розин, 2001-2011], Ю.М. Резник [Резник, 1999–2003]), герменевтические проблемы понимания её феноменов – И.В. Дмитриевская [Дмитриевская, 1994; 1995; 1998], В.Г. Кузнецов [Кузнецов, 1991], Е.Н. Шульга [Шульга, 2002-2012]), когнитивные проблемы реформирования системы образования – О.Е. Баксанский [Баксанский, 2011], жизненно важные проблемы оптимизации медицины –

В.И.Моисеев [Моисеев, 2007], аксиологические проблемы обретения нравственной устойчивости в информационном обществе (К.К. Колин [Колин, 2005-2013]) и интеллектуальной безопасности в условиях изощённых информационных войн (И.И. Быстров [Быстров, 2001), инфомедийной идеологической борьбы с религиозным фундаментализмом, сектантством, паранаукой – Э.П. Кругляков, М.В. Садовский, А.А. Ивин [Кругляков, 2006; Садовский, 2006; Ивин, 2007]. Бесконечный ряд других философских проблем тематизирует рефлексию перехода от традиционной культуры к электронной культуре в сквозном процессе глобальной интеллектуализации информационного общества – Э. А. Азроянц [Азроянц, 2002], А.Н. Чумаков [Чумаков, 1994-2007]) – от личностных гуманистических установок – В.А. Кувакин [Кувакин, 1998-2003], Ф.М. Цанн-кай-си [Цанн-кай-си, 2005; 2007] до общечеловеческих ноосферных регулятивов – Н.А. Сляднева [Сляднева, 2009], Д.Г. Смирнов [Смирнов Д.Г., 2008]).

В размышлении над подобными проблемами акцентируется роль основательной, философской рефлексии по поводу исследований искусственного интеллекта (ИИ). Востребованность философии ИИ подчёркивается традитивными суждениями: насколько в традиционной культуре важна роль естественного интеллекта как целостная способность упорядочивать чувственные данные, рационально рассуждать и разумно регулировать индивидуальное и социальное поведение, настолько в электронной культуре значима роль интеллектуальных компьютерных систем, реализующих подобные функции. Насколько человеческий разум обуславливает расцвет традиционной культуры, настолько креативно-эвристический потенциал интеллектуальных электронных технологий задаёт степень развития электронной культуры.

Актуальность философских исследований ИИ обусловлена конвергентным развитием перспективной материально-технологической инфраструктуры электронной культуры – НБИКС-комплекса (В.Л. Макаров, Д.И. Дубровский), все компоненты которого – нано-, био-, инфо-, когни-, социо-технологии – принципиально редуцируемы к (квази)алгоритмическим формам электронных преобразований. Компоненты НБИКС-комплекса наукоёмки, методология их интеграции предполагает сложнейшую меж- и кросс-дисциплинарную деятельность (В.И. Аршинов), эффективности которой способствуют интеллектуальные компьютерные технологии. Они создают условия организации плодотворного научного творчества, способствуют постановке и решению слабо формализованных задач (Д.А. Поспелов, Г.С. Поспелов), поддерживают разветвлённый трафик научных коммуникаций (д.пед.н. Т.В. Кузнецова), сложный семан-

тико-прагматический поиск информации (Я.Л. Шрайберг), многомерный анализ глобальных информационных ресурсов (О.Б. Сладкова, Н.В. Лопатина, С.М. Оленев, С.А. Чазова), их многокритериальный анализ и замысловатую экономику распределения информационных продуктов (А.Б. Антопольский, А.Б. Долгин). В целом, реализуя многообразие спектра интеллектуальных функций (Г.С. Осипов), компьютерные технологии неотъемлемы от инфраструктуры современных естественных и социо-гуманитарных наук, для синтеза которых (В.Л. Макаров, В.И. Якунин, С.С. Сулакшин, С.Г. Кара-Мурза, В.А. Твердислов¹) философия решает традиционные задачи методологической интеграции и мировоззренческой ориентации.

Актуальность философии ИИ обусловлена не столько макрометодологическими интенциями конвергентного синтеза мега-нано-технологий, сколько менее масштабными проектами электронной культуры, современный этап развития которой характеризуется бурным становлением электронных институций (А.Б. Антопольский). В новых социальных кластерах тесно переплетаются элементы традиционной и электронной культуры. Спонтанно возникают и постепенно начинают доминировать в общественной жизни э-музеи, э-библиотеки, э-образование, э-наука, э-искусство, э-государство, э-правительство, э-политика, э-бизнес, э-туризм, э-сообщества, э-семья, э-дом и пр. (префикс «э-» акцентирует на диффузии электронных технологий в социальные практики). Становление э-институций непосредственно обуславливается сложными базисными культурными сценариями (В.М. Розин). Философия ИИ призвана соразмерять общественные интенции с возможностями компьютерных технологий, осуществлять конструктивно-критическую экспертизу проектов э-культуры, обосновывать социокультурные программы формирования э-институтов, э-хозяйства, э-экономики, э-власти, э-общества, э-личности.

В наибольшей степени актуальность философского осмысления искусственного интеллекта как феномена современной культуры (Ю.Ю. Петрунин) и, в общем, как феномена постнеклассической рациональности (В.С. Стёпин) обусловлена повседневной жизнью. В ней укоренены интеллектуальные техноло-

¹ Современные идеи синтеза наук обозначены в пленарных докладах на Всероссийской научной конференции «Гуманитарные и естественные науки: проблемы синтеза», 3 апреля 2012 г., Президиум РАН, г. Москва: В.Л. Макаров «Об использовании компьютерных методов в общественных науках», Якунин В.И. «Новая гуманитаристика как производительная сила общества», Сулакшин С.С. «Проблема научности гуманитарного познания», Кара-Мурза С.Г. «Две ветви обществоведения: натурфилософская и научная», Твердислов В.А. «Активные среды как (методологическая) основа конвергенции гуманитарных и естественных наук».

гии, – от бытовых приборов, работающих в недоопределённой динамике внешней и внутренней среды, до глобальных социальных веб-сетей, организация которых предполагает селективную коммуникацию в соответствии со смыслами деятельности виртуальных акторов. В среде электронных технологий и коммуникаций по-иному воспринимаются идеалы социального бытия, трансцендентная проблематика, смысл жизни. Статические архетипы традиционной культуры сменяются канонами программируемой рациональности и, чем сложнее универсалии культуры как программы надбиологической жизнедеятельности (В.С. Стёпин), тем богаче и ярче проявляется феномены электронной культуры. Философская рефлексия по проблемам ИИ стала непосредственно соотноситься с проблемой интерсубъективности (Н.М. Смирнова, Д.И. Дубровский), в связи с чем требуется разработка философии искусственного интеллекта как когнитивистской версии социальной, персоналистической и экзистенциальной философии.

Существующие на данный момент варианты философии ИИ не отвечают масштабу этих актуальных задач. Ряд версий вписывает частные методологические концепции ИИ в несоизмеримо широкие теории, например, в общефилософскую онтологию и гносеологию (А. Сломан), категориально-методологический аппарат диалектического материализма (С.М. Шалютин, Г.Н. Поваров, А.К. Зыков, Л.С. Восков). Другие варианты, напротив, слишком узки – например, программистский (Дж. Маккарти), когнитивистский (М. Боден), нейрокомпьютеринговый (Ю.Ю. Петрунин, А.В. Савельев, М.А. Рязанов, П. и П. Чёрчленды), социально-эпистемологический варианты (М.Ю. Опёнков).

В связи с несоответствием существующих версий философии ИИ тенденциям развития интеллектуальной электронной культуры, диссертант выдвинул гипотезу, состоящую из двух положений. Во-первых, философия ИИ представляется междисциплинарной сферой коллективных знаний и практик философов, учёных и инженеров. Во-вторых, философия ИИ – автономная сфера философских исследований, определённая основным вопросом («Может ли машина мыслить?») и конституированная специфическим категориально-методологическим аппаратом. На данную гипотезу впервые обратили внимание Д.И. Дубровский, Н.М. Смирнова, А.Ф. Зотов, А.В. Иванов в ходе проведения I-ых грязновских чтений (философский факультет МГУ им. М.В. Ломоносова, 2003 г.). В последующем обсуждении проекта «Философия искусственного интеллекта» приняли участие В.А. Лекторский, В.К. Финн, К.К. Колин, И.И. Быстров, А.С. Нариньяни, В.В. Миронов, А.А. Гусейнов, А.А. Скворцов,

В.А. Кувакин, Ф.М. Гиренок, Д.Г. Лахути, Г.П. Путилов, Д.В. Быков, А.П. Огурцов, В.И. Аршинов, И.П. Меркулов, А.С. Карпенко, И.К. Лисеев, Н.С. Юлина, В.Г. Кузнецов, В.А. Глазунов, В.И. Самохвалова, Б.В. Бирюков, А.Д. Королёв, В.В. Трушков, Е.В. Демидова, Л.А. Филлипенко, Н.Х. Сатдинова и др. Результатом явилось проведение Всероссийской конференции «Философия искусственного интеллекта», 18-20 января 2005 г., г. Москва, Московский институт электроники и математики (МИЭМ), сопредседатели: В.Л. Макаров, В.А. Лекторский, Д.В. Быков, Д.И. Дубровский, Г.С. Осипов, уч. секретарь – А.Ю. Алексеев. Конференция вызвала широкий общественный резонанс и оказалась крайне востребованной в условиях бурного становления электронной культуры, так как научное мероприятие подобного масштаба не проводилось с 1965 г. Диссертант полагает, что, начиная с дискуссий данной конференции, наметился современный этап систематического развития отечественной философии ИИ. Решением конференции явилось создание в марте 2005 г. Научного Совета РАН по методологии искусственного интеллекта при Отделении общественных наук РАН – НСМИИ РАН, председатель – ак. В.Л. Макаров, сопредседатели – ак. В.А. Лекторский, проф. Д.И. Дубровский, уч.секр. – А.Ю. Алексеев. Междисциплинарный характер изучения ИИ в рамках мероприятий НСМИИ РАН, в научной и технической организации которых диссертант принимал непосредственное участие, убедительно подтвердил первое положение гипотезы. На протяжении семи лет тематика исследований значительно расширилась и конкретизировалась, количество исследователей выросло в дисциплинарном и региональном масштабах. Это показывает, что искусственный интеллект стал существенным атрибутом современной культуры и достоин самостоятельного направления в философской науке. Экспликации данного положения посвящена первая глава данной работы.

Второе предположение связано с выявлением и развитием методологического инструментария, придающего междисциплинарным исследованиям ИИ статус специфического философского направления. Обоснованию этого предположения посвящены вторая и третья главы.

Важность темы определена востребованностью для развития философии ИИ единого категориально-концептуального аппарата междисциплинарных исследований. Необходимо разработать применительно к рефлексии над проблематикой когнитивно-компьютерных технологий методологические средства, аналогичные общепhilosophическим категориям.

Эталоном междисциплинарных дискуссий по изучению базовых принципов создания, функционирования и применения в обществе компьютеров, которым приписывается (атрибутируется) «интеллект», принято считать тест Тьюринга (далее – ТТ). Показателем его международного признания является большое разнообразие различных версий, модификаций и интерпретаций, предложенных на протяжении шестидесяти лет осмысления его роли для философии, науки, техники, искусства, культуры. Тесты Тьюринга атрибутируют компьютерным системам разнообразный спектр витальных, психических, персональных, социальных феноменов, внося существенный вклад в их изучение. Однако отдельное и разрозненное применение частных тестов не обеспечивает многостороннее и многомерное изучение когнитивных феноменов в контексте гиперсложного пространства электронной культуры, поэтому необходимо объединить тесты в комплексном формате. Предполагается, что Комплексный тест Тьюринга послужит концептуальной основой интегрального междисциплинарного освоения проблематики ИИ: идентификации, спецификации, систематизации, дифференциации, унификации, координации и интеграции сложных знаний, методов и средств, касающихся проблематики ИИ.

Эмпирическую базу исследований составляют результаты современных междисциплинарных исследований по актуальным философско-методологическим проблемам искусственного интеллекта, организованные в рамках международных и всероссийских научных мероприятий НСМИИ РАН при непосредственном участии диссертанта с 2005 г. по 2008 г. в должности ученого секретаря, с 2008 г. по настоящее время – координатора научных программ.

Теоретической базой исследований являются работы по философии искусственного интеллекта, функционалистской, точнее - компьютеристской (computationalism) – парадигме мышления, современной аналитической философии сознания в рамках функционализма, физикализма, бихейвиоризма, дуализма, эпифеноменализма, экстернализма, народного и теоретического психологизма. Ориентации среди этих многочисленных направлений способствовала широко известная библиография Д. Чалмерса «Философия сознания». Крупный блок теоретических оснований составляют проекты искусственной жизни, искусственной личности, искусственного общества. Если последний проект достаточно широко освещается в отечественной литературе благодаря деятельности «Лаборатории искусственных обществ» (руководитель – ак. В.Л. Макаров), то проект искусственной личности и связанные с ним проблемы «другого соз-

нения» и «философских зомби» слабо представлены в русскоязычных переводах.

Основными теоретическими источниками являются многочисленные работы, посвящённые тесту Тьюринга, включая оригинальную статью «Вычислительные машины и интеллект» (1950 г.), в которой сформулированы категориальные идеи атрибуции «мышления» машине, принципиальный проект интеллектуального компьютера и типовой полемический стандарт междисциплинарных дискуссий по поводу интеллектуальных машин. Со многими частными тестами русскоязычные исследователи знакомятся впервые.

Особая роль в диссертационном исследовании отводится статье отечественного пионера когнитивной науки С.Н. Корсакова «Очерк о новом способе исследования посредством машин для сравнения идей» (1832 г.), впервые в полном объёме переведённую спустя 175 лет с французского языка на русский под научной редакцией диссертанта. Для раскрытия принципиальных возможностей реформирования современного компьютеринга революционными идеями С.Н. Корсакова использовались классические теоретико-алгоритмические модели, что выявило двойственность проекта Корсакова относительно традиционных принципов вычисления и его своеобразную постнеклассическую специфику: в отличие от машины Тьюринга как метафоры работы «математика», пребывающего в «рефлексии» автоматных состояний, обзоревающего и преобразующего «мир» посредством узкого канала символической репрезентации, в машине Корсакова действует художник, энергично «пробивающий» многочисленные «окна в реальность» для нерелексивного её «схватывания». Концептуальная двойственность проектов предопределила обоснование проекта Корсакова как протонейрокомпьютера и привлечение литературы, отражающей оппозицию «репрезентационизм/коннекционизм».

Ряд философско-методологических положений опирался на работы, посвящённые теоретическим и инженерным проблемам ИИ. Привлекались работы по проектированию информационных технологий, информационных систем, баз данных и баз знаний, лингво- и нейро-компьютерингу, по символической логике, теории алгоритмов и программирования, истории кибернетики и информатики. В этой рубрике широко представлены труды членов Российской ассоциации ИИ.

Степень изученности темы

Современная отечественная философия ИИ как система междисциплинарных исследований представляет широкий и разнообразный спектр темати-

ческих рубрик. Здесь присутствуют работы предельной теоретической общности, например, парадигма постнеклассической рациональности как контекст исследований ИИ (В.С. Стёпин, В.И. Аршинов, В.А. Глазунов). Социально-философская тематика прослеживается в работах К.К. Колина, В.И. Самохваловой, Н.М. Смирновой, В.К. Финна. Широко представлены профильные методологические исследования, например, в области нейрофизиологии (А.М. Иваницкий, Д.П. Матюшкин, В.С. Сергин), компьютерной системотехники (А.С. Нариньяни, В.Г. Горохов, Г.Г. Рябов).

Философию искусственного интеллекта как специфическую сферу философского знания изучали Алексеев А.Ю., Алексеева И.Ю., Катречко С.Л., Опенков А.М., Петрунин Ю.Ю., Поспелов Д.А., Савельев А.В., Уёмов А.И., Финн В.К., Шалютин С.М., Юлина Н.С., Boden M., Churchland P.M., Churchland P.S., Dennett D.C., Dreyfus H.L., Penrose R., Putnam H., Searle J.R., Sloman A., Stalker D. F., Wang H. и др.

Тест Тьюринга как особый предмет философских исследований изучали: Turing, A., Алексеев А.Ю., Вархотов Т.А., Васильев В.В., Дубровский Д.И., Зотов А.Ф., Ключева Н.Ю., Костикова А.А., Лещёв С.В., Смирнова Н.М., Деннетт Д., Маккарти Дж., Серль Дж., Хауэлл Р., Жижек С., Юлина Н.С., Saygin A.P., Cicekli I., Akman V., Bringsjord S., Bowie G., Caplain G., Carleton L., Cole D.J., Copeland B.J., Oppy G., Dennett D.C., Dowe D., Chalmers D.J., Barresi J., Block N., Clark T., Cohen L.J., Proudfoot D., Sylvan R., Crawford C., Crockett L., Dretske F., Dyer M., French R.M., Piccinini G., Gunderson K., Harnad S., Hauser L., Hofstadter D.R., Jacquette D., Karelis C., Lassague J., Leiber J., Lucas J.R., McCullough D., McDermott D., McGinn C., Michie D., Moor J.H., Rapaport W., Schweizer P., Shanon B., Thagard P., Watt S. и др.

В разработку методологических и эпистемологических проблем ТТ внесли вклад Алексеев И.Ю., Анисов А.М., Батура Я.Н., Бахтияров К.И., Березняк М.В., Бескова И.А., Боброва А.С., Богатых Б.А., Бодякин В.И., Болдачев А.В., Борисова Е.О. Бурцев М.С., Виноградов К.А., Виноградов Д.В., Ворожцов А.В., Горбушин Н.Г., Дубровский Д.И., Еремин В.М., Загоскина Н.А., Зорина М.А., Зорина Н.В., Зотов А.Ф., Каракозова Э.В., Кислов А.Г., Колин К.К., Кольяпин М.В., Коновалова Н.В., Кострикина И.С., Кочергин А.Н., Криницын Л.В., Крылов А.К., Кузнецов О.П., Кузнецов С.О., Лекторский В.А., Леонов А.М., Макаров В.Л., Махиборода А.В., Меркулов И.П., Милов В.А., Михайлов Д.В., Моисеев В.И., Нариньяни .С., Огурцов А.П., Павлов К.А., Павлов Ф.А., Пашинцев П.Е., Петрунин Ю.Ю., Плесневич Г.С., Побережный А.А., Ребещенкова И.Г.,

Редько В.Г., Савельев А.В., Сафонов Е.В., Сокол Л.М., Сорокина Е.С., Талкер С.Л., Тарасов В.Б., Тимофеева З.М., Ускова Е.В., Филатов В.П., Херсонюк Н.Г., Шклярик Е.Н., Янковская Е.А., Ярославцева Е.И. и др.

Теоретический контекст теста Тьюринга, и именно функционалистскую парадигму мышления разрабатывали Дубровский Д.И., Иваницкий А.М., Мцхветадзе А.В., Черниговская Т.В., Абрамова Н.Т., Алексеев А.Ю., Кураева Т.А., Антипенко Л.Г., Бугаков И.А., Гайдамачук О.В., Герасимова И.А., Тумасян А.К., Шпёррле М., Груздева М.Л., Заикин С.И., Карицкий И.Н., Кочергин А.Н., Ладов В.А., Лобастов Г.В., Лоскутова Е.В., Майленова Ф.Г., Макаручук М.М., Огурцов А.П., Панов О.В., Самохвалова В.И., Сатдинова Н.Х., Сердюков Ю.М., Серёдкина Е.В., Соловьев О.В., Бекова Е.А., Суворов А.В., Торосян А.Ц., Трушкина Н.Ю., Кундин Л.В., Юлина Н.С. и др.

Методологическую взаимосвязь исследований естественного и искусственного интеллекта изучали Дубровский Д.И., Холодная М.А., Булычёв И.И., Шутов Р.В., Баксанский О.Е., Смирнова Н.М., Юлина Н.С., Розин В.М., Кочергин А.Н., Кудряшова Т.Б., Кузнецов О.П., Черниговская Т.Б. и др.

Проблемы моделирования систем ИИ изучали Редько В.Г., Глазунов В.А., Чистоходова Л.И., Игнатъев М.Б., Катречко С.Л., Буркова И.А., Семенов Н.В., Харитонов А.Ш., Чумакова Е.И., Воронков Г.С., Глинский И.Л., Земскова О.О., Нежванованова И.Ю., Кольцов М.А., Ширнин О.Ю., Карпук Н.В., Рубашкин В.Ш., Чистоходова Л.И., Истратов В.А.

Логические и математические вопросы ТТ разрабатывали Меркулов И.П., Васюков В.Л., Карпенко А.С., Павлов С.А., Кузичев А.С., Бахтизин А.Р., Анисов А.М., Бахтияров К.И., Бодякин В.И., Махиборода А. В., Нечитайлов Ю.Ю. и др.

Социокультурные аспекты изучения искусственного интеллекта изучали Алексеев А.Ю., Булычёв И.И., Кирюшин А.Н., Артюхов А.А., Крючков В.Л., Маликова Я.С., Розов М.А., Баксанский О.Е., Кучер Е.Н., Бондаренко В.М., Бондаренко С.Б., Бубнов А.В., Восков Л.С., Зыков А.К., Ганжа А.Г., Гусева А.В., Демиденко Э.С., Дорохова О.А., Зотов А.Ф., Иноземцев В.А., Казакова А.Е., Карелин В.М. Катречко С.Л., Королёв А.Д., Кузьмин А.А., Лещев С.В., Никольский А.Е., Новиков Е.В., Петруня О.Э., Плужникова Н.Н., Поваров Г.Н., Пойзнер Б.Н., Соснин Э.А., Розин В.М., Савельев А.В., Светлов С.В., Скворцов А.А., Смирнова Н.М., Таланов В.М., Тетюев Л.И. Филиппенко Л.А., Черногорцева Г.В., Черняков А.Г., Дрозд А.Л., Кислюк А., Маклашевская Н.В., Болдачев

А.В., Тарасов В.Б., Косинова Т. А., Матанцева И.Ю., Мухина К.И., Сёмочкин М.А., Чижов И.Г., Тумасян А.К.

Методологические проблемы роботизации и исследования виртуальной реальности рассматривали Алексеев А.Ю., Алексеева И.Ю., Бондарь А.В., Петрунин Ю.Ю., Савельев А.В., Болнокин В.Е., Глазунов В.А., Чан Дык Хай, Ле Ван Хок, Ву Нгок Бик, Нгуен Минь Тхань, Глазунова О.В., Карпук Н.В, Данилевский И.В., Еремин В.М., Орлов А.С., Шевцов А.И., Ивашов Е.Н., Пак М.М., Панов А.В., Александров Е.В., Домась К.И., Игнатъев М.Б., Лесков Л.В., Лобовиков В.О., Мешков Н.А., Неусыпин К.А., Нгуен Кхан Кхием, Фам Вьет Куонг, Нехамкин В.А., Пронин М.А., Станкевич Л.А., Тимофеев А.И., Афонин В.Л, Макушкин В.А., Быков Р.Э., Борисов В.А., Гаврин Д.В., Кундин Л.В., Кирик Т.А., Дрозд А.А.

Интеллектуальные системы в гуманитарных науках изучали Лахути Д.Г., Финн В.К., Бениаминов Е.М., Воронов А.В., Гринченко С.Н., Гусакова С.М., Егоров А.Г., Еремин В.М., Зайцев П.Л., Зацман И.М., Зенкин А.А., Иванова Е.Ю., Капелько О.Н., Колин К.К., Лисин А.И., Михеенкова М.А., Мунин П.И., Рыжов В.П., Соколова И.С., Шемакин Ю.И., Нариньяни А.С.

Искусственный интеллект как фактор развития электронной культуры исследовали Алексеев А.Ю., Савельев А.В., Лещёв С.В., Шлыкова О.В., Лукьянова Н.А., Болнокина Е.В., Гейбатов С.Г. , Антопольский А.Б., Кудряшова Т.Б., Колин К.К., Конькова Т.А., Лесин С.М., Савельева О.А., Карпук С.Ю., Кузнецова Т.В., Кириллова А.А., Гук А.А., Фролова Ю.В., Середкина Е.В., Глазунов В.А., Ширинкин М.А., Хейло С.В., Иванова В.Г., Дрозд А.Л., Делицын Л.Л. , Орлов К.В., Родионова Л.М., Ключина А.А., Тенякова О.М., Ярославцева Е.И., Харламов С.Ю., Беляев Г.Ю., Иванова В.Г., Шульга Е.Н.

Проблему интерсубъективности («другого сознания») в контексте философии ИИ и теста Тьюринга разрабатывали Смирнова Н.М., Дубровский Д.И., Алексеев А.Ю., Зайцев А.И., Alexander P., Aune B., Austin J., Ayer A. J., Baron-Cohen, Cohen D.J., Buck R., Castaneda H., Everett T., Feigl H., Gallagher K., Glennan S.S., Hampshire S., Hyslop A., Jorgensen J., Reuter B.M., Locke D., Malcolm N., Melnyk A., Pap A., Peacocke C., Price H.H., Sagal P., Sober E., Spencer W., Ziff P.

Проблему философских зомби изучали: Алексеев А.Ю., Дубровский Д.И., Гарнцева Н.М., Нагуманова С.Ф., Кураева Т.А., Balog K., Bennett K., Blackmore Susan J., Braun D., Bringsjord S., Brueckner A., Byrne A., Chalmers D.J., Cottrell A., Dempsey Liam, Flanagan O., Polger T., Guzeldere G., Hill C.S., McLaughlin

B.P., Kirk R., Kraemer E. R., Levine J., Locke D., Marton, P., Melnyk A., Moody T., Murat A, Guzeldere G., Nagel T., Perkins M., Perry J., Squires R., Stoljar D., Thomas N., van Gulick R., Yablo S.

Проект искусственной личности представлен в работах следующих авторов: Алексеев А.Ю., Деев В.В., Серёдкина Е.В., Конькова Т.А., Куляница А.Л., Dennett, D., Pollock J., Barresi J., Bringsjord S., Brown M., O'Rourke J., Hobbs J., Massing W., Mulhauser G. R., Rickert N. W., Scholl B., Tirassa M. и др.

Проблематику «искусственных обществ» разрабатывали Макаров В.Л., Бахтизин А.Р., Истратов В.А., Конькова Т.А., Шамхалов Ф.И., Бесстремьянная Г.Е., Бобкова И.А., Редько В.Г., Бурцев М.С., Бахтизина Н.В., Цветоват М., Рullo M., Лавров А.А., Савельев А.В., Алексеев А.Ю., Bartle R., Burke T., Castronova E., Eiben A.E., Epstein J.M., Axtel R., Kennedy J., Eberhart R.C., Tesfatsion L., Tomlinson B., Blumberg B., Yee N. и др.

Объект исследования – искусственный интеллект как феномен современной культуры.

Предмет исследования – философский и методологический потенциал комплексного теста Тьюринга в развитии философии искусственного интеллекта.

Цель исследования – разработка категориально-методологического инструментария философских исследований искусственного интеллекта на основе комплексного теста Тьюринга.

Задачи исследования:

1. Изучение искусственного интеллекта как феномена электронной культуры.
2. Систематический анализ многочисленных версий теста Тьюринга.
3. Разработка комплексного теста Тьюринга как базового категориального аппарата философии искусственного интеллекта и как концептуального инструментария исследования когнитивных проблем электронной культуры.
4. Анализ способов приписывания компьютерным системам социальных, персональных и ментальных феноменов (экспликация дефинитной функции комплексного теста Тьюринга).
5. Философская экспертиза технологий построения интеллектуальных компьютерных систем, проектов электронной культуры и программ развития искусственного интеллекта (экспликация критической функции комплексного теста Тьюринга).

6. Разработка философско-методологических средств, пригодных для обоснования стратегии развития междисциплинарных исследований искусственного интеллекта, для модификации ориентиров научно-методологических исследований искусственного интеллекта и для реформирования фундаментальных принципов компьютерной технологии (экспликация конструирующей функции комплексного теста Тьюринга).

7. Позиционирование постнеклассического субъекта электронной культуры (экспликация конституирующей функции комплексного теста Тьюринга).

Методологическая основа исследования

Методология диссертационного исследования имеет комплексный междисциплинарный характер и опирается, во-первых, на аналитическую философию сознания и философию искусственного интеллекта, во-вторых, на общую методологию когнитивных, компьютерных наук и нейронаук, в-третьих, на философско-методологический императив экспликации научно-теоретического содержания в когнитивных структурах жизненных миров. В работе задействован ряд положений лингвистической философии, социально-философской феноменологии, философской герменевтики. Методология исследований опиралась на логические приемы критики и аргументации, принципы компаративистского анализа, методы интерпретации исторических текстов, парадигмы программирования и проектирования информационных систем, фундаментальную теорию алгоритмов, символическую логику. Базовой методологической основой являются методы концептуального анализа и синтеза, реализуемые средствами мысленных экспериментов.

Теоретической основой исследований являются работы по философии искусственного интеллекта, современной аналитической философии сознания в рамках функционализма, физикализма, бихевиоризма, дуализма, эпифеноменализма, наивного, народного и теоретического психологизма. Ориентации среди этих многочисленных работ способствовала широко известная веб-библиография «Philosophy of Mind» (Д.Чалмерс), в которой философии искусственного интеллекта отводится дисциплинарный подраздел.

Эмпирическую базу исследований составляют международные, всероссийские и региональные научные мероприятия Научного совета РАН по методологии искусственного интеллекта (НСМИИ РАН), организованные при непосредственном участии диссертанта в должности ученого секретаря (2005-2008 гг.) и координатора научных программ (по настоящее время). В контексте этих

исследований «философия искусственного интеллекта» представлена динамичной системой междисциплинарных проблем и подходов к их решению.

Научная новизна работы

1. Впервые в отдельное направление «философия искусственного интеллекта» выделяются методологические и мировоззренческие исследования компьютерного представления идеализированных моделей естественного интеллекта.

2. Продемонстрирована эффективность осуществления междисциплинарной координации философско-методологических положений, научно-теоретических методов и инженерно-практических приемов исследования искусственного интеллекта посредством рефлексии над тестом Тьюринга и его версиями как категориальными мысленными экспериментами.

3. Раскрыта роль комплексного теста Тьюринга как концептуального инструментария критического исследования когнитивных феноменов электронной культуры: на философско-методологическом уровне представлен метод обоснования стратегии развития искусственного интеллекта; на научно-теоретическом уровне продемонстрирована эффективность экспертизы компьютерных систем на предмет их «интеллектуальности»; на инженерно-практическом уровне показана возможность оценки целесообразности применения компьютерных технологий в социальных практиках.

4. Несмотря на абсурдность теста зомби и футурологическую отдаленность теста искусственной личности, продемонстрирована целесообразность их совместного использования для категориального анализа проблемы интересубъективности в контексте электронной культуры.

5. Акцентируется философско-методологическая значимость машины Корсакова, осуществлена интерпретация ее как формального определения коннекционистского алгоритма и предложен проект комплексной машины Корсакова-Тьюринга, объединяющий коннекционистскую и символическую парадигмы искусственного интеллекта.

6. Разработан критерий демаркации интеллектуальной компьютерной системы, задаваемый анализом ее логико-лингвистической организации в контексте трехмерной семантики, которая фиксирует концептуальную взаимосвязь приватного когнитивного феномена, его научной репрезентации и компьютерной реализации.

7. Предложена парадигма тестового функционализма (тестового компьютеризма), способствующая конституированию феноменологии электронной

культуры посредством исполнения постнеклассическим субъектом функций комплексного теста Тьюринга.

Основные положения, выносимые на защиту:

Концептуальный статус комплексного теста Тьюринга в философии искусственного интеллекта определяется совокупностью функций:

1. Комплексный тест Тьюринга способствует развитию философии искусственного интеллекта как сложной системы междисциплинарных исследований, являясь необходимым концептуальным средством предварительной междисциплинарной координации исследовательских программ в области искусственного интеллекта;

2. Комплексный тест Тьюринга служит основой понятийного аппарата философии искусственного интеллекта и определяет когнитивные явления электронной культуры, приписывая компьютерным системам витальные, психические, личностные и социальные феномены;

3. Комплексный тест Тьюринга обеспечивает обоснование концептуальной критики технологий искусственного интеллекта, формирует новые теоретические принципы построения интеллектуальных компьютерных систем и способствует реформированию фундаментальных основ компьютерной технологии;

4. Комплексный тест Тьюринга осуществляет концептуальную экспертизу проектов развития электронной культуры и обеспечивает категориальное решение проблемы интерсубъективности посредством совместного применения теста искусственной личности и теста зомби.

5. Комплексный тест Тьюринга способствует построению феноменологии компьютерного мира, обеспечивая постнеклассическое позиционирование субъекта электронной культуры.

Теоретическая и практическая значимость работы

Философское понимание проблем искусственного интеллекта и разработка методологии познания электронной культуры имеют важное мировоззренческое значение для развития информационного общества. Диссертационное исследование закладывает основы координации междисциплинарных исследований по онтологическим, гносеологическим, социокультурным, научно-теоретическим и инженерно-технологическим проблемам искусственного интеллекта.

Образовательное значение состоит в обновлении вузовских курсов по философии достижениями в постановке и в решении междисциплинарных проблем искусственного интеллекта. Положения работы целесообразно использовать для разработки новых и дополнения существующих учебных курсов бакалавров, магистрантов и аспирантов по истории и философии науки, методологии научного исследования, философии информатики, философии техники, социальной философии, философии психологии, философии культуры. Для развития образовательной стратегии исследований искусственного интеллекта целесообразно включать результаты работы в учебные курсы по направлению подготовки «Интеллектуальные системы в гуманитарной сфере».

Дефинитная функция комплексного теста обеспечивает рациональное обсуждение искусственных систем, в которых когнитивные феномены представлены в неразрывной связи с научными интерпретациями результатов их анализа и компьютерными проектами имитации этих феноменов.

Критический потенциал комплексного теста Тьюринга целесообразно использовать в социально-философской экспертизе государственных и общественных программ развития электронной культуры России.

Конструирующую функцию комплексного теста Тьюринга следует, в частности, использовать в методологии проектирования интеллектуальных компьютерных систем и, в целом, для разработки стратегии модернизации компьютерной инфраструктуры современной цивилизации.

Конституирующая функция значима для решения праксеологических задач постнеклассическим субъектом электронной культуры, так как комплексный тест Тьюринга способствует формированию функционалистской онтологии, в которой концепт «Я» выносится за скобки квазиалгоритмических детерминаций компьютерного мира, продолжая играть для него регулятивную роль.

Апробация работы

Основные идеи диссертации изложены в научных публикациях диссертанта – в монографиях, статьях, докладах, учебных программах. Диссертант отстаивает положение о тесной взаимосвязи научно-теоретической и научно-организационной деятельности. Поэтому идеи работы оказались востребованными для проектов международных, всероссийских и региональных мероприятий НСММИ РАН, основными из которых диссертант считает следующее: конференция «Философия искусственного интеллекта», МИЭМ, 18-20 января 2005 г.; ежемесячный семинар НСММИ РАН "Философско-методологические основы искусственного интеллекта", ЦЭМИ РАН, ИФ РАН, 2005-2015 гг. (руково-

дители: С.Н.Васильев, Д.И.Дубровский, В.А.Лекторский, В.Л.Макаров, диссертант – ученый секретарь и научный координатор): ежегодная конференция студентов, аспирантов, молодых учёных «Искусственный интеллект: философия, методология, инновации» (проводится с 2006 г.).

Впервые идея разработки философско-методологических основ искусственного интеллекта путем применения различных версий теста Тьюринга была предложена в проектах молодёжной конференции «Философия искусственного интеллекта», 20 мая 2004 г., МИЭМ; круглого стола «Функционализм и компьютерные науки», МИЭМ, 20 января 2005 г.; круглого стола «Философско-методологические проблемы когнитивных и компьютерных наук» на IV Российском философском конгрессе «Философия и будущее цивилизации, МГУ им.М.В.Ломоносова, г.Москва, 24-28 мая 2005 г. В докладе «Человек и компьютер: проблема различения естественного и искусственного интеллекта» на семинаре «Человек и мир», 24 июня 2004 г., Институт микроэкономики, вопрос А.Тьюринга – «Может ли компьютер мыслить?» – обозначился как «основной вопрос философии искусственного интеллекта».

Идея комплексного теста Тьюринга отражена в докладах: «Роль теста Тьюринга в исследованиях искусственного интеллекта» на заседании семинара НСММИ РАН «Философско-методологические проблемы искусственного интеллекта», 26 апреля 2006 г., ЦЭМИ РАН; «Комплексный тест Тьюринга и проблема сознания» на конференции "Философия сознания: аналитическая традиция. Третьи Грязновские чтения", философский факультет МГУ имени М.В.Ломоносова, 6 ноября 2009 г.; «Идея комплексного теста Тьюринга» на конференции «Тест Тьюринга: философские интерпретации и практические реализации», 19 ноября 2010 г., там же; «Роль комплексного теста Тьюринга в синтезе гуманитарных и естественных наук» на конференции «Гуманитарные и естественные науки: проблемы синтеза», Президиум РАН, 3 апреля 2012 г.

Комплексный тест как концептуальный инструмент познания явлений электронной культуры был предложен в докладах: «Когнитивные исследования в электронной культуре» на Второй конференции «Актуальные проблемы современной когнитивной науки», Ивановский государственный университет, г.Иваново, 15 октября 2009 г., в докладе на 50-м заседании семинара НСММИ РАН 16 декабря 2011 г., ИФ РАН, в проекте всероссийского круглого стола «Может ли компьютер читать?», 25 июня 2014 г., ИФ РАН.

Перспективы построения постнеклассической теории сознания на базе комплексного теста Тьюринга обсуждались в докладах: «Эволюционно-

эпистемологические особенности функционализма теста Тьюринга» 21 марта 2013 г. на круглом столе «Эпистемологические основы когнитивных исследований» в Российском государственном гуманитарном университете; «Функционализм машины Тьюринга vs функционализм машины Сеченова» на симпозиуме «150 лет “Рефлексам головного мозга”», 6 декабря 2013 г., Первый Московский государственный университет им.И.М.Сеченова; «Тестовый компьютеризм как высшая форма функционализма» на VII-ой конференции с международным участием «Актуальные проблемы современной когнитивной науки», Ивановский государственный химико-технологический университет (ИГХТУ), 17 октября 2014 г.

Идея о взаимной обусловленности исследований искусственного интеллекта и развития электронной культуры обсуждалась в докладе «Методологические контуры наук о культуре» на конференции «Новые пути наук о культуре», 28 мая 2009 г., Московский государственный университет культуры (МГУКИ). Данная тема нашла продолжение в докладах на всероссийских конференциях, проводимых в г.Лодейное поле Ленинградской области: «Электронная культура: трансляция в социокультурной и образовательной среде» (2007 г.), «Неопросветительство в малом городе» (2009 г.), «Электронная культура: феномен неопросветительства» (2010 г.), «Проблема интерсубъективности в электронной культуре» (2012 г.). Интерес научной общественности вызвала программа круглых столов «Искусственный интеллект и развитие электронной культуры России» и «Проблемы развития электронной культуры России», которые проводились, соответственно, в рамках XI и XV Международной научной конференции “Модернизация России: ключевые проблемы и решения” на базе Института научной информации по общественным наукам (ИНИОН) РАН, 16-17 декабря 2010 г. и 19 декабря 2014 г. Немалое значение для развития этой тематики имел проект работы секции «Электронная культура и искусственный интеллект» в рамках Московского форума культуры «Культура как стратегический ресурс России в XXI веке», 1 июля 2010 г., МГУКИ и Всероссийской конференции «Электронная культура: интеллектуальные технологии», 29 мая 2011 г., МГУКИ.

Тематика философских зомби в контексте методологии искусственного интеллекта обсуждалась в докладах «Проблема философских зомби и искусственный интеллект» (МИЭМ, 18 января 2005 г.), «Проблема философских зомби и искусственное общество» (МГУ им.М.В.Ломоносова, 20 мая 2005 г.). На круглом столе «Что чувствуют зомби?» в рамках международной конференции «Натуралистические концепции сознания», 25 мая 2013 года, СПбГУ, диссер-

тант осуществил общий обзор проблемы зомби в аналитической философии. Абсурдность философских зомби в исследованиях искусственного интеллекта отстаивалась в докладе «Философские зомби vs философские сомнамбулы» на VI-ой конференции «Актуальные проблемы современной когнитивной науки», г.Иваново, 18 октября 2013 г.

Идея использования машины Корсакова для реформирования теоретико-алгоритмических основ компьютерной технологии предложена на заседании лодейнопольского семинара 7 июля 2007 г. Далее идея обсуждалась в докладе «Семен Николаевич Корсаков (1787-1853 гг.) – основатель когнитивной науки» на IV-ой конференции «Актуальные проблемы современной когнитивной науки», г.Иваново, 21 октября 2011 года. В докладе «Электронное визуальное искусство в искусственном интеллекте (философские аспекты)», 7 октября 2009 года на V-ой конференции «Анимация как феномен культуры» во Всероссийском государственном университете кинематографии имени С.А.Герасимова (ВГИК), выдвинуто предположение о том, что машина Корсакова является формальным определением деятельности абстрактного художника, по аналогии с тем, что машина Тьюринга является определением работы абстрактного математика. В докладе «Ретроспектива коннекционизма» на XII конференции «Нейрокомпьютеры: разработка и применение», 18 марта 2013 г., Московский городской психолого-педагогический университет, проект С.Н.Корсакова предложено считать исходным пунктом парадигмы коннекционизма.

Концепция вычислимости на основе машины Корсакова-Тьюринга использовалась при построении объемной семантики терминов когнитивного словаря. 28 апреля 2014 года в рамках V-ой конференции ««Философия. Язык. Культура», Высшая школа экономики – обсуждался доклад «Объемная (3D) интенциональная семантика словаря искусственного интеллекта». В докладе «Объемная (3D) семантика когнитивных терминов» на VI-ой конференции «Актуальные проблемы современной когнитивной науки», ИГХТУ, 18 октября 2013 г., обсуждался критерий демаркации исследований искусственного интеллекта путем трехмерного (три-дименсионального) анализа значений слов, используемых в дискурсе когнитивной науки.

С презентациями проекта «искусственной личности» диссертант выступает, начиная с 1994 г. Философское содержание проекта отражено в докладах: «Искусственная личность: информационно-технологический подход» на конференции «Проблема человека: мультидисциплинарный подход», Институт микроэкономики, 1998 г.; «Проблемы искусственной личности: развитие меж-

дисциплинарного подхода к искусственному интеллекту» на конференции «Проблемы эффективности и безопасности функционирования сложных технических информационных систем», г.Серпухов, 19 апреля 2006 г.; «Когнитивно-компьютерные модели сознания и тест Тьюринга» на конференции «Актуальные проблемы математики, механики, информатики», 12 октября 2006 г., г.Пермь; «Сложные методологические проблемы сложных интеллектуальных систем» на конференции «Актуальные проблемы методологии, философии науки и образования», 9 февраля 2007 г., г.Уфа; «Проблемы проекта искусственной личности» на семинаре НСМИИ РАН, 3 июня 2009 г.

Отдельные положения диссертационного исследования привлекались для разработки материалов по учебным курсам: философский факультет МГУ имени М.В.Ломоносова: «Философско-методологические основы исследований искусственного интеллекта» (спецкурс для бакалавров); Ивановский государственный химико-технологический университет: «Социально-философские основания электронной культуры (программа для аспирантов)»; МГУКИ: «Философские вопросы искусственного интеллекта», «Социокультурные проблемы искусственного интеллекта», «Проблема сознания в электронной культуре», «Психосемиотические аналитические системы», «Интеллектуальные информационные системы», «Логико-математические основы проектирования компьютерных систем», «Информационно-аналитические технологии»; Государственный университет гуманитарных наук (факультет философии и политологии): «Социально-философские вопросы искусственного интеллекта»; Первый Московский государственный медицинский университет им.И.М.Сеченова: «Проблема сознания в электронной культуре».

Структура диссертации

Работа состоит из введения, трёх глав, заключения, библиографического списка литературы из 432 наименований (124 - на иностранных языках). В первой главе – два параграфа, во второй – три, в третьей – четыре параграфа. Общий объем диссертационного исследования составляет 490 страниц.

Глава I. Философия искусственного интеллекта: междисциплинарный поход

1.1. Анализ дискуссий по философско-методологическим проблемам искусственного интеллекта

Современный этап широкого и интенсивного развития отечественных философско-методологических исследований искусственного интеллекта начался с проведения Всероссийской междисциплинарной конференции «Философия искусственного интеллекта», состоявшейся 17-19 января 2005 г. на базе Московского государственного института электроники и математики (МИЭМ) под эгидой Отделения общественных наук РАН, Института философии РАН, философского факультета МГУ им. М.В.Ломоносова, Центрального экономико-математического института РАН (ЦЭМИ РАН), ВИНТИ РАН, Института программных систем РАН (ИПС РАН) и МИЭМ. *Сопредседателями конференции* выступили В.А. Лекторский, В.Л. Макаров, Д.В. Быков, Д.И. Дубровский, Г.С. Осипов. В роли *учёного секретаря конференции* выступил диссертант. Собственно идею проведения конференции он и выдвинул. Первоначально идея была одобрена Н.М. Смирновой, А.Ф. Зотовым. На I-ых Грязновских чтениях (философский факультет МГУ им. М.В. Ломоносова, 30 октября 2003 г.) Д.И. Дубровский крайне положительно отреагировал на данную идею, обсудил её с В.А. Лекторским, который приложил уникальные организаторские способности для привлечения ведущих специалистов ИФ РАН к составлению программы конференции и её PR-компании.

Проблематика философии ИИ вызвала оживленный интерес, неожиданный для организаторов данного мероприятия. Было заявлено 390 анкет участников конференции из 44 городов России. Среди них – 99 докторов наук, 87 – кандидатов наук. Междисциплинарный состав конференции характеризовался следующими специальностями: биологическими, медицинскими, психологическими, техническими, физико-математическими, филологическими, философскими, химическими, экономическими науками. Международное участие было представлено странами: Болгария, Вьетнам, Германия, Армения, Беларусь, Грузия, Украина.

До начала конференции на основе поступивших тезисов докладов в издательстве ИФ РАН был выпущен сборник материалов под редакцией В.А. Лекторского и Д. И. Дубровского. В нём представлено 133 статьи 152 авторов, из них 46 – доктора наук, 56 – кандидаты наук [Философия ИИ, 2005].

На конференции был сформирован круг методологических проблем философии ИИ, который в последующем получил незначительные изменения (пп.1.2). Рассмотрим основные направления отечественной философии ИИ, которые исследователи выдвинули в процессе подготовки и проведения конференции «Философия искусственного интеллекта».

1.1.1. Становление философии искусственного интеллекта

Эпистемологические проблемы искусственного интеллекта

В.Л. Макаров в работе *«Получение нового знания методом компьютерного моделирования»* обсуждал принципиальные вопросы, касающиеся роли и возможностей создаваемых человеком средств познания, «приборов для чтения книги жизни» [Макаров, 2006]). Эти приборы весьма многообразны, начиная с физических инструментов познания, таких как синхрофазотроны, микроскопы, телескопы, и заканчивая средствами изучения социальных явлений, измерения социальных процессов (опрос населения, институты голосования, Интернет-обработка и т.п.). Все эти разнообразные приборы подлежат классификации по аналогии с классификацией наук. Отмечая тот факт, что использование прибора (акт измерения) связан с феноменом «вмешательства» в изучаемое (измеряемое) явление (это особенно четко выражено принципом неопределенности Гейзенберга в физике), докладчик рассмотрел такой сравнительно недавно изобретенный прибор как компьютерное моделирование. В работе выяснялась проблема степени адекватности компьютерной модели и роль мысленного эксперимента. Приводя ряд примеров действующих компьютерных моделей и результатов экспериментов с ними, В.Л. Макаров обсудил вопрос: возможно ли считать результаты такого рода экспериментов новым знанием о реальности, причем именно о действительной, а не виртуальной реальности. Он положительно ответил на данный вопрос, хотя и отметил некоторые парадоксальные моменты и трудности разграничения подлинной и виртуальной реальностей.

Философия когнитивных наук и искусственный интеллект

В.А. Лекторский в работе *«Философия, когнитивные науки и искусственный интеллект»* [Лекторский, 2006] осуществил анализ новой ситуации, которая возникла в исследовании познавательной деятельности в связи с появлением (в конце 70-х гг. прошлого века) и интенсивным развитием т.н. когнитивной науки, пытающейся интегрировать психологию, философию, логику, лингвистику, исследования в области искусственного интеллекта и нейронауки. Было показано, что в этой связи понятия познания и интеллекта приобретают более

широкий смысл, чем это имело место в недавнем прошлом – вплоть до отождествления жизни с познанием. Вместе с тем возникает целый ряд острых философских проблем: природы ментальных репрезентаций как они понимаются в когнитивной науке; разработки новых подходов к пониманию интенциональности; соотношения понятий обыденной и классической психологии с понятиями когнитивной науки; возможностей и границ редукции при изучении психической жизни; объяснения познания как вычисления (по аналогии с работой компьютера) и др. Изучение сознания, которое ныне привлекает внимание множества специалистов как результат развития когнитологических исследований, заставляет обсуждать в новом контексте старые философские вопросы о единстве и множественности сознания, о природе рефлексии и интроспекции, о характере Я и др. В докладе были выделены 3 стадии в развитии разработок в области искусственного интеллекта как результат критики и самокритики этих исследований: 1) Классический вычислительный подход к пониманию познания; 2) Коннекционизм, рассматривающий познание как параллельно распределённую деятельность по переработке информации; 3) Представление познания с точки зрения теории динамических когнитивных систем. В последнем случае когнитивные процессы, тело познающего существа и окружающий внешний мир рассматриваются как три аспекта некоей единой деятельности, резкая дихотомия внешнего и внутреннего снимается, ставится под вопрос существование ментальных репрезентаций. В докладе также было показано, что в современных когнитологических исследованиях воскрешается в новой форме старая философская дилемма в понимании познания: подход Аристотеля и подход Декарта.

Проблема сознания в искусственном интеллекте

Д.И. Дубровский в работах на общую тему *«Искусственный интеллект, мозг и проблема сознания»* [Дубровский, 2005-1; 2005-2; 2006-1; 2006-2] подчеркивал, что теоретическое соотнесение искусственного интеллекта (ИИ) с естественным (ЕИ) – важнейшее условие новых разработок ИИ и вместе с тем одно из актуальных направлений исследования проблемы сознания. В западной аналитической философии преобладают редукционистские подходы к ЕИ, стремящиеся элиминировать качество субъективной реальности (СР), без которого понятие сознания теряет свой главный смысл. Задача же состоит именно в том, чтобы объяснить качество СР. В этом отношении функционалистский тип объяснения (противостоящий парадигме физикализма) сохраняет определенные возможности. Имеется в виду информационный подход к пониманию СР, кото-

рый позволяет теоретически корректно объяснить характер связи явлений СР с мозговыми процессами и их способность управлять телесными изменениями (а вместе с тем объяснение феноменов произвольности, свободного воления на основе принципа самодетерминации в деятельности головного мозга). Качество СР обусловлено особым типом информационных процессов в мозгу, который принципиально отличается от того, что имеет место в системах ИИ. Это качество СР – уникальный эффект биологической самоорганизации, найденный в процессе эволюции и резко повысивший приспособительные возможности усложняющихся живых систем. В рамках парадигмы функционализма (в частности на основе принципа инвариантности информации по отношению к физическим свойствам ее носителя) теоретически допустима возможность создания таких систем ИИ, которые, наконец, обретут главное отличительное свойство ЕИ (качество СР). Однако супероптимистические прогнозы на этот счет требуют критического осмысления, особенно в той части, когда нам обещают скорую замену биологического человека электронным.

Вопросы структуры сознания и подсознательного в связи с задачами моделирования в области ИИ анализировались в работах И.А.Герасимовой «Системно-иерархическая модель сознания и восточная метафизика» [Герасимова, 2005]; А.Ю.Алексеева и Т.А.Кураевой «Проблема зомби и перспективы проекта искусственной личности» [Алексеев, 2005-2; 2005-5], О.В. Панова «Применение объективных инструментов познания функциональной структуры бессознательного как возможность формирования новых принципов искусственного интеллекта» [Панов, 2005]; Н.Ю.Клюевой (Трушкиной), И.В.Михейкина и Л.В.Кундина «Решение проблемы сознания в рамках концепции сильного искусственного интеллекта» [Трушкина, 2005]. Проблема "квантового сознания" обсуждалась Л.Г. Антипенко «Новый взгляд на проблему искусственного интеллекта сквозь призму квантово-компьютерных технологий» [Антипенко, 2005]: ИИ не способен возвыситься до ранга естественного ИИ на основе классических компьютеров с их рекурсивными вычислениями, эта задача может решаться лишь на основе квантово-информационных вычислений. Указанная проблема под углом зрения синергетики рассматривалась И.А. Бугаковым в работе «Информационно-синергетический подход к сознанию как основа для создания искусственного интеллекта» [Бугаков, 2005].

Принципиальные вопросы моделирования функций естественного интеллекта связаны с пониманием природы психического отображения и управления. Они специально рассматривались в ряде докладов. Оригинальную точку зрения на природу психического отображения высказал И.Н. Карицкий в работе «Ос-

новные психологические принципы самообучения восприятию систем искусственного интеллекта» [Карицкий, 2005]. На примере восприятия была рассмотрена «начальная форма психического» как результат соотнесения ансамблей нейронных возбуждений от разных сенсорных систем, вызванных одним и тем же объектом. А.В. Мцхветадзе (Грузия) в работе «Искусственный интеллект и некоторые аспекты природы психики» подчеркнул, что психика возникает как отношение всего организма к определенным структурам мозга и различным внешним и внутренним раздражителям [Мцхветадзе, 2005]. Это отношение и создает качество идеального, т.е. отображения действительности в форме субъективной реальности. Соображения о психическом управлении высказали О.В. Соловьев и Е.А. Бекова (Украина) В работе «Об одном условии осуществления истинно-целенаправленных (субъективно-реализуемых) процессов» [Соловьев, 2005] показано, что всякий акт психического управления включает отображение вероятного будущего, которое детерминирует осуществляемую в настоящий момент реакцию. Особенность психического причинения состоит в том, что функциональные системы управляющих уровней мозга посредством своих семантических и прагматических (т.е. не сугубо физических, а именно информационных) факторов детерминируют объективно осуществляемую активность функциональных систем управляемых уровней.

Нейрокомпьютерные модели искусственного интеллекта

Одно из перспективных направлений исследований в области ИИ связано с нейрокомпьютерным моделированием когнитивных способностей и их эволюцией. В этой связи следует отметить работу В.Г. Редько «Задача моделирования когнитивной эволюции» [Редько, 2005]. Автор представил результаты исследований ведущих лабораторий мира в области нейрокомпьютерного моделирования адаптивного поведения, а также обсудил проект «Мозг анимата», направленный на разработку общей теоретической платформы для построения широкого класса нейрокомпьютерных моделей эволюции когнитивных способностей. Содержащиеся в докладе выводы, а также сам феномен генерации в искусственных нейросетях когнитивных метапрограмм, безусловно, представляют большой интерес для эпистемологии, поскольку касаются процессов эволюции человеческого познания. Интересные методологические проблемы нейрокомпьютерного моделирования когнитивных способностей были подняты также в работе В.И. Бодякина «Концепция построения искусственного разума» [Бодякин. 2005].

Следует, на наш взгляд, отметить работу С.О. Кузнецова по логическому анализу формальных понятий. Здесь применяется технический аппарат соответствия Галуа, теория решеток и операции над графами. Заслуживает внимания также работа Д.В. Виноградова, где средствами первопорядковой логики (и слабой логики второго порядка) формализуется синтез познавательных процедур, включающих индукцию, аналогию и абдукцию [Виноградов 2005]. В итоге реализуется ДСМ метод, суть которого будет раскрыта в следующем пп. (при анализе работ В.К.Финна). В работе В.Б. Тарасова «Синергетический искусственный интеллект» представлен систематический обзор основных исследовательских парадигм в области ИИ [Тарасов, 2005].

Вопрос о соотношении естественного и искусственного интеллекта в той или иной степени оказывается в поле зрения многих исследователей. Естественный интеллект, по-видимому, представляет собой некую интегральную характеристику когнитивных способностей. Его нельзя отождествлять с отдельными когнитивными способностями, например, с мышлением. Мышление тесно интегрировано с работой памяти (кратковременной и долговременной), с восприятием и распознаванием образов, знаково-символическое (логико-вербальное) мышление управляется символьным (вербальным) сознанием и т.д. В ряде работ высказывается мысль о единой когнитивной природе естественного и искусственного интеллекта. Адекватная эпистемологическая концепция естественного и искусственного интеллектов, скорее всего, должна базироваться на моделях переработки информации. Представляется также важным для ИИ развитие аппарата неклассических, в частности, многозначных логик, на что обратил внимание С.А. Павлов [Павлов, 2005; 2006].

Практические приложения теоретических исследований искусственного интеллекта

Г.С. Осипов в докладе *«Искусственный интеллект: немного теории и важнейшие приложения»* дал краткую обобщающую характеристику состояния исследований в области ИИ. Были выделены такие направления работ как представление и приобретение знаний, моделирование рассуждений и поведения, экспертные системы реального времени, динамические интеллектуальные системы, автономные интеллектуальные агенты, методы обучения интеллектуальных систем, задачи анализа текста, модели пользователя, человеко-машинный интерфейс. Было дано краткое описание приложения динамических интеллектуальных систем в задачах исследования дальнего космоса, управления маневрированием в открытом космосе, управления беспилотными самолетами.

тами и вертолетами: проекты WITAS (беспилотный вертолет), CADCON (автономная подводная лодка), Cyber Cars (автономные транспортные средства), МИНЕРВА (робот для выполнения социальных функций) и некоторые проекты интеллектуальных систем военного назначения.

Проблемы компьютерного моделирования естественного интеллекта

Данные проблемы были представлены А.П. Огурцовым «Достижения и трудности моделирования интеллектуальных актов» [Огурцов, 2005]. Представив в систематическом виде основные достижения в области ИИ, он отметил, что трудности современного компьютерного моделирования – это, прежде всего, трудности философии и психологии сознания, эвристические возможности которых отстают от запросов развития ИИ. Компьютерные модели пока далеки от биологической основы работы интеллекта и его эвристических механизмов. Существенный семантический пробел между естественным языком и языками программирования настоятельно требует разработки новых средств анализа естественного языка. В этих целях важную роль призваны играть многозначные логики, теория «нечетких множеств», новые разработки лингвистов, перспективы, открываемые нанотехнологиями.

Интеллектуальный интерфейс «человек-компьютер»

Н.С. Юлина в работе «Обретение разумности: «сократический диалог» и интеракция с компьютером» [Юлина, 2005] проанализировала особенности диалога (в форме устной речи), как человеческого способа коммуникации с его многомерными ценностно-смысловыми измерениями, по сравнению с узким когнитивным диапазоном интеракции человека с компьютером. Эти качественные различия и определяют основные трудности сближения искусственного интеллекта с естественным.

Методологические проблемы роботизации и исследования виртуальной реальности

В.А. Глазунов и М.Б. Игнатьев обращают внимание на множество реальных разработок робототехнических систем различного назначения: к 1989 г. только в СССР было произведено около 80 тысяч промышленных роботов, которые заменили свыше одного миллиона рабочих на тяжелых и монотонных работах. В настоящее время на планете функционирует свыше 10 миллионов роботов, их количество и качество будут расти, что требует серьезного экономического, социального и философского осмысления. Роботы господствуют

при исследованиях в космосе, под водой и других опасных средах, создаются системы боевых роботов. Системы виртуальной реальности превращаются в виртуальные миры, объединяя островки компьютеризации в глобальные «континенты».

В.М. Еремин [Еремин, 2005-1; 2005-2], М.Б. Игнатьев [Игнатьев, 2005] и Л.В. Лесков [Лесков, 2005] отмечают важность выявления структурированной неопределенности, которая может быть использована для управления на различных уровнях.

В.А. Глазунов и О.В. Глазунова [Глазунов, Глазунова, 2005] рассматривают вопросы моделирования творческих актов в робототехнике. И.В. Данилевский представил важные проблемы телепортации, которая, по мнению автора, возможна не только в квантовомеханических системах, но и в гуманитарных структурах [Данилевский, 2005]. В.О. Лобовикова (г. Екатеринбург) поставил проблему правового обеспечения применения роботов и аватаров [Лобовиков, 2005]. М.А. Пронин [Пронин, 2005] рассматривал эпистемологические проблемы виртуальной реальности, что методологически необходимо для строительства виртуальных миров. Л.А. Станкевич (г. Санкт-Петербург) обозначил тему развития проблемы человекоподобных роботов и обобщил собственный большой практический опыт: он создал команду, занявшую первое место на мировом чемпионате по игре роботов в футбол [Станкевич, 2005].

Социогуманитарные аспекты изучения искусственного интеллекта

ИИ как исследовательская программа сегодня представляет собой конгломерат различных дисциплин, связанный не столько общей задачей, сколько общим когнитивным истоком, попыткой ответить на извечный «фаустовской» вопрос: «что значит знать»? [Зотов, 2005]. В поисках ответа на этот вопрос исследователи стремятся воспроизвести («смоделировать») естественный интеллект (частично или полностью) искусственными устройствами – как идеальными (логические формулы и математические алгоритмы), так и «материальными» (счетные машины и информационная техника). Дальнейший прогресс в этом направлении привел к становлению множества самостоятельных дисциплин, объединенных глубокими историческими истоками и общим названием. И такое состояние дел оказывается как источником неопределенности базовых понятий, так и стимулом к широкому междисциплинарному синтезу. Предварительное условие конструктивного обсуждения поставленных проблем, отмечает А.Ф. Зотов, состоит в том, чтобы определить ключевые термины и выразить с

их помощью предметы и задачи каждой из дисциплин, а также – насколько это возможно – сформулировать метаязык, в терминах которого можно было бы выражать междисциплинарные контакты. Историко-философский, историко-научный и семантический анализ, подчеркнул он, способен, с одной стороны, избавить от напрасных надежд и тупиковых поисков, с другой – определить область плодотворных междисциплинарных контактов [Зотов, 2005].

Другой важной темой обсуждения, логически связанной с первой, оказывается «снятие» традиционной оппозиции «естественное – искусственное» и смысловые сдвиги этих понятий в контексте обсуждения философии искусственного интеллекта. В.М. Розин подчеркивает, что даже столь фундаментальная когнитивная активность, как восприятие, отнюдь не «естественная» и не «элементарная», но сложная культурно и семиотически опосредованная деятельность, укорененная в символической репрезентации наличных практик. Однако В.А. Кутырев (Нижний Новгород) отмечает, что попытка наложить печать «искусственного» на любые проявления человеческой жизни является теоретическим оправданием жестких социальных и педагогических технологий.

При обсуждении природы субъективной реальности как атрибута человеческого сознания в центре дискуссии оказалась проблемы корреляции когнитивной осведомленности и человеческого переживания, возможности моделирования эмоций в искусственных интеллектуальных системах. И.Г. Скотникова настаивает на том, что моделирование шести базовых человеческих эмоций в искусственных интеллектуальных системах можно считать уже решенной задачей [Скотникова, 2005]. Однако имеется неопределенность самого понятия «моделирование эмоций», отсюда убеждение в том, что сегодня нет более адекватного критерия сопоставимости человеческих и «машинных» эмоций, кроме эффективности выполнения функций, а он явно недостаточен. При обсуждении вопроса о природе субъективной реальности затрагивается проблема пределов возможной вербализации субъективных переживаний и механизмов их воздействия на знаково-символическую деятельность. При этом Н.М. Смирнова обращает внимание на то, что т. наз. субъективная реальность является не только коммуникативно опосредованной, но и нуждается в постоянной коммуникативной поддержке в повседневных социальных взаимодействиях.

Какова роль виртуальных коммуникаций в коммуникативной поддержке субъективной реальности? Каково воздействие виртуальных коммуникаций на систему социальных коммуникаций в целом? Каков культурно-антропологический статус опыта общения с коммуникативным партнером? Ка-

ково значение телесности в виртуальном общении? Вопросов куда больше, чем ответов. В качестве предварительного подступа к ним следует проявлять здравый критицизм в отношении технологической эйфории, проектов программирования сознания средствами новейших интеллектуальных технологий. Виртуальное общение имеет свою цену. Обогащая человека информационно, оно может привести к эмоциональному обеднению человека, притуплению чувства социальной ответственности. И об этой опасности философ должен постоянно напоминать обществу, дабы избежать «расчеловечивания» человека, его превращения в биологическую приставку к компьютеру.

Так, О.Э. Петруня вызывает к повышению социальной ответственности перед лицом новейших информационных технологий [Петруня, 2005]. Ибо сами по себе они не способны привести к возвышению общественной нравственности, о чем, увы, свидетельствует высокая посещаемость «неприличных» сайтов. Нужны серьезные культурные усилия, ибо работа в открытом информационном пространстве объективно требует обостренного чувства ответственности.

Проблема творчества в искусственном интеллекте

Одним из ключевых вопросов соотношения сознания и продукции ИИ является вопрос о творческой способности, полагает В.И. Самохвалова в работе «Предполагает ли машинное творчество возможность машинного вдохновения?» [Самохвалова, 2005; 2006]. Говорить о машинном творчестве преждевременно. Человек вкладывает в машину результаты своего творчества. Но изолированная технология, будучи продуктом творчества, сама не представляет творческого процесса. Нам до сих пор неизвестны глубинные механизмы, закономерности творческого акта. Поэтому они не могут быть заложены в компьютер, который воспроизводит лишь технологические операции.

Интеллектуальные системы в гуманитарных науках

Здесь можно условно выделить три основные группы: 1) Общие философские проблемы теории искусственного интеллекта, 2) Проблемы создания и использования интеллектуальных систем в гуманитарных исследованиях; 3) Информационные аспекты проблем искусственного интеллекта.

В первой группе относится, например, работа Д.Г. Лахути «О четырех кругах искусственного интеллекта и о дискуссии Поппера с Бернайсом» [Лахути, 2009].

Во второй группе явно выделяются темы: «Проблемы создания интеллектуальных систем для анализа социального поведения» (М.А. Михеенкова и

В.К. Финн) [Михеенкова, Финн, 2005], «Интеллектуальные системы в исторической науке: возможности и ограничения» (С.М. Гусаковой) [Гусакова, 2005]. Здесь рассматриваются научно-методологические вопросы создания интеллектуальных систем, основанных на использовании ДСМ-метода автоматического порождения гипотез и их практического применения при анализе социологической и исторической информации..

Третья группа включает, к примеру, следующее: «Естественные модели и искусственные системы» (Ю.И. Шемакин) [Шемакин, 2005]; «Поисково-оптимизационная картина мира и искусственный интеллект» (С.Н. Гринченко) [Гринченко, 2005]; «Философские проблемы информатики» (К.К. Колин) [Колин, 2005; 2006; 2010]; «Семиотические основания гуманитарной информатики» (И.М. Зацман) [Зацман, 2005]; «Квантовый ключ к теории сознания» (А.И. Лисин) [Лисин, 2005]; «Художественное творчество как информационный процесс» (В.П. Рыжов) [Рыжов, 2005].

В работах отмечается: 1) актуальность и важность организации и проведения дальнейших междисциплинарных исследований философско-методологических проблем ИИ и интеграции для этих целей усилий отечественных и зарубежных ученых, работающих в различных областях научного знания; 2) необходимость разработки *системы наиболее общих понятий по данной проблеме*, которая могла бы быть использована при проведении междисциплинарных исследований; 3) необходимость более тесной связи проводимых исследований с проблемами образования.

Выводы

Первоначальный этап становления отечественной философии ИИ раскрыл важность проблематики междисциплинарных исследований искусственного интеллекта для отечественной науки, образования и культуры в целом. Для того, чтобы дискуссии и методологические проекты носили не спорадический, а непрерывный характер, в принятой на конференции 2005 г. резолюции было решено создать *Научный совет Российской академии наук по методологии искусственного интеллекта (НСМИИ РАН)*. Основная цель Совета состоит в развитии указанных выше методологических направлений. Совету поручается следующая работа: 1) Организация постоянно действующего междисциплинарного семинара по философско-методологическим проблемам ИИ и систематическое проведение симпозиумов, «круглых столов» и иных форм научно-организационной деятельности по теоретическим, методологическим и социальным вопросам развития ИИ; 2) Подготовка и проведение Всероссийских конфе-

ренций по философии ИИ; 3) Создание сетевого электронного издательства по философским и методологическим проблемам ИИ, обеспечение его систематической деятельности, публикация в нем не только отечественных материалов, но и переводов работ зарубежных авторов; 6) Разработка системы рекомендаций по вопросам повышения качества теоретической и методологической подготовки студентов и аспирантов соответствующих высших учебных заведений и проведение в этом направлении непрерывной работы, включая издание специальных учебных пособий и др.

НСММИ РАН (<http://www.scmair.ru>) создан под эгидой Отделения общественных наук РАН. Первоначальный состав: председатель Совета – академик В.Л. Макаров, заместители – член-корр. РАН В.А. Лекторский и проф. Д.И. Дубровский, учёный секретарь – А.Ю. Алексеев). В состав членов Совета вошли видные философы и крупные учёные различных специальностей. Первое заседание Совета и постоянно-действующего семинара «Философско-методологические проблемы искусственного интеллекта» состоялись 17 марта 2005 г. Начиная с этого заседания, эстафета от спорадических и разрозненных отечественных исследований в сфере философии ИИ перешла к сложному междисциплинарному, межрегиональному и межвозрастному социальному институту, которым в настоящее время является НСММИ РАН.

На этапе становления философско-методологических исследований ИИ работа первостепенной значимости заключалась в *организации междисциплинарного взаимодействия специалистов различного профиля и квалификации*. Традиционные вопросы научной онтологии и теории познания применительно к проблематике ИИ, конечно, звучали. К ним например, относились: 1) *онтологические вопросы* – о теории интеллектуальной реальности как некоторой разновидности теории сознания (преимущественно, функционалистской ориентации). От главного ответа онтологической теории – в самом деле искусственные системы являются интеллектуальными либо им субъективно приписывается интеллектуальная способность человека – зависит решение 2) *теоретико-познавательных вопросов* – о методах познания интеллектуальных способностей, насколько полученные знания соответствуют действительности (в частности, о соответствии знаний как феномена субъективной реальности формализованным «знаниям» искусственной системы и пр.). Однако подобного рода вопросы имели второстепенное значение относительно научно-организационных проблем. Диссертант отнюдь не утверждает того, что философская методология ИИ тождественна социальной эпистемологии (которая трактуется в соответствии с [Лекторский, 2010]). Просто на исходном этапе открытия нового научного

направления научно-организационная работа представляется первичной относительно глубокой научно-теоретической проработки. Это объясняется тем, что не ясен предмет исследований: что, собственно, должна включать в себя философия ИИ. Здесь важна креативная генерация идей, не ограниченная до поры до времени жёсткими регулятивами той или иной научной школы.

Концептуальное поле философских исследований ИИ значительно расширилось и структурировалось в ходе работы НСМИИ РАН. Этому посвящены следующие два подпараграфа.

1.1.2. Анализ дискуссий на мероприятиях Научного совета РАН по методологии искусственного интеллекта (НСМИИ РАН)

Основные направления и пути развития отечественной философии ИИ раскрываются посредством анализа докладов и дискуссий в рамках заседаний научно-теоретического семинара НСМИИ РАН «Философско-методологические проблемы ИИ». Источниками являются так же дискуссии в рамках всероссийских и региональных мероприятий НСМИИ РАН (конференций, круглых столов, симпозиумов, рабочих заседаний). В научной организации этих мероприятий, в их инициации, в разработке тематических программ диссертант принимал и принимает непосредственное участие с момента создания НСМИИ РАН по настоящее время.

Комплекс философско-методологических исследований ИИ, сложившийся априори конференции «Философия искусственного интеллекта», в настоящее время расширяется по следующим направлениям: 1) в рамках постоянно действующего семинара «Философско-методологические проблемы искусственного интеллекта»; 2) в деятельности региональных отделений НСМИИ РАН; 3) в проведении конференций, симпозиумов, круглых столов и пр. научных мероприятий, иницируемых НСМИИ РАН либо самостоятельно (примером является вышеупомянутая ежегодная Всероссийская конференция студентов, аспирантов, молодых учёных «Искусственный интеллект: философия, методология, инновации»), либо в составе научных мероприятий, которые организуются отдельно от НСМИИ РАН, например, в рамках конференций Российской ассоциации искусственного интеллекта.

Постоянно действующий семинар «Философско-методологические проблемы искусственного интеллекта» учреждён НСМИИ РАН и проводится с марта 2005 г. не реже одного раза в месяц с целью интенсификации, систематизации, интеграции междисциплинарных теоретических и практических иссле-

дований в сфере общественных, гуманитарных, естественных и технических наук, разрабатывающих проблематику искусственного интеллекта; координации междисциплинарных проектов по актуальным философским, теоретическим и прикладным проблемам искусственного интеллекта; содействия в повышении уровня и качества образовательной подготовки школьников, студентов, аспирантов и молодых ученых, ориентированных на проблемы когнитивных и компьютерных наук; содействия преемственности поколений в развитии наукоемкого технологического комплекса страны за счёт популяризации отечественных научных школ и направлений, организации встреч ведущих учёных с молодыми исследователями и др.

Основной научно-практический результат работы семинара состоит в развитии междисциплинарного подхода к искусственному интеллекту и расширению информационно-коммуникативного поля взаимодействия специалистов и организаций, исследующих философские проблемы когнитивных и компьютерных наук.

Тематика семинара делится на два направления – научно-теоретическое и научно-практическое. Научно-теоретические мероприятия включают обсуждение докладов ведущих специалистов, изучающих философские, психологические, нейрофизиологические, информационно-технологические, лингвистические, программно-технические и другие аспекты проблематики методологии искусственного интеллекта. Научно-практические мероприятия (заседания НСММИ РАН) включают обсуждение проектов мероприятий и отчетов их выполнения.

То, что данный семинар является междисциплинарным и межрегиональным указывают следующие цифры, которые были представлены в регистрационной базе участников семинара (на 8 декабря 2011 г.). Всего представлено свыше 1000 учёных и специалистов из более 80 городов РФ, ближнего и дальнего зарубежья, а именно: Москва (594), Санкт-Петербург (27), Иваново (27), Владивосток (26), Уфа (25), Хайфон (19), Пермь (18), Киров (7), Новосибирск (7), Таганрог (7), Волгоград (6), Вологда (6), Воронеж (6), Томск (5), Белгород (5), Омск (5), Серпухов (5), Минск (4), Иркутск (4), Калининград (4), Челябинск (4), Хошимин (4), Харьков (4), Хабаровск (4), Тула (3), Екатеринбург (3), Тверь (3), Саратов (3), Тамбов (3), Луганск (2), Курск (2), Красноярск (2), Краснодар (2), Магнитогорск (2), Казань (2), Великий Новгород (2), Якутск (2), Мурманск (2), Обнинск (2), Ярославль (2), Мюнхен (2), Геленджик (2), Ставрополь (2), Ростов-на-Дону (2) и др. Всего участвуют представители из 12 стран, среди ко-

торых Вьетнам (23), Украина (17), Беларусь (16), Франция (3), Германия (2) и др. Из них – академиков РАН (6), членов-корреспондентов РАН (7), профессоров (118), доцентов (40), старших научных сотрудников (13). Всего – 158 докторов наук следующих специальностей: философских (66), технических (45), физико-математических (20), психологических (10), медицинских (5), биологических (3), химических (3), филологических (2), исторических (1), педагогических (1), социологических (1), экономических (1). Всего – 121 кандидат наук следующих специальностей: философских (54), технических (41), физико-математических (15), психологических (7), экономических (7), филологических (2), педагогических (1), биологических (1), медицинских (1), социологических (1), исторических (1).

Эти цифры являются убедительным аргументом в пользу того, что философия искусственного интеллекта в нашей стране окончательно состоялась как специфическая сфера междисциплинарных исследований, обладающая собственным предметом, особыми методами, специальными категориями.

Рассмотрим ряд направлений философско-методологических исследований, которые обсуждались в рамках семинара, исходя из анализа докладов, выступлений, сообщений на протяжении 50 заседаний (в сентябре 2015 г. состоялось 71-е заседание).

Проблема “сознание – мозг – искусственный интеллект” – главное направление междисциплинарных исследований искусственного интеллекта

Тематика инспирирована проблемой идеального и информационным подходом к изучению сознания, которые были впервые предложены 50 лет назад Д.И. Дубровским. В частности, "проблема идеального" состоит в расшифровке нейродинамических кодов психической активности и репродукции кодовых зависимостей в компьютерной системе. В связи с этой тематикой перекликаются идеи А.М. Иваницкого, К.В. Анохина; нейрофизиологические и нейрокомпьютерные концепции Д.И. Дубровского, В.Я. Сергина, Д.П. Матюшкина, А.Я. Каплана); исследования языка и мозга в контексте компьютерной метафоры (Т.В. Черниговская, К.В. Анохин, О.П. Кузнецов). Диссертант, однако, оспаривает доминантный статус проблемы «сознание-мозг-ИИ».

Проблематика ярко представлена в ряде работ и докладов сопредседателя НСММИ РАН *Д.И. Дубровского*. Собственно, упомянутый выше первый доклад, прозвучавший на Семинаре 17 марта 2005 г. на тему «Сознание, мозг, ис-

кусственный интеллект» был ориентирован на раскрытие главенствующей роли исследований сознания в философской методологии ИИ.

Разработка проблемы искусственного интеллекта (ИИ) необходимо связана с результатами исследования естественного интеллекта (ЕИ), который не ограничивается когнитивными функциями, представляет сознательную деятельность в целом. Здесь мы имеем дело с тем, что обычно именуют *проблемой сознания*. [Дубровский, 2002; 2003-1; 2003-2]. Сознание обладает специфическим и неотъемлемым качеством субъективной реальности (СР). Решение проблемы должно представлять собой теоретически корректный ответ вопросы: 1) Как связаны явления СР с мозговыми процессами, если первым нельзя приписывать пространственные и другие физические свойства, а вторые ими по необходимости обладают?; 2) Каким образом явления СР, которым нельзя приписывать физические свойства (массу, энергию), способны служить причиной телесных изменений, управлять ими?

Для решения этих вопросов Д.И. Дубровский предлагает широко известный "информационный подход к сознанию" [Дубровский, 1971; 1980; 1994-1]: 1) Информация необходимо воплощена в своем физическом (материальном) носителе, она существует лишь в определенной кодовой форме; 2) Информация инвариантна по отношению к физическим свойствам своего носителя, т.е. одна и та же информация может кодироваться по-разному; 3) В самоорганизующихся системах информация может выступать причиной изменений в ее субстрате и структуре, служить фактором управления; 4) Явления СР допустимо интерпретировать в качестве информации. Проблема сознания становится проблемой ИИ при сопоставлении мозга и компьютера. С этой позиции изучаются: дистанция между ними (у компьютера нет СР); способы кодирования информации у живых существ, человека; сложной самоорганизующейся системы; феноменологические характеристики естественного интеллекта; свобода выбора; перспективы компьютерных наук и информационных технологий и др. Формулируется стратегическая цель в решении проблемы сознания: выяснение существенных и необходимых свойств того типа самоорганизации, который создает представленность для системы информации в форме СР и способность оперировать ею (в «чистом» виде). Эта проблема имеет первостепенное значение для создания новых направлений разработки ИИ.

Проблематика сознания в связи с исследованиями ИИ продолжается в докладе на заседании семинара N 20 (30.05.2007) на тему «Информационный подход к проблеме сознание, мозг и перспективы искусственного интеллекта»

[Дубровский, Лебедев; 2007]. Докладчик считает, что проблема «сознание и мозг» - одна из фундаментальных проблем современной науки. Она имеет комплексный характер и включает существенные философские аспекты. Ее разработка способна стимулировать новые подходы к развитию искусственного интеллекта. В современных разработках ИИ возрастает значение тех результатов, которые получены при исследовании информационных процессов в головном мозге и биологических системах. Как свидетельствуют данные нейроморфологии и нейрофизиологии, в головном мозге переработка информации совершается одновременно, параллельно во многих различных по своим функциям структурах, результаты которой анализируются и выборочно интегрируются в зависимости от актуализованной цели, от хода решения задачи. Переработка информации в головном мозгу, выражающая текущую мыслительную деятельность, совершается отнюдь не по жесткой двоичной логической схеме. Скорее эта логика представляет собой многозначную логику, в которой число значений истинности есть величина переменная; при этом число значений истинности меняется в зависимости от характера решаемой задачи и, возможно, от разных этапов ее решения. В этой многомерной динамической структуре двоичная логическая схема лишь один из существенных моментов процесса переработки информации. Головному мозгу присущи развитые функции вероятностного прогнозирования, весьма оригинальные, эффективные способы сжатия информации и выборки нужных элементов из памяти, эвристического синтеза и другие операции, которые вряд ли допустимо приписывать современным компьютерам. Пока еще «компьютерные модели далеки от биологической основы работы интеллекта». Новый этап развития искусственного интеллекта требует максимально полного понимания работы биологических систем обработки информации и свойств организмов, обеспечиваемых этими системами. Именно таков магистральный путь сближения искусственного интеллекта с естественным.

Тема была значительно расширена на заседании № 15а, состоявшемся 30 ноября 2006 г. в рамках Симпозиума «Сознание и мозг», на который было представлено более двадцати докладов. Среди докладчиков – ведущие отечественные специалисты в области психологии, философии сознания, лингвистики. Обсуждались следующие доклады: Ю.И.Александров "От эмоций к сознанию"; Л.Г.Антипенко "Неклассический подход к решению классической проблемы «дух – материя, душа – тело, сознание – мозг»"; В.В.Васильев "Мозг и сознание: выходы из лабиринта"; Д.И.Дубровский "Зачем субъективная реальность" или "почему информационные процессы не идут в темноте?" (ответ

Д.Чалмерсу); А.М.Иваницкий "Физиология мозга о механизмах построения субъективного образа"; И.П.Меркулов "Сознание с точки зрения переработки информации"; В.Г.Редько "Как промоделировать сознание?"; В.Я.Сергин "Самоотождествление и сенсорно-моторное повторение как ключевые механизмы сознания"; В.Я.Сергин "Сознание как система внутреннего видения"; Т.В.Черниговская "Язык, мозг и компьютерная метафора" и др.

В каждом докладе акцент делался на следующих вопросах: 1. Как объяснить необходимую связь явлений сознания с мозговыми процессами? 2. Как объяснить каузальную функцию явлений сознания по отношению к телесным процессам? 3. Существует ли феномен свободы воли и если да, то как можно его объяснить с позиций деятельности мозга? 4. Каковы наиболее перспективные направления разработки проблемы "Сознание и мозг" (с учетом результатов психофизиологии, психоневрологии, нейролингвистики, нейроморфологии, психиатрии, когнитивных наук и др.)? Способны ли они оказывать существенное влияние на проблематику искусственного интеллекта и достижения в этой области?

С проблемой сознания в формате, по крайней мере, в том формате, который предлагает Д.И. Дубровский, тесно переплетается и включается следующая тема – нейрокомпьютерная тематика проблемы сознания, мозга и ИИ.

Нейрокомпьютерная проблематика сознания

Нейрокомпьютерная проблематика сознания прозвучала в ряде докладов. С докладом «**Нейронный подход к психофизиологическим основам интеллекта**» (48-е заседание семинара ФМП ИИ, 28 сентября 2011 г.) выступил Д.П. Матюшкин, содокладчик-оппонент: Д.И. Дубровский (Институт философии РАН) [Матюшкин, Дубровский, 2011]. Как психофизиолог, докладчик коснулся нейрофизиологического подхода к пониманию естественного интеллекта и внутреннего Я человека. Основой интеллекта человека является взаимодействие сложной системы нейронов мозга. Организация нейронной системы мозга, по мнению докладчика, должна быть иерархической, нейроны прямо связаны и сближены, а сами эти сети должны располагаться в мозге центрально, в равном удалении от его внешних границ и выступать основой целостности организма. Такие сетевые структуры и такие тормозные механизмы выявлены в ретикулярной формации и центральном сером веществе ствола мозга, в стриатумах, но более всего в ретикулярном таламусе. Стволовые и таламо-гипоталамические сети, представляют «низшую самость» (НС). Это низшее Я (сходное с активным Я Фрейда). НС обеспечивает подсознательные инстинктивные действия.

Но она, будучи малонейронной, включенной во врожденные механизмы, не может быть субстратом для накопления новых знаний (связей, ассоциаций). Так же имеется выполняющая эти функции надстройка над НС в виде переднего мозга с большими полушариями, которая представляет «высшую самость» (ВС). Далее докладчик при анализе системы НС+ВС, ссылается на известные концепции эго-системы Д.И.Дубровского (1986) и схему Дамасио (2000), имеет некую связь с данными Пенфильда (1958), полученными в исследованиях на человеке и со собственными данными Д.П.Матюшкина, полученными в опытах на животных (1956). ВС – это система аппаратов переднего мозга и среди них кора больших полушарий. Здесь очень много нейронов, которые могут быть обособленными и при этом потенциально связанными. Их сигналы – это коды-образы для системы КВАЛИА. Среди них особую роль играет нейронная проекция НС – то есть представление кодов потребностей и желаний организма, а так же кодов оценок данного индивидуума окружающими людьми. Это–ядро квалиа, и оно должно оцениваться как высшее Я (Эго) человека, определяющее его самосознание и самоидентификацию. Вокруг этого Я организуются прочие знания, то есть коды – сигнальных систем, в том числе второй сигнальной системы Павлова (1949) – коды слов (и жестов) обозначающих действия и предметы. Эти коды основы высшего человеческого знания - общечеловеческого «сознания» (по Симонову, 1981,1993). Но «квалиа» – вместе с тем и продукт взаимодействия массы нейронов ВС, то есть - это феномен многонейронного коллектива. Однако, активность ВС и тем самым "Я", получающие массу информации извне и изнутри организма, контролируется из НС, дающей разрешения на тот или иной вектор психической активности в соответствии с текущей доминантой НС. Если разрешения НС нет, то нет и осознания входящей информации, сознательного мышления и реализации сознательных программ. НС через восходящие системы разрешает активность того или иного сектора ВС. Таким образом НС и ВС постоянно взаимодействуют, обеспечивая вменяемость человека. И это имеет место и при поиске ответа на внешний стимул и при поиске нужного направления мышления при решении умственных проблем, так или иначе связанных с багажом знаний, с памятью, с жизненным опытом. На примерах докладчик убедительно продемонстрировал, что проблему интеллекта человека можно решать на основе нейронного подхода, не прибегая ни к квантово-механическим, ни к дуалистическим схемам.

Нейрокомпьютерная тематика проблемы сознания так же поднималась **В.Я. Сергиным** (Институт математических проблем биологии РАН) в докладе «**Нейронные механизмы осознания внешних и внутренних событий**» [Сер-

гин, 2011]. Основной вопрос, полагает докладчик, следующий: какой вид координированной нейронной активности мозга может порождать психически переживаемый феномен осознания? Убедительного ответа на этот вопрос пока еще нет, хотя он является ключевым для понимания любых форм сознания. Постулируется специальный вид координированной нейронной активности, который формирует процесс автоотождествления сигналов. Процесс автоотождествления состоит в передаче специфического паттерна возбуждения, порождаемого стимулом в нескольких областях коры головного мозга, опять на нейроны этих областей через автономные обратные связи. Совпадающие (тождественные) паттерны возбуждения - порождаемые стимулом и передаваемые по обратным связям - складываются на одних и тех же нейронных структурах, вызывая их интенсивную разрядку. Такой циклический процесс взрывообразно увеличивает интенсивность отображения стимула, что акцентирует его специфичность и создает наилучшие условия категоризации стимула распределенной долговременной памятью. Результат категоризации (свет, звук, запах) отображается специфическим паттерном нейронной активности, который включается в цикл автоотождествления, формируя доминирующее возбуждение. Тем самым, сенсорная категория представляется на входе в качестве отображения стимула. Внешний мир оказывается представленным субъекту не в физических характеристиках, а в сенсорных категориях: цвет, вкус, запах, прикосновение и т. п., что и составляет процесс сенсорного осознания. Явное представление стимула в сенсорных категориях означает, что процесс восприятия переходит с физиологического (объективного) на психический (субъективный) уровень.

К сожалению, ни один из сторонников нейрокомпьютерного подхода к решению проблемы сознания не раскрыл пути, как возможна компьютерная реализация их концепций. Здесь мы наблюдаем проблему, характерную для ИИ – вроде бы тематика тесно связана с моделированием мозговой активности, т.е. должна априори подпадать под исследования ИИ. На деле, это не так. Нейрофизиологические исследования, которые, несомненно, индуцируют некоторые идеи по поводу реализации/супервенции ментального и физического. Но не более того. Если нет проекта компьютерной реализации – это уже не проект ИИ, но некоторая другая сфера научной деятельности. Если есть некоторый компьютерный проект, но он объясняет ментальное лишь со ссылкой на онтологию психического, предложенного психологами, а не на собственную феноменальную онтологию – это не ИИ. Необходимы более строгие критерии принадлежности исследований тематике искусственного интеллекта, а не частные аспекты. Изучаемый в работе комплексный тест Тьюринга предоставляет строгий

критерий демаркации исследований ИИ от исследований, которые таковыми не являются.

Связь исследований естественного интеллекта и искусственного интеллекта

На заседании № 14, состоявшемся 1 ноября 2006 г. был заслушан доклад *О.П. Кузнецова* (г.Москва) на тему: **«Может ли искусственный интеллект развиваться независимо от исследований естественного интеллекта?»** [Кузнецов, Сергиенко, Холодная, 2006]. В качестве содокладчиков-оппонентов выступили известные психологи *Е.А. Сергиенко* и *М.А. Холодная*. Докладчик четко определил место ИИ среди когнитивных наук и указал на ряд его фундаментальных методологических проблем, связанные с междисциплинарностью знаний об интеллекте и с тем, что ИИ не имеет собственных средств для получения данных и верификации результатов. Верификация моделей ИИ – косвенная: модель ИИ должна соответствовать психологическим данным, быть адекватной на феноменологическом уровне. Отсюда – ответ на вопрос: «Может ли ИИ развиваться независимо от исследований естественного интеллекта?» выглядит так: в первом направлении (интеллектуальные технологии) может (и развивается), но при этом упускает много возможностей; во втором направлении (моделирование мозга) не может в принципе. Однако достичь окончательной цели в моделировании мозга – раскрыть механизмы его работы, т.е. показать, как он работает, - не может ни одна когнитивная наука в отдельности. И это во многом невозможно без использования подходов искусственного интеллекта. Однако компьютерная парадигма не достаточна для ИИ, в частности, голографическая модель памяти выходит за пределы этой парадигмы. То есть многие виды интеллектуальной деятельности в рамках компьютерной парадигмы либо не моделируются вовсе, либо моделируются с низкой эффективностью, которая неустранима при алгоритмическом подходе. Далее докладчик предложил ряд информационных принципов, практически отсутствующих в современной компьютерной парадигме: 1) несимвольные, аналоговые представления информации; 2) наличие механизмов, работающих со сходством вместо тождества и связанных с эффектами размывания и обострения образов; 3) сочетание малой глубины информационных процессов и высокой параллельности, что требует пересмотра традиционных теорий памяти; 4) распределенность информации, которая обеспечит надежность хранения информации. Эти и ряд других положений, требующие углублённого изучения естественного интеллекта, позволяют отрицательно судить о поставленном в наименовании доклада

вопросе – подытожил докладчик. Оппоненты О.П.Кузнецова – ведущие психологи страны, несомненно, поддержали докладчика – необходимы глубокие психологические проработки для того, чтобы ИИ достиг успехов.

Диссертант так же полагает, что в рамках традиционного компьютеринга невозможно развитие ИИ (проблематике нейрокомпьютинга и коннекционистской парадигме в связи с тестом Тьюринга посвящена Глава III). Однако не следует компьютерные онтологии воспринимать в форме моделей психического. Успехи ИИ как раз достижимы на бихевиоральном уровне изучения психических явлений без сопутствующих глубоких теоретических проработок по поводу «погружения» в сущность «интеллекта». Далее, когда внешне оцениваемое поведение рассматривается в контексте функционирования целостной системы, с учётом её состояний, а не в механистическом формате «входов/выходов», тогда на смену бихевиоризму приходит психологический функционализм. Данная форма ЕИ-исследований, в свою очередь, тесно коррелирует, более того, она возникла из концепции теста Тьюринга (Х. Патнэм, 1960, машинный функционализм – см. п.п. 2.1.1.). Отсюда вывод – изучение ЕИ следует осуществлять в контексте проработки и развития оригинального теста Тьюринга. А тест Тьюринга – это суть искусственного интеллекта. Так что и ЕИ-исследования и ИИ-исследования следует развивать совместно и параллельно. Для такого рода «параллельного» проведения исследований диссертант предлагает 3D-семантику представления проблемной области «психического» - ментальные термины рассматриваются совместно с нейрофизиологическими (шире – биологическими) терминами и инженерно-технологическими (п.п. 3.1.3) терминами.

Крайне важным в этой связи представляется вопрос проф. **Т.Б. Кудряшовой** (г. Иваново): **«Могут ли исследования естественного интеллекта развиваться независимо от исследований искусственного интеллекта?»**. Данный вопрос она поставила в 2011 г. в рамках конференции «Актуальные проблемы когнитивной науки», которая ежегодно в октябре проводится в г. Иваново. Апеллируя к проблематике когнитивной науки, изучающей механизмы психического и, в наиболее продвинутом варианте этой науки – к ИИ, который изучает и конструирует компьютерные механизмы психического, Т.Б. Кудряшова с той же аргументативной силой, как это прозвучало у О.П. Кузнецова, отрицательно ответила на поставленный ею вопрос. В самом деле, искусственный интеллект – это наиболее продвинутая модель естественного интеллекта, задающая условия, границы, возможности эвристического поиска сути и принципов разумной деятельности. Степень убедительности дово-

дов О.П. Кузнецова ничуть не больше степени убедительности Т.Б. Кудряшовой. Два десятилетия назад взаимодействие исследований ЕИ и ИИ философы называли «диалектическим». Сегодня, вместо «диалектики», удобнее использовать другие слова, например, «неоскептицизм». Поэтому можно подытожить дискуссию так: *«ИИ должен развиваться в рамках неоскептической методологии»*. Неоскептицизм предполагает в исследовании поиск предмета, противоположного изучаемому. Чем больше степень противопологаемости, тем ярче проявляется изучаемый предмет (П.Курц, В.А.Кувакин) [Курц, 1993; Кувакин 1998].

Когнитивная психология

На заседании № 10, состоявшемся 29 марта 2006 г. прозвучал доклад **Е.А. Сергиенко** на тему: «Истоки познания: революция в когнитивной психологии развития» [Сергиенко, Холодная, 2006]. Докладчик считает, что область когнитивного развития, в которой произошли принципиальные теоретические изменения, позволяет расширить представления об истоках познания, базовых механизмах образования категорий, иначе посмотреть на соотношение восприятия и действия, фундаментальной проблемы философии и психологии, иначе взглянуть на процесс развития субъектности человека. Онтогенетические исследования позволяют показать, что развитие самых высоко и сложно организованных уровней зависит от базовых, первичных ступеней и опирается на них. Далее докладчик выдвигает ряд тезисов: младенец – не сенсомоторный, а репрезентативный индивид, у которого категоризация событий и категоризация объектов происходит неравномерно. Развитие человека как субъекта происходит непрерывно в онтогенезе, что дает возможность рассматривать даже самого маленького ребенка как активного субъекта собственной жизни. Понимание мира и физического и ментального определяется уровнем развития субъектности. Проведённые исследования в области психологии развития позволяют внести ряд конструктивных положений в методологию искусственного интеллекта: Процесс становления спонтанных знаний опирается на универсальные организующие принципы, которые позволяют структурировать среду. Задавая базовые организующие принципы и правила оперирования с ними, процесс самоорганизации приводит к усложнению категорий и отношений между ними. Принципиально единство в организации восприятия и действия необходимо дополнить в виде восприятие-действие-репрезентации, что позволяет снять разобщение между ментальностью и активностью. Понимание мира и физического и ментального определяется уровнем развития субъектности. Критерии становления

субъектности дают возможность приблизить формализацию оснований субъективных моделей. Уровневое ограничение развития субъектности необходимо учитывать как в построении виртуальных моделей, так и реальном взаимодействии.

Проблема «язык-мозг»

Философские основания данной проблематики в контексте исследований лингвистической философии были озвучены на 12-м заседании Семинара, состоявшемся 31 мая 2006 г., где был заслушан доклад ведущего эксперта-аналитика ЗАО "Стратегия-Центр", доктора философских наук *М.В. Лебедева* на тему: «**О принципах моделирования языковых значений**» [Лебедев Черняк 2006]. Докладчик отметил продуктивность отдельных положений философии языка для построения систем искусственного интеллекта, в первую очередь, в теории и практики моделирования языковых значений. При этом на условия истинности и условия обоснованности значений накладывается ряд ограничений (со стороны холистического подхода), например, определение семантических отношений на лексическом уровне: синонимии, семантического сходства, антонимии, родо-видовых отношений; определение отношений между сверхфразовыми единствами (отношение "вопрос — ответ" при пользовательском запросе или соотношение между "основными идеями" текстов) и др. Поэтому во всех аспектах моделирования значений (лексических, сентенциальных, сверхфразовых) следует иметь в виду два уровня представления. Первый имеет жестко заданную структуру с априорно установленными темами и связями между ними. (например, это подразумевает наличие предопределенного рубрикатора с иерархией категорий, которая отражала бы общепринятую систему семантических категоризаций). Второй подвержен изменениям (в этом случае гипертекстовая структура представляет собой образованную в ходе анализа некоторого массива семантическую сеть, связи которой отражают актуальное семантическое строение текстов). Данные исследования моделей языковых значений следует положить в качестве теоретических оснований для проектирования интеллектуальных систем.

Несомненно, тематика лингвистического анализа крайне важна для методологии ИИ. Однако докладчик при обсуждении этой темы высказал следующее: «Дискета в дисководе — это аналог перманентного в имманентном». Такое высказывание лишает методологию ИИ статуса дисциплины, развивающейся в контексте философского анализа.

Другой подход, натуралистически ориентированный, представлен **Т.В. Черниговской** (г. Санкт-Петербург) в работах на общую тему: **«Язык, мозг и компьютерная метафора»** (см. [Черниговская, Анохин, Кузнецов, 2006; Черниговская, 2006]). Возникает фундаментальный, как она полагает, вопрос: «Мозг – зеркальный или прозрачный?» И считает, что данный вопрос особенно актуален для лингвистики и той части нейронауки, где исследуются когнитивная и, в частности, языковая компетенция человека. Причём одной из центральных методологических проблем современного этапа состояния этих наук – определение границ собственно антропологического списка таких компетенций, формируемых путём изучения языка. Список свойств человеческого языка, отличающий его от иных коммуникационных систем, следующий: ПРОДУКТИВНОСТЬ, ДВОЙСТВЕННОСТЬ, ПРОИЗВОЛЬНОСТЬ, ПЕРЕМЕЩАЕМОСТЬ, КУЛЬТУРНАЯ ПРЕЕМСТВЕННОСТЬ, СЛОЖНОСТЬ, ГИБКОСТЬ и пр. Однако не только язык с его сверхсложной организацией отличает нас от других биологических видов, причём является ли язык – центральным пунктом при этом – вот основной вопрос? Докладчица акцентирует внимание на интересных последних открытиях в антропологии, биологии, психологии, нейрофизиологии и др., в частности, на открытии т.н. зеркальных нейронов и вообще т.н. зеркальных систем. Эти системы дали совершенно новые подтверждения принципиальной важности по имитации и даже самого факта фиксации действий другого в нервной системе для когнитивного развития в фила- и онтогенезе и даже для возникновения языка и рефлексии как основ сознания человека. Зеркальные нейроны реагируют только на определённое действие, когда субъект делает что-то сам, когда видит это действие или слышит о нём. Это показывает, на основе чего развился мозг, готовый для функционирования языка и построения моделей сознания других людей, готовый для социального обучения и адекватного поведения в социуме. В связи с этим докладчик выделяет ряд свойств психических процессов, которые делают компьютерную метафору совершенно нерелевантной: 1) чрезвычайная роль контекста, а значит - возможность множественных трактовок сообщения и событий вообще; 2) избыточность и возможность многих путей для поиска одного и того же; 3) размытость, неточность, приблизительность описаний, не снижающая эффективность поиска в памяти и построения алгоритма поведения (то, что принято связывать с правополушарным типом сознания). Нельзя не согласиться: то, что просто человеку, сложно компьютеру и наоборот. Значит, мы что-то не то делаем, когда думаем, что моделируем человеческое мышление; 4) неадекватность аристотелевского типа мышления и даже искусственность его для

мозга, т.к. такому типу логики человека надо специально обучать; 5) множественность типов мышления, определяемых культурой и решаемой задачей. Не следует принимать за «норму», характеризующую наш биологический вид психические процессы здорового белого мужчины со средним образованием и интеллектом (подобно тому как не могут быть выведены типологические ментальные правила Языка на основе изучения английского языка). «То, что делает нас людьми — никакие абиссинцы с шумерами на своих счётах не отложат» — образно заключает Т.В.Черниговская.

На наш взгляд, основной вопрос состоит в несоответствии существующей парадигмы моделирования когнитивных феноменов реальным исследованиям мозга и языка.

Нетрадиционный компьютеринг

Тема модернизации принципов компьютеринга тесно связана с проблемами построения вычислительной техники, адекватной данным принципам. Одно из интересных направлений было озвучено в докладе Г.Г. Рябова «**Топологический процессор**» [Рябов, Суворов, 2006]. *Топологический процессор* (Т-процессор) является, подобно видеокarte, дополнительным устройством персонального компьютера или рабочей станции и связан с основным процессором шиной передачи данных в обе стороны. Он имеет собственную оперативную память большого объема. Т-процессор ориентирован на проведение метрических и топологических преобразований трехмерных растров (для отличия от традиционных «пиксельных» форм представления графической информации вводится новый термин — «вексельная структура»). Такие структуры являются исходными при решении задач в медицинской диагностике (магнитно резонансная томография), биологических исследованиях, распознавании образов трехмерной структуры, синтезе конструкций в строительстве и т.д. Т-процессор может быть одним из базовых инструментов *компьютерного зрения*. Докладчик предложил ряд математических моделей 3D растра и волновых алгоритмов метрических и топологических преобразований. На решетчатых моделях 3D растра предложен набор арифметических и логических операций как на *трехмерных операндах*, которые реализуются как на Т-процессоре, так и на основном процессоре компьютера или рабочей станции. Доклад представился крайне важным для развития теории и практики имитации дискретными процессами непрерывных волновых процессов. Как мы увидим в последующем, в п.п. 3.3.3 при построении инструментария, реализующего комплексный тест Тьюринга, идеи Г.Г. Рябова представляются крайне важными в плане приближения к идее

компьютинга, более адекватного проблематике интеллектуальных систем, нежели чем традиционный компьютеринг фон-неймановской парадигмы.

Однако в продолжении данного доклада, содокладчик В.В. Суворов сделал, на наш взгляд, некорректные методологические обобщения работ в модификации компьютерного базиса. С его точки зрения, проведенные исследования обеспечивают: 1) Формулирование и выдвижение проблемы выработки комплексного междисциплинарного подхода к исследованиям интеллекта в качестве фундаментальной проблемы современного этапа научно-технического прогресса; 2) Предложение, обоснование и развитие постнеклассических трактовок интеллекта и креативности. Придание интеллекту и креативности статуса объектов междисциплинарных исследований, не ограниченных психологическим содержанием; 3) Формулирование определений, критериев идентификации и атрибутов интеллекта и креативности, распространяющихся на явления в непсихических средах, включая нейробиологическую, искусственную (научно-техническую) и природную среды; 4) Представление креативности в качестве самостоятельного предмета объективных научных исследований; 5) Определение психического интеллекта в контексте единства орудий, креативности и потребности (цели) и т.д. У В.В. Суворова даже вышла книга на эти темы [Суворов, 2006].

Диссертант не способен в рамках этого единого доклада совместить положения инженерной методологии Г.П. Рябова и квазиметодологические обобщения и призывы В.В. Суворова. Как воксельная организация данных позволит решить проблему постнеклассической методологии построения интеллектуальных систем? Или, скажем, проблему креативности? По всей видимости, здесь налицо методологическая *ошибка скачкообразной генерализации результатов инженерной методологии до философско-методологического уровня.*

Тема компьютерных основ интеллектуальных технологий получило развитие в докладе «**Логика и техника: от релейно-переключающих схем до нанотехнологий**» (23 сентября 2011 г.), докладчик - В.Г. Горохов [Горохов, 2011]. Акцентируется внимание на том, каким образом интеллектуальная деятельность может быть представлена логически или логико-математически и обратились к первым этапам формализации вычислений, из которых выросла вся современная вычислительная техника. Первой моделью для реализации логических операций в современной технике стали релейно-переключающие схемы. И хотя использование релейно-контактных элементов для построения логических схем вычислительных машин не оправдало себя ввиду низкой надежности,

больших габаритов, большого энергопотребления и низкого быстродействия, релейно-контактные схемы начинают играть роль особых абстрактных объектов по отношению к их физически-конструктивной реализации в виде транзисторных схем. Они становятся посредниками между конструктивными схемами технических систем и их логико-математическим описанием. В этом смысле теория релейно-переключающих схем стала идеализированной моделью, хотя и основывающейся на анализе функционирования некоторых реальных объектов, но абстрагированной от их конкретного физического содержания. В структуре любой научной теории, - отмечает докладчик, наряду с концептуальным и математическим аппаратом важную роль играют теоретические схемы, образующие своеобразный внутренний скелет теории. Эти схемы представляют собой совокупность абстрактных объектов, ориентированных, с одной стороны, на применение соответствующего математического аппарата, а с другой - на мысленный эксперимент, т.е. на проектирование возможных экспериментальных ситуаций. Это - особые идеализированные представления (теоретические модели), которые часто выражаются графически (геометрически) или логически. В технике такого рода графические изображения играют еще более существенную роль, чем в естественной науке, поскольку одна из особенностей инженерного мышления заключается в оперировании схемами и модельными представлениями. Такие схемы являются результатом абстрагирования от качественной определенности протекающего через систему и преобразуемого ею естественного процесса (который лишь в частном случае будет физическим процессом). Увеличение вычислительных мощностей компьютеров сделало возможным определение не только макроструктур, но и с высокой точностью геометрической и электронной структуры больших молекул. В наносистемотехнике приводятся примеры реализации стандартных логических элементов «не», «или», «и» с помощью молекулярных переключателей и составления из них простейших логических схем. Еще одним интересным применением наносистемотехники, по мнению докладчика, является спинтроника, использующая в качестве ключевого элемента «спиновый клапан». Таким образом, независимо от физической, химической или даже биологической реализации в наносистемотехнике используется все та же логика, т.е. структура интеллектуальной деятельности, подлежащей автоматизации, что и в классической вычислительной технике.

Крайне интересен для целей нашей работы онтологический вывод, который делает В.Г.Горохов. В сущности, для современной науки и техники совершенно неважно, как, собственно говоря, выглядит действительная реальность. Важно лишь то, что ученый и инженер с ее помощью может правильно заранее

спланировать и реализовать свою деятельность и получить желаемые результаты. Этой цели и служат разнообразные графические представления и модели, которые развиваются сегодня и в наносистемотехнике. Хотя такому графическому представлению еще далеко до совершенства, поскольку оно является лишь первой попыткой построить абстрактную алгоритмическую модель нанопроцесса, называемого самосборкой наноструктур. Таким образом, как в научных теориях, так и в прикладных областях возрастает роль построения различного рода логических моделей, призванных, с одной стороны, объединять разрозненные дисциплинарные научные знания, а с другой – служить своего рода «проектом» будущих разработок, ориентированных на практическое применение.

Таким образом, мы видим, что логико-математическая абстракция (радиорелейная схема) объявляется базисом современной компьютерной технологии (ведь неважно, «как выглядит реальность»). Возникает вопрос – какая именно логико-математическая абстракция лежит в основе компьютеринга информационных технологий: элементарные цифровые электронные схемы, логические вентили, радиорелейные схемы или ещё что-то? Здесь очевиден синонимичный ряд терминов. Если интерпретировать логико-математическую конструкцию в буквальном смысле, т.е. если полагать, что способы электронных преобразований подпадают по априорные математические конструкции, присущие человеческому мышлению, тогда докладчик прав. Но тогда в наносистемотехнике мы не получим никаких качественных изменений компьютеринга, просто-напросто увеличится объем памяти, скорость вычислений. Логика электронных преобразований останется той же самой, как для нашего мира, так и для нано-мира.

Диссертант при изучении применимости радиорелейных схем в ИИ, склоняется в большей мере к «сантехнической логике», т.е. логике клапанов и вентилей канализации. В сантехническом формате удобно так же представлять и нанотехнологическую микрофизику. Сантехническую логику ментального успешно предложил в 1980 г. Дж. Серль при обсуждении «Китайской комнаты» (тест Серля – в п.п. 2.2.2). Сантехническая логика – это схема, более абстрактная в проблематике компьютерной реальности, нежели чем «радио-релейная логика».

Однако имеется и другой подход, не логический, а риторический. Он не вгоняет исследователей в жесткие рамки логических схем. Данный подход состоит в ином способе интерпретации фундаментальных конструкций. Фунда-

мент современного компьютеринга не является логико-математическим, он представляет риторическую фигуру – компьютерную метафору некоторой человеческой деятельности, в частности, метафору деятельности математика. Алгоритм работы математика, метафорически закреплённый в машине Тьюринга – суть фундаментальная метафора современного компьютеринга. Но она подлежит пересмотру, ведь можно использовать и другие компьютерные метафоры, например, метафору машины Корсакова, которая, в принципе, реализует как бы «алгоритм» работы художника. Об этой метафоре, важной для развития компьютерной науки и искусственного интеллекта будет доложено ниже в параграфе 3.3.

Фундаментальный ориентир на модернизацию принципов компьютеринга связан с модификацией оснований вычислительной математики (ВМ). Данная тема прозвучала в докладе А.С. Нариньяни **«Математика 21 – радикальная смена парадигмы: модель, а не алгоритм»**. Подразумевалась вычислительная модель (ВМ). Характеризуя текущее положение дел, докладчик противопоставил две альтернативные точки зрения: 1) «Традиционная»: ВМ – успешная, быстро развивающаяся область, предельно востребованная практикой и в основном отвечающая ее потребностям; 2) «Неоптимистическая»: ВМ находится в углубляющемся кризисе, оказываясь все более неадекватной в контексте растущих требований практики. В данный момент, полагает докладчик, у ВМ нет концептуальных идей выхода из этого тупика. Эти две точки зрения столь противоположны, что это не позволяет искать истину где-то посередине между ними. Область ВМ так велика, сложна и разнообразна, что выбор любой конкретной системы оценок неизбежно приведет к тому, что за ближайшими к ней деталями разглядеть суть проблемы будет практически невозможно. Кризис современной ВМ быстро углубляется, что вызвано ее неспособностью: решать основной объем текущих реальных задач, и иметь резерв развития, соответствующий темпу расширения пространства приложений.

Возможна ли революция в ВМ? – задается вопросом А.С. Нариньяни. И отвечает на него следующим образом: надо модифицировать представление об алгоритме и перейти к парадигме моделирования.

В отличие от алгоритма, как полагает А.С. Нариньяни, модель суммарно фиксирует (хотя и не так явно) связи всех параметров со всеми. Она является как бы портретом объекта моделирования, представляющим все доступные для формализации знания автора модели о предмете. Конечно, при этом возможно, что: 1) Знания неполны, плохо формализуемы, противоречивы и т.п., 2) Модель по разным причинам отражает не все знания автора, 3) «Портрет» в той или

иной степени далек от сходства, вплоть до полного искажения оригинала. Но аналогичные недостатки присущи и алгоритму. Наконец, одно из ключевых преимуществ модели – это то, что она работает как совокупность независимых составляющих. Это позволяет строить реальную модель коллективными усилиями многих экспертов, причем даже без прямой кооперации авторов модели, т.е. без знания каждого о вкладе остальных. При этом их знания далеко не всегда должны являться истиной в последней инстанции и могут быть противоречивы, это не мешает всей модели функционально быть производением этих знаний (конъюнкцией ограничений), а не их простой суммой. Что и способно обеспечивать результат, при котором адекватность модели, т.е. ее возможности в эффективности решения прикладных задач, может многократно превосходить компетенцию каждого из ее создателей.

Таким образом, для преодоления кризиса в ВМ необходима новая парадигма, в которой: 1) Спецификация задачи должна определяться ее моделью, а процесс решения от модели к результату стать достаточно универсальным, что позволит перейти от «островной» карты ВМ к «материковой», охватывающей большие пространства прикладных задач; 2) Аппарат определения задачи, т.е. аппарат моделирования, должен стать более адекватным для реальных проблем с учетом их естественных свойств; 3) Метод и специализированный алгоритм должны быть исключены из технологической палитры разработчика и пользователя ВМ.

На наш взгляд, отстаиваемые в данном докладе идеи проблематичны в силу неточности употребления терминов. Например, аппликативное программирование – это теоретико-категорное программирование, что достаточно давно практикуется профессором МИФИ В.Е. Вольфенгагеном. Но это – вычислительная, компьютерная модель, т.е. ВМ! Теория категорий, как известно, выражает многообразие не только алгебр, но и моделей (если в алгебраическую сигнатуру ввести символы компьютерных операций, то мы получим вычислительную математику). Т.е. ни о каких революциях в ВМ как предпосылках развития ИИ утверждать не стоит. Следует в полной мере использовать накопленный уже материал и ввести некоторое подобие единой системы классификации и кодирования методов ВМ, которая обеспечила бы унификацию и стандартизацию методов и средств вычислительной математики при программировании систем ИИ. Однако положения доклада крайне важны для формирования стратегии создания коллективной экспертной системы, в которой учтены «знания» экспертов и «знания» о «незнании экспертов». Однако кто или что отвечает за «знания» второго рода – машина-оракул или имеются некоторые эволюцион-

ные механизмы роста «знания», включая «знания о незнании»? Эти положения раскрываются в п. 3.3.

Искусственная жизнь

Тематику предложил **В.Г. Редько** в докладе «**Модели адаптивного поведения – задел исследований естественного интеллекта**» в виде эскизного плана исследований, направленных на решение проблемы происхождения интеллекта. План состоит из пунктов: А) Разработка моделей адаптивного поведения аниматов с несколькими естественными потребностями: питания, размножения, безопасности. Б) Исследование перехода от физического уровня обработки информации в нервной системе животных к уровню обобщенных образов. Такой переход можно рассматривать, как появление в "сознании" животного свойства "понятие". Обобщенные образы можно представить как мысленные аналоги наших слов, не произносимых животными, но реально используемых ими. В) Исследование процессов формирования причинной связи в памяти животных. Г) Исследование процессов формирования логических выводов в «сознании» животных. Д) Исследование коммуникаций, возникновения языка. Наше мышление тесно связано с языком, с языковым общением между людьми. Поэтому целесообразно проанализировать, - полагает В.Г. Редько, - как в процессе биологической эволюции возникал язык общения животных, как развитие коммуникаций привело к современному языку человека, как развитие коммуникаций и языка способствовало развитию логики, мышления, интеллекта человека.

Проект искусственной жизни – одна из наиболее обсуждаемых тем в мировой науке. Тематика формируется посредством комплексирования достижений из области биотехнологии и искусственного интеллекта.

Искусственная личность

На семинаре 3 июня 2009 г. обсуждался доклад диссертанта на тему «**Проблемы проекта искусственной личности**». В роли оппонента-содокладчика выступил **А.Р. Бахтизин** с темой содоклада «Искусственное общество в агент-ориентированных моделях». В докладе предлагались подходы к определению «искусственной личности» (ИЛ) как когнитивно-компьютерной системы, которая: 1) обладает «квазисознанием», т.е. функциональным подобием человеческой субъективной реальности; 2) функционально, поведенчески и/или физически неотличима от человеческой личности. Рассматривалось многообразие проектов искусственной личности, типовая архитектура искусствен-

ной личности в формате а) робототехнической и б) экспертной системы. Обосновывался интуитивно ясный тезис о существенном расширении поля междисциплинарных исследований ИИ в проекте искусственной личности. В ряду этих смежных проблем — проблема построения теста Тьюринга на персональное. Продолжая классификацию проблем сознания, начатую Д. Чалмерсом, проблемы проекта ИЛ предложено делить на 1) лёгкие проблемы, те, которые затрагивают только операциональные аспекты сознания; 2) трудные проблемы, которые связывают сознание с наличием «субъективного опыта» и самосознания; 3) самые сложные проблемы, состоящие в том, что у нас нет оснований считать искусственную личность невозможной, так как факт прохождения теста Тьюринга на субъективную реальность может служить основанием для приписывания ей сознания; 4) менее сложную проблему, которая заключается в расшифровке нейродинамических кодов ментальных феноменов. Аргументировался вывод о том, что проект искусственной личности является одним из наиболее перспективных проектов искусственного интеллекта. Он располагается «посередине» между проектом искусственной жизни, направленным на воспроизводство биологически-эквивалентных форм квазиорганической жизни, и проектом искусственного общества, призванным реализовать весь спектр социальной жизни среди искусственных агентов. Для первого проекта он задаёт ориентир создания наиболее высокоорганизованной формы искусственной жизни. Для второго создаёт базис, так как социальное вырастает из взаимодействия персоналогически активных искусственных агентов. «Может ли компьютер стать личностью»? Да, может, но на пути реализации проекта искусственной личности очень много трудностей и наисложнейших проблем. Данные проблемы решаемы, если быть убеждённым в позиции сильного ИИ: «все мы – компьютеры, но можем стать личностями и становимся ими». Подробно методологические проблемы проекта искусственной личности раскрыты в параграфе 3.1.2. Проект подвергается критике.

Искусственные общества

Данная тема впервые прозвучала 20 апреля 2005 г. в докладе *В.Л. Макарова, А.Р. Бахтизина и Н.В. Бахтизиной* «Моделирование социально-экономических моделей с помощью технологий искусственного интеллекта». Докладчики вполне справедливо полагают, что общественные науки, в том числе экономические, стоят на пороге масштабного и эффективного использования нового средства для получения знаний. Искусственное общество состоит из агентов, взаимодействующих между собой, отсюда компьютерные модели

называются агент – ориентированными моделями (agent – based models) – АОМ. Эти модели обладают следующими характеристическими свойствами: 1) неоднородность (агенты чем то отличаются друг от друга, что принципиально отличает эти модели от широко распространенных моделей экономики с агентом – представителем); 2) автономия (агенты действуют независимо друг от друга); 3) в явном виде задано или описано пространство, в котором действуют агенты (например, клетки на плоскости, сахарный ландшафт); 4) локальные взаимодействия; 5) ограниченная рациональность; 5) динамика неравновесная; 6) неявно используется математический аппарат и индуктивно-дедуктивный метод.

Искусственное общество – это агент – ориентированная модель, в которой изучаются свойства по крайней мере двух агентов, взаимодействующих между собой, похожими друг на друга и при взаимодействиях способных воспринимать друг друга подобными себе.

Сегодня проблематика искусственных обществ институализирована. В 2006 г. в Киото (Япония) состоялся Первый конгресс по социальному моделированию, организованный Pacific Asian Association for Agent –Based Approach in Social Systems Sciences (PEA) в кооперации с The North American Association for Computational Social Systems Science (NACHOS); The European Social Simulation Association (ESSAY); Kyoto University. По данной теме успешно защищаются диссертации.

В связи с тематикой "искусственных обществ" показательной представляется диссертация В.А. Истратова «Агентно-ориентированная модель поведения человека в социально-экономической среде» [Истратов, 2009-2]. Подробно остановимся на этой работе, так как она позволит подчеркнуть сложность атрибуции компьютерной системы социальных феноменов (см. п.п. 3.1.1) и важность построения 3D-языка, характеризующего данные феномены в контексте исследований феноменологии, биологии, информатики (см. 3.1.3). Модели социокультурных процессов, помимо наблюдаемого внешнего поведения человека и социума (которое фиксируется, в частности, в экономических показателях), включают различные формы репрезентации ментальной сферы субъективной реальности человека – живого существа, обладающего чувствами, разумом, волей и свободой, подчиняющегося природным законам и конституирующим социальные закономерности. Это положение справедливо подчёркивается автором [там же, с.3]. В работе реализована достаточно редкая попытка социального моделирования, выполненная в формате парадигмы постнеклассической междисциплинарности. Отчётливо прописывается способ вовлечения субъекта мо-

делирования (носителя ценностей, смыслов, знаний, психологических переживаний и волевых мотиваций) в процесс моделирования. В рамках такого подхода моделируемые сущности наделяются ментальными параметрами – теми, которыми владеет субъект моделирования. Данная интенция представляется неотъемлемой чертой методологии искусственных обществ. Поэтому крайне важным представляется анализ того, каким образом разработчик приписывает компьютерной системе ментальные свойства и, шире, социокультурные параметры. Значимость данного вопроса обусловлена во многом очевидным ощущением терминологической неловкости при изучении проблематики «искусственных обществ». Данное направление современных и перспективных исследований компьютерного моделирования активно спекулирует на достаточно широком материале психологических, биологических, социологических, политологических, демографических и др. исследований. Очевидно, что компьютерные системы не обладают психологическими намерениями, политическими целями, религиозными убеждениями и пр. Компьютерным системам не присуще состояние «счастья», «потребности», «склонности» и пр. Тем не менее, методология «искусственных обществ» повсеместно использует данные выражения, что вызывает сомнения в адекватности данного способа компьютерного моделирования.

В.А. Истратов уделяет достаточное внимание этому вопросу. Приписывание ментальных свойств в системах агентно-ориентированного моделирования (АОМ), как это контекстуально отмечается в автореферате, происходит двумя способами: 1) внутренние параметры агентов уподобляются параметрам человеческого сознания «воспринимать себе подобных» (В.Л. Макаров), «принимать решения» (Э. Бонабо), «любить» (А. Сломан); 2) параметры агента в максимальной степени очищаются от менталистской терминологии (т.е. компьютерные модели не «воспринимают», не «принимают решения», не «любят» и пр.). Эти свойства приписываются на этапе интерпретации результатов моделирования агентов, к которым применимы в основном логико-математические и кибернетические понятия, например, «множество агентов», взаимодействующих с минимальным управлением (Р. Аксельрод).

Позиция В.А. Истратова состоит в следующем. Вначале автор максимально «очищается» от менталистской терминологии, характеризуя АОМ в параметрах многочисленности, автономности, однотипности, логической связанности [там же, с.7]. Последовательно проводя эту линию, автор вводит крайне интересное понятие «надобности», которое, в отличие от понятия «потребности» имеет более чётко выраженный объективно-бихевиористский смысл [там

же, с.10] и поэтому в отличие от последнего понятия менталистски «менее нагружено». На этом этапе определения мы как бы погружаемся в традиционные модели теории однородных структур или нейронные сети.

Однако, избавившись от ментальной терминологии на начальной стадии, в последующем он наделяет агентов достаточно широким спектром психологических и социологических параметров – «настроением» [там же, с. 15] (в частности, «эйфорией» и «депрессией» [там же, с. 23]), «опытом» [там же, с. 16], «удовлетворённостью» [там же, с. 17], «авторитетом» [там же, с.18]. Осуществляется достаточно подробная персонологическая и физиологическая характеристика агентов [там же, с.16, 17]. И далее автор подводит нас к этической категории «счастье» [там же, с. 23].

Таким образом, возникает крайне интересный методологический ход в отношении к презентации и интерпретации моделей «искусственного общества»: на начальной стадии моделирования исследователь должен максимально очистить концептуальный базис от менталистских понятий. В дальнейшем, последовательно, шаг за шагом, вводит социокультурные параметры и отслеживает, на каких стадиях моделирования, при каких допущениях и ограничениях и в каком значении вводятся в арсенал моделирования или модифицируются социокультурные и, в частности, менталистские понятия. Несомненно, такая находка автора является интересным новаторским приобретением методологии «искусственных обществ» и доказывает методологическую компетентность предложенного подхода.

Однако, как полагает В.Истратов, в работе для подчёркивания этой идеи используются не совсем репрезентативные социокультурные концепции, например: этико-экономическая теория К. Грэма («экономика счастья», [там же, с.23], положения которой представлены в небольшой статье, что не предполагает фундаментального освещения затрагиваемой проблемы. Крайне невзрачными представляются исследования Д. Канемана в изучении человеческого опыта и человеческой памяти [там же, с.16] – этой теме посвящены обширные рубрики многочисленных конференций по когнитивным наукам, которые проводятся у нас в стране и за рубежом, обстоятельные статьи и монографии на эту тему насчитывают десятки тысяч библиографических позиций (в частности, широко известная у нас в стране классификация Д. Чалмерса – см. п.п. 3.1.3).

Подобного рода претензии можно предъявить и к иным источникам. Однако данное замечание снимается, если работу воспринимать не в формате законченной компьютерной модели «экономического счастья», а как *исходный*

прототип компьютерной системы, на базе которого в последующем предполагается организация учёта, систематизации, координации, интеграции социологических, психологических политологических, культурологических, этических, эстетических и широкого спектра других теорий социальных, гуманитарных и естественных наук при решении вопроса построения аппроксимативно адекватных моделей «искусственного общества». Работа написана стилистически ярко, удачно используются риторические фигуры для облегчения понимания трудных понятий авторской концепции АОМ. Например, применяются такие красочные метафоры, как «торец/фасад», «ресторанная отравка», характеризующие дистинкцию «потребность/надобность» [там же, с.10], «сытость/обжорство» для удобства анализа понятий текущей/долгосрочной склонностей [там же, с. 14] и др. Тем не менее, не совсем понятно, каким образом осуществляется приписывание денежных единиц показателю граничной цены пищи (формула 10) [там же, с. 20]. Числитель этой формулы измеряется в условных единицах (у.е.) надобности, которая, согласно формуле 1 [там же, с. 10], измеряется в условных единицах в диапазоне от -100 до 100. Единицей измерения знаменателя формулы 10 также является у.е. надобности. В итоге должна получаться безразмерная величина. Денежное выражение граничной цены пищи получается в результате некоторой функции приписывания стоимости, но данная функция явно не представлена, в силу чего возникает отмеченное затруднение.

В у.е. сложно измерить «счастье», «любовь», «разум» и пр. Если в начальном варианте 1960 г. машинный функционализм, сформулированный Х. Патнэмом в некоторой степени претендовал на теорию, подобную «теории счастья», то в комплексном тесте Тьюринга, как мы увидим далее в третьей главе, такого жесткого навязывания количественной объективной метрики нет – многое решает наблюдатель, субъективно атрибутирующий «счастье» изучаемым системам. Отсюда – возможность построения функционализма теста Тьюринга – тестового функционализма, который отличается от машинного функционализма тем, что доступ к машине Тьюринга как к конструкции, реализующей теорию «социального», осуществляется сквозь призму тестов Тьюринга, оценивающих её с позиции наблюдателя на предмет счастья, любви, содружества и пр. Но нормативного представления о компьютерных закономерностях «счастья» в тестовом функционализме нет!

Мы подробно остановились на анализе работы В.Истратова, так как это позволяет достаточно четко разграничить различные «измерения» объемной (трехмерной) семантики словаря искусственного интеллекта (см.пп. 3.3.3).

Антропоморфные роботы

В докладе «**Методологические вопросы робототехники**» (заседание N 19 от 25.04.2007 г. *В.А. Глазунов* рассмотрел методологические особенности научной робототехники с точки зрения подпадания ее под эгиду классической, неклассической или постнеклассической науки. Изначально робототехника может быть отнесена к классической науке, как и составляющие ее дисциплины – теория механизмов и кибернетика, однако каждой из этих дисциплин пришлось решать необходимые для их объединения задачи. Затем научная робототехника приобретает черты неклассической науки, причиной этому явились парадигмальные прививки, в частности, из физики. Далее робототехника во все большей степени получает характер постнеклассической науки - это обусловлено тем, что робот как техническое устройство наиболее полно отвечает принципам антропоморфности. Сделан вывод, что научная робототехника несет в себе тесно переплетенные проявления всех трех этапов развития науки: классического, неклассического и постнеклассического. Так же докладчик рассмотрел аспекты междисциплинарности робототехники с точки зрения петлеобразности развития составляющих ее дисциплин – теории механизмов и кибернетики. Для постановки мысленного эксперимента, характеризующего процесс междисциплинарного взаимодействия, разработана наглядная модель этого процесса. Установлено, что робототехника как научная система шире суммы составляющих ее частей – механики и кибернетики. Междисциплинарность обеспечила новый качественный уровень разрешения противоречия между противоположными требованиями автономности и гибкости проектируемых систем. Это произошло за счет достижения «синергии» (согласованного действия) приводных устройств.

Тема была развита в следующем докладе В.А. Глазунова по тематике антропоморфных и параллельных роботов. Содокладчиком выступил А.Ю. Алексеев. Рассматривались методологические особенности научной робототехники с точки зрения подпадания ее под эгиду классической, неклассической или постнеклассической науки. Сделан вывод, что научная робототехника несет в себе тесно переплетенные проявления всех трех этапов развития науки. В этой связи особо отмечают антропоморфные роботы, разрабатываемые в основном в Японии. Имеется робот для игры на флейте, для выражения эмоций. Важно, что решена весьма трудная задача двуногой ходьбы. Робот, как человек, стал «прямоходящим». Возможно, как и для человека, это станет новым этапом его развития. Так же в докладе были рассмотрены аспекты междисциплинарно-

сти робототехники с точки зрения петлеобразности развития составляющих ее дисциплин – теории механизмов и кибернетики. Установлено, что робототехника как научная система шире суммы составляющих ее частей – механики и кибернетики. Междисциплинарность обеспечила новый качественный уровень разрешения противоречия между противоположными требованиями автономности и гибкости проектируемых систем.

Интересны с точки зрения философии науки и техники робототехнические приложения к изучению научных революций. Докладчик разработал робототехнические модели для проведения мысленных экспериментов, позволивших выявить ряд свойств, характеризующих расположение точек бифуркации, течение процесса бифуркации и возможность управлять этими процессами. Установлено, что «катастрофичность» процесса бифуркации определяется несколькими факторами. Во-первых, это соотношение между внешними и внутренними силами. Во-вторых, это «траектория» точки приложения внешней силы, то есть закон изменения вектора внешних сил. В-третьих, установлено, что существуют ситуации, когда ни внешние, ни внутренние силы не способны в точке бифуркации определить направление дальнейшего развития, функцию этого выбора берут на себя случайные факторы.

В.А.Глазунов продемонстрировал модель, с помощью которой выявлены свойства, присущие научным революциям при рождении новой парадигмы. Этот процесс характеризуется появлением новой подсистемы, которая входит во взаимодействие с наличествующими подсистемами и затем внутри старой структуры вызывает «напряжения», этот период характеризуется хаосом. Для выхода на новую размерность необходимо воздействие, направленное «ортогонально» старой парадигме (здесь важными факторами являются междисциплинарность и парадигмальные прививки). Разработанная модель, по мнению докладчика, распространяется на герменевтические акты, а также на «моменты инсайта». Так же был рассмотрен многокритериальный подход к проблеме выбора научной теории. Установлено, что в этом процессе существуют, по крайней мере, три противоречивых критерия: это полнота соответствия теории опытным данным, простота теории и принадлежность теории к принятой парадигме. На всех этапах формирования парадигмы указанные противоречивые критерии обуславливают формирование паретовских множеств теорий. Каждый член этого множества имеет такой набор критериев, когда значение хотя бы одного из них лучше, чем у остальных членов множества. В качестве варьируемых параметров выступают принимаемые в рассмотрение научные факты. Набор и значение параметров может изменяться по мере получения новых данных из

опыта, при этом меняются и значения критериев, что может приводить к бифуркации и нарушению паретовского множества теорий. Теории, различающиеся на лингвистическом, семантическом или эмпирическом уровнях, при изменениях паретовского множества могут переходить из одного разряда различий в другие. В качестве обратного направления исследований – от научной методологии к робототехнике была рассмотрена взаимосвязь многокритериальности и эволюционизма в робототехнике, разработана соответствующая наглядная модель. Установлено, что при создании технических устройств и научных теорий все «жизнеспособные виды» должны располагаться на паретовских множествах. «Естественный отбор» осуществляется на основе двух главных критериев: простота и максимальная функциональность. Научное сообщество при оценке изобретения или теории руководствуется другими критериями – соответствие достигнутому научному уровню и степень превышения этого уровня. Каждое новое изобретение или научный результат изменяет существующее паретовское множество. Если существовавшие ранее члены множества не исключаются, а лишь дополняются новыми, то это – адаптационное изменение. Если же происходит исключение из множества каких-то его представителей, то имеет место бифуркационное изменение, и возможна смена парадигмы. В качестве примера уместно привел проекты «параллельных роботов» или роботов параллельной структуры. Эти устройства не имеют аналогов в природе и напоминают пространственные подвижные фермы.

Содокладчик А.Ю.Алексеев дополнил выступление сведениями об антропоморфных роботах Р. Брукса, которые представлялись для Д. Деннета инженерной реализацией гетерофеноменологической теории ментальных набросков. В свою очередь, позиция Д. Деннета важна для изучения эволюционно-эпистемологических механизмов ИИ (см. пп. 3.2.1).

Виртуальные миры

На заседании № 9, состоявшемся 22 февраля 2006 г. прозвучал доклад *М.Б.Игнатьева* (г. Санкт-Петербург) на тему: «**Лингво-комбинаторное моделирование виртуальных миров**». Докладчик исходил из того, что картина мира в каждую из эпох формируется на основе синтеза всех достижений во всех сферах человеческой деятельности. Современная, постнеклассическая наука должна иметь свои модели, среди которых особое место занимают лингво-комбинаторные модели, реализуемые на базе вычислительных систем для построения виртуальных миров. В рекурсивных структурах этих систем многократно реализуется принцип обратной связи, как для воспроизведения структур,

так и для порождения новых структур. Лингво-комбинаторное моделирование виртуального мира заключается в построении блоков 1) "населения виртуального мира" – агентов и аватаров; 2) намерений и целей агентов; 3) конструирования территориальной целостности; 4) производства виртуальных артефактов, в первую очередь, естественного языка; 5) формирования системы безопасности от компьютерных средств; 6) создания виртуальных финансов, средств эквивалентного обмена; 7) структуризации мира на «своих» и «чужих», т.е. определения внешних связей. Предполагаемая структура виртуального мира имеет конструктивные реализационные перспективы в современной компьютерной технологии.

Однако в докладе не совсем понятен термин «постнеклассическая наука». Непонятно позиционирование субъекта относительно компьютерной виртуальной реальности. В пп. 3.3.4 диссертант излагает собственную точку зрения на отношение субъекта к миру виртуальной реальности, которое подпадает под парадигму тестового компьютеризма.

Сетевой человеко-машинный интеллект

Для того, чтобы подойти к проблеме сетевого интеллекта, докладчики **М.Б. Игнатьев** и **А.Е. Войскуновский** акцентировали на современных представлениях об интеллекте естественном. Здесь привлекает компьютерная модель интеллекта, трехуровневая, в соответствии с которой выделяются три уровня – первый уровень, brain, аппаратное обеспечение мозга, второй уровень, mind, программное обеспечение мозга, третий уровень, consciousness, сознание. Уровни brain и mind работают непрерывно, они обеспечивают подсознание. Лишь небольшая часть подсознательных процессов выходит на уровень сознания, осознается человеком. Компьютерные структуры обеспечивают генерацию гигантского количества визуальной, слуховой и тактильной информации, которая, прежде всего, воздействует на подсознание людей, последствия этого воздействия не изучены. То, что осознается – решение новых задач, создание банков данных и знаний, разработка новых алгоритмов, игр и т.д., отражает лишь малую часть последствий от массового взаимодействия с компьютерами. В массовом порядке возникают феномены «телеприсутствия», или просто «присутствия» (Presence), т.е. ощущения присутствия в иной – не данной органами чувств – реальности (реальности компьютерной игры, дружеского кружка чаттеров и т.п.). Эффекты телеприсутствия частично совпадают, частично отличаются от эффектов мечтательности, сновидений, бреда, шаманского камлания, именуемых собирательно как измененные состояния сознания.

Интересны терминальные устройства на примере кибернетического велосипеда, который обеспечивает естественное взаимодействие человека с виртуальным миром посредством поворота руля велосипеда и вращения педалей, что позволяет осуществлять как одиночные так и групповые путешествия по различным виртуальным мирам, получать новые ощущения и впечатления. Другие терминальные устройства обеспечивают мобилизацию и развитие тактильных, сенсомоторных, визуальных, слуховых и вкусовых ощущений и впечатлений.

По сути, в контексте человеко-машинного распределённого интеллекта мы имеем феномен новой культуры, электронной культуры. Реальность творится совместно человеком и компьютерной системой. Диссертант данную форму культуры обозначил термином э-культура-2 (см.пп. 3.1.1).

Искусственный интеллект и философия информационного общества

Данная тема была открыта на заседании № 8, состоявшемся 25 января 2006 г. в докладе **К.К. Колина: «Россия и мир на пути к информационному обществу: философские, гуманитарные и технологические аспекты проблемы»** [Колин, 2006-1]. Докладчик охарактеризовал отличительные особенности современной глобальной информационной революции, национальные и международные программы информатизации общества, международные саммиты по проблемам информационного общества, перспективные средства и технологии для информационного общества, состояние и проблемы становления информационного общества в России и сделал вывод о том, что формирование информационного общества, основанного на знаниях - это главное направление развития цивилизации в 21-м веке. Возникают принципиально новые возможности и проблемы. Они требуют и философского осмысления и комплексного научного исследования. Примерами таких проблем могут служить: отставание общественного сознания от современных темпов развития цивилизации; глобализация сознания, обусловленная развитием процессов глобализации информационной сферы общества и новых видов коммуникаций; информационное неравенство людей в новой информационной среде общества; новая научная парадигма, связанная с переосмыслением роли информации в развитии природы и общества; новая философия образования в условиях становления глобального информационного общества.

Тема была продолжена на заседании N 16. 31.01.2007 в докладе **С.Н. Селетковой и Н.В. Днепровской «Мировые информационные ресурсы и их использование в экономике, науке и образовании»** [Селетков, Днепровская, 2007]. В докладе рассматривались состояние мировых информационных ресур-

сов и ресурсов знаний, основные проблемы, которые возникают у потребителя при поиске сведений из мирового информационного пространства, место и роль информатики в совершенствовании и развитии мировых информационных ресурсов и ресурсов знаний.

Социокультурный прогноз развития искусственного интеллекта

Тема была открыта на заседании 26.09.2007 в докладе А.С.Нариньяни **«Перспективы развития интеллектуальных информационных технологий нового поколения»** [Нариньяни, 2005-1; 2; 3]. Для возвращения России в число ведущих стран в области ИКТ возможен только один путь: поставить на реализацию комплекса стратегических инноваций, т.е. ключевых, качественно новых направлений прорыва, которые пока «скрыты за горизонтом», а завтра займут место единственно возможных на очередном этапе развития ИТ. Стратегические инновации – это те качественные преобразования, которые ведут к наиболее крутым поворотам в развитии технологий. Успех стратегических инноваций для их инициаторов многократно увеличивается эффектом внезапности. У большинства экспертов ощущение “вечности текущих устоев” сохраняется, как правило, и накануне самых крутых поворотов в ИТ: достаточно вспомнить программирование в кодах, казавшееся в начале 60-х годов основой компьютеризации на всю “обозримую перспективу” и через несколько лет навсегда ушедшее в историю. Крутые повороты развития – это глубокий кризис для «тяжеловесов», которым предстоит перестройка огромной, уже сложившейся индустрии, и уникальный шанс для сегодняшних аутсайдеров, первыми оценившим перспективу предстоящей революции.

Постнеклассическая рациональность и интеллектуальное информационное общество

Данная тема прозвучала в докладе В.С. Стёпина на 22-м заседании семинара 31.10.2007 [Стёпин, 2007]. Докладчик раскрыл типологию системных объектов и особенности саморазвивающихся систем. Современная наука и технологическая деятельность все чаще имеют дело со сложными саморазвивающимися системами. Их важно отличать не только от малых (простых) систем, но и от больших (сложных) саморегулирующихся систем. Новый тип рациональности, который сегодня утверждается в науке и технологической деятельности со сложными развивающимися, человекообразными системами, резонирует с древневосточными представлениями о связи истины и нравственности. Это, конечно, не значит, что тем самым принижается ценность рациональности, кото-

рая всегда имела приоритетный статус в западной культуре. Тип научной рациональности сегодня изменяется, но сама рациональность остается необходимой для понимания и диалога различных культур, который невозможен вне рефлексивного отношения к их базисным ценностям. Рациональное понимание делает возможной позицию равноправия всех "систем отсчета" (базовых ценностей) и открытости различных культурных миров для диалога. В этом смысле можно сказать, что развитые в лоне западной культурной традиции представления об особой ценности научной рациональности остаются важнейшей опорой в поиске новых мировоззренческих ориентиров, хотя сама рациональность обретает новые модификации в современном развитии. Сегодня во многом теряет смысл ее жесткое противопоставление многим идеям традиционных культур. Таким образом, на переднем крае научно-технологического развития, в связи с освоением сложных саморазвивающихся систем возникают точки роста новых ценностей и мировоззренческих ориентаций, которые открывают новые перспективы для диалога культур. А этот диалог, как сегодня считают многие, необходим для выработки новых стратегий жизнедеятельности глобализующегося человечества, для выхода из глобальных кризисов, порожденных современной техногенной цивилизацией. Темы, поднятые в докладе, крайне важны для культурологической параметризации проблематики сознания и искусственного интеллекта и выработки категориальных компьютерных универсалий, которые включают субъекта-деятеля в способ конституирования и конструирования сложных интеллектуальных компьютерных систем. Постнеклассическая тематика важна и для развития функционализма – функционализма теста Тьюринга – как базовая методологии изучения явлений электронной культуры.

Тема постнеклассической рациональности была продолжена на заседании N 18. 28.03.2007 в докладе **В.И. Аршинова «Синергетика как коммуникативная парадигма постнеклассического познания»** [Аршинов, Лепский, Рыжов 2007]. Докладчик считает, что синергетика с ее "нелинейной коммуникативистской парадигмой" делает наблюдаемым и узнаваемым то, что не наблюдаемо и неузнаваемо с позиций всех подходов к мозгу как системе, функционирующей "в норме" по преимуществу в состоянии равновесия, гомеостаза, более того, как системе, основная функция которой в том только и состоит, что бы этот самый гомеостаз сохранять и поддерживать. Наряду с традиционно объективированным представлением процесса (самоорганизации) паттерна (структуры), при котором параметры порядка подчиняют некоторые внешние по отношению к наблюдателю подсистемы (синергетика 1), и субъективированным представлением процесса распознавания образа, при котором параметр порядка

подчиняет некоторые внутренние детали образа в сознании (синергетика 2 — синергетика наблюдения), мы имеем здесь конструктивно-интеграционный процесс, в ходе которого один или несколько параметров порядка, которые подчиняют себе и внешне представленные подсистемы и внутренне репрезентированные свойства. Здесь мы имеем дело с тем, что мы предлагаем называть синергетикой 3 — синергетикой процессов управляемого конструирования человеком окружающей его среды на основе общих закономерностей самоорганизации космоса и его самого как (вместе с сознанием и самосознанием) как его неотъемлемой части.

Синтез естественно-научного и социо-гуманитарного знания в исследованиях искусственного интеллекта

Тему данного направления на примере медицины и социологии обозначил **В.К. Финн** в докладе «**Актуальные проблемы исследований искусственного интеллекта: прикладные задачи, методология и логика**» (5.11.2009 г.) [Михеенкова, Финн, 2005; Финн, 2007; 2008; 2009]). В докладе рассматривались базовые способности естественного интеллекта (ЕИ), покрывающие известные способности, изучаемые когнитивной психологией. Главный продукт ИИ, по мнению докладчика — интеллектуальные системы (ИС), т.е. компьютерная человеко-машинная система, такая, что она имитирует базовые способности ЕИ и имеет следующую архитектуру: $ИС = \text{Решатель} + [\text{База фактов (БФ)} + \text{База знаний (БЗ)}] + \text{Комфортный интерфейс (для пользователя)}$. ИС может работать в двух режимах — автономном и интерактивном. Последнее обстоятельство делает неприменимым известный тест А. Тьюринга для человеко-машинной системы и данные положения доклада крайне важны для нашей работы, так как В.К. Финн, по сути, формулирует требования к собственной трактовке теста Тьюринга, который мы обозначили как интерактивный тест Тьюринга (тест Финна) (см.2.3). В интеллектуальной системе выделяются три типа знаний: декларативные, процедурные и концептуальные. Последние представлены принципами конструирования ИС. В докладе характеризовался Решатель задач, образованный тремя подсистемами — Рассуждателем, Вычислителем и Синтезатором. Рассуждатель реализует синтез познавательных процедур — индукции, аналогии и абдукции, образующий правдоподобные рассуждения типа ДСМ. Кратко характеризуется ДСМ-метод автоматического порождения гипотез как пример интеллектуального анализа данных с машинным обучением в БФ. Важными составляющими ДСМ-метода являются правдоподобные рассуждения типа ДСМ (индукция + аналогия + абдукция) и формирование БЗ как квазиаксиоматической теории. Так же в докладе характеризовались практические

задачи, решаемые в ИС, реализующих ДСМ-метод, а также проблемы развития ИС. Обсуждались идеи Д. Маккарти о путях развития искусственного интеллекта и их сопоставление с работами Д.А. Поспелова. В заключение характеризовались основные продукты ИИ: интеллектуальные системы – когнитивные системы — интеллектуальные роботы, а также раскрыта фундаментальная роль ИИ для когнитивных исследований. Демонстрировались интересные результаты исследований в медицине.

Интересно то, что ДСМ-метод как метод генерации и получения нового знания в плане целевого назначения идеи коррелирует с докладами В.Л. Макарова, А.Р. Бахтизина [Макаров Бахтизин 2006] и с докладом Л.Н. Ясницкого [Ясницкий 2006, 2013]. Однако инструментарий принципиально отличен! У В.К. Финна – репрезентативная методология и технология, у других – нейрокомпьютерная. Результат и следовательно, роль инструментария эквивалентны – и те и другие обеспечивают получение нового знания, кодирование, интерпретацию, назначение терминов для неизвестных ранее кластеров, т.е. тех кластеров, на которые не было ранее направлено внимание исследователей. Репрезентативная и конненикционистская парадигмы в целевом плане совместимы, но не на уровне инструментария. В п.п. 3.3.3 будет представлен концептуальный проект комплексной машины Тьюринга-Корсакова, которая устраняет данный недостаток.

В последующем тема ДСМ-метода была конкретизирована в совместном докладе **В.К. Финна и М.А. Михеенковой «Интеллектуальный анализ данных и проблемы когнитивной социологии»** [Михеенкова, Финн 2010]. Интеллектуальный анализ данных (ИАД) рассматривается в докладе как реализация познавательного цикла «анализ данных – предсказание – объяснение» средствами интеллектуальных систем (ИС). Современные средства искусственного интеллекта и прикладной логики позволяют представить указанный цикл в виде схемы «индукция – аналогия – абдукция». В результате эмпирической индукции (основанной на установлении сходства имеющихся фактов) порождаются гипотезы о причинах рассматриваемых явлений; полученные гипотезы посредством вывода по аналогии переносятся на случаи с неопределенным исходом; абдукция объясняет начальное состояние базы фактов. Для решения проблем когнитивной социологии – получения нового знания на основе анализа эмпирических социологических данных с последующей формулировкой новых проблем – существенным оказывается реализация методов ИАД логическими средствами в ИС. Отмечается принципиальное отличие между статистическим ана-

лизом данных и интеллектуальным анализом данных. Первая методология имеет пассивный характер, вторая – активна. Она постоянно порождает социологические гипотезы и верифицирует их.

На наш взгляд, ДСМ-метод на сегодняшний день выступает ярким представителем символьной парадигмы ИИ. Поэтому при разработке гибридной репрезентативно-нейрокомпьютерной системы ИИ в качестве репрезентативной (символической, лингвистической) составляющей компьютеринга перспективно использовать именно ДСМ-метод.

Проблема синтеза естественнонаучного и социогуманитарного знания является достаточно древней, по-крайней мере, прослеживается с дильтеевских работ по разделению наук. В теме «физиков/лириков» возможны точки соприкосновения, как мы видели на примере выше представленных докладов. Однако возможны и философские недоразумения в этой сфере. В качестве примера приведём доклад **В.И.Моисеева** на тему **«Экспертные системы в медицине: методологические аспекты»**, с которым он выступил на 24-м заседании семинара 26.12.2007 г. В.И. Моисеев предложил медицину рассматривать в формате прикладной оптимологии [Моисеев, 2007]. Оптимология – наука об оптимизации, поиске наиболее оптимальных решений в различных областях. Главная задача медицины – оптимизировать деятельность организма человека, преодолеть разного рода неоптимальности его функционирования. Отсюда основная направленность в развитии медицинского знания – проведение в отношении к организму оптимизирующей деятельности и поиск разного рода критериев оптимальности в жизнедеятельности организма. Эффективность работы ЭС в медицине прямо зависит от уровня развития теоретических знаний в медицине. Тип рациональности медицинского знания носит междисциплинарный характер. Медицина не может быть отнесена не только к естественным, ни только к гуманитарным наукам. Проблема медицинского образа реальности («медицинской онтологии») как реальности, соединяющей в себе соматические и аксиологические (ценностные) элементы («аксиосоматика»). Отличие аксиосоматики от психосоматики. Аксиология мыслится как универсальный слой онтологий живого, не только обладающих психикой (связь с формулой «life is cognition» теории аутопозза, формула «life is evaluation» - «жизнь есть оценка (способность ценить)»). Субъекты – это жители аксиосоматических онтологий. Подобное направление развития теоретической медицины предполагает новый тип когнитивных технологий, имеющих существенно постнеклассическую ориентацию (включение субъектных структур в состав онтологических концептуализаций).

Аксиосоматический характер медицинской онтологии обнаруживает существенное сродство с онтологией самих экспертных систем, когнитивная технологичность которых обладает тем же единством полюсов соматического-технологического и ценностно-когнитивного, что и структуры медицинского субъектно-нагруженного типа реальности.

Несомненны онтологические корни подхода В.И. Моисеева – соловьевская философия всеединства, с канонами добра, истины и красоты, выраженные средствами логико-символического языка. Диссертант отрицательно относится к продуктивности такого рода моделей. Крайне метко выразился по этому поводу Ю.Ю. Петрунин на заседании конференции «Электронная культура: интеллектуальные технологии», 12 июня 2011 г.: «Библию, несомненно, можно формализовать логикой предикатов первого порядка. Вот только – зачем? Что дальше? Как это практически применить?» Показательным так же является высказывание В.И. Моисеева. В начале своего доклада он высказал интересную мысль – «я не являюсь специалистом ни в медицине, ни в построении экспертных систем, но по поводу методологии медицинских экспертных систем я могу многое конструктивное сказать. Я – философ». Такой априористский подход тоже верен, однако формализованная схоластика, по мнению диссертанта, существенно тормозит развитие ИИ.

Наряду с этим необходимо отметить, что В. Моисеев – один из немногочисленных исследователей, который обратил внимание на тест Тьюринга в период становления ИИ в современной отечественной науке, предложив уделять внимание при тестировании временным параметрам теста – различным отрезкам длины теста, а не длине теста «вообще» [Моисеев, 2005].

Нейрокомпьютер и общество

На заседании № 13, состоявшемся 27 сентября 2006 г. был заслушан доклад **Л.Н. Ясницкого** (г. Пермь) на тему: «**Искусственный интеллект и новые возможности метода компьютерного моделирования**». Докладчик охарактеризовал основные стратегии искусственного интеллекта и выделил нейросетевую парадигму как наиболее перспективную для решения задач оптимизации, прогнозирования и управления, распознавания и классификации образов и пр. При этом для построения компьютерных моделей с помощью нейросетевых технологий не требуется знание и использование фундаментальных законов природы, общества, человека (данное утверждение было подвергнуто критике со стороны оппонентов). Вместо этого, считает докладчик, нужно лишь подготовить обучающие примеры, содержащие статистические данные о предметной

области, ее свойствах и поведении. Если выборка оказывается репрезентативной, то нейросеть самостоятельно извлекает закономерности, необходимые для формирования математической модели. Методологические выводы докладчик базировал на достижениях Пермской школой искусственного интеллекта, успешно решающей ряд прикладных задач общественной значимости на базе нейросетевых технологий: прогнозирования потребления тепловой и электрической энергии организациями бюджетной сферы – медицинскими и образовательными учреждениями; в следственной практике МВД и ФСБ России – нейросетевые «детекторы лжи» намного превосходят традиционные полиграфы. Докладчик указал на современный кризис прикладной математики и непосредственно связал данный кризис с глобальными проблемами человечества, так как, например, от достижений математической науки напрямую зависит степень вероятности катастроф локального и глобального масштаба. Докладчик заключил, что применение методов искусственного интеллекта, в особенности, нейросетевых технологий, поднимет инженерные расчеты на уровень, соответствующий требованиям времени.

Это и другие положения доклада подверглись тщательному разбору и критике со стороны участников семинара. Показательным является мнение Л.Н.Ясницкого, которое он высказал при обсуждении доклада диссертанта в г.Пермь в рамках проведения семинара «Философско-методологические проблемы искусственного интеллекта» (рук. Е.В.Серёдкина) [Алексеев, 2007]. Л.Н. Ясницкий полагает, что если компьютерная система моделирует нейронную сеть, то она априори становится системой искусственного интеллекта, ведь эта система работает в соответствии с моделью искусственного интеллекта. Несомненно, здесь имеется не критичное понимание предметной области ИИ.

Выводы

Предложенная в данном параграфе концептуальная панорама философско-методологических исследований ИИ представляется в неструктурированном, немного хаотичном формате. - сегодня не стоит утверждать о возможности целостного охвата философской методологии ИИ и построения стройной системы философско-методологического знания ИИ. Подчеркнем динамику этого знания на примере. Так, в частности, на IV Российском конгрессе «Философия и будущее цивилизации», который состоялся 23-28 мая 2005 г. диссертант совместно с В.А.Лекторским и В.А.Глазуновым проводил **Круглый стол «Философско-методологические проблемы когнитивных и компьютерных наук»**. Мероприятие проходило в МГУ им. М.В. Ломоносова (24-25 мая), ЦЭМИ РАН (26 мая), МИЭМ (27 мая) [Философия и будущее цивилизации 2005]. В программе Круглого стола - около 50 докладов. Проблематика секции охватывала широкий круг вопросов по искусственному интеллекту, кибернетическим и синергетическим системам, когнитивным моделям, категориальному представлению информации, знаний, мышления, методологии обработки, хранения и преобразования информации в биологических, робототехнических, социальных системах. Например, М.Б. Игнатьев представил мир как информационную машину; Г.С.Воронков моделировал глобальные свойства живой и неживой природы; Бодякин В.И. изучал техническую реализацию и моделирование нейронных сетей; Кузнецов О.П. уделил внимание проблеме определения базовой категории искусственного интеллекта – понятию «знание»; Редько В.Г. – проблеме изучения когнитивной эволюции в направлении синтеза когнитивных и компьютерных наук, а также психологических теорий интеллекта. Наиболее оживлённые дискуссии вызвал доклад А.С. Нариньяни на тему «Перспективы развития интеллектуальных информационных технологий нового поколения», в котором затрагивались стратегические вопросы прорыва России на рынке информационно-коммуникационных технологий. Состоявшиеся дискуссии, как мы видим, с трудом укладываются в русло заявленных ранее тематик (рассмотренных в пп. 1.1.1).

Поэтому преждевременно говорить о дисциплинарной структуризации и систематизации философии искусственного интеллекта. Если в области искусственного интеллекта сложились достаточно устойчивые направления исследований, то в области философской методологии искусственного интеллекта такое направления пока только начинают формироваться. ИИ – это отдаленная перспектива. Поэтому методология ИИ может ассоциироваться с перспектив-

ными проектами естественнонаучного характера, например, с «искусственным мозгом». Либо с проектами гуманитарного плана, связанными с проблемами выражения «собственно человеческого» в среде компьютерных симулякров. В следующем параграфе рассматривается расширение деятельности НСМИИ РАН, в развитии, соответственно, направлений «Нейрофилософия» и «Философия творчества».

1.1.3. Итоги десятилетней работы НСМИИ РАН

Итоги развития междисциплинарных исследований искусственного интеллекта раскрываются в анализе дискуссий трех мероприятий. Первый класс дискуссий сложился внутри собственной проблематики методологии искусственного интеллекта. Эти дискуссии состоялись в рамках круглого стола "Философские исследования искусственного интеллекта: итоги и перспективы". Он состоялся 26 марта 2015 года в Институте философии РАН (ИФ РАН) в рамках юбилейного симпозиума, посвященного десятилетию деятельности Научного совета РАН по методологии искусственного интеллекта (НСМИИ РАН). На симпозиуме был подведен ряд научно-организационных и теоретико-методологических итогов десятилетней работы Совета, намечены перспективы на будущее. Руководителями симпозиума и круглого стола явились сопредседатель НСМИИ РАН академик В.А.Лекторский и координатор научных программ НСМИИ РАН А.Ю.Алексеев.

Второй класс дискуссий раскрывает особенности расширения сферы искусственного интеллекта в контексте натуралистического решения основных философских проблем. Дискуссии состоялись в рамках международного междисциплинарного семинара «Нейрофилософия», который проходит, как правило, на философском факультете МГУ имени М.В.Ломоносова (соруководители – А.Ю.Алексеев, В.Г.Кузнецов, А.В.Чусов).

Дискуссии третьего класса раскрывают особенности расширения сферы искусственного интеллекта на сугубо гуманитарную сферу человеческой деятельности – на феноменологию творчества. Обсуждения этой тематики проходили на круглом столе «Проблема творчества в компьютерном мире» (руководители – Ю.Ю.Петрунин и А.Ю.Алексеев) в рамках всероссийской научной конференции «Философия творчества» 9 апреля 2015 года, Институт философии РАН.

Представленные в данном параграфе дискуссии пока не опубликованы в широкой печати, поэтому представляют некоторую библиографическую цен-

ность. Аналитический обзор круглого стола "Философские исследования искусственного интеллекта: итоги и перспективы" издается в журнале «Философия науки и техники» в соавторстве с В.А.Лекторским. Обзор по поводу становления отечественной нейрофилософии издается в 11 номере журнала «Философские науки» за 2015 г. в соавторстве с В.Г.Кузнецовым, А.В.Савельевым и Е.А.Янковской. Обзор мероприятий конференции «Философия творчества» выходит в ближайшем номере «Вестник РГНФ» в соавторстве с Н.М.Смирновой.

Анализ собственной проблематики методологии искусственного интеллекта

Особый интерес участников круглого стола "Философские исследования искусственного интеллекта: итоги и перспективы" (26 марта 2015 г., ИФ РАН), вызывали те доклады, которые были заслушаны ранее, десять лет назад, так как появлялась возможность оценить тренд идей, которые выражались на первых заседаниях всероссийского научно-теоретического семинара "Философско-методологические проблемы искусственного интеллекта". Таких докладов было большинство.

Сопредседатель совета академик В.Л.Макаров в докладе «РОЛЬ МЕТОДОЛОГИИ В ИССЛЕДОВАНИЯХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА» подчеркнул, что проблематика искусственного интеллекта начинается с методологии искусственного интеллекта. Словосочетание «искусственный интеллект» завораживает и приманивает многих. Однако надо вначале понять, что это такое, определить главные понятия и методы.

Докладчик раскрыл ряд интересных перспектив в области компьютерного моделирования социокультурной сферы, получившее название «искусственные общества». Интригует то, что искусственные общества, как бы, в самом деле, "зарождаются" и "живут" в компьютере. Это программы самого различного назначения. Они связаны не только с общественно-значимыми проблемами моделирования процессов экономического развития страны, регионов, мира. Не лишены интереса задачи чисто любознательного плана. Например, можно посчитать, что бог, или кто-то другой, неправильно поступил, создавая двуполых людей. Что если произошла ошибка? Должно быть, к примеру, три пола. Разве не интересно узнать, какое общество получится из трехполых существ? Какой вклад вносит каждый пол в развитие общества? Почему бы не поэкспериментировать, общество ведь искусственное.

С «искусственным интеллектом» намного труднее, - считает В.Л.Макаров. "Интеллект", "мышление", "разум" – первичные понятия. "Общество" – вторичное. Философские исследования искусственного интеллекта неразрывны с междисциплинарным спектром всех иных исследований. К примеру, в дискуссиях по поводу кибернетического бессмертия участвует много специалистов (движение "Россия-2045"). Например, врачи. У них такой путь к бессмертию. Берём обычного человека. Постепенно заменяем его. В начале – руки, затем - ноги, потом – сердце, и, наконец, вместо мозга, имплантируем более надежное устройство. Возникает вопрос: получили ли мы искусственного человека? Наверняка это утверждать нельзя. Вот здесь и требуется философское осмысление. За решение такого рода вопросов и отвечают специалисты НСММИ РАН. Они задают главный - методологический - вектор технологии создания "искусственных существ". В подобных работах по тематике искусственных систем возникает много разнообразных проблем. Однако, начиная с методологии, мы поступаем правильно. Надо начинать с адекватного понимания понятий, среди которых самое главное: что такое «искусственный интеллект»? Прошедшие десять лет дискуссий укрепили эту убежденность - заключил В.Л.Макаров.

Со своей стороны диссертант отмечает, что исследования ИИ начинались именно с определения понятия "машинного интеллекта". Форма нормальной постановки данного вопроса, согласованного со "здравым смыслом", была предложена А.Тьюрингом в игре в имитацию естественного интеллекта. Но что такое "интеллект", никто обоснованно утверждать не может по причине того, что значения данного слова имеют "статистический" характер (А.Тьюринг). А раз так, то проблема искусственного интеллекта является вечной - философской - проблемой, но более сложной, нежели чем тематика, которая поднимается вопросом "что такое - интеллект". Ведь в сфере ИИ понятие «интеллект» надо как-то компьютерно реализовать, а не только его определить.

С алармистским докладом «ЭКЗИСТЕНЦИАЛЬНЫЕ ПОСЛЕДСТВИЯ НАСТУПЛЕНИЯ "ЖИДКОГО МИРА" КОМПЬЮТЕРНОЙ СРЕДЫ» выступила Т.В.Черниговская. Изучается термин «Liquid World», который не совсем внятно переводится с английского на русский язык: «жидкий», или «текучий мир». Однако термин хорошо отражает современную ситуацию, так как все мы уже сейчас оказались в некоей текучей реальности, или приближаемся к ней с большой скоростью. В таком мире человек имеет как бы три «мозга». Один мозг – его собственный, другой мозг – в руке у бога или чёрта, а третий мозг

представляет собой глобальную сеть, к которой человек подключается с каждым днем всё более и более постоянно. Следуют выводы.

Со "временем" непонятно. Во-первых, в социальных сетях реальные и виртуальные люди все более и более перестают разделяться. Ведь мы общаемся с людьми в чатах, скайпах, соцсетях и пр. И одновременно (в буквальном смысле) мы общаемся с реальными людьми. Граница между общением с людьми внутри и вне сети размылась. Во-вторых, люди переходят от идеи что-то купить/приобрести, т.е. сделать «своим», к идее аренды. Это означает: «взять» на некоторое время. Из этого и много другого следует временность и кратковременность вообще всего. Это – такая текучесть, когда «ничто не мое». Возникает феномен «невременного времени» («Timeless Time»).

Непонятно и с "пространством": то есть где мы вообще живем, где мы находимся? Мы можем спокойно включаться вот в этом пространстве и этой точке. То есть размылось и "время", и "пространство". Проблематичен онтологический статус "информации": она не только в ноутбуках и на флэшках, - сегодня она вообще в «облаках» находится. Генетическую информацию скрыть невозможно, какие бы юридические законы не принимались, а телефон дает возможность определить, где вы находитесь и что совершали.

В вариантах «умного дома» сами предметы начинают «общаться» друг с другом, например, самонаполняемый холодильник - с магазином. В технологиях э-медицины пациент и врач общаются друг с другом не лично, но с помощью устройств, которые собирают информацию о нашем организме: о состоянии ушного прохода, горла и т.д. Происходит реальный анализ состояния здоровья. Возникает вопрос: у нас что-нибудь свое остается? Хотя бы тело – оно наше? Или уже и наше тело является предметом манипулирования со стороны тех, кому надо им манипулировать?

Возникает проблема антропологической "текучести": люди-то вообще где? Куда мы попали? Нам, как людям, место в этом мире все еще отведено? И как мы сможем с этим справиться? – вопрошает докладчица. Отсюда - персонологическая неопределенность. Как быть с "Я", если человек - это «Homo Sapiens Autocreator» – «самосоздаваемый человек» (М.Маклюэн). Допустим, у меня заменили руку. Далее – заменили ногу. Печень заменили, сердце заменили, и, вообще, все заменили. В этом случае Я – все еще Я? Допустим, в мозг помещен чип, увеличивающий скорость и память. Я подключена к социальным сетям. И снова: Я – все еще Я? Где моя идентичность? И вообще, есть ли она? Где она начинается? Где наши границы? Я-то где? Что из всего этого мое? Или уже во-

обще ничего нет? Не снится нам ли все это? Или, в самом деле, весь мир перемешался? Виртуальное смешалось с реальным, физическое – с идеальным?

Помимо мировоззренческих проблем, докладчица затронула и ряд методологических проблемы ИИ. Например, проблему "инженерии знаний": в глобальной экспертной системе, в принципе, можно собрать все знания, которые человек способен сформулировать. Но такие знания особо и не нужны: они имеются в книжках, учебниках, энциклопедиях. Нужны знания гениального врача, а такие знания пока никак не собрать.

Диссертант полагает, что картина "Жидкого мира" требует философско-методологического подкрепления. В самом деле, мы уже сейчас живем в таком мире, образ которого обрисовала Т.В.Черниговская. Он возник во многом из-за развития ИИ и внедрения продуктов его исследований. Важно выработать фундаментальную позицию относительно Я и компьютерного мира. Поэтому в данной работе предлагается концепция тестового функционализма, которая позволяет рационально изучать и ориентироваться в компьютерном мире благодаря использованию концептуального инструментария комплексного теста Тьюринга, благодаря тому, что Я овладеваю инструментом проникновения в суть явлений компьютерного мира (пп.3.4.3).

Величковский Б.М. выступил с сообщением на тему "ФИЛОСОФИЯ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, КОГНИТИВНЫЕ НЕЙРОНАУКИ – ВЗАИМОДЕЙСТВИЕ НЕОБХОДИМО!". Изучение истории философии и когнитивной науки показывает: разные кризисы, сломы, научные революции связаны с тем, что за ними просто кроются различные философские основы. Помимо прочего, имеются и философские основы самой психологии. Это – философские основания: кантианство, фихтеанство, феноменология. В некоторые счастливые свои моменты когнитивная наука представляет своеобразную экспериментальную философию. Третья составляющая – искусственный интеллект – при всех своих успехах и удачах, в целом, очень неоднозначное направление. В целом искусственный интеллект ограничивается уровнем рационального, но не выходит на уровень "понимания", это более высокий уровень. Поэтому интеграция необходима, поскольку, например, невозможно объяснить развитие языков без постулирования наличия некоторого глубинного "языка". Сегодня крайне остро встает проблема свободы воли в связи с развитием нейрокогнитивных технологий. И надо еще учесть в связи с проблемой сознания, что до восемнадцатого века слова «сознание» и «совесть» в русском языке были синонимами. И в сербо-хорватском языке это до сих пор так. И прилагательное

«сознательный» имеет оттенок значения «совестливый». Учитывая все это, мы, возможно, должны расширить спектр проблем, который рассматривается в когнитивных нейронауках и в искусственном интеллекте. А в философии это было сделано уже столетиями назад.

Диссертант находит эти положения крайне важными для изучения проблемы компьютерной реализации ментальных феноменов и предлагает далее мнемоническую фигуру для фиксации социокультурного измерения понятия «сознание» - «философскую пентаграмму», которая отличается от «философского квадрата» Д.И.Дубровского возможностью фиксации различных культурных традиций употребления слова «сознание» (пп.3.1.1).

Финн В.К. в докладе "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ КАК НАУЧНОЕ НАПРАВЛЕНИЕ И ПРОБЛЕМЫ ТОЧНОЙ ЭПИСТЕМОЛОГИИ" выдвинул основное требование к исследованиям искусственного интеллекта: когда мы начинаем о нем говорить, то должны каким-то образом рационально охарактеризовать эту область. На заре своего развития, когда он возник в Стэнфордском университете под руководством Дж.Маккарти, с ИИ связывались романтические представления. Эта метафора использовалась как стимул для программирования различных умственных задач, в том числе, шахматы, в связи с автоматическим доказательством теорем. Со временем, искусственный интеллект превратился в научное направление. И как научное направление он может быть и должен быть охарактеризован в точных терминах.

Что такое современный искусственный интеллект? В.К.Финн определяет его как исследования, в результате которых мы имеем имитацию и усиление интеллектуальной деятельности посредством автоматизации познавательных процедур в компьютерных интеллектуальных системах. Следовательно, мы должны уточнить понятие «компьютерные интеллектуальные системы». Разумеется, искусственный интеллект занимается не только компьютерными интеллектуальными системами, хотя это главный его продукт. Он занимается также методами, которые используют современные компьютерные интеллектуальные системы.

В.К.Финн следующим образом определяет состав современной компьютерной интеллектуальной системы: это довольно сложная программная система, которая состоит из базы фактов, базы знаний, решателей задач и комфортного интерфейса для пользователя. Решатель задач - сложная и нетривиальная структура. Он состоит из трёх уровней: 1) Рассуждатель – это главное средство компьютерной интеллектуальной системы (инструмент её "мозга").

2) Вычислитель - реализует вычислительные процедуры. 3) Синтезатор, который объединяет рассуждение и вычисление, а также формирует стратегии для решения соответствующих задач (в первоначальном варианте «синтезатора» не было, был «интерфейс» - см. предыдущий пп.).

В.К.Финн полагает, что для методологии и философии интересна аналогия, которая имеет место между структурой компьютерной интеллектуальной системы и феноменологией рационального сознания. Традиционно сложилось два направления в ИИ, как утверждали и Д.А.Поспелов и известный специалист по когнитологии М.Боден: 1) алгоритмическое; 2) бионическое, основанное на идее о моделировании мозга. Однако основные результаты, носящие практический характер и допускающие обобщение на теоретическом уровне, относятся, всё-таки, к первому направлению - к направлению, когда создаются компьютерные интеллектуальные системы, имеющие алгоритмическое и логическое обеспечение.

Данное принципиальное утверждение В.К.Финна является базовым тезисом в защиту символической (или репрезентативной) методологии ИИ. Далее мы увидим, что тезис подобной силы выдвигает О.П.Кузнецов, но уже в защиту коннекционистского подхода.

Какая же структура соответствует феноменологии рационального сознания? По мнению В.К.Финна, сфера рационального сознания состоит из системы знаний, продуктивного мышления и субъективного мира личности. В компьютерной интеллектуальной системе имитируется и усиливается естественный интеллект (ЕИ). Причем второе - нетривиально и, как мы увидим далее, тест Финна принципиально отличается от оригинального теста Тьюринга во многом благодаря тому, что первый за счет интеракций ИИ и ЕИ усиливает ЕИ, в то время как тест Тьюринга лишь имитирует ЕИ. В какой мере отражается субъективный мир личности? Конечно, он в какой-то мере отражается в комфортном интерфейсе, но для современных систем это не является важным вопросом, хотя на самом деле, нужно создавать такие компьютерные интеллектуальные системы, которые в своём интерфейсе учитывали типы личности. Это - дело будущего, - полагает В.К.Финн. Важный вопрос в следующем. Для того, чтобы говорить, что компьютерные интеллектуальные системы с этой архитектурой являются интеллектуальными, необходимо, чтобы имитировались и усиливались определённые способности человека: 1) вывод существенных характеристик из данных; 2) способность к рассуждению; 3) способность отбора посылок, релевантных целям рассуждений; 4) целеполагание (цель, план действия в какой-то мере

должны отражаться в интеллектуальной системе); 5) формализация рассуждений и представление рассуждений в решателе задач.

Решение задач связано с формализацией эвристик, а эвристика реализуется в результате синтеза познавательных процедур. Примером синтеза познавательных процедур является следующее: индукция – это обнаружение некоторых сходств в данных, затем – аналогия, то есть использование полученных сходств для предсказания каких-то эффектов и, наконец, это абдукция Ч.Пирса, то есть принятие гипотез посредством объяснения начального состояния базы данных. Очевидно, - считает В.К.Финн, - синтез познавательных процедур соответствует элементарному познавательному процессу: анализ данных, предсказание, объяснение. И всякая компьютерная система, претендующая на интеллектуальность, обязана реализовать в Решателе элементарные познавательные процессы: анализ данных, предсказание, объяснение. Эта архитектура является носителем интеллектуальной системы.

Далее В.К.Финн раскрывает проблемы т.н. "точной эпистемологии". Такая эпистемология внутри исследований ИИ пока только на этапе становления: это научное понимание того, как следует формализовать и автоматизировать процесс познания. Процесс познания можно действительно реализовать и формализовать с помощью компьютерных интеллектуальных систем, о которых я говорил. И тогда возникает три проблемы точной эпистемологии.

Первая проблема точной эпистемологии – это формализация и классификация различных видов понятий. Она связана с главным разделом ИИ, в который входят: представление знаний, рассуждение и механизм открытия знаний. В связи с этим необходимо иметь развитые средства для формализации понятий. В рамках ИИ возникли новые представления о понятийных структурах: 1) теория фреймов; 2) квалификация различных понятий как индуктивных понятий, которые с помощью индуктивных определений могут быть реализованы; 3) понимание того, что имеются три типа знаний: первый тип знания – это декларативные знания, которые представлены в базах фактов. Второй тип знаний – это процедурные знания. И третий тип знаний – это концептуальные знания, которые направляют исследователя, давая ему возможность сформировать методологию построения интеллектуальных компьютерных систем.

Для целей настоящей работы, полагает диссертант, - крайне важно выделение третьего вида "знаний", которые представляются в конструкции самой компьютерной системы. А именно, помимо представления классических (для методологии проектирования ИИ) процедурных и декларативных знаний, в ба-

зу знаний требуется включать еще и концептуальные знания. Собственно, тесты Тьюринга, если их представлять в компьютерной системе, способы их идентификации, систематизации, координации, интеграции и пр. - суть "концептуальных знаний". В связи с этим диссертант полагает, что язык комплексирования тестов Тьюринга, предложенный далее в п.п.2.3 является носителем этих "знаний". Если так, то философско-методологические представления о комплексном тесте Тьюринга "не зависят", они включаются во внутреннее устройство системы, становясь, тем самым, непосредственным фактором ее функционирования. Собственно, такой подход коррелирует со схемой взаимосвязи философии и ИИ, который предлагал Дж.Маккарти: эпистемологические знания непосредственно инкорпорируются в структуру компьютерной системы (см. п.п.1.2). Отсюда - необходимость создания средств представления концептуальных "знаний". Эти средства далее рассматриваются в п.п.3.4.

Вторая проблема точной эпистемологии ИИ, - считает В.К.Финн, - это распознавание эмпирических закономерностей. Под эмпирической закономерностью мы должны понимать некую регулярность, которая сохраняется в последовательности расширяющихся баз фактов и которая выражена с помощью существенных терминов. Если мы распознали такую регулярность, то она может быть двух видов: это может быть эмпирический закон, если эта регулярность повторяется всегда в последовательности расширяющихся данных, и это может быть эмпирическая тенденция, если отклонений от этой регулярности бывает немного. Таким образом, если мы рассматриваем в познании эмпирическую закономерность, то мы можем пополнять базу знаний автоматически.

И, наконец, - полагает В.К.Финн, - важной проблемой точной эпистемологии является тест Тьюринга. Дело в том, - считает докладчик, - что компьютерные интеллектуальные системы имеют два режима: автоматический и интерактивный, а, следовательно, это открытое множество исходных данных, партнерское взаимодействие человека и машины. В случае интерактивного режима тест Тьюринга не работает. Диссертант полагает, что в интерактивности – суть оригинальности т.н. «теста Финна» (пп. 2.3.3).

Кузнецов О.П. в докладе "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И НАУКИ О МОЗГЕ – МЕТОДОЛОГИЧЕСКИЕ ПРОБЛЕМЫ" утверждает: в отличие от других когнитивных наук – наук о мозге - искусственный интеллект исследует не мозг как таковой, а средства решения интеллектуальных задач. Это задачи, которые человек решает не с помощью алгоритмов, а путем использования опыта, знаний, сообразительности и рассуждений.

О.П.Кузнецов является сторонником коннекционистского подхода. По поводу же оппонирующего, символьного подхода (такой подход он называет "компьютерная парадигма") утверждаются следующие основные положения: мозг – это биокомпьютер, и его когнитивные механизмы надо рассматривать как механизмы переработки информации; электронные компьютеры эффективнее биологических компьютеров: они в миллион раз быстрее, не ошибаются, не устают и т.д.; следовательно, мозг изучать не надо; искусственный разум будет создан другим путем – путем создания компьютерных программ.

Полемика между Б.М.Величковским, О.П.Кузнецовым и В.К.Финном впервые в отечественной науке отразила фундаментальный статус дискуссий между «символистами» и «коннекционистами». Об этом – в следующем пп., в связи со становлением нейрофилософских идей в ИИ.

Нейрофилософия: роль методологии искусственного интеллекта в натуралистическом решении основных философских проблем

В последние несколько лет в мировой научной литературе часто используется слово "нейрофилософия". Появились сайты, статьи, журналы, монографии, учебные пособия. В научных центрах и лабораториях разрабатываются проекты. В ведущих вузах мира преподаются учебные курсы. Сложно и невозможно однозначно определить их профиль: "нейрофилософия" проникла в сферу естествознания, в сферу социальных и гуманитарных дисциплин и, конечно, в технические (компьютерные) науки. В нашей стране несколько лет назад также возникло достаточно мощное научное движение, исследования которого объединились в рамках тренда "нейрофилософия". Оно сформировалось в ходе дискуссий на междисциплинарном научно-теоретическом семинаре «Философско-методологические проблемы искусственного интеллекта», организованного НСММИ РАН. Нейрофилософская проблематика возникла с самого начала работы семинара. На первом заседании 17 марта 2005 г. в докладе "Сознание, мозг, искусственный интеллект" Д.И.Дубровский обосновал стратегию развития исследований искусственного интеллекта: расшифровка кодов нейральных коррелятов психических явлений. На втором заседании с докладом "Компьютерное моделирование искусственных миров 20 апреля 2005 г. выступили В.Л.Макаров и А.Р.Бахтизин, раскрыв особенности проекта нейрокомпьютерного моделирования экономики страны. Изучение искусственного интеллекта существенно ограничивает нейрофилософскую тему, так как нацелено на инженерные вопросы компьютерной реализации когнитивных феноменов. Поэтому нейрофилософам "стало тесно" и в феврале 2015 года НСММИ РАН учредил

секцию "Нейрофилософия", руководитель которой стал В.Г.Кузнецов, д.ф.н. проф., зав.кафедрой философии и методологии науки МГУ им.М.В.Ломоносова. Формированию этой секции способствовал ряд мероприятий. Эти мероприятия нашли отражение в работах диссертанта: [Алексеев, 2014-4; 2015-1;2;3;5].

Нейрофилософии как самостоятельное направление в философской науке

В нашей стране систематическое обсуждение философской проблематики, обозначенное словом "нейрофилософия", началось с проведения 6 декабря 2013 г. юбилейного симпозиума, посвященного 150-летней годовщине со дня публикации статьи Ивана Михайловича Сеченова «Рефлексы головного мозга». Как отметил руководитель мероприятия академик В.А.Лекторский, цель симпозиума не ограничивалась изучением истории становления натуралистического подхода в философии сознания, хотя и это крайне важно, так как он доминирует в современной науке. Важно показать, что идеи, связанные с философской рефлексией над проблемами изучения психики и мозга, играют значительную роль в жизни человека и общества.

Прозвучали доклады, во многом определившие последующие нейрофилософские дискуссии. К.В.Анохин в докладе «Сеченовская программа построения объективной психологии: какой она могла бы быть сегодня?» реконструировал обобщенные концепции современных нейромоделей и при демонстрации результатов оптоволоконной регистрации экспрессии генов мозга с "некоторым изумлением для себя" отметил: для исследований мозга невозможно придумать пунктов более современных, чем те, которые сформулированы в сеченовской программе полутора вековой давности. Н.С.Чуприкова показала, что современная теоретико-познавательная проблема объяснения развития мышления из чувств была сформулирована в сеченовских работах. С.К.Судаков в работе «Рефлексы головного мозга. От мышечного сокращения до везикулярного транспорта» отметил, что И.М.Сеченов впервые показал возможность изучения психической деятельности человека путем исследования рефлексов головного мозга, т.е. субъективной реальности объективными методами. Н.С.Косицын в докладе «Ультраструктурные корреляты работы мозга» отметил, что достижением мировой нейрофизиологии явилось открытие и доказательство И.М.Сеченовым центрального торможения на конкретных структурах мозга. Ю.Ю.Петрунин подробно раскрыл, каким образом идеи И.М.Сеченова повлияли на психологию, кибернетику, искусственный интеллект. Г.И.Шульгина подчеркивала роль изучения процессов торможения и их связь с социодинамикой.

Е.А.Юматов ввёл важный методологический принцип возможности изучения психических процессов при помощи живых структур. А.Ю.Алексеев сравнил машину Тьюринга с машиной Сеченова и обосновал функционалистскую концепцию сознания, основанную на концепции последней. В.И.Бодякин применил сеченовскую идею самообучения для разработки проекта компьютера, способного пройти тест Тьюринга. И.А.Бугаков рассмотрел "сознание" как «сверхрефлекс» по порождению рефлексов. Д.Н.Кавтарадзе на примере образовательных имитационных игр раскрыл суть гипотезы о ретардации, т.е. о закреплении морфо-функциональных механизмов адаптации мозга. И.В.Степанян продемонстрировал результаты обнаружения латентной упорядоченности длинных нуклеотидных последовательностей. А.Н.Иноземцев отметил большое значение сеченовского метода различения психофизиологических функций в норме и патологии на примере управления эмоциональными реакциями. Г.Б.Крюков обобщил идеи о рефлексе до уровня человеко-машинной цивилизации. Э.И.Денисов подчеркнул продуктивность идеи формирования «тёмных чувств из внутренней среды организма» для "гигиены умственного труда". Н.А.Новосёлова-Савельева и А.В.Савельев на основе раскрытия особенностей функциональной нейроанатомии сетчатки глаз, подчеркнули роль инженерной методологии в изучении проблемы сознания. Ю.М.Хрусталёв осуществил обзор научной жизни И.М.Сеченова.

Отмечалось, что И.М.Сеченову принадлежит популярное изложение новейших для своего времени методов и теорий нейрофизиологии, манифест определенной научной методологии, исследовательской программы в психофизиологии и, наконец, философское обобщение научных исследований. Пожалуй, главным результатом работы симпозиума стала яркая демонстрация тесной связи между науками о мозге и философской наукой. Именно это пробудило интерес общественности к формированию "нейрофилософии" как специфической области взаимовлияния философии и нейронауки.

Нейрофилософия: междисциплинарный подход

Впервые в отечественной философии слово "Нейрофилософия" эксплицитно использовалось для именованя секции на конференции «Нейрокомпьютеры и их применение», 18 марта 2014 г., Московский городской психолого-педагогический университет.

А.Ю.Алексеев отметил, что "нейрофилософия" - это конвенциональное обозначение стратегического направления философской науки, концентрированно характеризующего современные натуралистические интерпретации одно-

го из «основных вопросов философии», тысячелетиями обсуждавшегося при решении проблемы отношений идеи/материи, духа/тела, психики/мозга. Сегодня используется новая терминология, например, субъективная реальность/нейтральная активность, когниция/реализация. Однако суть проблемы инвариантна относительно лексических значений. Натуралистический тренд обусловлен достигнутым уровнем развития наук о мозге, когнитивных наук, информатики, искусственного интеллекта. Нейрофилософия, во-первых, это философско-методологическая рефлексия над фундаментальными основами нейронаук, т.е. своеобразная прикладная философия, подобная "философии математики", "философии физики", "философии искусственного интеллекта". Во-вторых, нейрофилософия - это систематическая форма изучения мировоззренческих аспектов, опирающаяся как на категориальные знания о нейрофизиологических основах психических явлений, так и на компьютерные методы имитации, моделирования, репродуцирования мозговой, психической и социальной активности.

Ю.Ю.Петрунин обозначил третий аспект трактовки понятия нейрофилософии. В мировой философской литературе термин закреплен за частным вариантом философии сознания - элиминативным материализмом, предложенным П.С.Черчленд. Однако в рамках нейрофилософии могут успешно развиваться онтологические, гносеологические и методологические модели, весьма далекие от этого узкого направления.

Е.А.Янковская отметила: узость предмета нейрофилософии как частной теории сознания не позволяет использовать эвристический потенциал моделирования мозговой активности ни в мировоззренческих ориентирах, ни в методологии естественных, технических, гуманитарных и социальных наук. Следует расширить семантику термина, придать нейрофилософским исследованиям междисциплинарный статус.

Т.В.Черниговская отметила важную проблему "времени" как квалиа, т.е. параметра качества субъективной реальности, в связи с естественнонаучными способами изучения "биологических часов", которые продуцируются мозгом. Возникает ряд онтологических проблем "бытия времени" и эпистемологических проблем удостоверения его реальности: "время", в котором мы существуем, продуцирует мозг, и, в то же время, "время" – одно из качеств субъективной реальности - квалиа. Декан композиторского факультета, проф. Московской гос.консерватории им.П.И.Чайковского А.А.Кобляков доказывал, что найденные в музыкальном творчестве «субъектные» модели, основанные на трансмер-

ных отношениях, распространяются на сложную структуру бинарных оппозиций, лежащих в основе языка и мышления. И.А.Бугаков поддержал классическую философско-методологическую парадигму: главная задача нейрофилософии – поиск категориальных принципов и законов мозговой деятельности, например, таких как динамическое восприятие, пластичность, предвидение, фрактальность. А.В.Савельев доказывал большую степень общности и, следовательно, перспективы философского исследования «нейрософии». Н.С.Косицын рассмотрел результаты авторских микроэлектронных исследований дендротубулярного аппарата нейронов и заключил, что одна из главных задач в изучении мозга – исследование субстрата нейронных клеток, а не только связей между ними. Г.И.Шульгина изучала возможности имитации работы головного мозга с учетом спецификации процессов торможения. Д.Н.Кавтарадзе констатировал, что представления биологов об архитектонике мозга, структурно-функциональных единицах, нейронных ансамблях, пластичности нейрональных связей можно успешно интерпретировать в контексте гуманитарных наук. О.Е.Баксанский отмечал, что в условиях НБИКС-конвергенции "нейрофилософия" становится интегративной методологией конвергентных технологий. В.И.Михалюк и С.В.Колосова выявили типовые инварианты информационной нейросистемы. В совместном докладе С.В.Петухова, В.И.Свирина, И.В.Степаняна и Л.В.Хазиной сообщалось о моделировании молекулярно-генетической системы посредством "тензорчисел», которые обобщают комплексные и гиперкомплексные числа. В докладе Г.С.Воронкова и А.В.Чечкина отмечались две парадигмы нейросетевого моделирования: мозг из поступающей информации создает модель мира и мозг сам является нейронной моделью мира. В.И.Бодякин утверждал гипотезу функционального единства физического и когнитивного миров. В.В.Колушов и А.В.Савельев отмечали приоритет отечественных разработок в освоении проблемы «Нейрокомпьютеры и общество». С.В.Лещёв сообщал о принципах построения «нейронета». Л.Б.Логунова выдвинула т.н. "эпидемиологическую гипотезу" происхождения человека, сознания, социальности и представила ряд футурологических сценариев. В докладе В.Д.Цыганкова, И.В.Степаняна и С.К.Шарифова демонстрировался нейрокомпьютер «Эмбрион» как макроквантовый когерентный нейрокомпьютер не-фон-неймановской архитектуры. Д.С.Андреюк усмотрел аналогии между эволюционными подходами в биологии и экономике и предложил эволюционную схему изменения паттернов поведения людей в экономически-значимых группах. О.Б.Сладкова предложила мониторинг социокультурного пространства осуще-

ствлять средствами нейротехники. С.Ю.Карпук обсуждала проект нейрокомпьютерного моделирования «смысла» метафор.

Через год, 17 апреля 2015 года, там же, на XIII-ой всероссийской научной конференции «Нейрокомпьютеры и их применение» В.Г.Кузнецов представил доклад "История науки и обоснование нейрофилософии", в котором предлагалось признать за исходную позицию нейрофилософии минимальный проект (слабую версию) обоснования, опирающийся на переносе "по аналогии" на нейропсихологию принципов естественных наук (физики, химии, астрономии).

Дальнейшие дискуссии не только в рамках работы данной секции, но и в ходе последующих мероприятий продемонстрировали продуктивность понимания нейрофилософии в формате междисциплинарного подхода, а так же ее эвристический потенциал для философии ИИ. Используются категориальные концептуальные средства, обеспечивающие коммуникацию разнопрофильных специалистов, научно-популярный характер представления специфических проблем, формирование методологических ресурсов для нередукционистской идентификации, систематизации, унификации, операционализации, координации и интеграции методов и средств, связанных с проблематикой исследования мозга и реализации интеллектуальных функций.

1 октября 2015 года, там же, в докладе на конференции "От истоков к современности. 130 лет организации психологического общества при Московском университете", В.Г.Кузнецов обозначил взаимосвязь философии, психологии и когнитивной нейронауки в рамках решения проблемы сознания, которая стала доминировать не только в современной философии и психологии, но и в биологии, физиологии, медицине, во всем комплексе современных когнитивных нейронаук.

Проблемы междисциплинарного подхода в нейрофилософии выпукло проявились на Международном междисциплинарном симпозиуме "Актуальные вопросы нейрофилософии", 20-23 июня 2015 года в г. Судак (Крым, Россия). Общим лейтмотивом докладов звучала проблема междисциплинарной коммуникации.

На секции "Методологические вопросы нейрофилософии" изучалось соотношение нейрофилософии и нейронаук, нейрофилософские онтологические модели, мета-методологическая рефлексия над нейронауками, проблема НБИКС-конвергенции. Доклад Т.В.Черниговской «Конь и трепетная лань: ученый на стыке наук» посвящался проблеме междисциплинарных связей между гуманитарными и естественнонаучными аспектами изучения психических и

лингвистических феноменов. Предполагается, что в рамках нейрофилософии формируется структурированный терминологический аппарат, который учитывает специфику обеих областей и одновременно структурирует общее исследовательское поле. В.И.Аршинов утверждал, что различные исследователи в области нейронаук постоянно сталкиваются с явлением сложности, например, со сложностью нейронных сетей, поэтому нужна специфическая онтология сложности. В докладах Н.М.Смирновой, А.Ю.Алексеева, Е.А.Янковской, А.В.Чусова, А.В.Савельева, Т.В.Каштановой исследовались два типа проблематики. Прежде всего, для чего нужна нейрофилософия нейронаукам и нужна ли вообще? Во-вторых, рассматривались различные философско-методологические парадигмы, например, коннекционизм и феноменология, интерпретируемые в рамках нейрофилософского дискурса.

Секция «Естественнонаучные вопросы нейрофилософии» рассматривала нейробиологический фундамент психических феноменов. Изучались проблемы редукции ментальных феноменов к нейрофизиологическим процессам, интеграция естественнонаучных и гуманитарных исследований ментальных феноменов, стратегии исследования сознания, базирующиеся на результатах нейроэкспериментов, нейронаучные парадигмы, интеграция философских понятий в естественнонаучный дискурс при исследовании психических процессов, нейрофизиологическая основа специфических психических феноменов. Е.Н.Шульга подчеркивала невозможность полной редукции феноменов, описываемых «гуманитарными» средствами, к области нейрофизиологии. Тоус Джозеф Мария (Барселона, Испания) в докладе «На пути к теории проприоцептивного сознания» демонстрировал специфическую парадигму исследования сознания, исходя из которой ментальные феномены следует рассматривать как продолжение телесного, прежде всего, кинестетического опыта, который формирует в мозге "карты тела", задающие активность индивида. Е.В.Лосева отмечала роль нейроэкспериментов как базиса нейрофилософии: они позволяют устранить из философских исследований непродуктивную метафизику. Н.С.Косицын показал, каким образом общефилософские понятия, такие, как «сложность» и «структурность» способны интегрироваться в методологическую основу нейронаучных исследований. В докладах Г.С.Воронкова, Н.И.Шульгиной, к.ф.н. Ю.С.Моркиной, д.м.н. В.В.Гаврилова рассматривались различные философские проблемы от понимания языка животных до вопросов о нейральном базисе творческой деятельности.

Секция «Технические и инженерные вопросы нейрофилософии» была посвящена философско-методологическому анализу проектов, связанных с при-

менением результатов нейронаучных исследований к созданию технических систем. В докладе И.В.Степаняна подчеркивалась необходимость единой инженерно-методологической платформы. С.В.Лещев, А.В.Сидоренко, Ю.В.Гуров демонстрировали возможности анализа нейрокомпьютерных проектов.

Секция «Социально-гуманитарные вопросы нейрофилософии» затрагивала проблемы из области теории познания, онтологии, логики, методологии науки. В докладе А.А.Коблякова «Нейрофилософия и новые трансмерные отношения» «субъектные» трансмерные модели, обнаруженные в музыкальном творчестве, интерпретировались результатами магнитно-резонансной томографии. В докладе В.Л.Васюкова «Квантовая логика в контексте нейронаук» анализировалась возможность применения специфического логического аппарата, используемого для описания феноменов квантового микромира, для описания некоторых нейрофизиологических процессов. А.А.Юрасов в работе «Нейрокомпьютерная метафора ментального времени» рассмотрел онтологическую проблематику субъективного времени и сделал вывод о том, что модели времени необходимо строить с учетом феноменологии от первого лица, в рамках которой проявляется качество субъективной реальности, но с учетом соотношения с ее нейродинамическими коррелятами.

На секции "Нейрометафизика" раскрывались особенности "чистых" рассуждений, основанных на мысленных экспериментах из области нейронаук. А.Кузнецов исследовал теорию доступного сознания Н.Блока. А.П.Беседин раскрывал главную роль когнитивной нейронауки в проекте Д.Чалмерса по поводу единой науки о сознании.

Участники круглого стола «Роль нейрофилософии в конвергентной методологии НБИКС» подчеркивали, что нейрофилософские идеи являются концептуальной инфраструктурой этой методологии - их можно обнаружить во многих ее теоретических составляющих.

В ходе проведения симпозиума произошло структурирование исследовательского поля "нейрофилософии" в рамках пяти групп научных проблем: 1) что такое "нейрофилософия", каковы ее специфические характеристики и направления изучения; 2) традиционная психофизиологическая проблема: каким образом соотносятся (нейро)физиологические и психические процессы и как на базе материального субстрата нервной системы возникают ментальные феномены; 3) методологические вопросы применения метанаучных и логических категорий к анализу концепций, принципов и методов нейронаук; 4) методологические особенности применения результатов нейронаучных исследований в со-

циально-гуманитарных областях; 5) каким образом возможна интеграция разрозненных нейронаучных исследований?

Представленные выше мероприятия - круглые столы, конференции, симпозиумы - способствуют оперативному и целостному охвату разнообразных аспектов предметного поля "нейрофилософии". Однако детальная проработка важных проблем требует иной научной организации. Это осуществлялось на заседаниях постоянно действующего международного междисциплинарного семинара "Нейрофилософия". Раскроем проблематику нескольких заседаний.

На первом заседании, 12 ноября 2014 г., обсуждался вопрос: **что такое нейронаука?** А.Ю.Алексеев предложил формальный критерий нейронауки: науке приписывается префикс "нейро-", если в ее инструментарии преобладают коннекционистские методы исследования. Возможна слабая корреляция метода и предмета. Так, если мозговая активность изучается бихевиористскими методами и в результатах отсутствуют теоретические обобщения междисциплинарного статуса, то это наука "о нейро-", но не нейронаука. Если же, например, "социальное" представимо в формате нейросетевой спецификации, то это - нормальная нейронаука, в данном случае, нейросоциология. Возникает вопрос о спецификации этих методов. Если апеллировать к принципу работы современного нейрокомпьютера, то вопрос не решается. Этот принцип базируется на традиционной теории формальных нейронов, а она позволяет лишь имитировать коннекционистские процессы средствами символической парадигмы программирования. Т.е. "коннекционизм" редуцируется к "символизму". Однако имеется и нередуктивный способ спецификации. Это демонстрировалось в докладе А.Ю.Алексеева "Машина Корсакова (1832 г.) как формальное определение коннекционистского алгоритма". Содокладчиком-оппонентом выступил А.С.Михайлов. Оба независимо друг от друга в 2007-2011 гг. занимались переводом с французского и интерпретацией идей статьи С.Н.Корсакова "Karsakof S. Aperçu d'un procédé nouveau d'investigation au moyen de machines à comparer les idées. – St. Petersburg, 1832. 22 p., 2 pl. ("Очерк о новом способе исследования посредством машин для сравнения идей").

А.С.Михайлов в докладе "Теоретико-множественная интерпретация и визуализация работы машин Корсакова" доказывал, что вычислимость в смысле Корсакова полностью находится в границах основ дискретной математики. А.Ю.Алексеев для выявления специфики коннекционистского метода сравнивал машину Корсакова с машиной Тьюринга. Коннекционистская интерпретация следует из буквального толкования оригинальной статьи. Программирова-

ние машины Корсакова осуществляется не посредством некоторого языка программирования, а за счет соединения (коннекции) признаков предметов. Формирование коннекций в общем случае происходит субсимвольным, «дословесным» способом. Машину Корсакова нецелесообразно редуцировать к теоретико-множественному базису, т.е. к символьной интерпретации, так как элиминируется специфика нейрокомпьютинга. Разработка "настоящего" нейрокомпьютера, а не его эмуляторов, - это дело будущей технологии. Тем не менее, вопрос о "нормальной" нейронауке получает некоторое теоретическое разрешение.

Второе заседание состоялось 24 декабря 2014 г. в формате круглого на тему **"Что такое нейрофилософия?"**. Рассматривались вопросы истории возникновения понятия, соотношения нейрофилософии и методологии нейронауки, мировоззренческие аспекты нейрофилософии. В.Г.Кузнецов предложил прекратить споры об авторстве термина «нейрофилософия» и проследил историю этого направления с учетом того, что «философия нейронауки» и «нейрофилософия» различаются по предмету исследований. Период с 1960-х годов до середины 1980-х годов - это становление когнитивной нейронауки и формирование идейных предпосылок нейрофилософии. Следующий период до настоящего времени характеризуется дискуссиями, которые инициированы многочисленными публикациями супружеской пары Патриции и Пола Чёрчлендов. Их подход характеризует взаимоотношение философии, нейронауки и "народной психологии", из которой надо построить "научную психологию" путем использования аппарата биологических и искусственных нейронов. На современную "нейрофилософию" оказывают влияние открытия нейронаук, дискуссии между сторонниками физикалистского и информационного подходов в психосемантике относительно понимания когнитивной репрезентации, обсуждение каузальных и телеологических теорий ментального содержания и мн.др. Нейрофилософия нужна философам, которые начинают понимать бесперспективность проблематики сознания, не учитывающей комплекса нейронаук, так и нейрочеловеку нужна для обретения концептуального фундамента своим исследованиям.

В самом широком смысле термин «нейрофилософия» используется для обозначения специфического, совершенно самостоятельного раздела философской науки, являющегося междисциплинарным объединением комплекса направлений. Это организованная на общем философском уровне система, объединяющая подходы к изучению объектов самой различной природы, которые подобны мозгу в поведенческом, функциональном, структурном, семантиче-

ском и др. отношениях. Поэтому всегда следует выбирать контекст использования слова "нейрофилософия".

Продолжая дискуссию, К.В.Анохин отметил два аспекта значения слова "нейрофилософия". Первая трактовка - нейрофилософия как дисциплина. Здесь в научном плане обсуждать нечего, так как это - предмет для администрации вузов. Вторая трактовка - нейрофилософия как проблема. Здесь дискуссии крайне необходимы. "Нейрофилософия" - это ярлык, обозначающий спектр глубочайших проблем. Слово было удачно придумано и выпущено в мир Патрицией Черчленд. Однако не надо отождествлять "нейрофилософию" с тем вариантом философии сознания, который она предлагала. Например, генетика не отождествляется, более того, дискредитируется тем вариантом генетики, который предлагал У.Бэтсон, автор этого термина. Нейрофилософия как проблема достойна конструктивного обсуждения. Ярлык обсуждать не надо.

Очевидной проблемой нейрофилософии является соотношение мозга и сознания. В решении этой проблемы философские исследования переходят в область естественнонаучных исследований. Мозг не может быть понят без проблемы субъективного опыта и сознания. Но и сознание не может быть понято без изучения мозга. Однако предмет исследований шире, ведь в данной трактовке из субъективного опыта исключаются бессознательные когнитивные процессы, во многом определяющие поведение человека и животных.

Каков практический статус нейрофилософии, нужна ли она нейронауке? Если в высказывании А.Эйнштейна по поводу затруднений, которые физик испытывает в настоящее время, и которые вынуждают обращаться к философии чаще, чем об этом мог бы задуматься физик предыдущих поколений, заменить словом "физик" словом "нейроученый", то мы получим описание сегодняшней ситуации.

На круглом столе было много выступлений. Например, С.А.Бушев отметил параллели между биосемиотикой и нейросемиотикой. Н.И.Шульгина пыталась применить диалектико-материалистическую метафизику. Не без интереса участники заслушали высказывание Ю.И.Александрова: в "обычной" нейронаучной среде у термина «нейрофилософия» явно отрицательная коннотация. Нейрофилософией называют что-то, очень удаленное от эксперимента и факта и, в каком-то смысле, даже противоположное "настоящей науке". В отношении работ ряда нейроученых часто употребляется, как отрицательная характеристика, высказывание "все это - нейрофилософия". Намерение переделать такое мнение смелое, но не безнадежное.

В целом, дискуссия показала, что вопрос определения нейрофилософии можно соразмерить с одним из основных вопросов общей философии: "Что такое философия?". Нейрофилософ вынужден постоянно "кружить" вокруг этого вопроса, пытаясь найти ответ. А так как нет предельной определенности, вопрос становится философской проблемой.

Третье заседание семинара, которое состоялось 12 февраля 2015 г., обозначило важную методологическую функцию нейрофилософии - критическую функцию. С докладом на тему "Нейрокомпьютинг между наукой и лженаукой" выступил Ю.Ю.Петрунин. В докладе четко обозначена проблема применимости методов технических наук (нейрокомпьютинга) в социокультурной и гуманитарной сферах.

Содокладчик В.А.Кувакин (член Комиссии РАН по борьбе с лженаукой и фальсификацией научных исследований), детально охарактеризовал "лженауку" и призвал к свободной и независимой научной экспертизе, пример которой продемонстрирован докладчиком при анализе заявлений упомянутой в предыдущем пп. "Пермской школы нейрокомпьютеров" Л.Н.Ясницкого.

Противоположная к предыдущей, проблема применения методов гуманитарных наук в естественнонаучной сфере (нейронауках) прозвучала в докладе В.Г.Кузнецова "Герменевтика текста и возможность ее экстраполяции на предметную область когнитивистики" 26 февраля 2015 г. в Институте философии РАН. Рассматривалась классическая герменевтическая методология, ее универсальный статус для изучения проблемы интерпретации и понимания в различных сферах, включая естественнонаучные. Изучалась классическая герменевтика Шлейермахера, критика ее с позиции фундаментальной онтологии Хайдеггера и философской герменевтики Гадамера, критика этой критики и возврат к классической форме в контексте когнитивистики, в частности, в концепции Д.Деннета, который использует "интерпретацию" в сфере когнитивной психологии человека и в области производства артефактов. Расширительная, по отношению к классике, трактовка герменевтики возможна в этом варианте. Однако, герменевтический метод надо применять с осторожностью, вряд ли следует апеллировать к "меметике" Докинза, как это сделал Д.Деннет.

В докладе Н.И.Чуприковой "Дифференциально-интеграционный принцип развития в философии, психологии, биологии и нейронауке", который обсуждался 12 марта 2015 г., прозвучала проблема применения универсальных категорий в нейронауках. Философский категориальный аппарат позволяет подвести неупорядоченный хаос изучаемых явлений под единую систематизирован-

ную онтологию. Это необходимо для наведения порядка в сфере психологии, а также в исследованиях по проблеме развития, как исторического, так и психологического. Проблема заключается в отсутствии рационального определения категории «развитие». Поэтому для интегральной парадигмы нужны модели рациональной философии, позволяющие сформулировать всеобщий универсальный закон развития. Этот закон всегда предполагает переход от исходной неразвитой целостности к дифференцированным формам и проявляется в эмбриональном, психологическом, когнитивном и других видах развития, что подтверждается многочисленными эмпирическими данными. Оппонент А.В.Чусов усмотрел гипостазирование универсального закона развития. Сам по себе никакой закон непосредственно наблюдать нельзя. Следует, скорее, говорить о гносеологическом и методологическом значении закона, избегая его онтологизации.

Проблема "творчество и нейронаука" и, соответственно, относительно новый, нейрофилософский подход к анализу вечной проблемы творчества отчетливо прозвучали в пленарных докладах Всероссийской междисциплинарной конференции "Философия творчества", 8-9 апреля, ИФ РАН (руководитель - Н.М.Смирнова).

С позиций современного уровня нейробиологических исследований Т.В.Черниговская аргументировано критиковала устаревшие представления о (право-лево)полушарной асимметрии мозга, препятствующие исследованию творчества как целостного феномена, и наметила контуры междисциплинарного анализа "творчества" в контексте конвергенции естественных и социально-гуманитарных наук.

К.В.Анохин представил результаты новых исследований сложнейших когнитивных структур мозга, «ответственных» за память и творческое мышление. Проблема творчества рассматривалась в контексте оригинальной конструкции когнитом/коннектом. Когнитом - это система субъективного опыта, коннектом - архитектура нейронной сети организма. Совместное их рассмотрение - интереснейшая интерпретация "основного вопроса нейрофилософии". Формально оба члена диспозиции представимы гетерархическими гиперсетевыми структурами.

В.И.Аршинов интерпретировал концепцию К.В.Анохина с позиции "сложности" и, переключаясь с идеями Т.В.Черниговской, изучил проблему в контексте (пост)неклассической парадигмы сложности, в рамках которой реализуется процесс конвергенции знаний.

В.Г.Кузнецов в докладе "Синестезия: восприятие природы и связь с творчеством» раскрыл суть феномена цветозвукового "смещения чувств" и обосновал вывод о невозможности искусственного образования синестезии: человек не может изменить действие законов природы, даже если будет воздействовать на нейрокогнитивные механизмы чувственного восприятия.

Почему теме нейрофилософии диссертант уделяет столь большое внимание в связи с междисциплинарными исследованиями ИИ? Потому что тематика «искусственного мозга» и тематика «искусственного интеллекта» тесно коррелируют друг с другом. Показательно то, что все (!) докладчики упомянутого выше круглого стола "Философские исследования искусственного интеллекта: итоги и перспективы» (26 марта 2015 г., ИФ РАН) касались темы "Нейрофилософия и искусственный интеллект". Так, доклад на тему "Сознание, мозг. искусственный интеллект (10 лет спустя)" представил Д.И.Дубровский. Он напомнил, что на такую тему выступал десять лет назад на первом теоретическом семинаре НСММИ РАН. С тех пор наука сделала значительный шаг вперед, в силу чего эта тема стала еще более актуальной, чем прежде. Возросшая взаимозависимость между исследованиями мозга, сознания и разработками в области искусственного интеллекта ярко выражена в процессах конвергентного развития НБИКС. ставшего главным трендом развития науки и технологий в начале XXI века. Докладчик подробно остановился на двух фундаментальных взаимосвязанных направлениях: на исследовании "зеркальных" нейронных систем в головном мозге и на нейрокриптологических исследованиях ("чтение мозга").

В докладе "Искусственные общества" В.Л.Макаров, подводя десятилетние итоги работы, руководимой им лабораторией «Искусственных обществ», отметил то, что в последние годы для имитации социокультурных процессов, как реальных, так и виртуальных (например, для изучения эволюции трехполых существ), сотрудниками лаборатории применяется методология многоагентного программирования. Это - символная парадигма, оппонирующая исходной, нейрокомпьютерной парадигме. Однако «многоагентность» базируется на нейрокомпьютеринге, на методах параллельной обработки информации: в искусственном обществе, как и в естественном, агенты действуют совместно.

Т.В.Черниговская в докладе «Экзистенциальные последствия наступления "жидкого мира" компьютерной среды» обрисовала алармистские перспективы развития нейронаук. Изменяется онтологическая проблематика вследствие метаморфоз пространства и времени, в которых мы стали жить. Происходит модификация всех фундаментальных отношений, включая праксеологические и

антропологические. Поэтому необходимо живое участие философа в нейронауке и нейротехнологии.

В.Г.Редько в докладе «Моделирование когнитивной эволюции - перспективное направление междисциплинарных исследований» выделил роль методов и средств нейроинформатики в понимании процесса развития биологических и эпистемологических механизмов животных, человека, а также в построении подобных механизмов в искусственных системах.

В.Г.Кузнецов в докладе «Когнитивистская гипотеза и искусственный интеллект (истоки и взаимодействие)» сомневается в общепринятом мнении о востребованности наработок современной аналитической философии сознания для методологии нейронауки. Философия сознания - это постоянно растущий, невероятно пестрый конгломерат из мысленных экспериментов, которые все далее и далее двигают важнейшее философское направление в сторону метафизики. Нейрофилософия, в силу собственной натуралистической ориентации, позволяет этот порочный тренд исправить. Именно эти нейрофилософские идеи следует использовать в компьютерных моделях ИИ.

Ю.Ю.Петрунин в докладе «Методологические проблемы управления знаниями» обращает внимание на неклассические вопросы, обусловленные появлением новых социогуманитарных наук, таких как менеджмент, маркетинг, публичное управление. Требуется изучать нерациональные аспекты принятия решений, связанные не со знаниями, а с убеждениями, верованиями, ценностями. Нейрокомпьютерная парадигма снижает остроту проблемы моделирования трудно формализуемых составляющих управления.

А.Ю.Алексеев в докладе «Роль комплексного теста Тьюринга в современной философии искусственного интеллекта» раскрыл ряд нейрокомпьютерных идей, которые использовались в концепции тьюринговой игры в имитацию интеллектуального поведения. Как в оригинальном тесте, так и его различных модификациях, в частности, в тесте Черчлендов, идеи параллельного процессирования подменяют идеи нейрокомпьютинга. Перспективна концепция машины Корсакова-Тьюринга, объединяющая принципы нейро- и лингвокомпьютинга. Это позволяет формировать богатое репрезентативное содержание средствами механизма Корсакова в контексте автоматных переходов механизма Тьюринга.

Наибольший вклад в развитие нейрофилософии как методологии нейрокомпьютерной науки и как части методологии ИИ внесла дискуссия по поводу следующих трех докладов.

Б.М.Величковский в докладе «Философия, искусственный интеллект, когнитивные нейронауки – взаимодействие необходимо» обозначил связи, необходимые для развития комплекса НБИКС-технологий. Сегодня «чистых» когнитивных наук и нейронаук не существует. Методология искусственного интеллекта базируется как на исследованиях в области когнитивной психологии, компьютерной лингвистики, так и на исследованиях в области нейрофизиологии.

Тему продолжил О.П.Кузнецов в докладе «Искусственный интеллект и когнитивные науки: методологические проблемы». Он высоко оценил роль исследований мозга и вклад нейронаук в развитие искусственного интеллекта. Для обоснования предложил интересные методологические исследования концепции К.В.Анохина "когнитом/коннектом".

С резкой критикой этих двух положений выступил В.К.Финн. В докладе «Искусственный интеллект как научное направление и проблемы точной эпистемологии», опираясь на "логические каноны научной рациональности", он продемонстрировал инвариант интеллектуальной системы, которая состоит из интерфейса пользователя, базы знаний и решателя задач, включающего средства порождения и проверки гипотез о предметной области. Зачем нужны исследования мозга? Ведь с логико-эпистемологической позиции интеллектуальная деятельность не связана с мозговой активностью.

Таким образом, на круглом столе впервые в отечественной науке четко обозначилась оппозиция символьного и коннекционистского подходов.

Значительным событием для участников 12-го заседания семинара с общей темой «Нейрофилософия: решение основных философских проблем» явился доклад Патриции Смит Чёрчленд на тему "The Brains Behind Morality", который состоялся 18 июня 2015 года на философском факультете МГУ имени М.В.Ломоносова. Автор термина "нейрофилософия" коснулась темы, к которой стремится любое развитое философское направление, темы этики (возможно, полагая, что нейрофилософия достигла этого уровня). Последние книги П.Черчленд «Braintrust» (2011 г.) и «Touching the Nerve» (2013 г.) посвящены, преимущественно, вопросам этики и, в частности, происхождению социальных норм. П.Черчленд принадлежит к числу представителей натуралистического подхода к решению проблемы соотношения психики и мозга и предлагает связать поставленные в традиционной философии проблемы сознания с нейробиологическими ответами на них. П.Черчленд является лидером в современной нейроэтике. Однако в решении вопроса А.Тьюринга «Может ли машина мыс-

литель» наблюдаются пробелы в адекватной трактовке нейрокомпьютинга (см. далее критику «теста Черчлендов» в пп.3.3.3).

Не менее важным событием для участников семинара явился доклад основателя отечественной нейрофилософии - Д.И.Дубровского. Начиная с его статей 60-х годов и докторской диссертации (1968 г.). в отечественной философской науке поднимаются и решаются те же самые проблемы, которые сегодня в мировой науке поднимаются и решаются под брендом "NeuroPhilosophy" и «Brain-Reading»; эти проблемы широко и основательно разрабатывались в его монографии «Психические явления и мозг. Философский анализ проблемы в связи с некоторыми актуальными вопросами нейрофизиологии, психологии и кибернетики» (М.: Наука, 1971) и многих последующих его публикациях.

Д.И.Дубровский выступил на семинаре 29 сентября 2015 г. с докладом "Проблема сознания и нейрофилософия". Содержание доклада представлено в статье настоящего выпуска "Философских наук". Поэтому обозначим только ряд положений дискуссии.

В.Г.Кузнецов подчеркнул связь концепции Д.И.Дубровского с герменевтической методологией. Семантическое содержание явления субъективной реальности предполагает расшифровку мозговых кодов. Декодирование - это преобразование одного кода в другой, "непонятного" в "понятный", согласно Д.И.Дубровскому. В герменевтике декодирование есть средство достижения понимания. Герменевтические приемы не являются искусственно, сознательно привнесенными в психику человека. Это естественный, интересубъективный аппарат нашего сознания. Нейрочтение аналогично простому письменному чтению и его герменевтическому истолкованию. Существенно, что эта проблематика распространяется не только на явления психики, квалиа, но и на деятельность мышления в целом.

А.Ю.Алексеев предложил назвать "кодом Дубровского" взаимосвязь когнитивного феномена с его нейральными коррелятами и со способом кодирования информационной зависимости между ними. Эта внешне простая концептуальная конструкция, тщательно прорабатываемая Д.И.Дубровским на протяжении пятидесяти лет, в нейрофилософии играет роль, аналогичную той, которую играет тест Тьюринга в философии искусственного интеллекта. Е.А.Никитина обратила внимание на то, что информационный подход, предложенный докладчиком, имеет статус общенаучного методологического подхода и позволяет выработать общий язык, на котором обсуждают проблему «сознание-мозг» в философии, психологии, психофизиологии, нейронауке. А.Н.Кочергин и

А.В.Чусов отметили необходимость основательного критического обсуждения концепции Д.И.Дубровского. О.Э.Петруня предложил от объяснительной схемы перейти к прогностической схеме. В.Д.Цыганков предложил включить в теорию Д.И.Дубровского рассмотрение квантовых концепций нейрального кодирования. Э.А.Дейнека раскрыла специфику французской нейрофилософии, ее отличия от англоязычного варианта и показала, что информационный подход к сознанию, предложенный докладчиком 50 лет назад, правомочно считать концептуальным основанием самостоятельной отечественной (русскоязычной) нейрофилософии.

Наш краткий обзор становления в отечественной философии интересного направления "нейрофилософии" целесообразно завершить перечислением вопросов, которые, по сути, поставил Д.И.Дубровский в своем докладе.

Первостепенные задачи нейрофилософии:

1) методологическое обоснование и оценка способов формирования инвариантов явлений субъективной реальности как достаточно определенных объектов нейронаучного исследования и экспериментальных результатов, а так же, что особенно важно, их корреляции с мозговыми процессами;

2) на основе результатов п.1, исследование проблемы Я, произвольного действия, феномена свободы воли;

3) включение результатов п.1,2 в методологию конвергентного развития НБИКС.

Все эти задачи являются и задачами развития методологии ИИ. Это было заявлено Д.И.Дубровским и его последователями (к числу которых относится и диссертант) на самом начальном этапе становления философии ИИ. «Нейрофилософия» является брендом развития различных, порой, противоречивых направлений в сфере нейронаук. Так же и «Искусственный интеллект» выступает символом развития когнитивных и компьютерных наук. Он не претендует на дисциплину, но на систематическую постановку проблем для этих сфер научной деятельности.

Проблема творчества в компьютерном мире: роль методологии искусственного интеллекта в решении базовых гуманитарных проблем

Эта тематика обсуждалась на первой Всероссийской междисциплинарной конференции «Философия творчества». Конференция состоялась 8-9 апреля 2015 г. на базе Института философии РАН (ИФ РАН). Организаторами конференции явились сектор философских проблем творчества ИФ РАН и Научный совет РАН по методологии искусственного интеллекта (НСМИИ РАН).

Доклады, дискуссии и обсуждения на «Круглом столе» продемонстрировали философскую и научную неисчерпаемость проблематики теоретического анализа этого самого таинственного проявления человеческого духа.

Тематическая широта конференции - от логической составляющей творческой деятельности до сугубо прикладных аспектов разработки психологических методик анализа ее эффективности – привлекла к участию в конференции не только профессиональных философов, но и психологов, биологов, историков, культурологов, музыковедов, архитекторов, музейных работников, широкий круг специалистов в области общественных, естественных и компьютерных наук. Философский анализ творчества стал не только концептуальной основой междисциплинарного синтеза различных отраслей естественно-научного и социально-гуманитарного знаний, но и когнитивной матрицей исследования его сугубо прикладных аспектов.

Значительное место в работе конференции заняли доклады, сообщения, свободные дискуссии и обмен мнениями в рамках Круглого стола «Проблема творчества в компьютерном мире», организованного НСММИ РАН. Руководителями явились сопредседатель НСММИ РАН, заведующий сектором теории познания ИФ РАН, академик РАН *В.А.Лекторский*, к.ф.н. *А.Ю.Алексеев* и профессор кафедры стратегических коммуникаций факультета государственного управления МГУ им.М.В.Ломоносова, д.ф.н., к.т.н., проф. *Ю.Ю.Петрунин*. Последний открыл заседание сообщением «Компьютер и творчество: мифологемы искусственного интеллекта, вечного двигателя и философского камня». Он выявил, по меньшей мере, три взаимосвязанных значения этого термина, показав, что искусственный интеллект как антропологическая идея является мифологемой, отражающей важные черты современной техногенной культуры. *А.Ю.Алексеев* в докладе «Феноменология творчества в исследованиях искусственного интеллекта» обосновал суждение о том, что предметом исследований искусственного интеллекта «искусственный интеллект» никогда и не был. Это показывают исследования работ А.М.Тьюринга, в рефлексии над которыми, по мнению докладчика, проблематика творчества в компьютерном мире интеллектуальных технологий представлена наиболее рельефно, нежели чем тематика интеллекта. Докладчик также отметил, что исследования в области искусственного интеллекта базируются на изучении компьютерных способов имитации, моделирования и репродуцирования творчества и заключил, что креативный нейрокомпьютинг обеспечивает значительно большие возможности имитации продуктивной творческой деятельности в сравнении с лингвокомпьютингом.

Совместный доклад *С.В.Петухова* и *И.В.Степаняна* «Фибоначчи - ступенчатые генетические строи и музыкальная терапия» отмечен поиском сопряжения феноменологии генетического кодирования, формализмов инженерной теории помехоустойчивого кодирования и музыкальных пропорций, что, по мнению докладчиков, в перспективе открывает возможности услышать «мелодию» геномов животных и человека. Докладчики отметили, что достижения молекулярной генетики и биоинформатики привели к новому пониманию самой жизни: «Жизнь есть партнерство между генами и математикой»: каждый организм представляет собой алгоритмическую машину многоканального помехоустойчивого кодирования. В попытках найти ответ на вопрос, какая математика нужна для создания помехоустойчивой системы генетического кодирования, авторы проанализировали эвристические возможности формализмов инженерной теории помехоустойчивого кодирования, базирующихся на использовании математических матриц: проекционным операторам, задаваемым квадратными матрицами; матрицами Адамара; гиперкомплексными числовыми системами и др. Докладчики приходят к выводу, что в свете данных о музыкальной гармонии в молекулярной системе генетического кода музыка является не только инструментом генерации эмоций, но и – в существенной степени – принципом организации и языком живой материи. В этой связи выдающиеся композиторы предстают как исследователи структурных принципов организации живого вещества на основе развитой интуиции и способности ощущать биоинформатику своего организма.

В сообщении *Т.Б. Кудряшовой* (руководитель Ивановского регионального отделения НСМ ИИ РАН) «О роли памяти в человеко-машинном мире» обоснован вывод о теснейшем взаимодействии памяти и воображения в творческом процессе: условием творчества является интеллектуальное усилие по вспоминанию, активизирующее различные пласты сознания. В результате специфического взаимодействия памяти и воображения изначальный замысел (или «динамическая схема») уточняется и развивается в образ. Так, несмотря на различие языков теоретического описания, наметилось общее проблемное поле феноменологического анализа с нейрофизиологическим изучением роли памяти в творческом процессе, о котором в пленарном докладе сообщил *К.В.Анохин*. Их «общим знаменателем» является следующее: в мире компьютерных технологий нередко звучит предположение, что дальнейшее развитие человеческого интеллекта будет связано с разделением функций между интеллектуальными системами и человеком. Предполагается, что последний будет исполнять функции, требующие нетривиального, творческого решения, а интеллектуальные систе-

мы - более простые алгоритмизированные функции, а также рутинные операции, к числу которых нередко относят запоминание, хранение извлечение информации в нужный момент. Однако, передав функцию запоминания машине, человек рискует вместе с ней потерять и способность к творчеству, поскольку усилия по вспоминанию являются неотъемлемой составляющей творческого процесса. Таким образом, сохранение ведущей роли человека в сложных интеллектуальных технологических системах подразумевает развитие не только воображения, мышления и проч., но и способности к активному вспоминанию как неотъемлемой составляющей творческого действия.

В.Г.Кузнецов в докладе «Социализация личности в контексте традиционной культуры и компьютерного мира» показал, что основные характеристики творческого процесса социализации личности не зависят от того, в традиционной или электронной культуре они реализуются.

Доклад *Т.А.Кувалдиной* (Волгоградский государственный социально-педагогический университет) «Интеграция знаний и междисциплинарные связи: анализ понятий и проблем» посвящен креативному потенциалу интеллектуальных технологий в образовании, обоснованию инновационных учебных курсов междисциплинарного характера. Ею представлены примеры моделирования систем понятий различных учебных дисциплин (информатика, экология, психология, физика), в том числе — в их взаимосвязях, а также — примеры сравнительно-исторического анализа преемственных и перспективных связей между понятиями курсов информатики и экологии. Рассмотрены современные подходы к проблеме визуализации знаний в теории и практике обучения студентов и школьников и сопоставлены отечественные и зарубежные технологии использования интеллект-карт (карт знаний, ментальных карт, опорных схем, аналитических таблиц, разных наглядных средств обобщающего и систематизирующего характера). Визуализация знаний в современном образовании подразумевает требование закладывать в образовательную технологию компьютерные программы, которые моделируют креативные способности человека.

В докладе *О.Э. Петруни* «Творчество и проблема демаркации в компьютерном мире» показано, что проблема демаркации продолжает оставаться важнейшей эпистемологической проблемой, т.к. информационно-коммуникативная революция второй половины XX в. значительно обострила проблему дерационализации знания, особенно в гуманитарной области. Поэтому сегодня налицо опасность возникновения новой мифологии на базе пралогического мышления, способной нанести ущерб не только научной рациональности, но рационально-

сти вообще. Свойственная аналитической философии XX в. интенция к обособлению научных теорий от содержательной семантики философских онтологий не дала желаемого результата. Напротив, обнаружилось, что попытка проведения демаркационной линии отнюдь не означает полного разрыва научного и философского контекстов: философские онтологии обосновывают и конституируют предметные области различных областей наук, формируют их исследовательский горизонт и ядро методологий. *В.Э.Карпов* в докладе «Об одном методе генерации псевдолитературных произведений» раскрыл историю и особенности одного из основных направлений искусственного интеллекта, ориентированного на создание художественных произведений: литературных текстов, музыки, живописи. *А.В.Савельев* в докладе с парадоксальным названием «Прошедшее будущее технического творчества» дал высокую оценку забытого уровня отечественного индустриального и технического творчества. В подтверждение этой оценки докладчик привел примеры-иллюстрации некоторых удивительных технико-технологических прорывов недалёкого, но - как полагает докладчик - усиленно забываемого прошлого, великие изобретения которого многократно превысили самые смелые измышления фантастов.

Проект компьютерного редактора танца как алгоритмически воспроизводимого способа фиксации движений человеческого тела для решения хореографических задач представила *Т.Б.Бадмаева* в докладе «Танец в компьютерном мире». Компьютерный редактор танца реализует кинетографию - классическую хореографическую теорию, разработанную *С.С. Лисициан* в 1930-40 гг. и основанную на корпусе европейской танцевальной письменности, накопленной на протяжении последних 500 лет. Помимо культурологических источников, в этой теории широко используются математические и естественно-научные методы исследования закономерностей движения человека. Проблему соотношения вербального и невербального в творческой деятельности продолжил *А.Н.Лощилин*.

А.В. Чусов выступил с сообщением «О простоте и сложности творчества». Опираясь на ряд положений Платона, Канта, Гегеля, Маркса, Хайдеггера и Бенямина, он рассмотрел проблематику творчества как созидания нового в аспектах простоты и сложности новых объектов. Рассматривая онтологические вопросы творчества в контексте определения мира как «взаимодействия объектов», он поставил вопросы: какое место по отношению к существующему в мире занимает творчество? Как оно проявляется по отношению к уже существующему в мире? Докладчик показал, что представления, бытующие в компьютерном мире, не являются симулякрами – для их реализации необходимо их

культурно-антропологическое «присвоение». Поэтому целесообразно различать не только различные типы онтологий, но и различные типы «присутствия». Присутствие в мире следует, по мнению докладчика, рассматривать как присутствие не объектов, а вхождений объектов в мир. В мире объективно существуют мировые конstellации, когда объект не присутствует, но структура мирового места показывает его отсутствие. Обращаясь к проблематике свободы творчества, докладчик пояснил, что творчество присутствует не только тогда, когда появляется новый тип объекта. Творчество в собственном смысле слова предполагает новый воспроизводящийся уровень взаимодействий, и только в этом случае творчество в онтологическом смысле состоялось.

В докладе *Е.А.Янковской* «Цифровое неравенство и проблема творчества» дано общее определение цифрового неравенства: невозможность или ограниченность доступа к современным информационно-коммуникационным технологиям для различных социальных групп. Негативным следствием цифрового неравенства является социальная депривация, ограничение творческой самореализации. Помимо технических, территориальных и экономических проблем, ограничивающих доступ к информационным технологиям, существует социально-гуманитарная составляющая цифрового неравенства. Она представляет собой отсутствие социально-гуманитарных условий, поддерживаемой социумом мотивации для доступа к ИКТ. Сюда относятся такие составляющие цифрового неравенства, которые основаны на возрастной, гендерной, культурной, национальной, расовой и другой подобной дискриминации. Нужно отметить, что преодоление «внешних» составляющих цифрового неравенства не всегда приводит к преодолению его социально-гуманитарных аспектов, поскольку формальное наличие доступа к ИКТ еще не гарантирует того, что не будут возникать неформальные ограничения. Социально-гуманитарная составляющая цифрового неравенства в еще более значительной степени ограничивает творческую самореализацию, поскольку из-за нее интериоризируются негативные социальные установки, возникают «внутренние» границы для творческих проявлений.

Е.Н.Шульга в сообщении «Естественный/искусственный интеллект и творчество» показала, что взаимоотношение интеллекта и творчества – наиболее сложная проблема, требующая дополнительных эмпирических свидетельств. *М.А.Пронин* в докладе «Акт творчества – онтология события» рассмотрел проблему творчества как феномена виртуальной реальности в междисциплинарном контексте. В докладе «Проблема творчества и латентная реальность» *А.Д.Королев* утверждал, что мы живем в мире неявных знаний, и требуются

особые креативные методы выявления скрытого – «выговаривания умолчаний». В свою очередь, *В.И.Бодякин* предложил количественные параметры определения феномена «творчество» как меры восходящей эволюции самоподдерживающихся процессов.

Молодые, но уже имеющие международную известность компьютерный дизайнер *Т.Пожарев* (Сербия) и музыкант *Н.А.Попов* представили доклад «Особенности компьютерного моделирования смысла» (под научным руководством А.Ю.Алексеева). В докладе рассмотрен проект компьютерной системы, которая проходит комплексный тест Тьюринга на персональную осознанность, творческое мышление и, что немаловажно, на морально-правовое вменение. Знаменательно, что данный проект разработан на основе интерпретации «знаний» экспертной системы в процессе междисциплинарного взаимодействия специалистов и, в отличие от аналогичных зарубежных робототехнических систем, обладающих «квазисознанием», учитывает и «внелогические» (социальные) факторы моделирования смысла (особенности проекта представлены ниже при описании отечественного проекта «искусственной личности»).

В докладе А.Ю. Алексеева «Феноменология творчества в исследованиях искусственного интеллекта» основное внимание уделено анализу различных аспектов комплексного теста А. Тьюринга. Он отметил, что исследования в области искусственного интеллекта базируются на изучении компьютерных способов имитации, моделирования и репродуцирования творчества. При этом креативный нейрокомпьютинг обеспечивает значительно большие возможности имитации продуктивной деятельности в сравнении с лингвокомпьютингом. Так же доказывалось то, что предметом исследования искусственного интеллекта "интеллект" вовсе и не является. Данный предмет - "творчество", т.е. феномены творческой деятельности. В связи с этим, если, допустим, стратегией развития ИИ является разработка моделей имитации, моделирования или репродуцирования "сознания", то тактическая, ближайшая цель - это разработка моделей имитации, моделирования или репродуцирования. Данная идея детально раскрывается в п.п.3.2 данной работы.

Содержательным итогом работы «Круглого стола» стало признание того, что проблема творчества для исследований ИИ представляется более «живой», оперативной, нежели чем проблема «сознание-мозг-искусственный интеллект», изначально определенная как главенствующая на этапе становления НСМИИ РАН. Именно на решение проблем творчества следует направлять философские, научные и инженерные усилия, если мы желаем продвинуться в исследо-

ваниях «естественного» интеллекта и достичь ощутимых успехов в развитии искусственного интеллекта и электронной культуры в целом.

Выводы: философия искусственного интеллекта, трактуемая как сфера междисциплинарных исследований, за десять лет работы НСМИИ РАН шагнула как в сторону естественно-научной проблематики (нейрофилософия), так в сторону гуманитарных исследований (проблема творчества). В связи с этим «искусственный интеллект» представляется как символ, бренд, слово, которое задает проблемные области компьютерной реализации когнитивных феноменов, но не как научная дисциплина, с ее специфическим объектом, предметом, методами и инструментарием. Тем не менее, философию искусственного интеллекта можно и должно рассматривать как научную дисциплину. Такой статус этой философии придает базовый категориальный аппарат – тест Тьюринга, являющийся источником идей по поводу компьютерной реализации интеллекта и других когнитивных феноменов. В следующем параграфе роль теста Тьюринга изучается как с позиции внутреннего оформления этой сферы междисциплинарных знаний, так и с внешней стороны – в контексте электронной культуры.

1.2. Роль теста Тьюринга в современных исследованиях искусственного интеллекта

1.2.1. Становление проблематики теста Тьюринга в отечественных дискуссиях

В данном параграфе раскрываются общие представления о роли теста Тьюринга в аналитической философии искусственного интеллекта (Philosophy of Artificial Intelligence). Для изучения возможности трансляции в отечественную науку данной формы исследований ИИ, диссертант в апреле 2004 г. организовал **молодёжную конференцию «Философия искусственного интеллекта»** среди студентов Московского института электроники и математики. В работе конференции приняли участие видные учёные – Д.И. Дубровский, К.К. Колин, А.С. Нариньяни, Г.П. Путилов. В ходе конференции изучались вопросы: обусловленности функционалистской концепции мышления тестом Тьюринга; связи идей А.М. Тьюринга (основателя философии ИИ и Дж. Маккарти (основателя науки «Artificial Intelligence»); истории развития проблематики теста Тьюринга и его практических приложений.

Также представлены результаты научных мероприятий, посвящённых юбилейным датам – **55-летию теста Тьюринга** (Круглый стол «Философско-методологические проблемы когнитивных и компьютерных наук» на IV Философском конгрессе, 23-28 мая 2005 г., МГУ им. М.В. Ломоносова, МИЭМ, ЦЭМИ РАН, руководители – В.А. Лекторский, В.А. Глазунов, А.Ю. Алексеев) и **60-летию теста Тьюринга** (научная конференция «Тест Тьюринга: философские интерпретации и практические реализации», 28-29 октября 2010 г., философский факультет МГУ им. М.В. Ломоносова, руководители – А.А. Костикова, Н.Ю. Ключева).

Заслугу в становлении проблематики теста Тьюринга в отечественной философии искусственного интеллекта диссертант приписывает в некоторой степени и себе: за год до проведения указанной выше конференции ("Философия искусственного интеллекта", г. Москва, МИЭМ, 2005 г.) он в качестве тренировки организовал студенческую конференцию с тем же названием «Философия искусственного интеллекта». Конференция состоялась там же, 20 мая 2004 г. в Московском государственном институте электроники и математики (МИЭМ). В роли председателя конференции выступил проректор вуза **Г.П. Путилов**, Он предложил «основные вопросы» конференции как основные вопро-

сы философии ИИ: Что такое искусственный разум? Каковы границы расширения его возможностей? Что ждет нас в недалеком будущем в связи с развитием и внедрением современной теории искусственного интеллекта? Каковы дальнейшие перспективы человека?

В работе конференции приняло участие свыше 250 человек, в основном, студенты старших курсов МИЭМ. Было представлено около 80 тезисов докладов, многие из которых опубликованы [Философия ИИ, 2004]. Доклады, в основном, были построены на базе переводов статей англо-американских философов искусственного интеллекта (ИИ). Такой подход к изучению научной проблемы, по сути, автореферативное «списывание», был обусловлен, к сожалению, отсутствием в то время в отечественной науке серьезных наработок в области философии и методологии ИИ, по сути, полным неведением в этой философской проблематике (последняя конференция всесоюзного масштаба по проблеме философии ИИ состоялась свыше 30 лет назад!). С другой стороны, последнее десятилетие отмечалось небывалым всплеском интереса философов (в основном, англо-американской традиции) к проблематике ИИ – за этот период появилось несколько тысяч (!) достаточно крупных публикаций. Поэтому студентов, потрудившихся над докладами нашей конференции, по праву следовало бы считать пионерами, открывающими для нашей страны проблематику философии ИИ – очень важную проблему современной науки, культуры и техники. Достаточно высокий статус конференции был обеспечен благодаря гостям - Д.И. Дубровскому и К.К. Колину. Д.И. Дубровский акцентировал внимание на проблематике сознания и стратегической важности для страны развития философских исследований ИИ. К.К. Колин рассмотрел проблему формирования информационной технологии как самостоятельной науки о методах и средствах создания высокоэффективных информационных технологий (в узком понимании этого термина), определил отличительные признаки высокоэффективных технологий и основные принципы их проектирования, сформулировал общие критерии для оценки эффективности и социальной полезности интеллектуальной информационных технологий. Философия информатики, таким образом, тесно была связана с философией ИИ.

Функционалистская концепция мышления и тест Тьюринга

Особый интерес для становления методологической проблематики теста Тьюринга, представляет работа секции № 1. «**Функционалистская концепция мышления. Тест Тьюринга: pro et contra**». Впервые за несколько десятилетий в нашей стране зазвучала тема, посвящённая тесту Тьюринга (ТТ) (работы

А. Жукова «Базовые положения теста Тьюринга», Д. Комарова «"Может ли машина мыслить"? Полемический стандарт Тьюринга», С. Лизоркина «"Искусственный интеллект" Алана Тьюринга»)¹. Большинство работ студентов посвящено различным модификациям ТТ. Чтобы раскрыть важность этих пионерских студенческих докладов-сообщений следует вспомнить, каким образом в нашей стране протекала дискуссия после публикации в 1960 г. перевода статьи Тьюринга (1950). В отечественной философии и науке вопрос о возможности построения мыслящих машин вызвал ожесточённые дискуссии, в основном, на волне пан-кибернетизма. В 1963 г. академик А.Н.Колмогоров, в недалёком прошлом (до 1957 г.) ярый противник кибернетики как науки, пишет статью «Автоматы и жизнь», где утверждает о теоретической возможности воспроизводства автоматами всех видов человеческой активности и не только интеллектуальных, но и эмоциональных. Лозунгом А.И.Колмогорова стало следующее высказывание: «Всего лишь автомат? Нет, мыслящее существо!». На это Б. Бялык отвечает статьей «Товарищи, вы это серьезно?». В ответ академик С.Л. Соболев пишет работу «Да, это вполне серьезно!»². Завязалась серьезная полемика, в рамках которой обсуждались вопросы возможности/невозможности естественнонаучного определения таких понятий, как воля, мышление, эмоции и др.; противоречия/соответствия социализму кибернетических концепций «думающих» машин; исторического подхода к пониманию машины и техники в целом, согласно которому машины – это продукт общественно-трудовой деятельности человека, они не трудятся, трудится человек посредством её и пр. Возникли этические проблемы контроля над «интеллектуальными» машинами. Вскоре в нашей стране споры утихли. До конференции 2005 г. в отечественной печати работы, посвященные данной теме, встречались крайне редко. В англо-американской философии дискуссии, инициированные проблематикой теста Тьюринга, не угасали. Особо усилились дискуссии в середине 90-х гг. 20 века. Это обусловлено тем, что: 1) развитие компьютерной техники привело к реализации многих предположений Тьюринга; 2) в области методологии появились крупные теоретические наработки в области построения самообучающихся программ, реализации алгоритмов «как бы» волевой мотивации, «эмоциональ-

¹ Данные работы очень небольшие по объёму (2-5 стр.). Они, как правило, представляют попытку вникнуть в содержание работ, переводимых с англ. языка. Конечно, ни о каких «философских зомби» и «мозгах в бочке» студенты в ту пору (2002-2004 гг.) не могли слышать. В некоторых работах, тем не менее, имеется попытка вынести личное мнение по той или иной проблеме. Работы представлены в [Философия ИИ, 2004].

² См. исследования американского историка науки Л.Р.Грэхема: [Грэхэм 1991, С. 266-291]. Оригиналы статей в [Колмогоров, 1964, с.10]. Статьи Бялика Б. и Соболева С. – там же.

ного» и «неформального» поведения; 3) развернулись чисто теоретические дискуссии, обусловленные, во многом, оппонирующими статьями, в которых, как правило, опровергается оригинальная концепция Тьюринга и предлагается модифицированный тест (см. в [Философия ИИ 2004] следующие статьи: В. Морозов «Многообразие тестов Тьюринга»; А. Денисов «Тест Френча: Субкогнитивистское опровержение концепции Тьюринга»; И. Матанцева «Тест Блока: антибихевиористское опровержение тьюринговой концепции мышления» и И. Чижов «Тест Блока: нестандартные антибихевиористские возражения Тьюрингу»; А. Ласточкин «Тест Лавлейс: машина творить не может!»; Е. Романова «Наивная психология и инвертированный тест Тьюринга»; Д. Родионов «Тест Серля: интенционалистское опровержение концепции Тьюринга»; А. Никишев «Критика сильного искусственного интеллекта. Аргумент Гёделя»; А. Великанов и А. Макарычев «Тест Тьюринга и Д. Деннет».

В ряде докладов рассматривались вопросы практического программирования ТТ (работы Д. Зворыкина «Зачем вкладывать деньги в тест Тьюринга? Лойбнеровская премия»; А. Клопова «Программирование Теста Тьюринга»). Обсуждались вопросы о том, является ли ТТ тестом на интеллект – возможно, А. Тьюринг преследовал целью не определение интеллектуальности системы, а нечто иное (И. Рыбин «Социокультурные аспекты теста Тьюринга»; Ю. Цветков «Тест Тьюринга и паранойя»).

Тест Тьюринга – концептуальная основа не только искусственного интеллекта. Из рефлексии надо проблемой ТТ возникла новая мощная парадигма – функционализм, которая доминирует в современной философии сознания. Различным аспектам функционализма был посвящен ряд работ: (А. Алексеевой «Функционализм, физикализм и искусственный интеллект»; Д. Романова «Парадигма функционализма: как представить ментальное в нементальных терминах?»; И. Зайцева «Квалиа и парадигма функционализма»). Различные аспекты применения функционалистской концепции прослеживаются в ряде работ: Н. Маклашевской «Функционально-структурные аспекты понятия «потребность»»; Д. Соболева «Компьютер может мыслить (М. Минский)»; А. Чудакова «Функционалистский статус любви».

Итоги работы секции № 1 подвёл **Д.И. Дубровский**. В выступлении он акцентировал важность темы и отметил о концептуальной возможности расширения понятия теста Тьюринга [Дубровский 2004]. Это связано с важной теоретической проблемой современной философии сознания – проблемой диагностирования системы на предмет обладания ею субъективной реальностью.

Теоретические приложения теста Тьюринга

Работа секции № 2 *«Искусственный интеллект и «здоровый смысл»* продолжила тематику тестов Тьюринга и была посвящена, в основном, изучению работ выдающего математика, философа, программиста Дж. Маккарти, который впервые ввёл термин «искусственный интеллект». Большинство его работ посвящено применению народно-психологической позиции «здорового смысла» к созданию интеллектуальных систем. Здесь следует выделить проблемы поиска определения понятия ИИ: (В. Прасолова «Что такое искусственный интеллект?»; Е. Сименел «Зачем искусственному интеллекту философия?»); вопросы эпистемологии (И. Гаврилов «К вопросу эпистемологической адекватности репрезентаций»; Р. Горюнов «Псевдоволья»); проблемы логики (М. Красивская «Логические аспекты создания искусственного интеллекта»). Был представлен ряд докладов, посвящённых реализационным ориентирам позиции «здорового смысла»: М. Гришкин «Реализационные перспективы теории речевых актов»; М. Лапин «Экспертные системы, основанные на здоровом смысле»; Д. Родионов «Проблема дискурса искусственного интеллекта. Конструкторская позиция».

Историко-философская проблематика и тест Тьюринга

В работе секции № 3 *«Историко-философские перспективы компьютерного моделирования»* представлены исследования, посвящённые возможностям применения историко-философских знаний в ИИ. Рассматривались вопросы применения диалектики (А. Артюхов «Контентуальная модель смысла (А.Ф. Лосев)); схоластического понятия «виртуальность» (М. Сёмочкин «Виртуальная реальность и математика Н. Кузанского»); стоической логики (Т. Косинова «Искусственный интеллект и стоическая эпистемология»). Также были представлены исследования, основанные на работах не столь далёких от нас: изучение ИИ в контексте проблемы идеального, которая двадцать пять лет назад вызвала ожесточённые споры и определила оппозицию В.Э. Ильенкова и Д.И. Дубровского (А. Дробященко «Проблема идеального и искусственный интеллект»); рассмотрение лингвистической проблематики (М. Королёв «Проблема естественных видов в искусственном интеллекте»); о возможности применения интуитивистской метафизики (Я. Маликова «Интуитивистские ориентиры моделирования смысла (А. Бергсон)») и даже богословских текстов (М. Розов «Искусственный интеллект и святоотеческий опыт»). В историко-философском контексте прозвучал доклад А. Ивановой «Тьюринг и проблема вычислимости сознания».

Особой новинкой явился доклад Т.Кураевой «Зомби и искусственный интеллект». Студентка подготовил интереснейший доклад, связав воедино разноплановые темы ИИ, философии сознания и социально-философские аспекты. В настоящее время данную тематику Т.А. Конькова (Кураева) развивает в кандидатской диссертации, обучаясь в аспирантуре ИФ РАН.

Несмотря на явный интерес студентов (они в этот период изучали историю философии), результаты работы секции показали, что данной методологической тематике, посвящённой вовлечению историко-философских идей в компьютерную технологию, присущ «искусственный» характер попыток взращивания «древних» философских положений на аналитической почве. Философия ИИ должна базироваться на современных новейших достижениях философской мысли, связанных с научно-технологическими достижениями. Иначе дискурс получается расплывчатым и слабо связанным с существом вопроса. В итогах конференции звучал такой тезис: «На трудах древних греков искусственный интеллект не построить».

Практические приложения теста Тьюринга

Работа секции № 4 «Философия искусственного интеллекта и компьютерная технология была ориентирована на методологические вопросы применения ИИ, опять же в контексте ТТ. Рассматривались вопросы приложений ИИ в биотехнологии (Е. Александров, К. Домась «Бионика как направление робототехники»; М. Казанский «Квантовые компьютеры и квантовая механика»; М. Кольцов «Парадигма коннекционизма как методология нейрокомпьютерной технологии»; М. Пак, А. Панов «Методологические аспекты нанотехнологии»; И. Подопрigора «Естественно-языковый интерфейс: три подхода к моделированию “смысла”»), в социокультурной среде (А. Каданцева «Человек и компьютер: друзья или враги?», И. Михейкин «О распределении функций между человеком и компьютером в информационно-коммуникационных технологиях»). Ряд докладов отражал тему развития ИИ: А. Шулаков «История компьютерной технологии»; Е. Амелькин «Конкретизация термина “Искусственный интеллект”». Представлено два Интернет-навигатора по сайтам, на которых отражена проблематика философии ИИ (С. Колесников «Интернет-навигатор “Искусственный интеллект”»; О. Нестеров «Философия искусственного интеллекта в Интернет-среде»). Несколько докладов было посвящено соотношению феноменальных качеств (квалиа), «смысла», а также биологического с компьютерным (И. Смирнова «Клоны и «полуискусственный интеллект»; А. Бондарь «”Ква-

лия” как базовая категория виртуалистики»; В. Крючков «Что значит “Быть роботом”»?).

Таким образом, по энтузиазму и степени проявленной студентами активности можно судить о том, что молодым интересны идеи рефлексии над проблематикой ИИ. Душная атмосфера бессмысленного заучивания гегелевских категорий «марксианской» философии заменяется увлекательной игрой в мысленные эксперименты, которые предлагаются философией искусственного интеллекта. И от того, насколько будущие специалисты преуспеют в самостоятельной мыслительной работе над методологическими проблемами развития информационно-коммуникационной технологии, в создании собственных концептуальных и теоретических конструкций в области философии искусственного интеллекта, во многом зависит будущее нашей страны. Требуется немного – интеллектуальная свобода философствования, связанная с необходимостью профессионального роста. В настоящее время некоторые бывшие студенты стали кандидатами наук (М. Пак, А. Ласточкин), и несомненно, методология ИИ в этом им способствовала.

В последующем идея данной студенческой конференции диссертант использовал в рамках организации постоянно действующей **Всероссийской конференции студентов, аспирантов и молодых учёных «Искусственный интеллект: философия, методология, инновации»**. Предполагалось, что конференция будет проводиться ежегодно на базе центральных и региональных научно-образовательных центров. В настоящее время проведено восемь подобных конференций при непосредственном организационном и концептуальном участии диссертанта. При подготовке первой (г. Москва, МИРЭА, 2006 г.) и второй конференции (г. Санкт-Петербург, СПбГУ, 2007 г.) диссертант сформулировал следующие направления философских исследований (они остались неизменными на сегодняшний день). *Цель* конференций подобного рода: повышение творческого потенциала молодёжи в сфере прорывных информационных технологий; поддержка талантливой научной молодежи и содействие реализации ее способностей; повышение качества теоретико-методологической подготовки студентов, аспирантов и молодых учёных; интенсификация научно-исследовательской деятельности молодых специалистов; содействие преемственности поколений в развитии наукоёмкого технологического комплекса страны; организация междисциплинарных дискуссий по актуальным философским, методологическим и теоретическим проблемам искусственного интеллекта.

Тематика, по сути, воспроизводит тематику «взрослой» конференции: философско-методологические проблемы развития компьютерных и когнитивных наук; теоретические разработки новых направлений в развитии информационных технологий; эпистемологические, методологические и логические вопросы моделирования интеллектуальной деятельности человека; сознание, мозг, искусственный интеллект; концептуальные проблемы исследования виртуальной реальности; методологические проблемы роботизации; искусственный интеллект как фактор развития био- и нанотехнологий; искусственный интеллект в сфере медицинских, промышленных, военных и др. технологий; интеллектуальные информационные системы и технологии в образовании; интеллектуальные системы в естественных, технических, социальных и гуманитарных науках; социальное моделирование: искусственная личность и искусственное общество; искусственный интеллект и искусство; искусственный интеллект и этика; человек в информационном обществе; футурологические прогнозы развития искусственного интеллекта; современное общество знаний: философские аспекты.

В рамках 3-ей конференции студентов, аспирантов и молодых ученых (Москва, МИРЭА, 2009 г.) диссертант организовал секцию «Электронная культура и искусственный интеллект». Тематика данной секции затронула сферу современной культуры, её онтологию, эпистемологические особенности познания феноменов компьютеризованной реальности, её аксиологические, праксеологические, антропологические параметры. Данная секция оказалась разведочной в плане организации секции «Электронная культура и искусственный интеллект» в рамках Московского форума культуры (Москва, июль 2010 г.). В настоящее время данное направление раскрыто в тематике постоянно действующей конференции «Электронная культура: интеллектуальные технологии» (методологические аспекты рассмотрены ниже, в пп. 1.2.3).

Диссертант полагает – для освоения новой философской тематики следует "первыми в атаку бросать" студентов и аспирантов. Успешное проведение студенческих конференций 2004 и 2006-2012 гг. в плане подготовки «взрослых» научных мероприятий это подтверждает.

Роль теста Тьюринга в философской методологии искусственного интеллекта

На заседании № 11 научно-теоретического семинара НСММИ РАН «Философские проблемы искусственного интеллекта» состоявшемся 26 апреля 2006 г. был заслушан доклад диссертанта на тему: **«Роль теста Тьюринга в методологии искусственного интеллекта»**. Докладчик показал, что тест Тьюринга имеет фундаментальное значение для проблематики искусственного интеллекта и вычислительной парадигмы сознания, так как позволяет связать формальное с каузальным, т.е. теоретический уровень представлений о ментальных явлениях с реализационными средствами компьютерной техники. Отсутствие в нашей стране философской и методологической проблематики искусственного интеллекта непосредственно связано с забвением тематики теста Тьюринга – лишь только после сорокалетнего игнорирования её, в 2005 г. на конференции «Философия искусственного интеллекта» (г. Москва) была впервые на всероссийском уровне организована дискуссия. Докладчик охарактеризовал основные принципы теста Тьюринга, отличие данного мысленного эксперимента от эксперимента с машиной Тьюринга. Если последний даёт демонстративное определение понятия алгоритма, то первый – квалиалгоритма, который, в свою очередь, наиболее чётко определяет операционально-процедуральные процессы интеллектуальных компьютерных систем. Докладчик выделил крупные модификации теста Тьюринга – тесты Серля, Блока, Ватта, Френча, Брингсйорда и др. Данные модификации характеризуют различные особенности и проблемы реализации ментального в компьютерных системах. Развитие тематики теста Тьюринга обуславливает не реальные достижения в области технологии, скорее, критическое осмысление претензий когнитивных и компьютерных наук на воспроизводство не только ментальных, но и биологических, социокультурных, персонологических феноменов на базе компьютерной парадигмы. В рамках данного доклада диссертант впервые выдвинул идею о необходимости *интегрального, комплексного использования различных тестов Тьюринга с целью изучения социокультурных явлений компьютерогенного мира*.

Юбилейная тематика теста Тьюринга

По поводу проблематики теста Тьюринга следует отметить крупное мероприятие – научно-практическую конференцию с международным участием, которая проводилось на философском факультете МГУ им. М.В. Ломоносова 28-29 октября 2010 г. Организаторами конференции явились **Н.Ю. Ключева** и

А.А. Костикова. Конференция была посвящена 60-ой годовщине публикации статьи А. Тьюринга, а так же 30-ой годовщине публикации статьи Дж. Серля, в которой была изложена концепция одной из крупнейшей версии теста Тьюринга – «Китайской комнаты». Формат конференции предполагал работу студенческой секции и круглого стола с участием специалистов в области философии ИИ и философии сознания. На студенческой секции изучались следующие вопросы: роль теста Тьюринга в развитии философского дискурса о природе человеческого мышления, интеллекта и сознания; трансформация принципов человеческой коммуникации, опосредованной электронными средствами передачи сообщения в современном мире; Тест Тьюринга пройден: за и против; Тест Тьюринга и «китайская комната» Дж. Серля: имитирует или на самом деле мыслит?; Что имитирует компьютер: человеческое поведение и разумное поведение, проблема антропоморфизма ТТ и др. Тематика круглого стола «Тест Тьюринга: философские интерпретации и практические реализации» была следующей: комплексный тест Тьюринга, тест Тьюринга и китайская комната Дж. Серля, методологическая ценность теста Тьюринга для компьютерных наук, «игра в имитацию» как мыслительный эксперимент и др. Участниками круглого стола были *Васильев В.В.* («Показывает ли Китайская комната Серля неэффективность теста Тьюринга?»), *В.К. Финн* («Отношение к Тесту Тьюринга с точки зрения современных интеллектуальных систем»), *Алексеев А.Ю.* («Комплексный Тест Тьюринга»), ст.преп. кафедры философии языка и коммуникации, *Клюева Н.Ю.* («Методологическая ценность Теста Тьюринга для развития компьютерных наук»), *Потапов А.* («Искусственный интеллект: представимость и возможность») и др.

Следует отметить, что в целом данная конференция показала имманентность тематики теста Тьюринга философским исследованиям ИИ. Дискурс по поводу роли и статуса ТТ протекал в более спокойной среде, в условиях понимания того, о чем говорят докладчики и в атмосфере оживлённых дискуссий – не так, как это было шесть лет назад, во время студенческой конференции 2004 г. **Начиная с конференции 2010 г. в МГУ тематика теста Тьюринга в отечественной философии состоялась.**

1.2.2. Роль теста Тьюринга в философии искусственного интеллекта

В данном подпараграфе искусственный интеллект рассматривается как целостная многоуровневая система, образованная философскими (мировоззренческими и методологическими), научными (теоретическими и эмпирическими) исследованиями, а так же практикой инженерных разработок (предпола-

гающих конструирование работоспособных интеллектуальных систем в быту и общества) и способами применения их в быту (то, что сегодня обозначается термином юзабилити – способностью человека освоить компьютерные новации ИИ). Предлагаются различные формы философии ИИ: пропедевтическая, критическая, конструктивистская, экзистенциальная. Для изучения специфики философского знания рассматривается научно-методологический, научно-теоретический, научно-эмпирический уровень исследований ИИ. Предлагается учебный курс философии ИИ как самостоятельной образовательной и философской дисциплины. Основу курса составляют исследования теста Тьюринга и его версий.

При изучении вопроса роли и места философской методологии и мировоззрения в структуре междисциплинарных исследований ИИ следует позиционировать два подхода к философии ИИ: 1) философия ИИ (ФИИ) как методология научных исследований ИИ; 2) ФИИ как самостоятельная дисциплина, частично затрагивающая конкретные разработки ИИ в части общенаучной методологии.

Следует отметить, что проблематика ФИИ до 90-х годов в нашей стране велась крайне слабо. Но причина не в том, что «буржуазные» теоретические подходы к пониманию не только искусственного, но и естественного интеллекта не согласовывались с официальной марксистско-ленинской диалектикой. Причина проще – исследования ИИ не настолько были развиты, чтобы подвергаться философской рефлексии. В начале перестройки данная тема в нашей стране вообще заглохла по вполне понятным причинам. Сейчас развивается в основном благодаря энтузиазму исследователей философии ИИ. И это всё происходит то время, когда в англо-американской науке ФИИ отпочковалась в самостоятельную область философского знания.

Следует так же отметить, что до проведения конференции «Философия ИИ» (2003-2004 гг.) диссертант в период защиты кандидатской диссертации на схожую тему часто употреблял сочетание «Философия искусственного интеллекта», что вызывало недоумение у многих преподавателей и коллег по философскому факультету МГУ им. М.В. Ломоносова. В те годы диссертант выглядел в философском сообществе как представитель паранауки (или, если угодно, парафилософии) (поэтому ему пришлось в ряде работ защищаться от обвинений его взглядов в «паранаучности» [Алексеев 1998-1; 1998-5; 2000; 2001]). Начиная с 2005 г., после проведения конференции «Философия искусственного интеллекта» (пп. 1.1.1), такого рода философия обрела научную легитимность в

нашей стране. Среди отечественных ученых, философов и специалистов самых разных специальностей к данной проблеме проявился определенный интерес. После тридцатипятилетнего забвения методологических оснований ИИ конференция реанимировала данную проблематику, широко представленную дискуссиями советских времён – «Может ли машина мыслить?».

Итоги конференций показали с очевидностью, по крайней мере, для диссертанта, то, что философии ИИ не является прикладной философией, наподобие философии физики, биологии и пр. Такой взгляд представляется слишком узким, он ограничен логико-эпистемологическими рамками той или иной научной школы, практикующей разработку и внедрение интеллектуальных систем. ФИИ намного шире. ФИИ имеет непосредственный выход к социокультурным горизонтам, например, к проблематике сознания. В частности, ФИИ позволяет аналитически четко формулировать ключевые аспекты основного вопроса философии – о соотношении сознания и бытия – актуальность которого не снимается (в чем убеждают многие отечественные учебники), а повышается в условиях становления э-культуры. Только в современных условиях данная тема звучит по-иному, например, не в форме соотношения «сознание/мозг», но «сознание/мозг/компьютер».

На фоне подготовки, проведения и анализа конференции по ФИИ в МИЭМ (Московский университет электроники и математики), ГУГН (Государственный университет гуманитарных наук) и в Госуниверситете «Дубна» диссертантом параллельно преподавался экспериментальный курс "Философия искусственного интеллекта". Курс проходил в контексте изучения студентами проблем социальной философии. Первый курс, читаемый в техническом математическом вузе – в МИЭМ, был ориентирован если не на профессиональных программистов, то, по крайней мере, на специалистов ИТ. Поэтому здесь превалировала тематика методологических обобщений способов программирования. Второй курс (для ГУГН и университета в г.Дубна) был предназначен для будущих специалистов в области гуманитарных и социальных наук – философов, политологов, социологов, журналистов, психологов и др.

Философия искусственного интеллекта как прикладная философская дисциплина

Первый подход сложился в основном благодаря учебным курсам ИИ, где ФИИ выступает как часть методологии компьютерной науки в части, касающейся проблематики ИИ. В зарубежных странах на начало 2000 г. курсы ФИИ (The Philosophy of Artificial Intelligence) читались в большинстве зарубежных вузов в различных формах: 1) в рамках дисциплины Computer Sciences (см., например [Mohr 2000]); 2) в качестве введения в инженерии искусственного интеллекта [Perrot 2000]; 3) как часть курса по философии сознания [Wright W. 2001]. Форма преподавания ФИИ различна. Либо самостоятельный курс, читаемый в течение всего учебного года, либо четыре – десять лекций, предвещающих основной курс по программированию ИИ, либо, как в третьем случае, акцентирующих внимание на проблемах компьютерного (машинного) функционализма в рамках философии сознания.

Курс ФИИ, помимо программистов и математиков, всё более активно преподается социологам, психологам, биологам, физикам, политологам, журналистам и студентам многих других специальностей. В нашей стране такие курсы предусмотрены в рамках образовательной программы «Интеллектуальные системы в гуманитарной сфере», разработанной в РГГУ [Финн 2008] (см. пп.1.1.1, 1.2.1 (в резолюции конференции «Философия ИИ» 2005 г. предложена рекомендация Минобразованию РФ о внедрении данной специальности).

Так или иначе, философия ИИ – это мировоззренческо-методологическая «надстройка» над теоретическими исследованиями ИИ [Алексеев, 2005-6]. Показательным представляется мнение Дж. Маккарти – основателя искусственного интеллекта как научно-исследовательской дисциплины по поводу философии ИИ.

Подход Дж. Маккарти к философии искусственного интеллекта. Как широко известно, термин «искусственный интеллект» впервые был применён летом 1956 г. известным американским философом, логиком, программистом Джоном Маккарти (так же он разработал язык LISP, широко применяемый в программной инженерии ИИ). Данным термином он именовал научно-исследовательские работы, в руководстве которых участвовал. Также считается, что Дж.Маккарти впервые ввёл в обиход термин «философия искусственного интеллекта». Под философией ИИ он понимает специфическую область методологических исследований, подобную прикладным философиям, типа «фило-

софия биологии», «философия физики» (см. [McCarthy, 1969; 1987; 1998], особенно - [McCarthy, 1995-1]). Специфику философии ИИ Дж. Маккарти представлял следующим образом: любая реализация системы ИИ требует концептуально-программной реализации философских положений, в основном, эпистемологического содержания. Программная реализация означает кодирование той или иной философской теории или концепции, характеризующих способ познания и репрезентации окружающего мира. То есть философия ИИ – это *философия программиста*. Маккарти так и утверждает - начало философствования по проблеме ИИ следует начинать с изучения какого-нибудь языка программирования: «Если вы хотите изучать ИИ и не знаете, с чего начать, начните с изучения математической логики. Читайте больше и психологии и физиологии нервной системы. Выучите какие-нибудь языки программирования, например, Java или C++, они сейчас хорошо востребованы» [McCarthy, 1995-1].

Философия искусственного интеллекта – философия профессионального программиста.

Философия искусственного интеллекта как самостоятельная философская дисциплина

На взгляд диссертанта, о философии ИИ следует говорить как о *самостоятельной философской дисциплине*. Для того, чтобы ФИИ выполняла роль самостоятельной философской дисциплины, следует обратить внимание на носителя способности к философской рефлексии. Для первого варианта построения ФИИ несомненно это – программист, способный к философским обобщениям. Для второго варианта ФИИ уточним понятие «программист». Техногенный (точнее, компьютерогенный) стиль жизни современного человека обуславливает расширенную трактовку этого понятия. Сегодня под программистом понимают не узкого специалиста, способного кодировать алгоритмы на некотором языке программирования. Любой современный человек, «заброшенный» в компьютерогенную культуру, вынужден овладевать навыками программирования. Массовому характеру способности программировать благоприятствует развитие человеко-машинного интерфейса. Если понимать ещё шире, программирование – это неотъемлемая способность каждого человека, связанная с биологическими и социальными функциями постановки задач и регулируемого поиска путей их решения. Поэтому ФИИ может представлять интерес не только для специалиста технического или математического профиля, но и для тех, кто интересуется естественно-научными и общественно-гуманитарными сферами знаний.

Философия искусственного интеллекта – философия непрофессионального программиста.

Вслед за традиционным делением знания «философия-наука-практика» будем выделять три уровня изучения ИИ: философский, научный, инженерный. Ниже представлены пункты курса ФИИ, выраженные в формате концептов. Курс одинаковый для студентов разных специализаций, однако если будущих специалистов в сфере компьютерной технологии он «поднимал» к высотам философской рефлексии, то будущих специалистов гуманитарного профиля он «опускал» с проблемами технологии.

Унификация учебных курсов обусловлена тем, что они основаны на проблематике теста Тьюринга. Данная метафора компьютерного интеллекта понятна и для будущих специалистов-гуманитариев и для специалистов естественных наук. Т.е. благодаря тесту Тьюринга философия ИИ состоялась и как прикладная философия и как самостоятельная дисциплина.

Диссертант с 2008 г. читает учебные курсы по проблематике искусственного интеллекта будущим специалистам по прикладной информатике (в сфере культуры и искусств) в Московском государственном университете культуры и искусств. Тематика теста Тьюринга распределялась между следующими дисциплинами:

Информатика (информационные системы и технологии) – одна лекция, которая характеризует постнеклассическую парадигму построения информационных систем и именно «человекомерное» видение на информационные технологии обеспечивает тест Тьюринга.

Интеллектуальные информационные системы [Алексеев, 2008] – данный курс предназначался для изучения моделей представления знаний и по сути воспроизводил тематику стандартных учебных курсов по ИИ [Осипов, 2009; Гаскаров, 2003; Ясницкий, 2005]. Акцент в выборе учебного пособия делался на работе Дубровина [Дубровин, 2008], так как диссертант принадлежит, пожалуй, к наиболее сильной отечественной школе и традиции ИИ, сложившихся в стенах 27 ЦНИИ МО (диссертант в этом заведении работал в 1987-1996 гг. и специализировался в программировании интеллектуальных информационных систем военного назначения). Представление о широте и глубине направлений данной школы может составить коллективная работа [Интеллектуальные информационные системы, 1991]¹. Тест Тьюринга задаёт основные требования,

¹ 27 ЦНИИ МО в середине 2000-х был уничтожен как институциональная научная единица.

которым должны удовлетворять компьютерные системы, реализующие продукционный, фреймовый, семантико-сетевой, формально-логический подходы к представлению формализованных «данных» и «знаний» в современной программной инженерии.

Социокультурные вопросы искусственного интеллекта [Алексеев, 2010] – курс полностью базируется на концепции теста Тьюринга. Изучаются вопросы следующего плана – может ли компьютер мыслить, жить, любить, творить, понимать и пр.

Психосемиотические аналитические системы [Алексеев, 2012] – наиболее интересный курс, по мнению диссертанта. Объединяет в единое целое презентативную и коннекционистскую парадигмы, экстернализм и интернализм. Основан: 1) на метафизических теориях психического (сознания) – идеализме, физикализме, дуализме, бихейвиоризме, и – самое главное – на функционализме (компьютеризме); 2) метафизических концепциях семиотики; 3) когнитивной психологии; 3) когнитивной лингвистике; 4) компьютерной психологии; 5) компьютерной лингвистике. Тест Тьюринга инкорпорирован во все рубрики данного курса – например, физикалистская концепция погружается в предложенный диссертантом вариант функционализма – функционализм теста Тьюринга.

Информационно-аналитические технологии [Алексеев, 2011] – тест Тьюринга служит пропедевтическим целям и задает основные требования к интеллектуальным технологиям представления информации в современных электронных библиотеках и аналитическим информационным технологиям в плане реализации функций «понимания» текста на синтаксическом, семантическом и прагматическом уровнях.

По всей видимости, наибольший интерес с позиции диссертационного исследования представляет тематика курса, который диссертант преподавал в рамках специального курса «Социально-философские вопросы искусственного интеллекта» для философов и политологов Государственного университета гуманитарных наук в 2005-2006 гг.

Студентам предлагался следующий учебный план философии ИИ.

Пропедевтическая философия искусственного интеллекта

Уровни изучения ИИ: философский (мировоззренческий/методологический), научный (теоретический/эмпирический), практический (инженерный/бытовой). *Инженерно-технологический уровень*: краткая история инженерии ИИ: машины Р. Луллия, Г. Лейбница, С.Н. Корсакова, Ч. Бэббиджа, А. Тьюринга. Направления ИИ: игры, доказательство теорем и автоматизация рассуждений, решение проблем и планирование, техническое восприятие, понимание языка и перевод, экспертные системы, обучение, робототехника и системы виртуальной реальности. *Научно-теоретический уровень*: полисемия слова «интеллект». Понятие артефакта. Понятие интеллекта. Основные значения слова «искусственный интеллект», принятые исследователями ИИ. Конвенциональный статус теорий ИИ. *Методологический уровень*: от психофизиологической проблемы «сознание/мозг» к психотехнологической проблеме «мысль/чип». Мысленный эксперимент (МЭ) как базовый метод ФИИ. «Мельница» Лейбница как прообраз современных МЭ в области ИИ. Различие подходов в трактовке МЭ: реалистический, номиналистический и риторический. Метафора как средство прагматологической убеждённости в условиях онтологической неопределённости понятий «интеллект», «знание», «сознание» и т.п. *Экзистенциальный уровень*: ФИИ как рефлексия над проблематикой вопроса «Что значит «мыслить»?». Фундаментальный статус данного вопроса относительно иных философских и научных вопросов. Понятие критической и положительной философии искусственного интеллекта.

Критическая философия искусственного интеллекта

Основной вопрос ФИИ: «Может ли машина мыслить?». Тест Тьюринга (ТТ) и функционалистская парадигма мышления. Преодоление онтологической неопределённости понятий ИИ посредством лингво-операционалистского определения интеллекта. Базовые положения ТТ. Метафора Тьюринга «Человек-компьютер». Многообразие способов интерпретации ТТ: традиционная (эпистемологическая), биологическая, гендерная, социокультурная, психологическая и др. Основные темы дискуссии по поводу ТТ: теологическая, антисциентистская, креационистская, «от первого лица», «от технологического несовершенства», экстрасенсорная и др. Современные классификации ТТ: от «наивного ТТ» (лингво-семантическая тождественность машины и человека) к тотальному ТТ (персонологическая идентичность с учётом микрофизической неотличимо-

сти). Сложность идентификации интеллектуальности системы. Принцип «вскрытие покажет».

Критика сильного ИИ. Понятие сильного ИИ. Футурологические проекты сильного ИИ: иммортология, пост- и сверхчеловечество, технологический реинкарнационизм, кибертеология, пострелигия «силиконового человечества» и др. Мысленный эксперимент Дж.Серля «Китайская комната» как критика сильного ИИ. Основные направления дискуссии по поводу теста Серля: системологическое, робототехническое, нейрофизиологическое, с позиции «чужого разума», семиотикосинтаксическое. Интенциональность и вычислимость. Машина мыслить может, но понимать не может. Человек как мыслящая и понимающая машина. Актуальные интерпретации теста. Современный этап развития когнитивной науки как попытка разрешения проблемы Дж. Серля. Понятие когнитивной науки. Соотношение компьютерной и когнитивной наук. Структура когнитивной науки и направление её развития: от когнитивной «структуры» к парадигме «мыслящей телесности». Примеры когнитивных интерпретаций общества и культуры и их метафорический статус.

Критика слабого ИИ. Понятие слабого ИИ. Народная психология и наивно-психологистская установка приписывания ментальных свойств системам, в т.ч. компьютеру. Тест Лавлейс: компьютер творить не может. Тест Френча: МЭ «Тест чайки», игры «ассоциация слов» и «рейтинг неологизмов». Социокультурная обусловленность интеллекта. Проблемы интеллектуального шовинизма и тоталитаризма. Тест Ватта: инвертированный тест Тьюринга как преодоление наивно-психологистской установки. Проблемы параллелизма и иерархии отчетов от первого лица. Нечёткость границы между слабым и сильным ИИ.

Критика глобального ИИ. Понятие глобального ИИ. Метафоры глобального ИИ: репрезентативная (ИИ как всемирная экспертная система) и коннекционистская (ИИ как «мозг» человечества). Онтологический статус общественного сознания. Мифы ноосферы и алгоритмов цивилизационного развития. Функционалистская парадигма социокультурных систем. Тестирование социальной системы на предмет «ментальных свойств»: сложность преодоления биохейвиоральной парадигмы. Антибиохейвиоральная критика глобального ИИ и психофункционализм Н. Блока. Тест и машина Н.Блока как критика репрезентативной трактовки глобального ИИ. МЭ «Китайская нация» как критика коннекционистской трактовки глобального ИИ.

Критика формального ИИ. Понятие формального ИИ. Вычислительная концепция разума и понятие формальной системы. Общие идеи математизации

логики и логизации математики. Анализ соотношения «мысль-слово-число». Истинность и вычислимость. Гёделева нумерация. Теорема Гёделя о неполноте. Р. Пенроуз и Н. Лукас: приложения теоремы Гёделя для критики формального ИИ. Опровержения гёделевской критики. Алгоритм и квазиалгоритм: машина Тьюринга как формализация понятия алгоритма, тест Тьюринга как формализация понятия квазиалгоритма. «Квазиинтенциональность» компьютерной системы как преодоление проблемы формального ИИ.

Критика функционалистской парадигмы мышления. Многообразие трактовок функционализма. Анализ парадигм мышления в контексте теорий сознания: панпсихизм, спиритуализм, физикализм, бихевиоризм, дуализм, идеализм, материализм, элиминативизм, ментализм, двухаспектная теория, феноменология и др. Неочевидность метафоры «человек-компьютер» Тьюринга. Отрицательное решение основного вопроса философии искусственного интеллекта («Может ли машина мыслить?») на основе риторико-метафорической трактовки МЭ.

Положительная философия искусственного интеллекта

Методологические проблемы ИИ. Определение положительной философии искусственного интеллекта. К проблеме программирования философских понятий. Методологические проблемы фреймов, нейросетевых структур, динамических и многоагентных систем. Логико-эпистемологические задачи ИИ. Проблемы интеграции коннекционистской и репрезентативной парадигм. Проблемы компьютерного моделирования «смысла»: от «данных» и «знаний» к «смыслам». Социокультурные особенности моделей смысла.

Практические перспективы ФИИ. Кибернетическая, синергетическая, генетическая парадигмы ИИ. Квантовая парадигма сознания и квантовые компьютеры. Голографическая модель сознания как концептуальная основа голографических баз «смыслов». Микромир и нанотехнология. Основные идеи симбиоза «человек-компьютер».

Информационный подход к сознанию. Основные положения информационного подхода к сознанию (Д.И. Дубровский): проблема идеального, анализ соотношения «сознание-мозг-искусственный интеллект». Особенности нередуктивного функционализма Д.И. Дубровского. Понятие субъективной реальности. Критика редукционистских и когнитивистских подходов к трактовке субъективной реальности. Принцип инвариантности информации по отношению к физическим свойствам ее носителя. Перспективы информационного подхода к решению психотехнологической проблемы.

Технологический статус проблемы философских зомби. Определение философских зомби. Проблематика зомби и проекты искусственной жизни, искусственной личности и искусственного общества. Зачем человеку сознание? Зачем члену общества осознание общности? Принцип «несущественности сознания». Люди и зомби. Роботы и зомби. Киборги и зомби. Концепция псевдосознания. Мысленные эксперименты «Земля зомби». Дискуссии по поводу принципа «несущественности сознания». Научно-теоретический и инженерно-технологический статус проблематики философских зомби: социологические, политологические, антропологические, аксиологические, праксеологические и иные приложения. Проекты искусственной личности, искусственного общества, искусственного мира. Практико-преобразующие ориентиры проектов.

Экзистенциальная философия искусственного интеллекта

Остановимся подробнее на данном уровне изучения проблематики ИИ [Алексеев, 2006-6]. На экзистенциальном уровне подчеркивается необходимость привлечения проблематики ИИ для получения, в первую очередь, определённости понятия естественного интеллекта. Неизбежно в изучение вовлекаются связанные с «интеллектом» понятия: «знание», «вера», «убеждение», «смысл», «ценность», «личность», «общество», «культура» и др. Всплывают коварные вопросы фаустовского типа: «Что значит «знать»?» и производные от них суждения сократического типа, например: «Я не знаю, что значит “знать”». Следовательно, я не знаю того, что ничего не знаю». Возможны когнитивно-моралистские рассуждения о том, что знание – это определённая телесно-разумная архитектура человека, из которых следуют этико-когнитивные проекты «мыслящей телесности», бурно развивающиеся сегодня в когнитивной науке.

Конкретным примером задач экзистенциального уровня проблематики ИИ может послужить систематизация познавательных ситуаций, предложенная Д.И. Дубровским в ряде работ в связи с анализом феномена веры [Дубровский, 1994; 2002]. Выделяется четыре типа ситуаций: 1) *Знание о знании*, означающее факт рефлексивности присущего субъекту знания, отображение наличного, субъективно переживаемого знания, несущего одновременно и ценностное отношение и фактор активности; 2) *Незнание о знании*, выступающее как момент процесса познавательной активности, в котором всегда присутствуют неосознаваемые субъектом компоненты упорядочивания, оценки, эвристики и пр. и которые в последующий момент могут быть осознаны; 3) *Знание о незнании*, означающее наличие проблемы и потребности в расширении сферы познаватель-

ной деятельности, что порождает состояние эвристической напряженности, стремление решить проблему, а так же веру в возможность её решения и веру в себя, способного её решить; 4) *Незнание о незнании* – ретроспективно выявляемая в самой абстрактной форме допроблемная ситуация, в которой пока не определился новый объект незнания, в силу чего познавательная активность отсутствует, готовясь проявиться на стадии предпроблемной ситуации, при переступании субъектом черты полного незнания о незнании. Д.И. Дубровский считает, что каждый отдельный человек и каждая человеческая общность, вплоть до всего человечества, одновременно находится в этих четырёх гносеологических ситуациях и, так как каждая такая ситуация проникнута целостной духовно-практической активностью, то неизбежно «знания» и «незнания» окрашены экзистенциальной экспрессией, заданной стремлением к подлинным смыслам и ценностям.

Несомненно, подобного рода задачи – прерогатива мировоззренческой функции философии ИИ. Их решение приводит не только к критическому анализу моделей «знаний» и различных проектов интеллектуальных систем, но и к антропологическому призыву перехода в состояние «беспокойства духа», особенно, в ситуации «незнания о незнании» – незнания человеком самого себя. Собственно, здесь и следует задействовать мощь ИИ – без ИИ познать ЕИ невозможно (тезис Т.Б. Кудряшовой, см. пп. 1.1.2).

Философско-методологический уровень исследований искусственного интеллекта

В связи с исследованием роли и места философской методологии в структуре исследований ИИ особый интерес, на наш взгляд, представляет диссертационная работа Н.Ю. Ключевой (Трушкиной) на тему «Отношение теоретических концепций и компьютерных моделей в исследованиях искусственного интеллекта» [Трушкина, 2006]. Это – одно из немногих диссертационных исследований, в которой явно прописывается философско-методологическая тематика ИИ. В целом работа посвящается осмыслению интегрального, сквозного процесса взаимосвязи трёх сфер знания, изучающих интеллект (мышление): 1) философских оснований, принципов, категорий, положений; 2) частных теорий, понятий, концепций гуманитарных, социальных и естественных наук; 3) прикладных компьютерных моделей искусственного интеллекта (ИИ). Предметом исследования являются воспроизводящиеся в философских и научных сообществах концепции интеллекта, его строение, функции, субстратная основа, ключевые попытки его моделирования в конкретных компьютерных моделях ИИ, а

так же формирование интегрального методологического подхода, комплексизирующего теоретические положения и практические наработки в области ИИ.

Н.Ю. Клюева решает по сути две основные задачи: 1) проводит анализ теоретических концепций интеллекта, в рамках которых даётся обстоятельная характеристика философских понятий об интеллекте в динамике их историко-философского развития, а так же предлагает пример применения научных концепций интеллекта на базе частных психологической и нейробиологических теорий, 2) осуществляет анализ связи компьютерных моделей интеллекта с вышеназванными теориями.

В главе I (здесь и далее до конца пп. нумерация глав и параграфов в соответствии со структурой работы Н.Ю.Клюевой) «Анализ теоретических концепций интеллекта» Н.Ю. Клюева: 1) Анализирует представления о структурах интеллекта в истории философии (от античности до начала 20 в.). При этом она выделяет понятие о мышлении как поиске истины (включая формальные методы её постижения), использует идею локализации интеллекта в человеке, формирует представление о структурах интеллекта в рамках классической логики, а так же выделяет математическую интуицию как элемент мышления, выходящий за сферу чисто формальных процедур и операций мыслительной деятельности. Отличает данную главу строгая сквозная систематизация теоретических концептов интеллекта. Благодаря ей чётко прорисовываются основные теоретические концепции мыслителей и школ по поводу представлений о мышлении и интеллекте. Создаётся систематизация теоретических положений, применимая к сфере ИИ. Строится структура как теоретической преемственности взглядов на интеллект, так и их отмечается концептуальная полифония. 2) Далее Н.Ю.Клюева справедливо замечает, что философские и научные концепции оказывают опосредованный характер на компьютерные модели и приводит пример выработки частной научной концепции интеллекта в контексте одной из психологических концепций – гештальтпсихологии. Продолжается сквозная типология понятий интеллекта, раскрывая преемственность философских и научно-научных представлений.

В главе II осуществляется анализ ряда классических компьютерных моделей интеллекта. Исследуется классическая логическая парадигма, представленная в программе Логик-теоретик и Общий решатель задач. Так же осуществляется анализ эвристических моделей, реализованных в экспертных системах и программах-имитаторах ELIZA и A.L.I.C.E, реализующих тест Тьюринга. Для раскрытия связи положений гештальтпсихологии и теории фреймов подробно

описываются их общие и отличительные параметры, показывается понятийная и операциональная аналогия между первой и второй. Наряду с этим раскрывается связь между теоретическими представлениями о биологическом нейроне (п.п.1.4.) и его компьютерным аналогом (п.п.2.3). В целом две представленные главы показывают связь между теоретическими концептами и компьютерными аналогами, что представляет собой новое направление в методологическом освоении проблематики искусственного интеллекта.

Исходя из структуры работы, отчётливо выделяются лишь два уровня – теоретический и прикладной. Тем не менее, исходя из содержания первой главы, в которой рассматривается уровень философских учений об интеллекте и частные научные теории, можно сделать вывод о *трёхуровневом делении всей системы моделирования ИИ*. И это представляется *методологически важным приемом*, тем более что в работе проводится чёткое разделение философского и частно-научного знания. На первом уровне – уровне философских обобщений создаётся категориально-понятийная сетка представлений об интеллекте (мышлении). На втором уровне – уровне т.н. теорий среднего уровня – используются теоретические концепты, опирающиеся на философские понятия и в то же время являющиеся производными различных гуманитарных, социальных, естественных наук, изучающих частные аспекты интеллектуальной деятельности. На третьем – прикладном уровне разрабатываются конкретные программно-информационные модели ИИ, опирающиеся на второй уровень.

Особо интересны исследования Н.Ю. Ключевой по поводу соотнесения и взаимосвязи ранее обозначенных теоретических концептов гештальт-психологии (п.п. 1.2) и конкретных моделей ИИ, представленной в теории фреймов (п.п.2.2). Делается вывод о частичном изоморфизме между этими моделями. Данная мысль подтверждается на примере других моделей, а именно, при соотнесении нейробиологических теорий (относительно внутренней структуры нейронов и ансамбля нейронов) с коннекционистскими (Н.Ю. Ключева называет их «структурными») моделями компьютерной технологии. Таким образом, идею о частичном изоморфизме она осуществляет как на базе репрезентативного ИИ (символьного ИИ, «старого доброго» ИИ – имеются различные терминологические модификации), так и на базе арепрезентативного, коннекционистского подхода. Это показывает полноту и целостность исследований на базе классических парадигм ИИ. В работе Н.Ю. Ключевой явно не раскрываются редукционистские переходы между тремя уровнями, однако из контекста можно реконструировать утверждение о невозможности программы чистой редукции теоретического уровня исследования интеллекта к его компьютерным мо-

делям. Это положение Н.Ю. Ключева обозначает термином *частичный (неполный) изоморфизм* (изоморфизм между человеческим и машинным интеллектом). Уровень естественного интеллекта в качестве примера фиксируется в психологических и нейробиологических исследованиях, уровень машинного интеллекта – в компьютерных моделях ИИ. Интересен взгляд Н.Ю.Ключевой на то, что способы и формы построения компьютерных моделей при инициальной опоре на те или иные теоретические положения в дальнейшем начинают «развиваться» по собственным законам. Это положение в указанной работе неоднократно подчёркивается. Представление об «автономном» развитии компьютерных моделей, теоретический источник которых трудно реконструировать, так же подчеркивает антиредукционистские интенции. Антиредукционистское предпочтение подчёркивается ещё и тем фактом, что неоднократно отмечается (особенно в Главе 2) хаотичный, фрагментарный характер применения теоретических концептов (не философских концепций, а теорий «среднего уровня») в методиках построения компьютерных моделей. Отсутствует единая база для субординации, координации и интеграции компьютерных моделей, в результате чего они начинают выполнять разрозненные функции, не удовлетворяя критерию целостности интеллектуальной деятельности. И, как считает Н.Ю. Ключева, наивысшим интегратором для обеспечения систематического единства компьютерных моделей выступает та или иная философская концепция. Однако в силу разнообразия философских взглядов на интеллект не возможна некая «Великая Теория искусственного интеллекта». Релятивизм по поводу представлений об интеллекте на самом высоком уровне – философском – неизбежен. Крайне важным представляется акцент Н.Ю. Ключевой на внеинтеллектуальных и даже бессознательных составляющих мыслительной деятельности. Механизмы интуиции, которые она раскрывает на базе богатого материала (математики Пуанкаре (пп.1.1.5), «инстайте» гештальтпсихологии (пп.1.2.1) и др., по мнению Н.Ю. Ключевой не подлежат компьютерной реализации (хотя имеется ряд когнитивных учёных – А.Сломан, Дж.Маккарти, в принципе, А.Тьюринг, и др.), которые считают обратное.

Однако следует отметить, что Н.Ю. Ключева предполагает возможность исследований, репрезентирующих всю область знания в целом («Введение»), выражая интенцию комплексного подхода между теоретическими положениями и практическими компьютерными моделями. Тем самым она впадает в противоречие, утверждая до этого множество альтернативных философских понятий о мышлении (интеллекте), которые изложены в представленном в работе каталоге теоретических концептов интеллекта. По всей видимости, имелись в

виду междисциплинарные исследования, ориентированные на ту или иную теоретическую парадигму, которая задаётся, в свою очередь определённым философским направлением, востребованным в силу определённых общественных потребностей. Тем не менее, в работе Н.Ю. Ключевой отчётливо прослеживаются приложения теоретических положений диссертации к новейшей актуальной теме развития отечественной философской мысли – к проблематике философской методологии искусственного интеллекта, которая сегодня охватывает междисциплинарные исследования социальных, гуманитарных, естественных и технических наук, включая и корпус методологических знаний о сквозном процессе построения компьютерных моделей в схеме «философия – научные теории среднего уровня – компьютерные модели» Этому, собственно, и была посвящена работа Н.Ю. Ключевой. И мы подробно на данной работе остановились по причине убежденности в том, что подобного рода работы формируют специфический стиль философско-методологических исследований ИИ.

Как мы видим, в частности, по представленной работе, научный уровень исследования ИИ тесно соприкасается с философской методологией ИИ. В настоящее время данные уровни исследований представлены, как правило, синкретично, так как общенаучная методология, которой пользуются практикующие специалисты ИИ, часто отождествляется с философской методологией. Но критерий демаркации между философско-методологическим и научно-методологическим знанием имеется и, пожалуй, самый главный – способность философской методологии соприкоснуться с проблематикой философского мировоззрения. Другие параметры – неспособность общенаучной методологии выразить фундаментальные отношения – онтологические, гносеологические (эпистемологические – для научного дискурса), аксиологические, праксеологические, культурологические отношения – представляются второстепенными критериями демаркации общенаучной методологии от философской методологии.

Мы подробно остановились на работе Н.Ю. Ключевой (Трушкиной), так как, на наш взгляд, данная работа крайне показательна – какого рода исследования, какой степени общности и предметной направленности они должны быть, чтобы их рубрицировать как исследования в области философской методологии искусственного интеллекта.

Научно-методологический уровень исследований искусственного интеллекта

На данном уровне изучения программно-инженерные разработки подводятся под определение системы, которое принимается и конвенционально закрепляется в коллективе разработчиков в качестве *ad hoc* понятия интеллектуальности. Следует отметить несколько определений интеллекта (искусственного интеллекта, интеллектуальной системы), предложенных специалистами в этой области [Pei, 1995] (см. также [Алексеев, 2006-6]). Искусственный интеллект – это: 1) способность настолько быстро находить правильное решение в обширном пространстве поиска, что для стороннего наблюдателя данная способность кажется априорной (*Ленат и Фейгенбаум*, 1991); 2) идентификация значимых проблем обработки информации и решение этих проблем (*Марр*, 1977); 3) методы достижения целей в сложных ситуациях и проблем, представленными в форме ситуаций и не зависящие от того, кем они решаются – человеком, марсианином или компьютерной программой (*Маккарти*, 1988); 4) способность компьютеров решать сложные проблемы (*Минский*, 1985); 5) характеристики, обнаруживаемые во всех человеческих действиях и проявляющиеся в любой реальной ситуации стремиться к системной цели посредством адаптации с некоторой скоростью к выявляемым требованиям среды определённой степени сложности (*Ньюэлл и Саймон*, 1976); 6) способность достижения «лучшего» за приемлемое время (*Шэнк*, 1991); 7) способность информационной системы адаптироваться к среде в условиях нехватки знаний и ресурсов (*Янг*, 1995); 8) наука о том, как усилить человеческие возможности в решении сложных задач (*Г.С. Осипов*, 2005); 9) системы, работающие на знаниях и включающие логический решатель задач (*В.К. Финн*, 2005); 10) квазиалгоритмическая реализация Теста Тьюринга (*А.Ю. Алексеев*, 2006-1;5).

Имеется ещё ряд определений ИИ. Очевидно, что их большое количество. Однако никто не может дать определение «интеллекта» на «все случаи жизни» и все они заслуживают внимания с известной долей осторожности.

Инженерно-методологический уровень исследований искусственного интеллекта

Инженерно-методологический (программно-инженерный) уровень – это уровень разработки прототипов и показательных проектов интеллектуальных систем и интеллектуальных информационных технологий, которые служат презентантами целого класса инженерных конструкций, запущенных после проведения опытно-конструкторских работ в производство и используемых в быту, а так же которые предназначены для эвристического прироста теоретического знания в сфере ИИ.

В инженерии ИИ методологические исследования распределены по следующим направлениям [Mohr, 2000]:

Игры (Игра в шахматы, Samuel, 1963);

Доказательство теорем и автоматизация рассуждений: «Логик-теоретик» Ньюэлла, 1956;

Решение проблем и планирование: а) «Универсальный решатель проблем» (Ньюэлл, Шоу и Саймон, 1963); б) программа СТРИПС Нильсона и Файка, 1971, созданная для управления мобильным роботом, названным Шейки (Shakey);

Восприятие: а) анализ письменного текста; б) распознавание изображений (image processing); в) анализ сцен; г) слежение за движением (motion tracking); д) распознавание лиц (face recognition);

Распознавание речи: а) распознавание отдельных выражений (isolated-word recognition); б) распознавание связной речи (continuous-speech recognition): программа HEARSAY II, 1975 г.);

Распознавание образов (pattern recognition);

Понимание естественного языка и машинный перевод (Natural language understanding and machine translation): программа Винограда SHRDLU (начало 1970 г.);

Экспертные системы: а) символическая логика (symbolic mathematics): программа Мозеса МАКСИМА (MACSYMA, 1977 г.); б) медицинская диагностика (medical diagnosis) МУСИН Шортлиффа и Бачмена (MYCIN, 1976); в) химический анализ: DENDRAL Фейгенбаума и Леденберга, 1964 (Feigenbaum and Lederberg); геологическая экспертная система: PROSPECTOR Дуды и Харта (Duda and Hart, 1979);

Обучение: 1) программа открытий (discovery programs): автоматический математик Лената (Automated Mathematician, 1977); нейронные сети (neural networks);

Робототехника.

На инженерно-методологическом уровне трудно дать рациональное определение понятию искусственного интеллекта. Здесь прослеживается зависимость «интеллектуальности» системы от инструментария её разработки. Так, если программа создана на Прологе, то она поспешно объявляется интеллектуальной в силу признания за Прологом возможностей построения интеллектуальных систем. Либо если моделирование предметной области осуществляется нейросетевыми средствами, то уже в силу этого программная система возводится в ранг интеллектуальных. С подобной точкой зрения диссертант столкнулся в дискуссии с Л.Н. Ясницким (см. выше п.п. 1.1.1).

Выводы: Предложенное в данном пп. деление исследований ИИ на философско-методологический, научно-методологический и инженерно-методологический уровни позволяет более четко специфицировать задачи, возникающие при конструировании систем, способных пройти тест Тьюринга (пп.3.3).

1.2.3. Роль теста Тьюринга в исследованиях электронной культуры

В современном научном дискурсе используется ряд терминов, смыслы которых позволяют соединять феномены культуры с информационно-компьютерными технологиями (ИКТ)¹. «Киберкультура» акцентирует внимание на алгоритмически управляемых способах регуляции жизни. «Цифровая культура» апеллирует к универсальности математических форм выражения объективной и субъективной реальности. «Информационная культура» предлагает формат информационных процессов, систем, технологий. В «семиотической культуре» процессы творения, функционирования и трансляции феноменов культуры отождествляются с динамикой знаковых систем. «Компьютерная культура» предлагает метафору индивидуальной и общественной жизни в виде программы обработки «социокодов». В «медиакультуре» культурологические

¹ Традиционно аббревиатура ИКТ применяется для обозначения информационно-коммуникационных технологий. На наш взгляд, здесь имеет место тавтология. Любая информационная технология включает в себя коммуникативный процесс установления синтаксической, семантической и прагматической общности между участниками информационного процесса. Однако не каждая информационная технология представима в компьютерном формате. Например, принципиально не решаем вопрос компьютерного моделирования смысла сообщения.

феномены погружаются в контекст виртуальной реальности, порождаемой мультимедийными технологиями. «Нейрокультура» переносит модели мозговой активности на социум, риторически связывая способ функционирования культуры с нейрофизиологическими функциями, либо фактически утверждая возможность построения нейросоциума на базе технологий нейроинтерфейсов и нейронета.

В данной работе изучается *«электронная культура»* («э-культура», «e-culture», «eCulture»). С конца 90-х гг. прошлого века этот термин часто употребляется в научном и общественном дискурсе. При употреблении термина «э-культура» явно предполагается то, что данная форма культуры не отождествляется со способами и средствами ИКТ и не редуцируется к ним. Данное положение явно зафиксировано в программном документе Нидерландского совета по культуре *«От информационно-коммуникационных технологий к э-культуре»*, 2003 г.: *«Э-культура – это новая цифровая сфера социокультурной деятельности человека, новые формы выражения человека посредством электронных технологий и интернет, новые роли и статусы институтов культуры и искусств под влиянием технологий»*¹.

По проблемам электронной культуры в настоящее время проводятся масштабные научные конференции, семинары, симпозиумы. Тематика мероприятий специфицируется в соответствии с представлениями организаторов и участников научных мероприятий о данной форме культуры. В нашей стране широко известна ежегодная международная конференция «EVA»². В 1990-е годы в её рамках изучались, как правило, проекты развития электронных визуальных искусств. Сегодня тема конференции существенно расширена до уровня изучения э-культуры как интегрального феномена современной культуры. С 2002 года в г.Сочи проводится ежегодная международная научно-практическая конференция «Электронный век культуры». Исходной темой являлись проблемы э-библиотек. Сегодня тематика так же приобрела расширенный формат. С 2002 года в г.Лодейное поле Ленинградской области проводятся ежегодные научные мероприятия, связанные с тематикой гуманизации э-технологий и развитием

¹ *From ICT to E-culture*. Advisory report on the digitalisation of culture and the implications for cultural policy. Netherlands Council for Culture, 2003 (http://www.cultuur.nl/files/pdf/advies/E-cultuur_engels.pdf)

² Например, в г. Москве 19-21 ноября 2013 года состоялась 15-я ежегодная международная конференция «EVA 2013 Москва: Информационное общество, культура, образование». Тема конференции – «Библиотеки и музеи в цифровой среде: диалог и сотрудничество»; 22 ноября 2013 года состоялась 12-я Международная научно-практическая конференция «Электронный век культуры» (<http://eva.rsl.ru/ru>)

специфических форм э-культуры в малом городе («неопросветительство», сайт «Неопросветительство»: <http://www.neo-prg.ru>). С 2007 г. в г.Астрахани проводятся ежегодные конференции на темы э-правительства, э-демократии, цифрового неравенства. С 2005 г. Научный совет РАН по методологии искусственного интеллекта (НСМИИ РАН) организует мероприятия, посвященные интеллектуализации ИКТ в рамках решения двух общих вопросов: 1) электронная культура – стратегический ресурс развития технологической и духовной сфер общества; 2) междисциплинарные исследования искусственного интеллекта – ключевой фактор развития электронной культуры¹.

Подобного рода дискуссии инициируют создание отделов, департаментов, центров при министерствах культуры различных стран. Особенно активно в сфере э-культуры работают западноевропейские организации, так как они, собственно, и явились пионерами изучения данного феномена. В нашей стране в 1991 г. был создан «Центр по проблемам информатизации сферы культуры» («Центр ПИК») для осуществления научно-исследовательской и инженерно-технической деятельности при решении проблем применения новых информационных технологий в сфере культуры России.

В вузах создаются выпускающие кафедры электронной культуры. В государственных, общественных и коммерческих организациях начинают функционировать информационно-аналитические центры электронной культуры. Столь широкого применения не находил ни один из выше приведённых ИКТ-ориентированных терминов «культуры».

Термин «электронная культура» проник в отечественную философскую литературу. В журнале «Вопросы философии» в 2013 г. в № 8 была опубликована статья Л.В.Баевой «Электронная культура: опыт философского анализа». В начале работы автор справедливо полагает, что детерминантой э-культуры являются высокие информационные технологии. Однако, далее приводятся общие и пространственные рассуждения о понятии информации, затрудняющие понимание

¹ Например, в рамках Московского форума культуры «Культура как стратегический ресурс Российской Федерации в XXI веке» (1 июля 2010 г.) состоялась работа секции «Искусственный интеллект и развитие электронной культуры России». Тематика секции была продолжена на XI Международной научной конференции «Модернизация России: ключевые проблемы и решения», 16-17 декабря 2010 г., ИНИОН РАН. Изучались вопросы: анализ основных направлений и тенденций развития электронной культуры в мировой и отечественной практике; «электронная культура» в федеральной программе «электронной России»; междисциплинарные исследования искусственного интеллекта – главный фактор интенсивного развития электронной культуры; образовательные программы «искусственный интеллект» в технических и гуманитарных вузах; искусственный интеллект – метатехнология НБИК-комплекса; компьютерные технологии аналитики; развитие мобильных сервисов в России.

того, что такое «информационная технология». Отталкиваясь от частной дефиниции э-культуры, предложенной А.Рончи в 2009 г., автор статьи формулирует собственное определение, согласно которому электронная культура представляет собой совокупность результатов творчества и коммуникации людей в условиях внедрения IT-инноваций, характеризуется созданием единого информационного пространства, виртуальной формой выражения, дистанционной технологией, либеральностью контента. Собственно философское осмысление феномена э-культуры, по мнению Л.В.Баевой, заключается в подведении «школьных» философских дисциплин под данную дефиницию. В связи с этим выделяются онтологический, социально-философский, антропологическо-аксиологический, этический и эстетический «ракурсы» этого определения.

Поскольку трактовка понятия «э-культура» крайне обширна, полисемична, многомерна, то, с нашей точки зрения, неправомерно тормозить философскую работу из-за узости исходного определения. Это в некоторой степени послужило поводом данной работы.

Первый этап «философского осмысления», исходя из традиции философско-лингвистического исследования, состоит в изучении различных контекстов употребления слова «э-культура». Вариации концептуального многообразия термина фиксируются достаточно чётким контекстом его употребления¹. После установления границ, в которых употребляется термин, наступают этапы онтологического, эпистемологического, аксиологического, праксеологического и других видов анализа. Если данной схемы не придерживаться, то просто-напросто нарушается первый логический закон $a=a$, когда исследователь начинает путать значения слов из-за различных способов их употребления.

Постнеклассический метод формирования контекстов термина «э-культура»

В соответствии с широко распространенными методологическими приемами изучения сложных социокультурных явлений, впервые предложенными В.С.Степиным в конце 1980-х годов, выделим три фундаментальных метода: классический, неклассический и постнеклассический.

Классический метод пассивен. Исследователь «объективно» изучает явления культуры по аналогии с тем, как изучаются явления природы. Э-культура накладывает специфику на данный метод: появляется возможность анализиро-

¹ Данный подход предлагался в [Алексеев, 2009-7; 2011-4].

вать её феномены собственными средствами, в частности, осуществлять информационно-аналитическую работу с текстами электронных библиотек и сайтов, отражающих тематику э-культуры. Такой анализ поддерживается программами поиска по ключевым словам, компьютерными средствами фактографического, концептографического, контентного анализа и пр. По результатам информационного поиска формируются кластеры, которые в последующем категоризируются и, наконец, контекстуализируются. Категориальные кластеры – это самые верхние совокупности терминов в иерархии определения степеней смысловой близости/дальности. Контекст – это поверхность, разделяющая эти классы общих понятий. Если далее проводится информационный мониторинг, то поисковые образы, на основе которых сформировались кластеры, выступают в роли индикаторов феноменов «э-культуры» для интернет-навигаторов.

Неклассический метод активен. Феномены культуры таковы, каковы методы их концептуального конструирования. Эти методы надо создать или, по крайней мере, проделать работу по выбору требуемого метода из совокупности наличных с учетом соответствующей модификации. Исследователь формирует дименсионалистское (многомерное, многоаспектное) представление предмета. Показательным примером неклассического метода является организация и проведение междисциплинарной конференции, на которой сталкиваются мнения, взгляды, концепции представителей различных научных школ и направлений. Исследователь (например, координатор конференции) стремится расширить предметное поле дискуссий и усилить конфронтацию между участниками. При столкновении взглядов, идентификации общих, особенных и различающихся значений, координатор создает своеобразный метаметодологический каркас, который систематизирует частные методологии и подходы к изучаемому явлению. Неклассический исследователь не проникает в суть предметной области. Он «скользит» по многообразию идей, концепций, представлений. Траектории «скольжения» формируют контексты употребления терминов.

Примерно по данной схеме в рамках деятельности НСММИ РАН был инициирован проект постоянно действующей всероссийской конференция «Электронная культура: интеллектуальные технологии»¹. В данном словосочетании термин «интеллектуальные» употребляется не случайно. В принципе, по стандартным функциям ИКТ невозможно судить об их роли в культуре. Они

¹ Проект постоянно действующей всероссийской конференции «Электронная культура: интеллектуальные технологии» был запущен в июне 2011 г. по инициативе НСММИ РАН: www.ec-ai.ru (председатель: академик В.Л. Макаров).

амбивалентны относительно целей и намерений их применения. Например, ИКТ, обслуживающие цели ЕГЭ, внедряют формальный численный подход к оценке когнитивных способностей учащихся. Вряд ли они вносят позитивный вклад в культуру, скорее, напротив. Из-за этой электронной технологии э-культура самоуничтожается: креативные люди вымирают и в будущем некому будет развивать новую форму культуры, по крайней мере, в нашей стране. По поводу же «интеллектуальных ИКТ» ни у кого нет сомнений в том, что они являются детерминантой э-культуры. Благодаря наукоёмкой междисциплинарной инструментальной поддержке и разумной регуляции процесса внедрения и использования ИКТ, эти технологии, в силу определения, стимулируют творческую деятельность человека, способствуют сохранению и развитию социокультурных традиций, согласованы с ними и не имеют негативных последствий. В целом, проводимая конференция посвящена рациональному внедрению и использованию инноваций в сфере культуры и искусств. Стоит задача: исследовать социокультурные параметры инновационных стратегий государственных и общественных программ модернизации России в переходный период расширения пространства традиционной культуры возможностями электронных технологий. Выявление среди участников различных контекстов употребления слова «э-культура» возможно благодаря междисциплинарным дискуссиям по актуальным философским, научным и практическим вопросам развития интеллектуальной электронной культуры, а так же организации взаимодействия по этой проблематике между деятелями науки, культуры, искусства и бизнеса, представителями законодательной и исполнительной власти.

На этапе сбора заявок и проведения ряда секций выявилось разнообразие трактовок термина «э-культуры» и связанных с этим термином сети терминологических ассоциаций. С «э-культурой» отдельные ученые и представители научных школ связывают следующее: инновационные возможности инструментария ИКТ; стратегии лидерства государства в сфере высоких технологий с учётом сохранения и умножения традиционных культурных ценностей; программы государственной модернизации; инновационный потенциал в сфере культуры и искусств; инструментарий моделирования социокультурных явлений; аналитические технологии в социокультурной сфере; интеллектуальные системы в обществе; когнитотехнологические достижения; искусственные общества; виртуальные миры; информационную культурологию; инновационные технологии в образовании; технологии информационной аналитики в управлении обществом; виртуальный музей; электронную библиотеку; умный дом; смарт-обучение; вы-

сокие технологии в современной культуре; социокультурные приложения ней-рокомпьютинга; дизайн, декоративно-прикладное и экранное искусство, поддерживаемое средствами ИКТ; количественные методы «измерения» явлений культуры; неопросветительство в малом городе; электронную филантропию; инновационные проблемы э-политики и др.¹.

Как мы видим, имеется большое разнообразие мнений и представлений о том, что такое «э-культура». По сравнению с классическим методом, неклассический подход значительно увеличивает количество интерпретаций этого термина, причем в пропорции «нелинейной дурной бесконечности»: каждое новое значение порождает новую сложную сеть ассоциаций. Источников терминологического разнообразия бесконечно много. Так, упомянутая выше ежегодная конференция EVA, проводимая в Москве, собирает тысячи докладов. Далее эта конференция движется по всему миру, каждый месяц останавливаясь в столице или крупном городе той или иной страны – в Лондоне, Париже, Пекине и пр. Очевидно, что никакая автоматизированная система контент-анализа (классический подход) не справится с выделением простых и составных ключевых слов, характеризующих «э-культуру». Неклассический подход вообще исходит из презумпции релятивности восприятия феноменов э-культуры и способов их познания. Причем состав, структура, содержание комплекса субъективных представлений темпорально зависимы, они являются функцией от момента и интервала их изучения.

Поэтому предлагается обратиться к феноменологическому изучению электронной культуры, включенному в постнеклассическую исследовательскую парадигму.

Постнеклассический метод конститутивен. Это метод смыслового оформления феноменов культуры с позиции *Я* как носителя, источника, потребителя э-культуры. Вряд ли такой субъект имеет трансцендентальный статус, скорее, он находится в поле интересубъективных взаимодействий. *Я* не только вырастает и живет в культурной среде, имманентно разделяя её ценности и смыслы. Он активно влияет и, более того, формирует её. Если в условиях традиционной культуры такое воздействие осуществить сложно, то в э-культуре, благодаря её ИКТ-инструментарии, трансляция субъективных смыслов, значе-

¹ В обсуждении принимают участие ведущие академические учреждения и вузы страны. В мае 2011 г. было проведено пленарное заседание и заседания секций. Результаты конференции требуют оформления, однако из-за отсутствия минимальной финансовой поддержки

ний, ценностей осуществляется мгновенно и массово. Например, используя ресурсы известной э-библиотеки eLibrary (РИНЦ), каждый автор имеет возможность донести свои идеи до широкой публики, не впадая в зависимость от стоимости бумаги, логистики распространения тиража и коммерческого интереса издателя.

Наиболее очевидным положением является то, что предметы традиционной культуры представляются, воплощаются, имплементируются, реализуются средствами ИКТ: в э-носителях информации и алгоритмах функционирования э-технологий. Такой контекст употребления термина «э-культура» мы обозначили термином «репрезентативный контекст» [Алексеев, 2010-1]. Новые отношения по производству информации и знаний, обусловленные развитием ИКТ, приводят к модификации старых и к формированию новых социальных структур: институциональный контекст. В институтах э-культуры модифицируются традиционные человеческие отношения по поводу власти, авторитета, распределения ценностей: политико-экономический контекст. Трансформируются и исчезают старые ценности, на смену которым приходят новые: аксиологический контекст. Нередуцируемость э-культурных феноменов к ИКТ в первую очередь проявляется в симультанных, естественно-искусственных способах восприятия и познания мира: виртуалистский контекст. Новое восприятие действительности обуславливает становление нового человека, человека электронной культуры: антропологический контекст. Несмотря на значительную трансформацию природы человека под непосредственным или опосредованным воздействием ИКТ, э-культура должна сохранять свою идентичность, по крайней мере, в своих высших формах проявления, среди которых доминируют разум, развитый интеллект, рациональность: ноологический контекст. Рациональность предполагает критическое познание возможностей реализации социокультурных феноменов средствами ИКТ: эпистемологический контекст. И, наконец, каково бытие *Я* и общества в мире электронной культуры: онтологический контекст.

Кратко обозначим выделенные контексты термина «э-культура».

1. Репрезентационный контекст

Ким Вельтман, один из пионеров определения «электронной культуры», под э-культурой понимает репрезентацию средствами ИКТ традиционных объ-

ектов культуры и искусств [Вельтман, 2002]. Под данное определение подпадают: э-книга как представление традиционного издания на электронных носителях информации; э-библиотека как системным образом организованная совокупность э-книг с локальным или сетевым режимом доступа к ней; виртуальный музей как коллекция оцифрованных объектов культуры и искусства; мультимедийная реконструкция исторического памятника; «семантическая сеть» в среде интернет, систематизирующая предметы культурологии, археологии, этнографии, искусствознания, историографии и пр. В репрезентативном контексте интересно онтологическое суждение специалиста Центра «ПИК» Н.В. Браккер: **«Существовать – значит быть представленным в интернете»**¹. В самом деле, если, например, отсутствует сайт организации, то её как бы и нет. Отсутствуют научные труды в РИНЦ – ученого вроде бы и не существует. Возможно, этот ученый велик и знаменит, но, начиная с 2014 г., при отсутствии трудов в РИНЦ и Scopus, российского научного гранта этот ученый не получит. Картезианскую формулу бытия (*cogito ergo sum*) уже сегодня следует переписать как *e-re-praesento ergo sum*, где «e-re-praesento» переводится как «электронная репрезентация».

2. Институциональный контекст

В рамках этого контекста изучаются социальные институты в условиях применения ИКТ. Социальная кластеризация новой формы культуры отличается от структуры традиционной культуры. Однако в смысловом плане различия несущественны, так как э-институты (термин предложен А.Б.Антопольским) выполняют функции, аналогичные функциям социальных институтов традиционной культуры. К э-институтам относятся: э-образование (*e-learning*), э-политика (*e-polity*), э-наука (*e-science*), э-торговля (*e-commerce*), э-туризм, э-медицина и др. (М. Шварц) [Schwarz, 2006]. По поводу э-институций возможны различные трактовки. Например, принципиально различаются трактовки проекта «э-науки». В Великобритании э-наука – это логистическая система организации научных исследований: продуцирования информации, её сложной обработки и разветвлённого распределения среди больших групп менеджеров и учёных исследовательских лабораторий, крупных университетов, правительственных учреждений. В США под э-наукой понимается «киберинфраструктура», обеспечивающая централизованное использование крупных суперкомпьютер-

¹ Суждение высказано в докладе на конференции EVA-2008 (Москва), <http://conf.cpic.ru/eva2008/rus/info/default.asp>

ных центров для организации сложных вычислений при решении задач ядерной физики, наук о жизни, социологии и пр.

Отметим факт того, что уже сегодня многие э-институты непосредственно инкорпорированы в организацию повседневной жизни.

3. Политико-экономический контекст

Новую окраску обретают идеи диалектического и исторического материализма (К.Маркс, Ф.Энгельс, В.И.Ленин), например, такие: культура – область производства материальной и духовной жизни; человек – субъект исторического развития, творимый культурой и творящий её по своему образу и подобию как целостное действительное существо в процессе трудовой деятельности; в процессе труда складываются объективные производственно-экономические отношения, выражающие отношения людей через их отношения к средствам производства (отношения собственности); при общественной собственности члены общества равны по отношению к средствам производства и между ними существуют отношения сотрудничества; на основе частной собственности между людьми устанавливаются отношения эксплуатации человека человеком; люди, лишённые средств производства, неизбежно оказываются в экономической зависимости от собственников средств производства, что предопределяет отношения господства и подчинения между ними. В итоге, эти отношения детерминируют культуру.

В условиях э-культуры отношения господства и подчинения проявляются в рамках проблемы т.н. «цифрового неравенства»: невозможностью или ограниченностью доступа к ИКТ для определенных людей и социальных групп («Digital Divide», термин введен в употребление в 1995 году в результате исследований государственной администрации телекоммуникаций и информации США). В условиях информационного общества доступность к ИКТ означает «владение средством производства». Имеются значительные различия в доступе к ИКТ среди людей с различным доходом, уровнем образования, расовой и гендерной принадлежностью и др. Так, всего лишь треть населения Земли имеет постоянный доступ к сети Интернет. Остальные две трети как бы выпадают из всемирной э-культуры.

Выделяются две компоненты цифрового неравенства: 1) отсутствие базисных составляющих физического доступа к ИКТ: экономически, технологически, территориально обусловленное цифровое неравенство; 2) отсутствие социально-гуманитарных условий для доступа к ИКТ, обусловленное возрастной,

гендерной, культурной, национальной, расовой дискриминацией. *Экономическая составляющая цифрового неравенства* определяется тем, что субъекты не способны полностью включаться в существующую экономическую систему, если у них нет доступа к современным ИКТ. *Технологическая основа цифрового неравенства* обусловлена отсутствием базовой инфраструктуры, подходящей для развития ИКТ. *Территориальная основа цифрового неравенства* обусловлена постоянной или временной невозможностью развития ИКТ на определенных территориях в силу их географического положения. *Возрастная основа цифрового неравенства* базируется на дискриминации по возрастному признаку: у людей старшего поколения в силу возрастных особенностей снижена потребность в новых технологиях и способность к их освоению. *Гендерной основой цифрового неравенства* является дискриминация по признаку половой идентичности: женщины в меньшей степени, чем мужчины, способны к освоению ИКТ, и в меньшей степени в них нуждаются. *Культурная, национальная, расовая основы цифрового неравенства* базируются на ксенофобии: представители определенных сообществ из-за своих психобиологических особенностей не способны должным образом использовать ИКТ.

При преодолении цифрового неравенства возникает противостояние двух целей: 1) эмансипация социальных групп и отдельных индивидов, 2) усиление экономической эксплуатации. Эмансипирующий потенциал информационных технологий заключается в предоставлении возможностей построения индивидуальной стратегии работы с информацией и знаниями. Важно не только потреблять, но и создавать информационные ресурсы. Современные средства автоматизации проектирования ИКТ и стандарты в области юзабилити (пригодности для пользователя), в принципе, позволяют реализовать новый тип участника ИКТ-процесса: пользователь-разработчик. Для этого субъекта важным становится аспект сотрудничества. Поэтому возникают массовые движения за открытость знания и открытое программное обеспечение. Однако возникает конфликт между знанием как коммерческим ресурсом и знанием как общественным благом.

Технологии обеспечения цифрового неравенства значительно усиливают традиционные способы экономической эксплуатации. Так как государство (перефразируем классика политэкономии), суть машина для поддержания господства одной социальной группы (правящей) над другой посредством установления и усиления цифрового неравенства, то вряд ли оно заинтересовано в развитии э-культуры. В нашей стране данное утверждение находит подтвер-

ждение, по крайней мере, в «Проекте Основ государственной культурной политики»¹. В этом проекте нет ни одного слова про государственные обязательства в сфере информатизации культуры. Интересно, что «поддержка и развитие благоприятной для становления личности информационной среды» (см. пп. 4 «Проекта») возлагается на самый нищий в нашей стране институт культуры – на библиотеку (!?)

Кто в условиях капитализма владеет электронной технологией, тот, посредством установления и усиления цифрового неравенства, формирует э-культуру.

4. Виртуалистский контекст

Электронная культура – качественно новая социокультурная реальность, не имеющая естественных аналогов. Приложения оцифрованной информации не сводятся к инструментальному использованию технологий. Компьютеры, мультимедиа и интернет открывают новые формы представления и выражения человека. Изменяются роли и статусы традиционных институтов культуры и искусств. Например, виртуальный музей – это не просто коллекция графических образов культурного наследия, сохраненная и воспроизведённая на компьютере (для репрезентативного контекста это была именно коллекция). Напротив, это новая форма интерпретации исторической реальности, которую «проживает» посетитель музея (А. Рончи) [Ronchi, 2008]. Утверждается новое качество восприятия действительности. Известный э-культуролог О.В. Шлыкова акцентирует на феномене мультимедийной реальности и отмечает два типа определения э-культуры: 1) это – новый тип культуры трансформации, культуры «метаморфозы», обусловленный развитием современного этапа электронных коммуникаций; 2) это – не техногенное явление, но социокультурный феномен, влияющий на ментальность, культуру в целом, способы восприятия, трансляции социальной информации [Шлыкова, 2009].

5. Аксиологический контекст

Так как феномен э-культуры – это симультанная производная от естественных способностей человека и ИКТ-средств, то в ней появляются новые и коренным образом модифицируются социальные отношения, культурные ценности, регулятивы, нормы поведения. В этих отношениях начинают доминиро-

¹ См. официальный сайт для обсуждения проекта: <http://obsudiproekt.ru>. Проект вынесен на обсуждение 23 мая 2014 года, которое завершилось 30 сентября 2014 года.

вать компьютерно-технологические стереотипы деятельности, как полагают Джон де-Хаан и Франк Хьюсман: технологические знания составляют существенную и неотъемлемую составляющую э-культуры, её новые ритуалы и нормы¹. К.К.Колин дополняет: «Электронная культура – это новая часть современной культуры, которая формируется на основе использования средств и методов информатики, информационно-коммуникационных технологий, связанных с их развитием семиотических систем общества, а также с изменением мировоззренческих и нравственных универсалий информационного общества» [Колин, 2011].

6. Ноологический контекст

Развитие электронной культуры обусловлено разработкой и внедрением интеллектуальных ИКТ. Укрепим это утверждение двумя продуктивными суждениями: 1) насколько в традиционной культуре важна роль естественного интеллекта как целостная способность упорядочивать чувственные данные, рационально рассуждать и разумно регулировать индивидуальное и социальное поведение, настолько в э-культуре значима роль интеллектуальных компьютерных систем, реализующих подобные функции; 2) насколько человеческий разум обуславливает расцвет традиционной культуры, настолько креативно-эвристический потенциал интеллектуальных электронных технологий задаёт степень развития э-культуры.

В этой связи интересна метафора, высказанная в 1980-е г. Ю.М. Лотманом: «Культура является интеллектуальным “аппаратом” цивилизации, своеобразным “искусственным интеллектом”»². В контексте традиционной культуры данное утверждение имеет лишь риторический статус. Однако в э-культуре оно не лишено онтологической притязательности. Предполагается, что интеллектуальные технологии, благодаря наукоёмкой междисциплинарной инструментальной поддержке, стимулируют творческую деятельность человека, способствуют сохранению и развитию социокультурных традиций, согласованы с ними и не имеют негативных последствий.

На наш взгляд, электронная культура – это культура, в которой духовная и материальная составляющие *форматируются* технологией искусственного интеллекта – технологией извлечения, репрезентации, хранения, обработки, пе-

¹ Jos de Haan, Frank Huysmans, 2002.

<http://www.scp.nl/english/publications/summaries/9037700926.html>

² Цит. по [Финн, 2007, с.30].

редачи «знаний». Ноологическая составляющая любой высокоразвитой культуры очевидна. Так же не вызывает сомнений то, что в основе любой технологии лежит *интеллектуальная* деятельность, человеческий интеллект. Тогда, по аналогии, в э-культуре возрастает роль электронного имитатора и усилителя интеллектуальной деятельности – искусственного интеллекта (ИИ). В новых условиях значительно расширяются идеи полувековой давности по поводу компьютерного моделирования свойств и способностей мышления, памяти, интуиции, сознания, творчества, обучения и пр. Современные исследования ИИ сфокусированы на проблемно-ориентированных интегративных областях. С начала 21 века институционально закрепляются следующие направления ИИ: «искусственная жизнь», тесно связанное с исследованиями биотехнологии; «искусственная личность», акцентирующее внимание на моделировании морального поведения х-систем, которые внешне оцениваются как обладающие свободой воли самосознающие индивиды; «искусственное общество» - компьютерный фундамент имитации большого разнообразия феноменов общественной жизни, начиная от материально-производственных отношений и завершая религиозными верованиями; «искусственный мир», которое сегодня недостаточно развито, однако заявлен ряд претензий со стороны т.н. «цифровых философов» по поводу компьютинга «вселенной». Данным исследованиям не только в силу исторических причин их становления присваивается ярлык «искусственный интеллект». В основе проектов – творческая деятельность человека, естественные интеллектуальные способности которого поддерживаются и усиливаются инструментарием электронной технологии искусственного интеллекта.

7. Антропологический контекст

Э-культура – это культура человека нового вида, е-Номо. Э-человек сохраняет биологическую принадлежность к *Homo Sapiens*, но качественно отличается от него за счет симбиоза с продуктами стремительно развивающихся сверхвысоких технологий НБИКС. Среда «э-обитания», естественный «э-отбор», «роботы внутри нас», «э-свобода» и пр. – это параметры э-культуры э-человека. Э-человек появится при жизни основной части нынешнего поколения (А.С. Нариньяни) [Нариньяни, 2005-2]. Несомненно, антропологический контекст намного важнее всех остальных, если считать верным кантрианское положение о том, что квинтэссенцией всех возможных исследований является ответ на вопрос «Что такое человек?». Философская антропология как методология информационной антропологии (К. Колин) предполагает необходимость проведения масштабных онтологических, эпистемологических, аксиологических и

праксеологических исследований э-культуры. Главным «инженером» э-технологий становится эксперт в сфере гуманитарных и социальных наук.

8. Эпистемологический контекст

Для рационального познания феноменов э-культуры предпочтительно воспользоваться принципом древних: «подобное поверять подобным». Применительно к исследованиям э-культуры это означает изучение категориальных оснований ИКТ и принципов построения компьютерных машин, реализующих эти технологии. Принципы современного компьютеринга редуцируемы к машине Тьюринга (1936 г.). В свою очередь, принципом исследования социокультурных феноменов предлагается считать базовые идеи, предложенные в тесте Тьюринга для доказательства возможности компьютерного «мышления». Тест Тьюринга является компьютерной дефиницией «интеллекта» и формальным определением понятия «искусственный интеллект», по аналогии с тем, что «машина Тьюринга» является формальным определением понятия «алгоритм». Тест существенно модифицирован в многочисленном ряду его последующих версий – *частных тестов Тьюринга*. Это – компьютерные дефиниции многообразных физических, витальных, ментальных, персональных, социальных феноменов, таких, как «интеллект», «жизнь», «творчество», «познание», «понимание», «любовь», «свобода», «общество» и пр. (Дж.Маккарти, Д.Деннетт, Н.Блок, К.Колби, Дж.Серль, Р.Френч, Ю.Генова, С.Ватт, А.Сломан, С.Брингсйорд, Дж.Лукас, Р.Пенроуз, Дж.Мур, Дж.Баресси, С.Харнад, П.Швайзер, Дж.Поллок, Р.Кирк, Д.Чалмерс, В.К.Финн, 1950-2013 гг.). Логично интегрировать частные тесты в комплексном формате [Алексеев, 2006-1; 2013-3]. *Комплексный тест Тьюринга* – с одной стороны, способ интеграции частных тестов. С другой стороны – это формальное определение понятия искусственного общества, и, более конкретно, электронного общества [Алексеев, 2011-3].

В эпистемологическом плане комплексный тест Тьюринга играет важную роль. Во-первых, он осуществляет функцию рациональной *дефиниции* социокультурных феноменов: «сознания», «другого», «личности», «Я» и пр. Это исключает смысловой дискомфорт при употреблении ИКТ-ориентированных понятий, таких как «машинное мышление», «компьютерное сознание», «человек-компьютер», «любовь робота» и пр., так как определяются феномены х-систем, субстрат которых не существенен или совсем не известен. Например, приемлемо тестировать интеллект, сознательность, любовь, нравственность и пр. неко-

торого социального института, в ходе тестирования общаясь с ним, как с личностью.

Второй функцией является *критика проектов электронной культуры*. Предлагается деструктивно-конструктивный формат аргументированного решения вопросов «Может ли компьютер мыслить, понимать, ощущать, жить, любить, осознавать, быть личностью, Я и пр.?». Например, легко доказывается неинтеллектуальность экспертных компьютерных систем. Парадигма их построения основана на универсально-эпистемологическом подходе, с позиции всезнающего разработчика, однако необходимо включать в эти системы эволюционно-эпистемологические механизмы, которые моделируют рост знаний у ребенка [Алексеев, 2013-2]. Схожим образом элегантно решается вопрос об иррациональности государственных решений, вынуждающих пользоваться главными «компьютерными вирусами-мемами» современного российского вузовского образования: компьютерными средствами оценки параметров ментальной сферы человека. Степень подготовленности абитуриента оценивается автоматизированной системой единого государственного экзамена; знания студента – средствами федерального Интернет-экзамена профессионального образования, авторитет преподавателя – э-библиотечными программами многочисленных индексов цитирования¹.

В-третьих, комплексный тест Тьюринга осуществляет *функцию конструирования* новых принципов ИКТ-технологий на фундаментальном и прикладном уровнях. Например, обосновывается стратегия переориентации технологии ИИ от проблематики моделирования феномена сознания на решение более прагматичных задач имитации, моделирования и репродуцирования творческой деятельности [Алексеев, 2013-2]. Компьютерное моделирование феномена творчества, а не иллюзорного феномена сознания представляется направлением «главного удара» в доктрине развития интеллектуальных ИКТ и согласуется с

¹ См. критику неразумного использования средств ИКТ для оценки ментальных феноменов, впервые представленную на страницах журнала «Полигнозис» в [Алексеев, 2009-1]. Продолжение критики на базе комплексного теста Тьюринга непосредственно переносится на анализ нерационального использования ИКТ для политического обслуживания современной «демократии». Каким образом простой арифметический подсчёт голосов за имидж лидера средствами автоматизированной системы выборов способен выразить параметры социального идеала? ИКТ обладают значительно более мощным, некалькуляционным потенциалом для э-политики: дистанционной организацией выборной кампании, введением системы экспертных оценок, распределением голосов по конкретным пунктам программы политических лидеров вне зависимости от их харизматических компетенций, как правило, навязываемых со стороны СМИ; формированием структуры представительной власти посредством сетевых сообществ и др.

современными представлениями о социокультурной значимости творческой деятельности [Самохвалова, 2007].

Прикладной статус конструирующей функции заключается в разработке и обосновании принципиальных проектов компьютерных систем. Категориальный статус таких проектов сравним с машиной Тьюринга. Например, предложенный автором проект комплексной машины Корсакова-Тьюринга на абстрактном теоретико-алгоритмическом уровне интегрирует коннекционистскую и репрезентативную парадигмы ИИ. Это позволяет кардинально реформировать современный компьютеринг и изменить облик э-культуры, так как в её развитии равноправные роли получают «математик» (метафора машины Тьюринга) и «художник» (метафора машина Корсакова) [Алексеев, 2012-2].

Комплексный тест Тьюринга является концептуальным инструментарием организации «чистых» дискуссий по поводу рационального познания феноменов э-культуры и развития ИКТ-проектов: к аргументации не примешиваются идеологические, политические, экономические, финансовые, религиозные, морально-этические, эстетические и иные соображения. Способности познания э-культуры формируются собственными концептуальными средствами э-культуры.

9. Онтологический контекст

Онтология э-культурного мира эксплицируется на трех уровнях. Первый уровень интуитивно связывается с достижениями электронных технологий: э-культура – это всё то, что посредством этих технологий способствует сохранению и умножению материальных и духовных благ человека, общества, человечества. Определение в высшей степени абстрактно и референт для нашего термина подобрать невозможно.

Вторая дефиниция конкретнее: предполагается наличие некоторой культурологической концепции, проекта, программы, которые потенциально или актуально реализуются высокими технологиями. Так как high-tech сегодня связываются с комплексом нано-, био-, инфо-, когни-, социо-технологий, то правомочно такое определение: э-культура – это культура, существование и развитие которой обусловлено технологиями интегративного комплекса НБИКС. Комплекс НБИКС представляется сложным конгломератом «электронных» технологий – целенаправленных и регулярно воспроизводимых способов использования электронных устройств. Вряд ли возможно четкое определение «электронного устройства». Первый «патент на изобретение» такого устройств

приписывается Фалесу (VII век до н.э.): он описал опыты по способности янтаря (с др.-греч. ἤλεκτρον – «электрон») притягивать легкие предметы. С этих опытов предлагается составлять летопись становления э-культуры.

Электронные устройства nano-технологий – это нанороботы. Несмотря на то, что размеры такого робота сопоставимы с размерами молекулы, он обладает достаточно широким спектром функций поведения, обработки, передачи информации и исполнения специфических программ. В биотехнологиях нанороботы обеспечивают манипуляцию формами жизни на клеточном уровне, обуславливая изменения на тканевом, организменном, популяционно-видовом, биогеоценотическом и биосферном уровнях витального. Нано- и микроэлектронная техника составляют современный базис реализации информационных технологий как электронных преобразований информации, представленной на ее электронных носителях. В когнитивных технологиях подобные репрезентации отражают и конструируют отношения психического/нейрального. В социальных технологиях результаты НБИК-технологий комплексно используются для регуляции индивидуальной и общественной жизни.

Третий уровень употребления термина «э-культуры» самый конкретный. На нем концепт экзистенции *Я* вписывается в некоторую концепцию э-культурного бытия, т.е. в картину высокотехнологичного мира. Требуется в способы функционирования технологий, обеспечивающих и воспроизводящих *мою* э-культурную жизнь, включить *мое* сознание с его невероятно сложным комплексом мыслей, переживаний, болей, намерений и пр. Такое требование принципиально не согласуется с онтологическим тезисом репрезентативного контекста («e–re–praesento ergo sum») и самое поверхностное феноменологическое исследование альтер- и интерсубъективных отношений в контексте э-культуры раскрывает пустую формальность этого тезиса.

Возникает невероятно сложная философская задача смысло-жизненного плана. По всей видимости, её следует начинать с поиска аналогов решения подобных, но более простых задач в контексте традиционной культуры. Одна из таких задач является доминирующей в современной философской науке: «проблема сознания». Ранее на страницах журнала «Полигнозис» была выдвинута гипотеза о том, что в контексте э-культуры психофизиологическая проблема трансформируется в психотехнологическую проблему. «Проблема сознания» приобретает значительно расширенную трактовку, так как вполне корректно поднимается интригующий вопрос, который затруднительно было поставить в контексте традиционной культуры: можно ли феномен сознания *реализовать*

посредством электронной технологии? [Алексеев, 2010-2]. Под «реализацией» понимается: 1) имитация, заменяющая человека в тех или иных сферах социокультурной жизни; 2) моделирование, осуществляемое для изучения индивидуальных и общественных форм сознания; 3) репродуцирование, воспроизводящее функции сознания на искусственных системах; 4) креацирование, продуцирующее такие психические феномены, которых нет в природе.

Какие теории сознания приемлемы для решения данной проблемы? В принципе, все имеющиеся на сегодняшний день, так как в новых условиях происходит переинтерпретация этих вечных идей [Лекторский, 2006]. При наличии некоторых смысловых связей с технологией – на концептуальном уровне их всегда можно обнаружить с той или иной степенью онтологических притязаний – приемлемыми оказываются идеализм, физикализм, бихевиоризм, спиритуализм, эпифеноменализм, дуализм, дуаспектизм и многие другие. Например, в последнее время стала модной концепция панпсихизма – наиболее древний отголосок античной натурфилософии. Эти теории способны многое описать и объяснить. Однако в условиях прохождения сурового фильтра компьютерной методологии и инженерии, они оказываются несостоятельными.

На наш взгляд, для решения новой проблемы сознания наиболее адекватным является функционализм в высшей форме его соответствия реалиям компьютерного мира – *компьютеризм*. В противовес главным оппонентам (физикализму и бихевиоризму) функционализм предлагает более комплексный формат изучения психофизической реальности: психический феномен (боль, мысль, желание, убеждение) причинно зависит не только от внутреннего состава и строения системы (физикализм) и не только от внешней поведенческой диспозиции (бихевиоризм), но и от функции (роли), которую данный феномен выполняет в когнитивной системе в целом. Компьютеризм данную схему воспроизводит в контексте компьютерных форм представления, обработки и использования информации.

Нет возможности подробно останавливаться на критике нефункционалистских парадигм сознания и анализе современного функционализма аналитической философии сознания. Такой анализ ранее был достаточно подробно проделан [Алексеев, 2014-11]. Поэтому кратко отметим невероятное разнообразие «функционализмов», предложим метод их компьютеристской переинтерпретации и сформулируем идею тестового компьютеризма, как версию функционализма, в наибольшей степени соответствующей рассмотренному выше эпистемологическому контексту употребления термина «э-культура».

Разнообразие функционализмов

Среди базовых функционалистских парадигм выделяют (Дж.Левин): 1) *функционализм (далее – Ф.) машинных состояний (машинный Ф.)*: феномены сознания суть функции программы работы машины Тьюринга; 2) *психофункционализм*: методология когнитивной психологии, задающая исследовательскую программу описания и объяснения ментальных состояний и процессов в виде логико-предикативных «сущностей», играющих роли в осуществлении поведения; 3) *аналитический Ф.*: продуцирование, анализ, манипуляция и трансляция тематически-нейтральных функционально определенных терминов, обеспечивающих корреляцию психологического и нейрофизиологического языков; 4) *Ф. тождества функциональных состояний*: более чёткое выражение предыдущей лингвистической идеи посредством использования двумерного логического каркаса, одно измерение которого характеризует внутренний феноменальный мир человека, второе – возможные миры научных гипотез и теорий.

На рост разнообразия Ф. влияет конкретизации т.н. «реализатора» – функциональной системы, которая фактически осуществляет (реализует) социальное, психическое, нейрофизиологическое, компьютерное и др. состояние или событие. В данном случае выделяются компьютерный Ф., физический Ф., химический Ф., биологический Ф., нейрофизиологический Ф., психический Ф., персоналистический Ф. социальный Ф. и иные разновидности Ф-ов.

Различая трактовки «функции» получаем: *информационный Ф.*, представляющий функцию в формате информационно-кодовой причинной зависимости психического от нейрофизиологического (Д.И.Дубровский); *Ф. состояний* (классический функционализм Х.Патнэма); *контентный Ф.*, принимающий во внимание содержательные характеристики функциональных ролей реализаторов; *телеофункционализм*: функция как «быть полезным для чего-то»; *контекстный Ф.*, например, для физико-химических процессов контекстом выступают способы функционирования биологических систем.

В зависимости от подхода к трактовке «сознания» выделяются *квалиативный Ф.*, подчеркивающий трудности функционального описания и объяснения квалиа (Д. Чалмерс); *акцессный Ф.*, описывающий способы доступности феноменального содержания сознания (Н. Блок); *проективный Ф.*, изучающий проблему сознания с позиции приватного ощущения «какого быть» (Т.Нагель); *нарративный Ф.*, рассматривающий сознание как гетерархически организованную систему биографических суждений о личности (Д.Деннет) и пр.

При учете параметров множественной реализации некоторого психического феномена, выделяют (Т.Полгер): *слабый Ф.*, если физическая организация realizатора незначительно отличается от человеческой; *инопланетный Ф.*, когда realizатор принципиально отличается от человеческого (Х.Патнэм); *стандартный Ф.*, если природа realizатора не определена; *радикальный Ф.*, который в качестве realizатора принимает любую высоко организованную систему, вне зависимости от параметров её субстрата.

Помимо этого, выделяют *метафизический (онтологический) Ф.*, предназначенный для обоснования онтологических утверждений типа «быть в ментальном состоянии значит быть в функциональном состоянии относительно входа/выхода системы, других ментальных состояний и системы в целом»; *интенциональный Ф.*, объясняющий способность функционального состояния обладать интенциональным содержанием; *семантический Ф.*: рассматривающий функцию как отношение между знаком и его референтом; *теоретический Ф.*: функционалистская теория – это теория, у которой все термины специфицированы как функциональные дескрипции; *методологический Ф.*, изучающий любую систему в терминах функций и функциональных ролей.

Предлагаются интегральные формы представления функционализмов. Например, *континуальный функционализм*, непрерывно фиксирующий дескрипции Ф-ов, принимая во внимание их беспредельное разнообразие и нелинейную дурную бесконечность (А.Алексеев); *библиографический Ф.*: предполагающий возможность построения единого словаря, на базе которого возможно систематическое, как в библиотеке, описание различных Ф-ов (Дж.Левин); *обобщающий Ф.*, фиксирующий т.н. «главное функционалистское отношение», например, как полагает Т.Полгер, отношение множественной реализации, в корреляции с которым осуществляется последующая интеграция Ф-ов.; *редукционный Ф.*, сводящий все Ф-ы к одной архетипичной идее, т.е. к машинному Ф. (Х.Патнэм); *антиредукционный Ф.*, который предполагает создание средств, обеспечивающих интеграцию Ф., не привязывая их к конкретной разновидности Ф.

Компьютеристская переинтерпретация функционализмов

Выделим два смысла употребления слова «компьютеризм»: узкий и широкий. В узком смысле, это – самостоятельная функционалистская парадигма, в наибольшей степени аутентичная исходным идеям функционализма, т.е. машинному Ф. Отличие в том, вычислительная концепция сознания обогатилась

идеей содержания этих состояний в виде теорий репрезентации и «внутреннего языка мысли» (Дж.Фодор) [Horst, 2011].

В широком смысле «компьютеризм» означат метаметологическую парадигму репрезентации функционалистских и нефункционалистских теорий сознания в компьютерном формате. В самом деле, любая теория – это аргументированная дискретная последовательность лингвистических комплексов, значения которых образуют возможные логические миры. Особых концептуальных трудностей не представляет процесс программирования такой теории и осуществления её интерпретации в рамках техники виртуальной реальности. *Любую разновидность функционализма можно считать компьютеризмом.* В результате такой переинтерпретации получаем, например: психокомпьютеризм, если психологическая теория посредством известной методики Рамсея-Льюиса представлена в виде Пролог-программы; аналитический компьютеризм, лингвистические средства которого включают словари психологических, нейрофизиологических и тематически нейтральных терминов и т.д.

Поэтому компьютеризм представляется универсальной парадигмой для теорий сознания – настолько универсальной, насколько общим является компьютерное представление мира. Самый главный контраргумент против компьютерного универсализма – аргумент Геделя – в нашем случае не срабатывает, так как здесь не поднимаются абстрактные математические проблемы программирования аутореферентных суждений «вообще». Нас интересуют ограниченные предметные области – конкретные теории сознания с конечными словарями, с использованием ограниченного функционала инструментальных средств программирования и пр. Именно в этом смысле компьютеризм является общей метатеорией сознания. Собственно, уже машинный Ф. представлялся Х.Патнэму универсальной парадигмой, так как предполагалась возможность описать мои ментальные состояния как машинные состояния «мышления», предметом которого являются теории сознания, включая собственную теорию машинного Ф. Компьютеризм обладает еще большей степенью общности, так как дополняет «состояния» моего «сознания» контентом этих состояний.

Однако какова роль *Я* в этой запрограммированной реальности, где любая теория сознания представима в компьютерном формате и, следовательно, э-технологична?

Для выяснения этого предложим новую форму функционализма, которую мы назвали «тестовый компьютеризм». Эта форма компьютеризма потребует

детального исследования на протяжении всей данной работы, так как для осуществления требуется концептуальный инструментарий – комплексный тест Тьюринга. Тестовый компьютеризм будет представлена в конце диссертации (п.п.3.4) при изучении конститутивной функции комплексного теста Тьюринга.

Выводы. Приведем ряд общих соображений по поводу перспектив развития отечественной э-культуры в рамках обозначенных в работе контекстов. У репрезентативного, институционального, виртуалистского и онтологического направлений нет перспектив, сравнимых, например, с перспективами развития э-культуры Вьетнама. Бесперспективность обусловлена государственной стратегией т.н. «инновационного» развития ИКТ в сфере культуры, согласно которой проекты должны самофинансироваться. Однако отмеченные направления требуют мощных фундаментальных исследований и соответственно, основательных *государственных* вложений, не только на уровне гиперсложной конвергенции НБИКС, но и при создании новой э-институции, например, библиотеки РИНЦ – главной научной э-библиотеки страны, государственная поддержка которой несколько лет назад по неизвестным причинам была приостановлена.

В рамках политико-экономического направления вряд ли возможны успехи: частная собственность на ресурсы и средства ИКТ с неизбежностью ведет к увеличению и закреплению цифрового неравенства.

В некотором смысле «перспективны» аксиологическое и антропологическое направления. Здесь решаются «философские» проблемы новых форм «духовного бытия человека и общества», осуществляется экспликация «нравственных универсалий» в контексте «ноосферы» с учётом «стратегий устойчивого развития цивилизации». Философов в нашей стране много: у нас самая мощная в мире философская организация – Российское философское общество. Поэтому имеется обширное поле спекулятивных дискуссий на тему некоторого перспективного е-Номо в контексте e-Culture, который в 2045 г. превратится в трансгуманоида постчеловеческой расы, что, по сути, является программой широко известного общественного движения «2045».

Перспективно эпистемологическое направление. В рамках его создаются и применяются мысленные компьютерно-ориентированные тьюринговые тесты. Комплексный тест Тьюринга применим для изучения очень широкого спектра феноменов э-культуры. Он экономичен. Не требует никаких финансовых затрат на проведение технологических, социологических, политологических, эко-

номических, идеологических и др. работ. Предлагается «сесть в кресло», посредством отлаженных тьюринговых тестов поразмышлять о категориальной проблематике э-культуры и нарисовать критически выверенную ясную концептуальную картину путей развития э-культуры. К сожалению, поверхностный эскиз такой картины, набросанный нами в ряде отмеченных выше работ по поводу перспектив отечественной э-культуры, эксплицирует то, что негативные сценарии её развития уже реализованы. Они набирают обороты в институтах э-политики, э-медицины, э-образования, э-науки. Россию ожидает без-э-культурное будущее.

С другой стороны, россиянину не угрожают онтологические проблемы «борьбы» с квазиалгоритмическими формами э-культурного бытия по причине отсутствия такового. Возникает сладостный эффект «свободы от неведения». Можно легко переключаться на иные формы культурной жизни, к примеру, петь и танцевать.

Выводы

В первой главе в целом решалась общая задача представления искусственного интеллекта как феномена электронной культуры и, соответствующее масштабам этой задачи, представление философии ИИ как способа фундаментальной рефлексии над проблемами современной культуры.

В строении междисциплинарных исследований ИИ чётко выявлены следующие уровни: 1) *философский*, который обеспечивает рефлексию над а) мировоззренческими ориентирами современного человека, связанные с модификацией традиционных представлений о смысле жизни, свободе, социальном идеале; б) методологическими способами организации междисциплинарных и трансдисциплинарных исследований когнитивных феноменов; 2) *научный уровень*, на котором а) выдвигаются теоретические конструкции (идеи, модели, системы) и б) исследуются прототипы когнитивно-компьютерных систем – эмпирические модели, которые верифицируют и эвристически обогащают теоретические положения; 3) *практический уровень*, на котором а) проектируются, внедряются, развиваются инженерные системы и технологии и б) используются компьютерные технологии в быту и повседневности. Система когнитивных исследований целостна, так как на бытовом уровне действует конкретный индивид – носитель мировоззрения и поэтому уровень 3.б замыкается на уровень 1.а. Система междисциплинарных исследований вследствие подобного рода «логического замыкания» становится **целостной системой**.

Философия искусственного интеллекта является динамичной системой отношений между проблемами философии сознания, нейрофилософии, когнитивной науки, нейронауки, информатики и компьютерной техники. Такую многомерную философию неправомерно считать незначительной частью метафизических исследований «философии сознания», как полагает Д.Чалмерс, отведя философии искусственного интеллекта область чисто теоретических изысканий [Chalmers, Bourget, 2010], оторванную от научных и технических проблем искусственного интеллекта. Чтобы от метафизических спекуляций перейти к практическим ориентирам этой философии, требуется раскрыть роль теста Тьюринга как базисного концептуального элемента философско-методологических исследований ИИ, выявить функции категориальных концепций компьютерной имитации когнитивных феноменов в понимании и развитии электронной культуры.

Философия искусственного интеллекта представляется двояко: 1) как прикладная дисциплина, конкретизирующая общеполитическую проблематику

в контексте частной методологии научной и профессиональной деятельности программиста; 2) как специальная сфера философской рефлексии, активизирующая выработку мировоззренческих ориентиров и выполняющая интегральную методологическую функцию в ходе изучения проблематики философии ИИ и применения полученных философских знаний в научных исследованиях, в профессиональной инженерной практике и в жизнедеятельности.

Философия ИИ проявляется четко как философская сфера знаний в более широком контексте – в контексте современной культуры. Таким образом, если в первом параграфе главы проблематика философии ИИ по сути отождествлялась с междисциплинарными проблемами исследований ИИ в рамках сквозного процесса взаимодействия специалистов на уровнях «философия-наука-инженерия», то во второй параграфе внимание акцентировалось на "внешней" для данной системы знаний среде исследований – на электронной культуре. И собственно в контексте изучения электронной культуры возникла важная эпистемологическая, онтологическая и праксеологическая проблематика, связанная с субъектом культуры, с Я. Задача построения философии ИИ, таким образом, стала практической задачей, в отличие от ее теоретического статуса в контексте междисциплинарного дискурса науки и техники, относящихся к ИИ.

В философских мировоззренческих исследованиях главная задача состоит в нахождении фундаментальных культурных универсалий, эпистемологических средств, моделей, метафор, концепций – всего того, на чём, собственно, базируется представление мира. В философских исследованиях когнитивной науки и технологии главная задача состоит в выявлении фундаментальной компьютерной метафоры, посредством которой человек воспринимает и понимает окружающий мир – природу, отношения с другими людьми, самого себя. Философия ИИ предлагает такую концептуальную конструкцию – тест Тьюринга. Он востребован, начиная с эпистемологического уровня исследования э-культуры: познание электронной культуры невозможно осуществлять традиционными методами. Нужен специальный концептуальный инструментарий философского познания явлений компьютерного мира. В роли этого инструментария, собственно, и предлагается применять тест Тьюринга, используемый совместно с его различными интерпретациями.

В следующей главе тест Тьюринга раскрывается в комплексном формате.

ГЛАВА II. ОТ ОРИГИНАЛЬНОГО ТЕСТА ТЬЮРИНГА К КОМПЛЕКСНОМУ ТЕСТУ ТЬЮРИНГА

Введение

Идея теста Тьюринга инспирирована проблемой интерсубъективности. В контексте э-культуры данную проблему характеризует пример, который автор назвал «Девушка по переписке» [Алексеев, 2006-1].

Молодой человек ищет спутницу жизни и обращается к интернет-службе знакомств. Задаёт параметры запроса – возраст, рост, объёмы, цвет волос, знак зодиака, привычки и пр. В ответ – ряд кандидаток. Начинает переписываться. Особенно нравится девушка Элиза. Переписка длится более полугода – в выборе нельзя ошибиться! Беседы ведутся на самые разные темы: семьи, детей, секса, быта, отдыха, зарплаты, экономики, политики, литературы, китайской кухни и пр. Играют в шахматы с переменным успехом. Девушка поражает объемом знаний, лёгкостью в общении, изысканностью стиля. И глаза у неё выразительные. Наконец молодой человек не выдерживает и назначает встречу на площади у памятника. Приезжает с букетом роз. Ждёт. Час, два, три. Никто не приходит. Вернувшись ни с чем, пишет гневное письмо. В ответ – «Прости, милый! Я – Электронная Лиза, компьютерная программа»¹.

Молодому человеку не повезло – и поделом ему. Во-первых, он не потрудился укрепить адекватность своих представлений действительному положению дел, не задумался об интерсубъективной проблематике в контексте компьютерной технологии. Во-вторых, он слишком доверился наивной феноменологии восприятия *другого*, атрибутировав телекоммуникационному собеседнику параметры человека. В-третьих, он не приложил усилий эпистемологического удостоверения общения с человеческой личностью, а не с программой, реализующей функции философских зомби.

¹ На выбор имени «девушки» повлияло название широко известной программы «понимания» естественного языка, – ELIZA, разработанной Дж. Вейзенбаумом в 1966 г. [Weizenbaum, 1974; 1976; 1996]. Программа состояла всего из двух сотен операторов на языке Бейсик и незатейливого алгоритма «понимания» задаваемых ей вопросов. При этом ELIZA настолько умело имитировала речевые способности собеседника, что секретарша Вайзенбаума при отсутствии шефа очень любила с нею «поболтать». В основе концепции программы лежала психотерапевтическая техника диалога с пациентом. Сегодня программы, у которых в основе диалога лежит такая же модель, принято называть «ELIZ’а-подобные программы» [Weizenbaum, 1996].

В шутку замечено, что в Интернете выдавать себя за человека может не только программа, но и животное, например, натренированная собака. Питер Стейнер в знаменитой карикатуре «В Интернете никто не знает, что ты – собака!» (1993 г.) [New Yorker, 1993] изобразил пса, сидящего перед компьютером и нажимающего на клавиши. «И в самом деле!» – восклицает В. Рапопорт по поводу этой карикатуры [Rapoport, 2000, Р. 469]: «В Интернете никто не знает, кто с кем на самом деле общается. Это неведение может привести к самым серьезным последствиям в реальной жизни. Если я ввожу номер кредитной карточки через Web-сайт для покупки книги, то неизвестно, куплю ли я эту книгу или выдам свой секретный код мошеннику. Компьютерная безопасность определяет способ социального благополучия и защиты личности, и даже жизнь или смерть. Если моя дочь общается посредством чата с лицами противоположного пола примерно её возраста, то кто знает, может быть, она общается с серийным убийцей. Обычно, правда, мы предполагаем такие опасности и полагаем, что общаемся не с мошенниками, убийцами, собаками. Или – как в нашем случае, с компьютерами».

Утверждения В. Рапопорта (2000 г.) сегодня устарели на фоне экспоненциального развития информационно-коммуникационных средств, сопровождаемого, в свою очередь, громкими судебными делами по поводу преступлений, «соучастниками» которых выступают компьютеры. По всей видимости, в недалёком будущем воплотится и сценарий «Девушка по переписке». Тогда актуальнейшей проблемой станет поиск методов и средств распознавания отличий между человеком и компьютерной системой.

Впервые подобный сценарий представил научному сообществу А. М. Тьюринг в небольшой статье «Вычислительные машины и интеллект». Статья была опубликована в 1950 г. в авторитетном журнале «Mind» [Turing, 1950] и стала самой цитируемой в литературе по философии сознания, когнитивной науке, искусственному интеллекту [Akman, Blackburn, Р. 391]. Широкие обсуждения и жаркие дискуссии по основному вопросу статьи «Может ли машина мыслить?» породили новейшую философию аналитической традиции – философию искусственного интеллекта. Изучению и развитию сценария А. Тьюринга в контексте электронной культуры посвящена данная работа.

2.1. Идея оригинального теста Тьюринга

2.1.1. От машины Тьюринга к тесту Тьюринга:

истоки конструирующей функции комплексного теста Тьюринга

Исследователи теста Тьюринга, изучая его возможности и предлагая собственные версии, как правило, начинают с фиксации ключевой идеи знаменитой статьи Алана Мэтисона Тьюринга «Вычислительные машины и интеллект». Идея состоит в замене неясного вопроса «Могут ли машины мыслить?» на более простой вопрос, который целесообразно применять в контексте изучения «машинного мышления». Сложность исходного вопроса возникает при попытке внятной дефиниции «интеллекта». Это – более сложная проблема, нежели чем, скажем, «самая сложная проблема сознания», позиционированная Н. Блоком и, далее, Д. Чалмерсом при объяснении отношения феноменальное/физикальное. В этих объяснениях мы *как бы* обладаем априорными дефинициями «феноменального», «физикального». Но что такое «интеллект» как фундаментальная внутренняя способность разумного человеческого существа (*homo sapiens*), благодаря которой мы владеем этими дефинициями? Вот самая тайная из тайн. Кстати, для А. Тьюринга не менее трудной оказалась проблема дефиниции «машины». В самом деле, что такое «машина»?

А.Тьюринг поступает по-инженерному прямо и конкретно. Он упрощает проблему, применив простую *антиэссенциалистскую стратегию*, в соответствии с которой надо исключить поиск глубинной сущности понятий. Это – неблагодарная и бесперспективная работа. Для инженерного успеха в построении интеллектуальных машин не нужны метафизические изыскания. Вопрос о мыслящих машинах следует перевести в практическую плоскость изучения: «Можно ли создать машину (точнее, универсальную цифровую машину), которая будет способна подражать интеллектуальному поведению человека?». Предмет вопроса переносится в контекст *игры в имитацию* (ИВИ) интеллектуального поведения. ИВИ имеет концептуальный статус, т.е. она представлена в формате мысленного эксперимента. Участниками игры выступают и люди и компьютеры. Если для стороннего наблюдателя (судьи), который не имеет возможности наблюдать игроков, процесс диалогового общения с компьютером не отличим от общения с человеком, то *компьютер может мыслить*. Всё очень просто.

Большинство специалистов в области когнитивной и компьютерной наук полагают, что для философии, теории и практики искусственного интеллекта данная статья явилась основополагающей. Соответственно, и идея теста Тью-

ринга, представленная в ней, выступает фундаментом для всего последующего строения ИИ. Это так. Однако, на мой взгляд, изучение данной работы, несомненно, крайне важной, но взятой отдельно и изолированно от других тьюринговых работ, не позволяет в должной мере осветить идею теста и оценить её роль для становления философии ИИ. Например, при использовании идеи Тьюринга в контексте методологических исследований проблематики сознания неправомерно утверждается, что «тьюринговый тест на интеллект искусно преодолел всё бескрайнее философское болото, возникшее из проблемы отношений разума и тела (mind-body problem)» [French, 1990, P. 55]. Оригинальный тест в «болото» отношений феноменального/физикального даже и не засасывало, поскольку А. Тьюринг не был философом и никогда не причислял себя к их числу. В вязкую тину метафизических словоплетений, из которых состоит современная философия сознания, попали интерпретации теста, но не оригинальный тест. Э.Ходгис, известный биограф А. Тьюринга отмечает следующее: «Алан Тьюринг, наверное, посмеялся бы над тем, чтобы назвать себя не только великим философом, но и просто философом. Он считал себя математиком» [Hodges, 2010]. Тем не менее, версии тьюринговой идеи имеют важные философские приложения. Поэтому совершенно правомерно в 1996 г. тот же Э. Ходгис публикует книгу «Тьюринг» в известной серии «Великие философы» [Hodges, 1997].

Статья Тьюринга стала одной из наиболее цитируемых работ в современной философской литературе. Этот факт объясняется просто. А. Тьюринг перенёс достаточно строгое математическое понятие вычислимости и алгоритма на традиционные проблемы души и тела, свободы воли и детерминизма. Вечные мировоззренческие вопросы получили современную трактовку в эпоху электронной культуры, в алгоритмизированном и компьютеризированном мире. Причем трансформация «вечных идей» в компьютерном и телекоммуникационном контексте осуществилась не путем высоких теоретических изысканий, но посредством экспликации обыденного процесса общения между людьми. Некоторые из людей (возможно, все) на поверку оказываются машинами, играющими роли людей.

Так как А. Тьюринг не был философом, то перед осуществлением философских интерпретаций теста следует обратить внимание на математические работы. Именно в них кристаллизовывалась идея теста на машинный интеллект. И собственно математиком, а не философом был А.М. Тьюринг, по крайней мере, он сам в этом признавался. Математическая линия зарождения идеи

теста явно прослеживается в концептуальной ветви, произрастающей из знаменитой конструкции, предложенной А. Тьюрингом в 1936 г. [Turing, 1936] и получившей в 1937 г. собственное имя «машина Тьюринга».

Машина Тьюринга – инженерная формализация интуитивного понятия «алгоритм»

Из курса школьной информатики известно, что алгоритм – это фундаментальное понятие математики и программирования. Имеется государственный стандарт на понятие алгоритма, и данную дефиницию не отменила переориентация страны с социалистической формы хозяйствования на капиталистическую. В соответствии с ГОСТ 19.004-80 (он сегодня заменён), «алгоритм — это точное предписание, определяющее вычислительный процесс, ведущий от варьируемых начальных данных к искомому результату». В данном понятии имплицитно фиксируются следующие свойства алгоритма: *детерминированность* – причинная предопределённость действий в каждом возможном случае и однозначность результата работы при заданных исходных данных; *массовость* – возможность варьирования исходных данных в определенных пределах и, соответственно, пригодность алгоритма для решения некоторой совокупности задач определённого класса; *результативность* – для любых допустимых исходных данных через конечное число шагов работа должна завершиться; *дискретность* – разбиение алгоритмического процесса на отдельные элементарные этапы, возможность реализации которых очевидна, а результат выполнения каждого элементарного этапа определён и понятен.

Это – интуиции по поводу понятия алгоритма. В теории алгоритмов выделяется несколько подходов к формализации этого интуитивного понятия. На наш взгляд, классическими подходами представляются математическая, лингвистическая и инженерная трактовки. Другие интерпретации – В.А. Успенского, А.Н. Колмогорова и, скажем, популярная сегодня интерпретация Г.Чайтина, сводимы к классическим подходам. Помимо этого, в исходных трактовках имеются предпосылки для репрезентации знаний в современных компьютерных системах, за исключением, конечно, нейрокомпьютинга и, соответственно, коннекционизма.

Математическая трактовка понятия алгоритма связывается с исследованиями Алонзо Чёрча в области теории вычислимых функций. Отделив понятие функции от понятия множества посредством введения операции функциональной абстракции (λ -операции), Чёрч построил логическое исчисление (λ -исчисление) и в 1936 г. выдвинул идею, что понятие алгоритмической вычис-

лимости выражается посредством частично определённых рекурсивных функций [Барендрегт, 1985]. На основе λ -исчисления в середине 20-го века был создан первый язык программирования интеллектуальных компьютерных систем Lisp (его реализовал Джон Маккарти [McCarthy, 1960]). Сегодня Lisp имеет мощные приложения в области аппликативного, комбинаторного, функционального и других парадигм программирования, которые ориентированы на построение «чистых» программных инструкций, абстрагированных от среды реализации. Недавно фирма Microsoft, известная инертностью к нормальным инновациям, включила в состав Visual Studio язык функционального программирования F#.

Лингвистическая трактовка понятия алгоритма предложена в конце 1940-х годов советским математиком Андреем Андреевичем Марковым (мл.) в теории нормальных алгорифмов [Марков, 1984]. Нормальный алгорифм применяется к словам определённого алфавита и задаётся посредством дискретной последовательности элементарных шагов переписывания строк, определяемых формулами подстановки. Данный подход сегодня активно используется в одном из классических способов программирования интеллектуальных систем – в продукционном программировании.

Инженерная трактовка понятия «алгоритм» непосредственно связывается с машиной Тьюринга. В состав машины Тьюринга (МТ) входят разделённая на ячейки бесконечная в обе стороны лента (возможны МТ, которые имеют неограниченное множество лент бесконечной длины), а также управляющее устройство, способное находиться в одном из множества состояний. Число возможных состояний управляющего устройства конечно и точно задано. Управляющее устройство работает согласно правилам переходов, которые, собственно, и представляют алгоритм, реализуемый данной машиной Тьюринга. Каждое правило перехода предписывает машине, в зависимости от текущего состояния и наблюдаемого в текущей клетке символа: стереть старый символ, оставить его без изменения, записать в эту клетку новый символ, остаться в предыдущем состоянии, перейти в новое и/или переместиться на клетки влево или вправо. Некоторые состояния машины Тьюринга помечаются как терминальные, и переход в любое из них означает конец работы, остановку алгоритма. Таким образом, понятие алгоритма эксплицируется способом функционирования машины [Turing, 1936]. В работе 1950-го года понятие «интеллекта» эксплицируется почти аналогично – способом функционирования машины, имитирующим интеллектуальное поведение.

В принципе данные три трактовки алгоритма эквивалентны. Факт эквивалентности выражен в знаменитых тезисах, которые доказать невозможно, но они представляются как очевидные положения. Тезис Чёрча – Тьюринга в самой общей форме гласит, что любая интуитивно вычислимая функция является частично вычислимой, или, что равнозначно, может быть вычислена некоторой машиной Тьюринга. В свою очередь, «принцип нормализации» фиксирует, что любой нормальный алгоритм, в смысле А.А. Маркова, эквивалентен некоторой машине Тьюринга и, наоборот, любая машина Тьюринга эквивалентна некоторому нормальному алгоритму. А раз так, то и алгоритм Маркова эквивалентен частично рекурсивной функции Чёрча, и, соответственно, машине Тьюринга.

Несмотря на общепринятое мнение о равнозначности трактовок понятия «алгоритм», машина Тьюринга стоит на вершине *эвристической компетентности*. Во-первых, машину Тьюринга легко вообразить в формате мысленного примера, в отличие от частично-рекурсивных функций и формальных грамматик, где требуется дополнительная конкретизация механизма, выполняющего исчисления и комбинирующего текст. Во-вторых, и это – главное, в машине Тьюринга двуаспектно совмещены два плана – формальный (символы на ленте) и каузальный (конкретное устройство чтения/записи символов, переходящее из одного состояния в другое).

Поскольку в машине Тьюринга выполняется соотношение *формальное/каузальное*, отсюда недалеко до идеи функционализма – доминирующей парадигмы современной философии сознания. Для философии сознания главным вопросом был и остаётся вопрос отношения «идея/материя», «дух/тело», «психика/мозг», «феноменальное/физикальное». Функционализм обозначил ряд новых дихотомий: «функциональное/реальное», «когниция/реализация», «формальное/каузальное». Машинный функционализм породил современный компьютеризм (computationalism), в котором «формальное» – это программная дескрипция, «каузальное» – физический процесс функционирования компьютера. Компьютеризм распространился на когнитивные феномены ментального, витального, персонального, социального содержания. Затем недалеко оказалось и до Социокультурные феноменов: свободы, бессмертия, смысла жизни, общественного идеала. Как видим, простой замысел инженера получил колоссальный резонанс во всех сферах современного компьютеризованного мира – мира экультуры.

**Функционализм машины Тьюринга – основная «функция» машины
Тьюринга в философии искусственного интеллекта**

Становление и развитие функционалистской парадигмы сознания и мышления обусловлено развитием кибернетики, теории информации, семиотики, системных и структуральных исследований. Функционализм – одно из главных теоретических событий аналитической философии 20-го века [Дубровский, 1994, С. 21]. Функционализм предлагает особый подход к проблеме «сознание-мозг-компьютер». При решении данной проблемы ищутся ответы на вопросы: что такое психическое, что такое мысль, где мысль существует, что делает мысль мыслью, что делает боль болью? Ответы на эти вопросы связываются с функциональной организацией целостного организма.

Картезианский дуализм утверждает, что ментальное («боль», «мысль», «переживание» и пр.), пребывает в особой духовной субстанции. Бихевиоризм идентифицирует психику с поведенческой функцией. «Боль», «мысль» – это диспозиции поведения. Они не имеют внутреннего, интенционального содержания. Физикализм в своей основной версии отождествляет психическое с мозговыми процессами. «Боль», «мысль» – это физические процессы, описываемые нейрофизиологическими терминами, референты которых либо соотносятся с народно-психологическими терминами, либо, по утверждению элиминативистов, не соотносятся с ними, но лишь с прогрессивно уточняемыми научными понятиями.

Функционализм утверждает, что психические явления – это, во-первых, отношения между чувственными данными, поведенческими реакциями и психическими состояниями. Эти функциональные отношения выполняют специфические роли в составе психики как целостной системы. Во-вторых, это – функции, реализуемые материальными, в частности мозговыми процессами. «Мысль», «боль» – это функциональные состояния нейромозговых коррелятов или других систем, субстанционально различных, но функционально инвариантных мозгу. В *исторически первой и оригинальной версии функционализма* такие состояния концептуально отождествляются с состояниями машины Тьюринга. Данная версия функционализма получила название *«функционализм машины Тьюринга»*. Часто употребляются синонимы: *«машинный функционализм»*, *«функционализм состояний машины Тьюринга»*, *«функционализм машинных состояний»*.

Концепция функционализма машины Тьюринга (ФМТ) имеет давнюю историю и многих последователей. Недавно вышли две крупные историко-философские работы на эту тему: Роберта Харниша «Сознание, мозг, компьютер: историческое введение в фундаментальные основы когнитивных наук»

[Harnish, 2001] и Роберта Камминса: «Сознание, мозг, компьютер: основы когнитивных наук. Антология» [Cummins, 2000]. В указанных антологиях наиболее почётное место отводится основателю машинного функционализма – Х. Патнэму, который, кстати, впоследствии радикально отказался от своих «тьюринговых» взглядов на проблематику сознания [Пассмор, 2002, с. 106–107].

Идея функционализма машины Тьюринга и её экспликация

Впервые концепция функционализма машины Тьюринга была предложена Х. Патнэмом в небольшой статье «Сознание и машины» [Putnam, 1960]. Сегодня машинный функционализм раскрывается следующим образом (цит. по [Levin, 2004]; см. также [Cummins, 2000]): «Обладающее сознанием существо можно рассматривать как машину Тьюринга (идеальный компьютер с конечным числом состояний), работа которой специфицируется множеством инструкций машинной таблицы или программы, каждая из которых имеет следующую форму. Если машина находится в состоянии S_i , получает входной символ I_j , то она перейдёт в состояние S_k и вырабатывает выход O_l (для конечного числа состояний, входных и выходных символов)».

В данном суждении чётко обозначается перспектива машинного функционализма – в способе реализации психических явлений, инвариантном относительно их материального носителя. «Значения» функции – феномены психики, сознания, самосознания, мышления и пр. – можно получить при различных «аргументах». В область определения такой «mind/brain»-функции входят, например, электрохимические мозговые процессы. Кстати, первый «функционалистский» проект искусственной личности (при данном понимании «функции», определённой на этих процессах) впервые предложила в 1818 г. М. Шелли в образе демона Франкенштейна [Шелли, 1990]. В область определения «функции» входят не только близкие нам нейральные системы животных и других живых существ, но и воображаемая тина марсиан, космологические системы Соляриса и пр. Наконец, аргументами «функции» являются дискретные электронно-цифровые преобразования логических вентилей – редукционный базис современного компьютеринга. Обширность области определения «психической функции», задание других функций относительно результатов предыдущих реализаций, входов и выходов системы, и, в конечном итоге, определение ролевой структуры психических функций в составе целостной системы (организма) позволяет сформулировать один из главных принципов функционализма – *возможность множественной реализации психического*. По словам Н.Блока, в такой перспективе программа изучения психики как функциональной организа-

ции системы исключительно человеческого индивида приобретает либеральный характер и дополняется «реализаторами» самой различной, нечеловеческой и небиологической природы.

Поскольку имеется многообразие способов физической реализации одной и той же машины Тьюринга, то тождественные ментальные феномены могут быть присущи как разным людям, так и, скажем, инопланетянам, у которых нет человеческого мозга, но имеются состояния мозговой тины, эквивалентные логическим состояниям машины Тьюринга [Putnam, 1960]. Через пятнадцать лет в работе «Философия и наша ментальная жизнь» [Putnam, 1975] Х. Патнэм более чётко формулирует «антропологическую» позицию машинного функционализма: 1) человек – это машина Тьюринга; 2) психологические состояния человека – это состояния машины Тьюринга или дизъюнкции состояний машины Тьюринга. В последующем, пытаясь совместить социокультурную проблематику с машинным функционализмом, Х. Патнэм посчитал концепцию функционализма глубокой ошибкой. В итоге он скатился к солипсизму в решении вопроса множественной реализации тождественных психических состояний на различных физических субстратах: ментальные состояния – это мои личные состояния и они предполагают знание глубоких социальных отношений между людьми, которые невозможно описать ни машинными состояниями, ни дизъюнкциями машинных состояний [Пассмор, 2002, С. 107]. Сознание – сугубо личностный феномен, и ни о какой компьютерной реализации *моего* сознания не может быть и речи.

Несмотря на скептицизм позднего Х. Патнэма, исходные оптимистичные идеи машинного функционализма сегодня активно эксплуатируются в функционалистских теориях сознания, в общем, и компьютеристских теориях, в частности [Horst, 2003].

Рассмотрим две превосходные архетипичные иллюстрации функционализма машины Тьюринга, предложенные Н. Блоком более тридцати лет назад в работах «Проблемы функционализма» [Block, 1978] и «Функционализм» [Block, 1980]. В первой работе предлагаются мысленные эксперименты, которые сегодня стали классическими: «Гомункульная голова» (1978 г.) и «Китайская нация» (1978 г.). Вторая работа имеет номинативную аналогию со знаменитой «Китайской комнатой» Дж. Серля¹. В этих работах достаточно ясно изла-

¹ Дж. Серль, кстати, никак не упоминает о «Китайской нации» в работе «Разумы, мозги, программы» [Searl, 1980], в которой предлагает мысленный эксперимент «Китайская комната». В последующем комментаторы «Китайской комнаты» восстановили справедливость, указав, что отмеченная работа Н. Блока является её концептуальной предшественницей [Cole, 2009].

гается идея машинного функционализма и устанавливается его связь с теоретическим функционализмом (психофункционализмом) посредством методики Рамсея – Льюиса.

Мысленный эксперимент «Гомункулная голова». В статье «Проблемы функционализма» Н. Блок предлагает эксперимент, в котором участвуют гомункулы – невидимые маленькие «человечки», сопровождающие и реализующие психическую деятельность человека, тем самым несколько отличаясь от фаустовских «гомункулов» – средневековых алхимиков, которые стремились создать искусственных личностей посредством манипуляций со спермой, кровью, навозом и пр. В работе Н. Блока «гомункул» употреблён в смысле У. Пенфилда как элемент функциональной организации моторных и сенсорных зон в коре головного мозга. Однако, на наш взгляд, предпочтительнее было бы употребить термин «демон Сэлфриджа», чем «гомункул Пенфилда»: «пандемониум Сэлфриджа», дуалистически эксплицирующий перцептивные способности [Линдсей Норман, 1974, С.78], очень похож на «гомункулную голову» Н. Блока. В «блоковской голове» (именно так иногда называют гомункулную голову) выражена простая версия машинного функционализма: психические явления описываются таблицей машины Тьюринга. Каждое ментальное состояние отождествляется с некоторым логическим состоянием, которое, в свою очередь, характеризуется структурой и содержанием таблицы, т.е. содержимым ячеек при корреляции их с другими ячейками и с таблицей в целом¹. Вход таблицы, которую предлагает Н. Блок, специфицирует импульсы нейронов органов чувств, выход – импульсы в моторно-двигательных нейронах.

Строение гомункулной головы таково: помимо «пандемониума Сэлфриджа», архитектура её очень схожа с платоновской пещерой. Напомним, что Платон применил образ пещеры для прояснения того, каким образом люди причащаются идее, которая бросает свет истины на стены пещеры. Пещерные люди Платона воспринимают отражения. Так же и гомункулы Н. Блока истине в глаза не смотрят. Им этого и не надо. Главное для них – осуществление действий, а не постижение истины.

Для представления схемы «деятельности» гомункулов предлагается вообразить тело, внешне неотличимое от человеческого. Голова, само собой, внешне так же неотличима. Хотя для функционализма это условие необязательно. Го-

¹ Любую версию функционализма можно редуцировать к спецификациям столбцов, строк и отношений между ними: реляционная схема – универсальная модель любой предметной области.

лова – как скала, в которой много-много самых разных пещер. Внутреннее строение этой «головы» забавно отличается от человеческой. Нейроны органов чувств соединены с системой освещения пустот в голове. В этих пустотах, как в пещерах, и живут гомункулы. В пещерные стены вмонтированы клеммы, подсоединяемые к моторно-выходным нейронам. Каждый гомункул должен выполнить простую задачу – реализовать «клетку» машинной таблицы, ячейки которой обозначают некоторое психическое состояние человека. На самой обозримой, центральной стене пещеры устроена доска объявлений. На доску прикрепляются открытки. В открытках написаны инструкции: какие действия должен предпринимать гомункул при получении данной открытки. Предположим, что внешнее относительно головы воздействие на чувственные нейроны открывает тот или иной вход. Освещается какая-нибудь открытка, на которую попал луч света. На этой открытке (в адресной части) написано «G», т.е. она адресована гомункулам, которые занимают квадраты G (*G-мэнам*, как они сами себя могут назвать). Допустим, освещается открытка I, т.е. активизируются нейроны входа, обозначаемого посредством «I». Один из G-мэнов имеет возможность прочесть то, что написано на открытке I, так как она освещена. На открытке в содержательной части написано «M». Таким образом, когда в полость головы проникает луч света, соответствующий входу I, читается открытка с адресом «G», G-мэн нажимает выходную кнопку «M» и вызывает моторно-двигательные механизмы M.

Гомункулная машина Тьюринга может быть представлена конечной совокупностью четвёрок или пятёрок компонентов (пять составляющих требуется, если выход делится на две части – для изменения состояния системы и выдачи результата): 1) текущего состояния; 2) текущего входа; 3) следующего состояния; 4) следующего выхода. Каждый гомункул в определённый момент времени имеет задачу, соответствующую одной четвёрке (пятёрке). Как мы видим, особо изощрённого интеллекта и творческих усилий G-мэну не надо, да и активизироваться он может крайне редко – только тогда, когда луч света падает на адресованную ему открытку. Однако, несмотря на низкий уровень интеллекта каждого из гомункулов и при условии организации адекватной машинной таблицы, «сообщество» гомункулулов способно реализовать самую сложную психическую деятельность, на которую только способны люди. Это происходит из-за того, что система гомункулов полностью моделирует логические состояния человека, т.е. функциональная организация гомункулов эквивалентна функциональной организации человеческой психики.

Очевидно, что в «гомункульной голове» воспроизведены более ранние идеи Р. Кирка, в которых на роль гомункулов претендовали лилипуты. В работе, послужившей началом современных дискуссий по проблеме философских зомби – «Зомби против материалистов» [Kirk, 1974], – Р. Кирк описывает следующий эксперимент. Лилипуты атаковали Гулливера и захватили его голову. Получив доступ к нервной системе Гулливера, они отсоединили его нервы, стали отслеживать входную информацию, поступающую от афферентных нервов и посылать выходную информацию к эфферентным нервам. Поведение Гулливера при этом ничем не отличается от его поведения в исходном состоянии. Ему присущи все поведенческие диспозиции бывшего Гулливера, однако он ничего не ощущает, не чувствует, не владеет никаким феноменальным опытом. Гулливер стал зомби. Далее сценарий усложняется – голова Гулливера возвращается к исходному состоянию, лилипуты подключают нервы туда, куда следует, и покидают «поле боя». Возникает вопрос: «Станет ли Гулливер ощущающим существом?» По-видимому, нет, так как сомнительно, что чувства вернутся при восстановлении исходного состояния, ведь нам не известны причины, закономерно связывающие факты сознания с физическими фактами. В этом мысленном примере помимо всего прочего поднимается проблема определения пространственно-временных и материально-энергетических границ сопровождения психическими явлениями физических процессов. Но здесь – арена битвы философских зомби с искусственными личностями (см. пп.3.1.2).

Конечно, мысленные примеры очень похожи. Но очевидной новацией Н. Блока относительно «Гулливера» Р. Кирка – и это для нас крайне важно! – явилось привлечение идей машины Тьюринга. Мы существенно приблизились к машинному функционализму, в частности, и к компьютерной проблематике сознания, в общем.

Мысленный эксперимент «Китайская нация». В следующем мысленном эксперименте Н. Блока, получившем в исследованиях философии сознания собственное наименование «Китайская нация», изучается номологическая возможность функциональной организации сознания. Возникает претензия на своеобразное метафорическое «измерение» количества гомункулов (агентов, нейронов), востребованных для реализации психической деятельности человека. Идея такого рода квантификации поднимает вопрос о мощности множества элементов, из которых состоит машинная таблица, и «силы» устройств, предназначенных для её обработки. Допустим, полагает Н. Блок, достаточно миллиарда гомункулов, и предлагает рассмотреть население Китая, которое, как извест-

но, достаточно многочисленно. Пусть каждый китаец снабжён мобильным телефоном для связи с каждым другим китайцем и искусственным спутником Земли (мы «воспользуемся» современными телекоммуникационными средствами, в отличие от Н. Блока, который применял рацию). Движения тел управляются радиосигналами, а сами сигналы посылаются в соответствии с инструкциями, которые китайский народ принимает со спутника. Сигналы со спутника доступны всем китайцам. Инструкции выдаются таким образом, что действующие по ним китайцы функционируют подобно отдельным нейронам. Средства радиосвязи играют роли синапсов. То есть в целом китайский народ дублирует нейродинамическую организацию человеческого мозга какого-то конкретного китайца.

Китайца мучают и избивают, снимая при этом томограмму мозга. Сигналы с нейронов больного мозга передаются через спутник соответствующим китайцам-нейронам. Каждый китаец, в соответствии с распределением болевых паттернов нейронов в мозгу индивида китайца-эталона, резко переходит в состояние боли, например, ранит себя кинжалом. Возникает вопрос в стиле социологического реализма – будет ли «китайская нация» в целом фактически испытывать боль, если в боли пребывают те китайцы, на которые указывают нейроны мозга индивида-эталона? Если имеется метафизическая возможность этого, то тогда психические качества (квалиа и феноменальные качества) оказываются соразмерными с функционалистской парадигмой.

Функционалистский ответ таков: да, китайская нация *может переживать квалиа*, например боль. Однако при этом каждый отдельный китаец не осознает того, что его нация испытывает боль, он «не видит за деревьями леса». Существо размером с нейрон (например, гомункул), находящееся в человеческой голове, может ошибаться в том, что психические феномены присущи только ему, а мозг в целом эти феномены не сопровождают. Система в целом, репрезентирующая *полный человеческий функциональный дубликат*, становится полным субъектом всех человеческих ощущений, или, по крайней мере, хотя бы дифференцированных ощущений, квалиа, как доказывает Шумейкер в [Shoemaker, 1975]. То есть система «китайской нации» фактически испытывает боль. Если такой подход верен, то тогда *ментальное имеет функциональную природу*, следовательно, феноменальные качества и квалиа специфицируются таблицей машины Тьюринга. Н. Блок – в оппозиции, но не в плане метафизических выводов, согласно которым индивидуальные сознания детерминируют либо не детерминируют общественное сознание. Метафизическое отношение к «китайской нации» возникает тогда, когда она рассматривается как система го-

мункулов, где каждый китаец – гомункул. Отсюда – неверные индуктивные заключения, которые В. Люкан назвал «гомункулизмом» [Luscan, 1987]. Н. Блок занимает оппозицию с формальной точки зрения, которая особенно ярко проявилась через два года в его концепции «психофункционализма» [Block, 1980] и сегодня считается классической функционалистской теорией сознания. Согласно этой концепции, функциональная интерпретация «психического» имеет континуальный характер. Поэтому репрезентативная мощность автоматной таблицы МТ слишком мала, посредством её не выразить структурного многообразия психических состояний.

Психофункционализм представляет закономерный этап развития функционализма МТ в плане акцента на более тонких моментах функционалистской трактовки сознания, нежели чем те, которые были предложены Х. Патнэмом в «грубой» машине Тьюринга. Подход Н. Блока сегодня имеет большое значение для проблематики сознания в общем и для развития идеи теста Тьюринга в частности, например, для создания «нового теста Тьюринга», реализуемого посредством так называемой «машины Блока». Психофункционализм и сопровождающие его блоковские новации заслуживают специального внимания.

Следует отметить, что мысленные эксперименты Н. Блока не только развивают тематику функционализма МТ, но и имеют важные последствия для организации дискуссий по поводу социологического функционализма, соотношения репрезентативизма и коннекционизма, вычислимости ментальных феноменов на социальных системах и пр. Поднимаются традиционные философские проблемы реализма, концептуализма, номинализма. Главное то, что социокультурная тематика окрашивается идеями компьютеризма и, по сути, становится темой электронной культуры.

Тезис эквивалентности теоретического и машинного функционализмов

Попытаемся экстраполировать на проблематику сознания рассмотренный ранее тезис об эквивалентности различных способов формализации интуитивного понятия «алгоритм»: машины Тьюринга, частично-рекурсивных функций Чёрча и нормальных алгорифмов Маркова. Имеется большое разнообразие вариаций идеи о том, что сознание – это функция мозга-компьютера. Выбор конкретной разновидности функционализма представляется делом вкуса исследователя или выбором парадигмальной установки научной школы. Мои личные убеждения принадлежат инженерной интерпретации проблематики «сознание-функция компьютера» в силу признания того, что конструкторские формы научной деятельности в практическом отношении более значимы, чем теоретиче-

ские, особенно когда теория аргументируется мысленными экспериментами метафизики.

В философии науки и техники многие являются сторонниками инженерного подхода. Например, полёты самолётов и космических кораблей стали возможны благодаря инженерному искусству, а не теоретизированиям по поводу полётов. Математик, возможно, унифицировал бы проблематику «мышления» посредством λ -исчисления – и здесь мы видим мощные инструментарии аппликативной комбинаторики. Лингвист предпочёл бы «интеллектуальные» функции рассматривать как текстовые преобразования и т.д. Имеются интересные комбинированные методы, например, современный вариант теории алгоритмов, предложенный Грегори Чейтиным. Алгоритм раскрывается посредством языка LISP – языка обработки текстов, представленных в формате списков, прекрасно подходящего для λ -исчислений [Chaitin, 2003]. Короче, всё дело во вкусе исследователя. Наша скромная задача состоит в том, чтобы показать равнозначность различных подходов к функционалистской трактовке отношения «сознание/мозг (компьютер)», и, показав эквивалентность теоретического функционализма функционализму машины Тьюринга, предпочтение отдать второму как более демонстративному варианту, поскольку в практическом, инженерном плане он более перспективен. Однако в конце данного параграфа попробуем отказаться от машинного функционализма как от способа решения психофизической проблематики. Он очень далёк от проектов построения реальных систем, которым можно приписать параметры Я, сознания, интеллекта. А раз так, то значимость всего современного функционализма подвергается сомнению, так как корни всех разновидностей функционализмов – в машинном функционализме. Поэтому делается вывод о необходимости реформирования функционализма. Для этого предлагается поставить субъекта функциональной системы в постнеклассическую позу и обозначить перспективы нового, тестового функционализма. Чёткая экспликация тестового функционализма будет предложена в следующей работе автора.

Раскроем тезис функционалистской эквивалентности путём демонстрации семантической тождественности языка, на котором описываются положения теоретического функционализма и языка описания работы машины Тьюринга. В этом нам поможет простая схема, предложенная Н. Блоком в небольшой работе «Функционализм» [Block, 1980a]. Работа, кстати, часто цитируется, причём иногда без ссылки на оригинал (например, [Вильянуэва, 2006, С. 138 – 140], хотя это представляется нам казусом переводчика). Индекс цитирования данной работы позволяет утверждать, что блоковская демонстрация функционалист-

ской парадигмы стала обще употребимым объяснительным каноном, раскрывающим особенности машинного функционализма.

Вначале Н. Блок обращает внимание на сравнения, которые постепенно позволяют сформировать представление о сознании как функции мозга. Например, «карбюратор» – функциональный термин. То, что им обозначается, смешивает воздух с топливом и подает полученную смесь в камеру сгорания. Так как карбюратор – инженерная конструкция, предназначенная для выполнения специальных функций, то термин «карбюратор» априори функциональный термин. Иное дело с другими терминами. Возьмем, например, термин «почка» (человеческая). В ряде случаев это – функциональный термин, так как обозначает то, что поддерживает требуемый химический баланс в организме, очищает кровь, выводя из неё вредные вещества. Имеются и другие способы употребления данного термина. Номинативный термин «почка», призванный называть обозначаемые предметы, ссылается на реальный объект в организме человека. Дескриптивный термин «почка» обозначает орган, находящийся там-то и там-то, причём на реальную почку он конкретно не указывает. Диспозиционный термин «почка» указывает, например, на ментальное содержание переживания боли, в случае, скажем, острого нефрита.

Функциональная трактовка любой предметной области настолько универсальна, насколько универсален способ употребления функциональных терминов. А этот способ, в самом деле, обладает большой степенью общности. Так, согласно некоторым стандартным логическим системам, к функциональным терминам можно свести все другие термины. Например, номинативный термин предметной области – это функция с пустым аргументом. Дескриптивный термин получается путем подстановок функций в аргументы исходных функций и последующих функциональных преобразований. Диспозиционный термин может быть выражен предикатором, который обозначает функцию, определённую на отношении, и значение которой классическая логика соотносит с двумя объектами «истина» и «ложь». Неклассическая логика имеет большее количество значений предикаторов, включая, например, квантово-логические дескрипции, предполагающие состояние вероятностей «пребывания» значения в зависимости от инструмента означивания.

Таким образом, функционалистская трактовка любой предметной области представляется достаточно универсальной. Вследствие столь большой общности при употреблении слова «функционализм» следует весьма тщательно обозначать контекст его употребления. «Функция» в функционализме машины

Тьюринга (МТ-представимая функция) отличается от функции, которая фиксируется бихевиористом в схеме «реакция-стимул». В бихевиоризме используется, как правило, диспозициональный термин, обозначающий намерение, желание, боль, уверенность и прочие пропозициональные отношения, которые «объективно» (точнее, контринтроспективно) оцениваются по входу и выходу системы. МТ-представимая функция отличается от функции в физикалистской схеме, когда утверждается отношение тождественности ментального и нейрального состояний – для отождествления требуется, по крайней мере, формальное отличие между двумя рядами феноменов. Отношение между этими рядами порождает функциональное представление. Для дуализма данное функциональное представление не требует последующего физикалистского редуктивного отождествления. Для идеализма редукция требуется, но, в отличие от физикализма, нейральное отождествляется с ментальным. Чем же тогда функционализм МТ отличается от функциональных представлений нефункционалистских парадигм сознания? Тем, что МТ-представимая функция включает сведения о *состоянии* машины Тьюринга, в котором МТ оказалась в результате входного стимула и предыдущего состояния в контексте учёта всех других ментальных состояний и выходных реакций системы в целом. Согласно патнэмовскому функционализму, *логическое состояние МТ есть психическое состояние*.

Имеются вариации в трактовке физической реализации логического состояния. При разных аргументах (разных физических реализациях) имеем одно логическое значение (ментальное состояние), но не наоборот. В алгебре данное отображение принято называть «функциональным». Правда, в классической математической теории отображений накладывается ещё ряд ограничений по поводу сюръективности/инъективности/биективности и пр. Для нас эти усложнения не существенны. Остаётся заметить, что, возможно, апелляция к математической трактовке понятия «функции» собственно и послужила поводом назвать логику функционирования машины Тьюринга функционализмом МТ.

Вот, собственно, основная идея функционализма машины Тьюринга. В электронной культуре идея крайне притягательна. Появляется возможность отождествлять психику с алгоритмом, схемы нейральной динамики – с про-

граммой, а компьютер – с мозгом¹. Возникает соблазн решать посредством компьютерной технологии не только частные проблемы, связанные с материально-экономическим достатком, политическим влиянием, социальным положением человека и пр. Компьютерная экспансия расширяется на мировоззренческие проблемы смысла жизни, смерти/бессмертия, общественного идеала и др.

Сегодня имеется много вариантов «функционализмов». Их количество равно количеству «функционалистов» и их оппонентов, помноженному на количество интервалов времени, когда исследователи придерживались определённой версии функционалистской идеи. Среди функционализмов немалой степенью общности обладает теоретический функционализм. Он, по сути, выражает народно-психологические концепции терминами функционализма МТ. Покажем равнозначность этих форм функционализма, предложив тезис эквивалентности:

Теоретический функционализм и машинный функционализм семантически эквивалентны.

Проиллюстрируем тезис примером Н. Блока. Способ функционалистского описания ментальных феноменов специфицируется автоматной таблицей машины Тьюринга, распознающей чётное/нечётное количество единиц во входной последовательности:

Состояния Входы	S ₁	S ₂
1	«Нечетный» / S ₂	«Четный» / S ₁
0	«Нечетный» / S ₁	«Четный» / S ₂

У автомата два состояния – S₁ и S₂, два входа – «1» и «0», и два выхода – слова «Четный» или «Нечетный». Таблица описывает две функции. Первая от входа и выхода, вторая – от входа и состояния на следующем этапе работы автомата. Каждый квадрат кодирует два условных предложения, специфицирующих выход и следующее состояние при текущем состоянии и входе. Например,

¹ Формула «сознание/мозг» = «программа/компьютер» поддерживается Дж. Серлем, Н. Блоком. Однако она не совсем корректна. Программа – это специфически оформленный текст, реализуемый компьютером и выражающий идеальную, мыслимую конструкцию – алгоритм. Также и в психофизической проблематике мы видим различия между «сознанием» как: 1) феноменальным приватным качеством, 2) способом научного (и ненаучного объяснения) «сознания» и 3) механизмом мозговой активности. То есть здесь также просматриваются не два, а три члена отношения.

левый верхний квадрат кодирует конъюнкцию двух предложений: если машина находится в S_1 и видит «1», то она говорит «Нечетный» (т.е. увидела нечётное количество раз символ «1») и переходит в S_2 . Правый верхний квадрат показывает, что если автомат находится в S_2 и видит «1», то говорит «Четный» и возвращается к S_1 ¹. Левый нижний квадрат показывает, что если автомат находится в S_1 и видит «0», то говорит «Нечетный» и остается в S_1 . Автомат должен начинать с состояния S_1 , так как, если его первый вход – «0» и он – в S_2 , то он неверно скажет «Четный». Данный недостаток Н. Блок исправил в следующей таблице:

Состоя- ния Входы	S_1	S_2
1	«Нечетный» / S_2	«Четный» / S_1

Как и в предыдущем случае, автомат имеет два состояния (S_1 и S_2) и два выхода («Четный» и «Нечетный»). Различие в том, что вместо двух инпутов имеется всего один вход – «1», другие символы он не воспринимает. Как и прежде, в каждой ячейке таблицы имеем две функции, одна от входа и выхода, а другая – от входа и следующего состояния. Каждый квадрат кодирует два условия: состояние входа («1») и выхода («Четный»/«Нечетный»). Левый квадрат специфицирует, что если автомат находится в S_1 и видит «1», то говорит «Нечетный» (указание, что он видел нечетное число единиц) и переходит в S_2 . Правый квадрат показывает, что если машина находится в S_2 и видит «1», он говорит «Четный» и возвращается в S_1 . Этот автомат проще предыдущего, но выполняет ту же задачу, избегая ложного объявления нечетного числа. Теперь зададим вопрос: «Что такое S_1 »? Ответ: S_1 – это отношение, полностью задаваемое таблицей, т.е. определённое в контексте восьми отношений для первой машины и четырёх – для второй машины, по два отношения на каждую клетку. Дадим явную характеристику S_1 для второй машины:

Находиться в S_1 = быть в первом из двух состояний, которые связаны друг с другом, а также с входами и выходами следующим образом: пребывая в первом состоянии, при «1» на входе переходить во второе со-

¹ Термины, используемые для перевода с английского языка, звучат несколько странно для тех, кто знаком с теорией автоматов. Например, вместо «видеть» следовало бы употреблять «считывать», не «говорить 'Нечётный'», но «записать 'Нечётный'». Мы оставили оригинальную терминологию, которую использовал Н. Блок. Она подчёркивает, что понятие автомата применяется к ментальным состояниям, а не к техническим средствам обработки символов.

стояние и выдавать «Нечетный»; пребывая во втором состоянии, при «1» на входе переходить в первое состояние, выдавая «Четный».

Далее Н. Блок применяет формат функционалистских суждений, инвариантных относительно предметной области. Формат был предложен Д. Льюисом [Lewis, 1965] (см. также в статье Дж. Левин [Levin, 1994]), однако, в отличие от наших представлений о возможности представления любого объекта и отношения предметной области функциональными терминами, здесь мы видим некоторое ограничение, по крайней мере, индивидуальные переменные и предикаторы функционально не выражаются:

Находиться в S_1 = быть x таким, что истинно $\exists P \exists Q$ [Если x находится в P и видит «1» на входе, то переходит в Q и произносит «Нечетный»; если x находится в Q и видит «1» на входе, он переходит в P и произносит «Четный»].

Остается сделать ещё один шаг – конкретизировать предметную область, т.е. перейти к проблематике сознания. Для этого предположим некоторую теорию ментальных процессов, которая характеризует все возможные соотношения между состояниями, входами и поведенческими реакциями. Пусть «боль» будет типовым ментальным процессом. Тогда можно сказать, что если вы сидите на стул, в котором торчит гвоздь, то это причинит боль и желание закричать «Ой». Функционалист следующим образом определяет «боль», применительно к представленному выше одновходовому автомату:

Пребывать в боли = быть в первом из двух состояний, которые связаны друг с другом, а также с входами и выходами следующим образом: сидя на стуле при наличии в нем гвоздя (вход), вскочить (перейти во второе состояние) и закричать «Ой».

Здесь Н. Блок конкретизирует широко известную методику Рамсея–Льюиса (см., например, в [Вильянуэва 2006, С.113 – 120]), специфицирующую ментальные состояния в терминах теоретического функционализма, т.е. посредством некоторого варианта языка символической логики, в данном случае логики предикатов первого порядка.

Пусть T – психологическая теория с n ментальными состояниями, из которых, допустим, 17-е является «болью». Тогда T можно задать следующим образом: F_1, \dots, F_n – функциональные термы, которые характеризуют n ментальных событий; i_1, o_1 и т.д. – это характеристики входа/выхода:

Чувствовать боль = быть x таким, что:

$\exists F_1, \dots, \exists F_n [T(F_1, \dots, F_n, i_1, \dots, o_1, \dots)]$ и x находится в F_{17} .

Подобным образом функционализм *описывает ментальное нементальными (формальными) терминами*, которые характеризуют психические состоя-

ния без всякого упоминания о психическом содержании, обусловленном реализациями этих состояний.

Приведённые таблицы и их «ментальные» интерпретации демонстрируют следующее.

1. Психические состояния – то же самое, что и состояния машины (машины Тьюринга). Эти состояния задаются отношениями к другим состояниям, к входной стимуляции и выходной реакции. Например, боль – это такое состояние, которое *предопределяет* крикнуть «Ой» в контексте учёта всех возможных состояний.

2. Поскольку ментальные состояния – это состояния машины, то метод определения состояний автомата пригоден для объяснения и предсказания психических способностей человека. Ментальные состояния можно более-менее полно задать в терминах логико-математического языка, описывающего психические состояния, чувственные данные и поведенческие реакции.

3. Функционализм устраняет главный недостаток бихевиоризма, так как, помимо внешне оцениваемого поведения, раскрывает внутренние структуры и механизмы, которые реализуют данное поведение.

4. Единичное функциональное состояние может быть представлено различными способами. Например, S_1 может быть механическим состоянием, состоянием электронной схемы, состоянием мозга и т.д.

5. Так как S_1 может быть реализовано различными способами, то неверно утверждение, что психическое состояние S_1 – это исключительно состояние нейронных клеток мозга: S_1 также является и электронным состоянием.

6. Так как первая и вторая таблицы взаимозаменяемы, то отсюда следует правомочность утверждения дизъюнкции логических состояний: идентичное ментальное свойство, качество, состояние может быть представлено разными таблицами.

Представленные Н. Блоком положения позволяют явно сформулировать ключевое положение функционалистской концепции сознания (и мышления):

Существо может обладать сознанием (интеллектом), не имея мозга, т.е. психическое причинно связывается не только с состояниями мозга, но и с состояниями иного материального субстрата.

В приведённых блоковских утверждениях мы чётко видим редукцию теоретического функционализма к машинному функционализму. По аналогии с теоретико-алгоритмическими тезисами Чёрча – Тьюринга – Маркова, тождест-

во различных разновидностей функционализма можно обозначить тезисом Н. Блока:

Спецификации теоретического и машинного функционализмов эквивалентны.

Таким образом, парадигма функционализма в общем случае характеризует ментальное (и интеллектуальное, см. [Block, 1981, Р. 8 – 10] в формальных терминах функциональных отношений между значениями на входе, значениями на выходе и другими состояниями целостного организма. Сторонники машинного функционализма полагают, что реализация данной формальной системы вполне достаточна для имитации, моделирования, более того, фактического воспроизведения интеллектуального, ментального, психического. Главное – аппроксимативно полно осуществить формализацию психических феноменов и использовать мощные компьютерные средства. В принципе, под общий тезис машинного функционализма подпадает и информационный подход Д.И. Дубровского: сложная система кодовой зависимости между нейральными состояниями и феноменами сознания это суть функциональная составляющая самоорганизующейся системы. Чтобы постичь тайны сознания, самосознания, личности, *Я* с целью произвольной компьютерации (манипуляции) состояниями собственного и «другого» сознания, следует расшифровать нейродинамические коды по аналогии с расшифровкой ДНК. Для самых сложных операций кодирования/декодирования информации в принципе достаточно формализма машины Тьюринга. Но хватит ли у машины Тьюринга мощности?

По всей видимости, поспешный марш-бросок через «объяснительную пропасть» «когниция/компьютер» не совсем верен. На перевале перед пропастью следует рассмотреть более конкретные варианты анализа когнитивных процессов. ФМТ даёт лишь слишком обобщённую систему спецификаций – спецификацию состояний. А как быть, скажем, с «событиями» или «содержанием» ментальных состояний? Необходимо более чётко определяться с функциональной феноменологией сознания и, параллельно данной работе, искать способы построения инструментария, реализующего данные функции. Здесь возникает необходимость привлечения более поздней конструкции (относительно машины Тьюринга) – теста Тьюринга. Тест Тьюринга существенно дополняет функционализм (и теоретический и машинный) *феноменологией наблюдателя*, даёт концептуальную возможность построения системы функций, на основании которых специфицируется интеллект и, в более общем варианте, уже на базе

привлечения нашей идеи (идеи комплексного теста) – специфицируются ментальные, но и биологические, персональные, социальные феномены.

Во-вторых, – и это главное, – машина Тьюринга двуаспектно фиксирует связь между *формальным* (символы на ленте) и *каузальным* (устройство управления). Возможность такого отношения способствовала идее Х. Патнэма об отождествлении психических состояний с логическими состояниями машины Тьюринга. Анализ психических состояний ведётся от первого лица либо от третьего лица при условии *эмпатического сопереживания* [Polger 2004]. Динамика психических состояний и смена логических состояний машины Тьюринга аналогичны не только по внешним поведенческим результатам, но и по формальному внутреннему строению реализационных механизмов («мозга», с одной стороны, и «компьютера», с другой). Программа, запущенная на машине Тьюринга, определяет порядок перехода из одного состояния в другое, продвижение головки по ленте, запись символа на ленту в соответствии с предыдущим состоянием и символом, считываемым с неё. Эти состояния Х. Патнэм назвал «логическими состояниями машины». Они описываются в логических (формальных) терминах. Логические состояния отождествляются с психическими состояниями человека. Более точно, в метафоре А. Тьюринга логические состояния машины отождествляются с интеллектуальными состояниями математика при решении алгоритмической задачи, например, деления одного числа на другое. Логические состояния реализуются физическими процессами электронных схем, тем самым постулируется каузальная связь логического и физического (технического).

Из этого следуют три классических функционалистских тезиса: 1) *Тезис изофункционализма*: Я (другой) – машина Тьюринга и мои (его) психические состояния тождественны логическим состояниям машины Тьюринга [Putnam, 1960]; 2) *Тезис множественной реализации*: психические состояния могут быть реализованы человеческим мозгом, компьютером, тинтой марсианина, если они имеют функциональную организацию, подобную функциональной организации психики сознательных существ [Putnam, 1967]; 3) *Тезис ментальной причинности*: каузальные отношения между когнитивными феноменами – это функциональные отношения.

Тезисы достаточно сильные. С 1980-х гг., пытаясь совместить машинный функционализм с социокультурной проблематикой и с достижениями естественных наук о мозге, Х. Патнэм снимает эти тезисы. Солипсическое решение вопроса множественной реализации возникло из-за того, что ментальные со-

стояния – это *мои* личные состояния, укоренённые в *социальных* отношениях между людьми. Не из-за того, что приватные когнитивные феномены невозможно ни репрезентировать машинными состояниями, ни описать алгоритмическими дескрипциями. Сложны и необъяснимы каузальные связи между социальным, психическим и биологическим.

В параграфе «Компьютерная психология и теория интерпретации» [Putnam, 1983] принципиально утверждает, что когнитивная теория не способна стать теорией ментального содержания, в основном, из-за проблем реализации. Физическая реализация психического, даже при наличии опосредующих функциональных систем между ними, т.е. построение своего рода «психологической физики» представляется утопией, наподобие того, как и социальная физика О. Конта оказалась лишь мечтой, – пишет Х. Патнэм в работе «Функционализм: когнитивная наука или научная фантастика?» [Putnam, 1997, Р.47]. Собственно, и когнитивная наука как таковая не состоятельна, так как более-менее прочный базис создаётся функционализмом. Другие психофизические парадигмы даже и не достигают методологических границ, гарантированных функционализмом. Поэтому наука бессильна в разгадке тайны человека – как возможны психика, сознание, Я.

Не известно, изменились бы взгляды Х. Патнэма, если бы он был знаком с информационным подходом к психофизической проблеме, предложенным Д.И. Дубровским. Каузальная причинность психических феноменов от нейральных носителей очевидна благодаря широко известным фактам нейрофизиологических исследований. Однако причинная зависимость когнитивных феноменов коррелирует, но не имеет прямой зависимости от преобразования энергии и вещества, как это объясняется, например, в концепциях «научного материализма». Природа причинности – информационная. Поэтому при условии расшифровки кода носителя информации – нейрофизиологической системы человека – имеется возможность искусственной реализации феноменов сознания в силу принципа инвариантности информации относительно её носителя. Принцип инвариантности – это аналог тезиса множественной реализации. Но имеется существенное отличие: чётко обозначен «механизм» (информационный), обеспечивающий «множественность реализаций». Поэтому причинную связь можно репрезентировать, более того, реализовать на небиологических носителях. Главным условием для инвариантной реализации – расшифровать коды высшей нервной деятельности человека и, далее, воспроизвести информационно-кодовые зависимости на субстрате, отличном от человеческого мозга. Т.е. стоит задача создания искусственного мозга (К.В.Анохин). В связи с этим

отметим, что в 2012 г. на одном из семинаров по проблеме сознания в ИФ РАН, данная концептуальная конструкция получила собственное название «Код Дубровского». С таким наименованием согласились Н.С. Юлина, В.И. Аршинов, В.С. Стёпин и др. известные философы. К сожалению, работы Д.И. Дубровского не известны за рубежом из-за отсутствия переводов на английский язык.

Вернёмся к антисциентистским настроениям позднего Х. Патнэма. Сегодня, в условиях развития НБИКС, они не популярны. Активно эксплуатируются ранние оптимистичные идеи машинного функционализма. *Компьютеризм* (computationalism) обогатил оригинальный функционализм репрезентационизмом и коннекционизмом. Первая парадигма ввела в когнитивную науку контент машинных состояний, интенцию, семантику. Вторая – квазиалгоритмическую парадигму «вычислений», так как чистый нейрокомпьютинг с трудом вписываются в традиционный формат вычислительных теорий [Алексеев, 2010-2]. Компьютеризм пошёл далее решения психофизической проблемы. Проблема стала частным моментом когнитивистской экспансии на физические, биологические, персональные, социальные объекты: когнитивно-компьютерные системы формально инвариантны относительно субстратных составляющих технологий из комплекса НБИКС, что собственно, даёт право когнитивной науке стать во главе конвергентной методологии данного комплекса. Функционалистские проблемы уникальности множественной реализации и ментальной причинности погрузились в тень радужных проектов типа «Россия-2045». Таким образом, эгологический тезис «Я – машина Тьюринга» не только не утратил силу, но и существенно усилился.

Тест Тьюринга – инженерная формализация интуитивного понятия «интеллект»

В отличие от машины 1936 г., предложенной для формализации интуитивного понятия «алгоритм», «машина Тьюринга» 1950 г. приобрела более конкретные очертания. Это – универсальная цифровая вычислительная машина (УЦВМ). Прототипом её послужила аналитическая машина Ч. Бэббиджа. Кстати, все современные широко распространённые компьютеры функционируют в соответствии с данным прототипом. УЦВМ лишь частично представима машиной Тьюринга, т.е. обеспечивает формализацию иного понятия, нежели чем понятие «алгоритм».

Для построения машины Тьюринга 1936 г. использовалась метафора работы математика при осуществлении вычислений. По всей видимости,

А. Тьюринг описывал свою собственную работу математика-криптографа. В машине 1950 г. метафора другая. Идею этой машины А.М.Тьюринг разъяснял, исходя из возможности выполнять все операции, которые могли быть выполнены *человеком-компьютером* (буквально «человеческим компьютером»¹). Человек-компьютер следует твердым правилам и не может от них отклоняться, однако сами правила меняются от одной ситуации к другой. Правила изложены в «Книге правил». Книга – это совокупность страниц, т.е. перфокарт. На каждой странице – своя собственная лента или совокупность лент (из предыдущей конструкции машины Тьюринга). Человек-компьютер, в зависимости от сценария своей деятельности, листает книгу и следует правилам, которые в ней изложены. Для человека-компьютера не ограничена поставка бумаги, на которой он осуществляет свои вычисления. В книге правил фиксируются сценарии, которые в общем случае не связаны с преобразованиями чисел. Для вычислительных операций предлагается вспомогательный калькулятор, хотя наличие его не существенно.

Компьютер состоит из трёх частей: 1) *память* – хранилище информации в виде совокупности перфокарт неограниченной длины; 2) *исполнительный модуль* – это часть «человека-компьютера», выполняющая элементарные вычислительные операции; 3) *блок управления*, который принудительно выполняет команды (инструкции) из Книги правил. Таким образом, Книга правил – это усложненная «таблица команд» машины Тьюринга. Блок управления отслеживает, какая команда (инструкция) удовлетворяет условию запуска и установленному порядку выполнения команд. В нормальном режиме работы он выполняет команды в последовательном порядке. В других случаях встречаются инструкции безусловных и условных переходов. Также могут встречаться команды, задающие *вероятностный «скачок»* по таблице команд. В этом случае человек-компьютер, как полагает А. Тьюринг, проявляет «свободу воли», причём невозможно определить, наблюдая за операциями такого компьютера, имел ли место случай или же имел место закон, ведь эффект случайности может быть реализован генератором случайных чисел.

Прототип УЦВМ – аналитическая машина Бэббиджа – это полностью механическое устройство. УЦВМ – это электронное устройство, причём элемент-

¹ В оригинале тьюринговой статьи используется выражение «human computer» – человеческий компьютер. Исходя из контекста суждений А.Тьюринга, «human computer» – это конкретное понятие. Поэтому перевод «человек-компьютер» предпочтительнее. Помимо прочего, выделяется компьютерный аспект антропологических представлений, наряду с другими аспектами, номинально закрепленными в историко-философской литературе: человек-животное, человек-растение, человек-машина и др.

ная база может быть весьма экзотической, скажем, память компьютера может быть построена на принципах сохранения акустических сигналов. Поэтому для человека-компьютера возможны не только дискретные, но и непрерывные способы функционирования¹.

В конструкции 1950 г. мы видим следующие новации: 1) возможна неопределённость действий (вероятностные скачки по таблице команд); 2) имеется возможность реализации уникального сценария, который после своего выполнения исключается из книги правил; 3) бесконечные ленты, которые бесконечно долго можно читать и столь же долго заполнять символами. То есть здесь мы имеем дело с другим понятием «алгоритма», нежели то, которое поддерживалось машиной Тьюринга, так как проявляются следующие свойства: *индетерминированность*, *сингулярность* (может решаться лишь одна уникальная задача), *нерезультативность* (процесс может и не иметь остановки); *непрерывность*. С другой стороны, устройство человека-компьютера позволяет реализовать и стандартные свойства алгоритма – определённость, массовость (регулярность), результативность, дискретность.

Возникает понятие *квазиалгоритма* – «как бы» алгоритма. Термины, входящие в определение алгоритма (например, упомянутое выше определение по ГОСТ), получают модальную нагрузку и дизъюнктивную окраску. Например, квазиалгоритм – это неточное или недоопределённое предписание, либо описание, задающее или характеризующее вычислительный или исчислительный процесс, ведущий от варьируемых или уникальных исходных данных к искомому или неизвестному результату, либо от фактического результата к неизвестным исходным данным. Процесс, в принципе, может не завершиться. Возможны *клинчи* – такие состояния, когда процесс не прекращается в результате однозначности как минимум двух выборов. Собственно программа может быть набором *клюджей* – модулей, которые, как ни странно, работают, хотя программист не способен ни повторить процесс их создания, ни объяснить, почему они всё-таки работают.

С учётом выше изложенного примем краткую дефиницию:

Квазиалгоритм – это модально определённый алгоритм.

¹ Современная элементная база способна создавать логические вентили, весьма различными и даже экзотичными способами. Помимо традиционных, собственно дискретных вентилей на полупроводниковой базе, имеются квантовые вентили (базовый элемент квантовых компьютеров), вентили на базе биологических молекул (биокомпьютеры), фононные вентили (фонон – квазичастица, представляющая квант колебательного движения атомов кристалла). Последние вентили сложно отнести к дискретным устройствам.

Таким образом, тьюринговский цифровой компьютер 1950 г. – это формализация интуитивного понятия «квазиалгоритм». Для этого понятия не будем искать аналоги из математического и лингвистического подходов – инженерная трактовка наиболее демонстративна.

После того, как мы отследили математическую ветку, которая привела А. Тьюринга к идее «мыслящих машин», и, немного забегая вперед, сформулируем компактный тезис:

Искусственный интеллект – это квазиалгоритмическая реализация теста Тьюринга.

2.1.2. Игра в имитацию интеллектуального поведения: истоки дефинитной функции комплексного теста Тьюринга

Значение термина «тест Тьюринга»

Термин «Тест Тьюринга» появился в середине 1970-х годов – свидетельствует Дж. Лассаж в статье «Какую разновидность Теста Тьюринга подразумевал Тьюринг?» [Lassegue, 1988]. По всей видимости, впервые данный термин ввёл известный робототехник и редактор журнала «Машинное мышление» Д. Мичи в 1974 г. По крайней мере, Дж. Мур в статье 1976 г. «Анализ Теста Тьюринга» употребляет «тест Тьюринга» как вполне устоявшийся в научном сообществе термин [Moog, 1976].

Авторы статьи в известной Стэнфордской web-энциклопедии Г. Оппи и Д. Доу полагают, что термин «тест Тьюринга» («The Turing Test») в настоящее время используется в трех смыслах [Oppy, Dowe, 2008].

1. Чаще всего «Тест Тьюринга» используется для обозначения тезиса, высказанного А. Тьюрингом для решения вопроса «Может ли машина мыслить?». Как полагает Тьюринг, вопрос «может ли машина мыслить?» является совершенно бессмысленным в контексте рационального обсуждения. Однако вопрос проясняется, если его переиначить: «Может ли цифровой компьютер играть успешно в определённую игру – «игру в имитацию» («The Imitation Game»)?».

2. Выражение «тест Тьюринга» иногда используется для обозначения биохевиоральных тестов на наличие сознания, мышления, интеллекта в контексте исследований проблемы сознания.

3. Реже данное выражение используется в теоретических дискуссиях по определению логически достаточных условий на наличие сознания, мышления, интеллекта [Block, 1981].

На взгляд автора, имеются и иные значения термина «тест Тьюринга», непосредственно связанные с междисциплинарными исследованиями искусственного интеллекта. Исследователи философии искусственного интеллекта и компьютерно-ориентированных направлений философии сознания высоко ценят тест Тьюринга за возможность организации рациональных дискуссий по поводу идентификации и реализации интеллектуальных систем различной природы. Он позволяет избежать вопросов, пустых и даже вредных с инженерной точки зрения, таких как «интеллектуальная сущность системы».

Дадим априорное рабочее определение «интеллекта» (в кантианском стиле), которое мы уже приводили вскользь в предисловии. Будем полагать, что примерно такого понятия придерживался А. Тьюринг, когда рассуждал об интеллектуальных машинах:

Интеллект - это интегративная способность упорядочивать чувственные данные, обоснованно рассуждать и нормативно регулировать поведение.

Более конкретные теоретические суждения об интеллектуальности систем вызывают справедливые нарекания, приводят к неопределённости этого понятия и его непригодности для конструктивной аргументации. Независимо от метафизических достижений в обосновании «сущности» рассудка и разума современная инженерия интеллектуальных систем впечатляет. Техническая мысль согласуется с тьюринговой антиэссенциалистской стратегией идентификации интеллектуальности систем, различающихся по субстратным основаниям. Инвариантами относительно интеллектуальной способности, по Тьюрингу, являются люди и компьютеры¹.

Простая – потому гениальная – идея А. Тьюринга состоит в следующем: интуитивное понятие «интеллекта» можно конструктивно формализовать и реализовать посредством вычислительных машин в сценарии игры в имитацию интеллектуального поведения. Компетентный наблюдатель – судья этой игры – интуитивно понимает, что значит действовать «разумно», скажем, он пользуется при этом нашим априорным определением «интеллекта». Тестируя вербальное поведение неизвестной *x*-системы, т.е. задавая некоторые вопросы и получая на них ответы, судья оценивает её на предмет интеллектуальности. Известно, является ли эта *x*-система человеком или компьютером. Судья принима-

¹ Сегодня физические инварианты «интеллекта» понимаются намного шире электронных УЦВМ. Это системы самой различной природы. Под тьюринговое тестирование попадают неорганические, биологические, социальные, глобальные, планетарные системы. Главное, чтобы их функциональная организация соответствовала сложности функциональной организации мыслящих существ.

ет решение о наличии или отсутствии в ней каких-либо форм чувственной, рас-судочной и разумной активности. Если *x*-система – это компьютер, то возникает инженерная задача её программирования. Предполагается возможность создания программы, способной ввести судью в заблуждение и вызвать у него иллюзию, что партнёром по диалогу является человек. Если результаты тестирования человеческой и компьютерной систем неотличимы, то утверждается, что компьютер способен играть в имитацию интеллектуального поведения, следовательно, способен мыслить.

То, что в контексте такой игровой ситуации компьютер «мыслит», не вызывает недоразумений. «Сущность» человеческого мышления не известна и, по крайней мере сегодня, она непостижима. Аналогичная ситуация сложилась в авиации: универсальной теории полёта нет, однако, как ни странно, самолёты и космические корабли летают¹. Что такое «интеллект» – конкретно и вразумительно сказать невозможно, можно только воспользоваться общим априорным понятием, как мы предложили выше. Независимо от этой лингвистической проблемы искусственного интеллекта, сегодня компьютеры успешно распознают образы, выигрывают в интеллектуальных состязаниях, советуют при принятии решений, самостоятельно добывают полезные ископаемые в местах, несовместимых с человеческой жизнью и пр. Инженерная мысль и программистская практика значительно опережают теорию интеллекта, если таковая вообще возможна.

Изучим *антиэссеналистские истоки тьюринговой концепции*. Э. Ходгис, известный биограф А. Тьюринга, убеждён в том, что проблема интеллектуальных машин имела для автора теста паранаучную подоплёку – проблемой машинного интеллекта он стал заниматься в 1941 г. после знакомства с теологической книгой Д. Сейерс «Разум для создания» (1941г.) [Sayers, 1941]². В книге проводится аналогия между человеком-творцом (в частности, сочинителем романов и пьес) и библейским учением о Троице. Любое человеческое творение состоит из Идеи (замысла человека-творца), Энергии (процесса записи идеи и реального «воплощения» на материальном носителе) и Силы (воздействия, ко-

¹ Сравнение истории ИИ с развитием авиации имеет давнюю традицию. Недавно эту идею аккумулировал Б. Уитби в выражении «аэродинамика интеллекта» в контексте социокультурных исследований ИИ. На русском языке издан перевод работы [Уитби, 2004]. К сожалению, перевод выполнен неквалифицированно (например, метод продукций – это, оказывается, «метод постановки»?!). Такая издательская работа вредна.

² Дороти Ли Сейерс (1893 – 1957) – широко известная в период между Первой и Второй мировой войнами английская писательница, автор детективов, поэтесса, драматург, эссеист, переводчик.

торое произведение оказывает на аудиторию при её воспроизведении). Д. Сейерс полагает, что «троица человеческого творения» аналогична богословской Троице: Отец, Сын и Святой Дух.

В тьюринговой игре в имитацию нет таких конкретных аналогий, возможно, конечно, была некоторая «духовная» инспирация. Поэтому, по всей видимости, знаменитый биограф не прав. Так, в библиографии к своей статье Тьюринг ссылается, во-первых, на математические работы, в настоящее время ставшие классическими: А. Чёрча «Неразрешимые проблемы элементарной теории чисел» [Church, 1936], К. Гёделя [Godel, 1931], С. Клини «Общерекурсивные функции натуральных чисел» [Kleene, 1935] и на свою собственную [Turing, 1937], во-вторых, на технические работы: А. Лавлейс, в которой дано описание аналитической машины Беббиджа и Хартри «Вычислительные инструменты и машины» [Hartree, 1949], и в-третьих, на философские работы: С. Батлера (главы из его работы «Иерихон», названные «Книга о машинах» [Butler, 1865], Дж. Джефферсона «Разум механического человека» [Jefferson, 1949], Б. Рассела «История западной философии» [Russell, 1940]. Никаких паранормальных источников мы не прослеживаем.

Тем не менее, вопрос остаётся прежним: как можно создать такую машину, в которой *интеллектуальное возникает из совершенно неинтеллектуальных машинных операций*? В машинном функционализме, как мы видели, вопрос имеет расширенную трактовку: *как ментальное появляется из нементального (функционального)?*

В «до-тестовой» аргументации, т.е. до статьи 1950 г., такая «возможность» имела онтологические предпосылки [Turing, 1948]: 1) интеллект производится мозгом; 2) мозг представим работой конечного автомата; 3) любая функция мозга не только вычислима, но и реализуема конечной машиной, например машиной Тьюринга, у которой длина ленты ограничена, а то и вообще отсутствует (когда фиксируется лишь пребывание в некотором ментальном состоянии); 4) всевозможный спектр вычислимых функций, определённых на основе машины Тьюринга с бесконечной лентой, представляет лишь теоретический интерес, на практике следует учитывать финитность нервной системы — А.Тьюринг даже дает приближённое значение ёмкости компьютерной памяти, необходимой для имитации работы мозга: 10^9 бит.

Но такой подход, идентифицирующий мышление с мозговой деятельностью, всё-таки не приемлем для инженера-математика: слишком много метафизически-тёмного. Наконец, в статье 1950 г. А. Тьюринг совершает *бихевиораль-*

ный поворот. Несомненно, механическое лишено интеллекта. Возможности машины ограничены простыми и повторяющимися задачами. Однако правомочны два противоположных суждения: 1) машины предназначены для выполнения любого задания, которое может быть выполнено человеком, если он будет следовать своим инструкциям точно и абсолютно *бездумно*; 2) машины могут соревноваться с человеком в любой области интеллектуального труда. Если принимается первое, то тогда спор не имеет места. Человек работает как машина – и это стандартная схема работы человека на производстве или, скажем, при осуществлении преподавательской деятельности в современном бюрократизированном вузе. Здесь разум не нужен и вреден. Для второго суждения требуются конвенциональные оценки интеллекта по его проявлениям в лингвистическом поведении.

В итоге, посредством бихевиоральной игры в имитацию интеллекта, понятие «интеллект» полностью отделяется, во-первых, от квазирелигиозных убеждений о создании интеллектуальных машин как продолжения «творения Божья», и во-вторых, от метафизических притязаний компьютерного репродуцирования мозговой активности.

Лингвистический бихевиоризм выбран. Ответ на вопрос «Может ли машина мыслить?» теперь не заключается ни в выявлении «сущности» мышления, ни того, что подразумевается под словом «машина». А. Тьюринг предлагает поиграть в игру в имитацию интеллектуального поведения.

Этапы игры в имитацию

Игра в имитацию (**ИВИ**) происходит между мужчиной (**М**), женщиной (**Ж**) и судьей **С** (Interrogator)¹, ни пол, ни гендерная ориентация которого не имеют значения. Цель судьи – определить, кто из игроков женщина, а кто – мужчина. Задаются условия игры: 1) игроки изолированы от судьи непроницаемой перегородкой (стеной), чтобы по внешним признакам он не смог догадаться, с каким игроком имеет дело; 2) общение между судьей и игроками осуществляется посредством телетайпа (телеграфа); 3) реализуется вопрос-ответный режим: судья пишет вопрос, предаёт его по телетайпу, игроки письменно отвечают судье и передают ответы опять же по телетайпу; 4) вопросы могут быть на любую тему: от математики до поэзии, от погоды до шахмат; 5) женщина говорит правду и только правду, доказывая, что она – женщина, т.е.

¹ Interrogator можно перевести как «наблюдатель», «следователь», «опрашивающий». Нам мой взгляд наиболее подходящим словом является «судья», ведь у нас «игра», а того, кто оценивает игру, принято называть судьей. В дальнейшем чаще будет употребляться данный термин.

она помогает судье разобраться, кто из них кто; 6) мужчина – обманщик, его цель состоит в том, чтобы доказать судье, что он – женщина. Хотя мужчина, по желанию, время от времени может выражать истину.

На первом этапе игры в имитацию (**ИВИ1**) судья имеет дело с людьми, которых он не видит. Он должен определить пол игрока по серии вопросов, которые задает неизвестному собеседнику по ту сторону стены.

На втором этапе игры в действие вступает машина (компьютер) **К**. По задумке Тьюринга, новой постановкой вопроса «Может ли машина мыслить?» будет вопрос: «Что произойдет, когда машина будет играть вместо **М**¹. Будет ли **С** ошибаться столь же часто, как и в игре, где участниками являются только люди?» (см. 3-й абзац в [Turing, 1950, P. 434]). Если судья будет ошибаться столь часто, как и в **ИВИ1**, то Тьюринг считает, что **К** демонстрирует интеллектуальные способности.

Данная ситуация изображена на рис. 1, который предложен в [Saygin, Cicekli, Akman, 2000]:

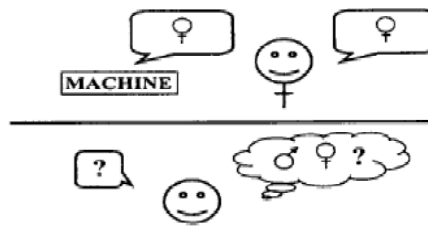


Рис. 1. Схема второго этапа игры в имитацию

Отметим следующий момент, который нам кажется важным для построения комплексного теста Тьюринга. Для изучения мысленных экспериментов способ графического изображения, приведённый на этой пиктограмме, не совсем удачен. Он нагляден, однако его трудно привести в соответствие с единой системой классификации и кодирования частных тестов. Использование этой системы позволит осуществить интегральное представление, программирование и применение частных тестов. Поэтому в следующей главе мы будем использовать формульно-символьный, а не графический способ записи теста Тьюринга и его различных модификаций.

¹ У Тьюринга мужчина обозначается символом А, женщина – В, компьютер – С. Судья никак не обозначается. Так как у нас своя символика, на мой взгляд, совершеннее тьюринговой по причине способности формализовать многообразие частных тьюринговых тестов, то мы будем подставлять в цитаты свои обозначения, не оговаривая их при этом.

Со сценарием **ИВИ2** всё понятно. В самом деле, если **К** успешно играет роль **Ж**, то он, в самом деле, может мыслить, по крайней мере, мыслить по-женски. После достаточно подробного обсуждения устройства компьютера **К** как универсальной цифровой вычислительной машины А. Тьюринг задаёт вопрос по поводу «интеллекта» иначе, нежели чем в первом варианте: «...вопрос «Могут ли машины мыслить» должен быть заменён следующим: можно ли вообразить цифровой компьютер, который способен хорошо играть в игру в имитацию (т.е. хорошо подражать человеку – А.А.)? Можно задаться более общим вопросом: «Вообразимы ли дискретные цифровые машины, которые способны играть в эту игру?» (см. 2-й абз. в [Turing, 1950, P.442].

И далее сценарий неожиданно становится очень запутанным: «Обратим внимание только на компьютер **К**. Не кажется ли достоверным то, что, увеличив его память и скорость до необходимых величин и снабдив его подходящей программой, компьютер **К** удовлетворительно сыграет роль игрока **М** в имитационной игре; *роль **Ж** при этом играетя мужчиной?*» [там же, курсив – А.А.].

То есть возникает новый сценарий – **ИВИЗ**. Женщина выбывает из игры. Вместо неё остаются «подражатели» – женщине подражает и компьютер и мужчина. Если компьютер, играя до этого роль **М**, по-прежнему обманывает, доказывая, что он – **Ж**, то что тогда делает **М**? А. Тьюринг не говорит ни о каких изменениях. Вспомним, что по начальным условиям (см. 2-й абз. в [Turing, 1950, P.434], роль мужчины заключается в том, чтобы доказать, что он женщина, но при этом он обманывает судью. Сейчас же, играя роль женщины, он должен доказывать, что он – женщина, но ведь он мужчина и должен обманывать. То есть, что тогда получается – мужчина поменял пол и стал женщиной? Или же он остался мужчиной, но тогда он может и обманывать и утверждать правду одновременно, т.е. противоречить самому себе, как известный лжец из парадокса Эвбулида. Короче говоря, совершенно непонятно, что имел в виду Тьюринг. Возникает какая-то абсурдная ситуация. Это достаточно тонкий момент, на который не часто обращают внимание комментаторы, переходя сразу к подражанию компьютером бесполого человека. Более тонкие наблюдатели – ими оказались в большинстве случаев социологи, занимающиеся гендерными проблемами, тщательно муссируют этот момент (по поводу Гендерного ТТ см. ниже).

«Зачем суетиться с женщиной, мужчиной и заменой?» – задаются вопросом Сэйджин, Цицекли и Акман [Saygin, Cicekli, Akman, 2000]. Ведь это значительно затрудняет понимание статьи. Неужели нельзя было сразу заменить пару

«женщина-мужчина» на «человек-машина», как это сделали дальнейшие исследователи теста Тьюринга.

За этими недоопределённостями ИВИ, как они полагают, скрывается методологическая честность. Игра в имитацию как игра между полами, несмотря на то что она многими считается непонятной, – тщательно проработанная идея. Игра предоставляет объективную основу для сравнения: женщина выступает в роли нейтральной стороны. Поэтому оба обманщика – **М** и **К** – могут быть *оценены по степени правдивости в своей лжи*. Необходимостью *нейтральной* стороны и объясняется то, почему в ИВИ три игрока, а не два – человек и компьютер. Именно благодаря третьей, нейтральной составляющей, судья может оценивать участников, играющих различные роли.

На мой взгляд, все эти трюизмы и противоречивые интерпретации по поводу трансформации полов не существенны. Можно полагать, что человек – это двуединство полов, условно разделённое эволюцией. У интеллекта с сексом каузальных связей мало. Можно так же оправдать запутанный сценарий предвидением одного из главных аргументов против мыслящих машин – «математического возражения» с позиции теоремы Гёделя, которая, как известно, является математическим выражением выше приведённого парадокса лжеца. Если **К** может играть роль «собственной роли», то значит, он преодолевает математическое возражение. Возможно, это так.

В любом случае разнообразие форматов ИВИ не мешает дискуссиям о тесте Тьюринга, в котором участвуют бесполое люди и компьютеры. Собственно, сам А. Тьюринг в последующих частях статьи «Вычислительные машины и интеллект» игнорирует транссексуальные проблемы и акцентирует на вопросе: «Может ли машина общаться на естественном человеческом языке?» (см. пример «беседы» в [Turing, 1950, P. 434]: вопросы о поэзии, математике и шахматах вряд ли можно считать гендерными).

Игра в имитацию и канонический тест Тьюринга

А. Тьюринг заменяет вечный философский вопрос «Могут ли машины мыслить?» на «игру в имитацию интеллекта». Это – вариант лингвистической игры¹. Игра состоит из трёх сценариев (ИВИ1 – ИВИ3) и последующих модификаций (ТТ1 – ТТ4).

¹ Неизвестно, был ли А. Тьюринг знаком с концепцией языковых игр Л. Витгенштейна (1945 г.). В «играх» этих мыслителей можно отметить много общих моментов, например, отказ от определения понятий, поиск значения слов в поведенческих сценариях, смысловой плюрализм.

ИВИ1. Играют три человека – мужчина (**М**), женщина (**Ж**) и бесполой судья (**С**)¹. Игроки располагаются в разделённых непроницаемой перегородкой комнатах, в одной – **С**, в другой – **М** и **Ж**. Диалог между ними ведётся средствами телетайпа. Цель **С** – определить посредством вопросов пол игроков. В начале игры **М** и **Ж** – неизвестные **X** и **Y**. По условиям игры **Ж** говорит правду и только правду, помогая судье, а **М** может говорить и правду и ложь, стремясь обмануть **С** путем подражания **Ж**. Игра заканчивается, когда **С** утверждает одно из двух: «**X** – это **М**, а **Y** – это **Ж**» либо «**X** – это **Ж**, а **Y** – это **М**».

ИВИ2. Место **Ж** занимает компьютер **К** – универсальная цифровая вычислительная машина, на которой запущена программа типа анализа и синтеза связного текста. Возникает вопрос: «Будет ли **С** ошибаться столь же часто, как и в игре, где участниками являются только люди?». Этот вопрос и заменяет исходный – «Может ли машина мыслить?». Если судья будет ошибаться столь часто, как и в ИВИ1, то **К** демонстрирует интеллектуальные способности.

ИВИ3. Играют **К** и **М**. Мужчина **М** заменяет **Ж**. Цель **С** остаётся прежней – определить пол игрока. В такие игры, по мнению Тьюринга, можно будет играть в 2000 г. При наличии памяти 10^9 бит, достаточной скорости операций и соответствующей программы, у **С** будет не более 70 % правильной идентификации игроков, кто из них – **М**, а кто – **Ж**².

ТТ1: Играют человек **Ч** и компьютер **К**. Преследуются цели: для **К** – имитация **Ч**, для **Ч** – доказательство человечности, для **С** – решение вопроса: «**X** – это **Ч**, а **Y** – это **К**» либо «**X** – это **К**, а **Y** – это **Ч**».

ТТ2: Играет только **К**, цель которого обмануть **С** путем подражания человеку. Цель **С** состоит в различении искусственного и естественного интеллекта (кто с ним играет – **Ч** или **К**?).

Собственно варианты ТТ1 и ТТ2 в настоящее время принято называть *тестом Тьюринга* (ТТ). А. Тьюринг не предлагал этот термин. Это послужило поводом едкого утверждения о том, что А. Тьюринг стал наиболее известным в среде философов «компьютерной ориентации» за то, что никогда и не предлагал [Lassegue 1996]. В самом деле, при буквальном прочтении статьи, слово

¹ Даются обозначения, отличные от оригинала в целях лучшего восприятия. Также слово «Interrogator» было переведено «судьей» вместо принятого в [Тьюринг, 1960] громоздкого перевода «кто-нибудь задающий вопросы».

² Здесь следует возразить Дж. Лассажу [Lassegue, 1996], который указывает всего на два сценария игры в имитацию. Если следовать буквальной трактовке статьи, к чему призывает Дж. Лассаж, то можно достаточно утвердительно сказать, что Тьюринг описывает две фигуры в подразд. 1 своей работы, а третья фигура введена в конце подразд. 5 при описании универсальной машины.

«тест» используется всего четыре раза, три из которых приходится на случай (особо интересный для нас), связанный с анализом возможности идентификации «другого сознания».¹ Однако, в защиту «теста Тьюринга» заметим, что, исходя из этических правил научной деятельности, не принято называть собственное творение своим именем. Вспомним, что А. Тьюринг не назвал собственным именем знаменитую «машину Тьюринга»: предложенная им дескрипция интуитивного понятия алгоритма, опубликованная в 1936 г. получила собственное наименование в журнале «Journal of Symbolic Logic» в 1937 г. [Lassegue, 1996].

Сформулируем ещё две формы теста Тьюринга, которые сегодня участвуют в междисциплинарных дискуссиях по искусственному интеллекту. Первая имеет некоторое отношение к тестированию интеллекта как главного атрибута *Homo sapiens*.

ТТ3: Судья тестирует некоторую систему **X** (**X** пребывает за тьюринговой стеной). Цель **C** состоит в определении человечности **X** либо в отсутствии у **X** такой характеристики.

Следующая трактовка выводит игру в имитацию за пределы дефиниции интеллектуального.

ТТ4: Судья тестирует некоторую систему **X** (**X** пребывает за тьюринговой стеной). Цель **C** состоит в определении наличия/отсутствия у системы когнитивных феноменов самого различного содержания: витальных феноменов (представляется ли система **X** живой или она мертва); ментальных феноменов (обладает ли система эмоциями, ощущениями, желаниями, убеждениями и пр.), персональных феноменов (можно ли оценивать **X** как личность, обладающую, по крайней мере, моральным вменением), социальных феноменов (**X** – это общество или бессвязный набор агентов)?

Несомненно, ТТ4 – наиболее общая трактовка игры в имитацию. Такая формулировка обычно фигурирует в современных когнитивных исследованиях.

2.1.3. Полемический стандарт теста Тьюринга: истоки критической функции комплексного теста Тьюринга

Тьюринг во время написания статьи «Вычислительные машины и интеллект» [Turing, 1950], несомненно, чётко осознавал, что некоторые его идеи бу-

¹ Дж. Лассаж ошибается, утверждая, что А.Тьюринг употребил слово «тест» два раза. На самом деле «тест» употребляется трижды в подразд. 6.4 при опровержении «Аргумента от “сознания”» и в подразд. 6.9 при опровержении «телепатического» аргумента.

дуг встречены враждебно, и, опережая, ответил на некоторые возражения, которым, как он полагал, его статья будет противоречить. Некоторые из них он обсуждал в более ранней статье 1948 г. (в [Turing, 1969]).

Отвечая на поставленный им же вопрос «Может ли машина мыслить?» Тьюринг неявно задаёт типовую структуру проведения дискуссий по данной теме, а также по другим темам, связанным с проблематикой искусственного интеллекта: 1) предмет возражения; 2) довод в пользу возражения; 3) ответ на возражение. В последующем исследователи ТТ, как правило, воспроизводят эту структуру и предлагают дискуссии (воображаемые и реальные) при описании частных тестов. Если же дискуссия явно не представляется, её можно реконструировать по многочисленным статьям оппонентов и пропонентов.

А. Тьюринг ограничился девятью положениями воображаемой дискуссии для защиты тезиса «Машина может мыслить!»). Эти положения предлагается назвать «*полемическим стандартом Тьюринга*». Предполагаемое возражение, опровергающее тезис Тьюринга, обозначено символом «–», довод в пользу – символом «+». Здесь нарушен порядок следования возражения, предложенный А. Тьюрингом в оригинальной статье. Возражения «теологическое», «антисциентистское» и «телепатическое» приведены в конце списка, как менее значимые для проблематики искусственного интеллекта. Некоторые положения Я сформулировал немного иначе, нежели чем в первоисточнике, но смысл от этого не поменялся. Каждый пункт тьюринговой дискуссии снабжён комментарием.

Математическое возражение

– Имеется несколько теорем, доказывающих ограниченность дискретных машин. Самая знаменитая из них – теорема Гёделя. Она показывает, что в замкнутой логической системе достаточной мощности обязательно найдется утверждение, которое нельзя ни доказать, ни опровергнуть, находясь в рамках этой системы.

+ Хотя и установлено наличие ограничений на мощность каждой конкретной машины, однако это утверждение берётся без всякого доказательства того, что человеческому интеллекту такие ограничения не присущи. Также возражения, построенные на таких теоремах, считаются не требующими доказательств, как и факт того, что машины, обладающие интеллектом, не ошибаются. Однако отсутствие или наличие ошибок не есть требование к мышлению [Turing, 1950, P. 445].

Математическое возражение апеллирует к гёделевской теореме и тьюринговым исследованиям описания формальных систем собственными средствами формализации. В контексте изучения интеллекта как активной человеческой способности данные исследования призывают «подняться» на уровень мета-языка и «над» алгоритмической рациональностью. Тогда будут разрешены ло-

гико-математические парадоксы, возникающие при аутоформализации, само-описании, самоидентификации и, в конечном счёте, вычислительной самореализации. Не только машине, но и человеку подобные парадоксы сложно разрешить. Например, парадоксы лжеца, брдобрея, теоретико-множественные парадоксы и пр. требуют от человека дополнительных ограничений – различать метаязык и объектный язык, фиксировать временные параметры относительно моментов утверждения о языковых и внеязыковых сущностях, вводить типы и дескрипторы в средства описания и пр. Такие достаточно жёсткие ограничения предлагает Б. Рассел, впервые за несколько тысячелетий достаточно близко подошедший к рациональному разрешению парадоксов. Предложенные Б. Расселом ограничения, позволяющие человеку не преодолевать парадоксы, а просто не впадать в них, ничуть не лучше условий традиционного программирования, в которых применяются строго типизированные типы данных.

Многие современные интеллектуальные и неинтеллектуальные системы программируются на базе нетипизированных языков, в которых отслеживание языковых уровней осуществляется за счёт учёта шагов рекурсивного определения термов, функторов и предикаторов предметной области. Всё это требует программной поддержки описания средств и способов программирования. Конечно, для преодоления парадоксов в общем случае представляется необходимым интуитивный выход за пределы формализованной рациональности. Но данная необходимость присуща не только компьютеру, но и человеку, затрагивает его внелогические, в частности, риторико-метафорические средства «самостной» дефиниции. Хорошей иллюстрацией является пример с Мюнхгаузенем, вытаскивающим себя из болота за волосы (или за усы, в разных источниках по-разному).

В последующем данное возражение стали называть аргументом Гёделя – Лукаса – Пенроуза [Lucas, 1961; Lucas, 1996]. В нашей стране, несомненно, основным специалистом по данному вопросу является Е.М. Иванов – философ из Саратова. В его работах достаточно подробно разъясняется роль теоремы Гёделя в анализе возможностей машинного мышления [Иванов, 1999].

Прекрасный рецепт «снятия» математического парадокса предлагает В.Л. Васюков – надо перейти к такой логической системе, где теория Гёделя не работает. Очевидными преимуществами обладает паранепротиворечивая логика. В ней не имеет смысла доказательство путем сведения к абсурду, принятое в рамках классической логики, – данный метод был основой доказательства гёделевой теоремы о неполноте.

Возражение с позиции сознания

– Чтобы быть разумной, машина должна обладать сознанием (т.е. осознающей себя), чувствовать удовольствие от успеха, расстраиваться от неудач и т.д.

+ Следует использовать вежливую условность того, что все люди могут мыслить. Обычно по отношению к другим людям не принимается солипсизм. Также ТТ можно использовать для оценки факта запомненной информации, т.е. «действительно ли кто-то понимает что-то или заучил это как попугай. Загадки сознания следует решать не прежде, чем мы сможем ответить на вопросы о мышлении и, в частности, о мышлении машин» [Turing, 1950, P. 446 – 447].

Крайней точкой развития данного мнения выступает солипсизм. Единственный способ *действительно* узнать, мыслит машина или не мыслит – это *быть* машиной. Тем не менее, согласно данной точке зрения, единственный способ узнать мыслит или не мыслит другой человек – *быть* этим другим человеком. А это невозможно. Здесь возникает знаменитая проблема «другого сознания», значимость которой увеличивается по мере развития э-культуры [Алексеев, 2008-3]. Решение опирается на шаткий аргумент «по аналогии». Умозаключение «по аналогии» считается необоснованным и в большинстве случаев запрещённым логическим приёмом. Поэтому, чтобы не впасть в солипсические галлюцинации, А. Тьюринг предлагает вежливо согласиться с тем положением, что другие люди, как и «я», обладают способностью сознавать, и, следовательно, мыслить.

Таким образом, проблематика компьютерного сознания, по Тьюрингу, вырастает из проблематики ИИ. На мой взгляд, здесь возможны параллельные пути решения, в частности, на пути создания проекта искусственной личности, в рамках которой «интеллект» и «сознание» тесно переплетены. С другой стороны, убедительные концептуальные исследования (например, Фланагана и Полджера [Flanagan, Polger, 1995; 1998] аргументируют отсутствие корреляции между интеллектом и сознанием (см. далее Тест зомби).

Возражение о различных невозможностях

– Машины никогда не смогут сделать *x*, где *x* – любая человеческая способность – чувство юмора, способность влюбляться или обожание клубники.

+ Такая критика зачастую представляет собой замаскированный довод о сознании. Некоторые *X* неуместны в контексте ТТ как тесте на интеллект, например такие, как способность делать ошибки и любить клубнику [Turing, 1950, P. 448 – 450].

Как замечает Тьюринг [Turing, 1950, P. 449] такая критика обычно представляет собой замаскированный довод о сознании, т.е. вытекает из предыдущего возражения.

На самом деле здесь мы имеем не только *x*-невозможность, но и *x*-неизвестность. Сам человек до конца не раскрыл своих способностей. По всей видимости, если это окажется возможным, то это будет некий постчеловек. Конечно, компьютеры не способны реализовать некую *x*-возможность. Они не понимают юмор, не любят, не строят поэтических метафор и сравнений и пр. Но это и не нужно. Компьютеру должна быть отведена своя роль, «человеку – человеково, компьютеру – компьютерово». Не нужна машина, которая участвует в конкурсах красоты среди девушек. Очень чётко было подмечено данное возражение молодыми студентами МГУ им. М.В. Ломоносова в 1960-е годы, когда тема машинного интеллекта была новой и притягательной. Родилось стихотворение, которое приписывается Роману Солнцеву (записано со слов В.И. Самохваловой):

*Ненависти нету между нами
Умная, хорошая, прощай!
Не моргай зелёными глазами,
Не губи поэзию – считай.*

Возражение леди Лавлейс

– Машины не способны к творчеству, не могут сделать ничего нового и не могут удивить нас. А если и удивляют своей способностью к мышлению, то в силу приписывания им этой способности со стороны человека.

+ Машины удивляли Тьюринга достаточно часто, как он это пишет. Признание чего бы то ни было удивительным требует достаточного «созидательного мыслительного процесса» независимо от того, кто – человек, машина или что-то еще – является автором этого удивительного и оригинального [Turing, 1950, P. 451].

В данной рубрике отрицается и утверждается креативная способность компьютерной системы. Возражение отводится, так как данная способность непонятна, невозможно найти подлинного автора идеи, не очевидны критерии оригинальности. К тому же результаты компьютерной обработки информации способны поразить любое воображение. Подробно аргументация рассмотрена при изучении теста Лавлейс.

Довод о неформальном поведении

– Интуитивно очевидно, что невозможно составить систему правил, которая описывала бы поведение индивида в каждой вообразимой ситуации.

Если бы каждый человек имел набор правил на все случаи жизни, он был бы не лучше машины. Таких правил нет. Следовательно, люди не могут быть машинами.

+ Очевидно, что полный свод таких законов нельзя представить, всегда может быть упущено какое-то правило. Но, наблюдая за действиями машины, также невозможно предсказать её поведение [Turing, 1950, P. 452].

Этот аргумент тесно соприкасается с аргументом Лавлейс. Данный довод не проходит, так как поведение компьютера, решающего сложные задачи, предсказать невозможно, возникает слишком много алгоритмически неразрешимых вариаций. В принципе, имеются достаточно проработанные теории анализа программ, работающие посредством выделения из текста программы некоторых базовых алгоритмических инвариантов (схем программ). Однако они позволяют предсказать поведение программы при наличии чётко фиксированных значений фактических параметров. В интерактивном варианте, а именно такой предполагается в ИВИ, практически невозможны предсказания по поводу поведения программы: по текущим состояниям программы сложно предсказать ответы на еще не заданные вопросы [Turing, 1950, P. 453]. И, наконец, по условиям игры, судья не имеет возможности осуществлять анализ программ, ведь он общается с x -системой.

Возражение по поводу непрерывности нервной системы

– Невозможно смоделировать поведение нервной системы с помощью дискретной машины, так как нервная система непрерывна.

+ Деятельность непрерывной машины можно представить в дискретной форме таким способом, что судья не заметит этого в условиях ИВИ.

Возражение опровергается возможностями аналогово-цифровых и цифро-аналоговых преобразований. Аналогично мультимедийной дигитализации аудио-видео-текстовых форм, в системах ИИ оцифровке подвергаются формы и способы интеллектуальной деятельности.

Возражение от боязни и неприязни («голова в песке»)

– В основе возражения лежит отвращение к идее о мыслящих машинах по причине того, что последствия появления и распространения таких машин будут ужасны. Большинство убеждено, что одной из главных особенностей человека является его способность к мышлению и эту особенность было бы неприятно разделять с машинами.

+ Возражение даже не стоит опровергать, а сторонникам следует найти какое-либо утешение, что-то вроде переселения душ [Turing, 1950, P. 444].

Большинство людей убеждено, что люди «особенные» среди других существ. Особенность заключена в способности к мышлению. Поэтому homo

sapiens неприятно разделить эту способность с «железкой» – с машиной. Но научно-технологический прогресс не остановить! Знаменитой мифологемой является борьба компьютеров с человечеством, прекрасно продемонстрированная в фантастическом фильме «Терминатор». За это актер А. Шварценегер был избран сенатором Калифорнии.

По поводу двух последующих тем полемики стандарта – теологического и телепатического возражения – имеется мнение, что они представляются лишними и не вписываются в научный дискурс. Так, авторы известной статьи в Стэнфордской энциклопедии вообще не стали их комментировать [Orru, Dowe, 2011]. Мне хочется восполнить этот пробел. Более того, я убеждён в важности этих возражений – они выражают мировоззренческие интенции создания мыслящих машин. Первый – теологический аргумент – поднимает вопросы, связанные с творением и бессмертием. В рамках компьютерной тематики – это компьютерный креационизм и компьютерная танатология соответственно. Второй – телепатический аргумент – акцентирует на сверхчеловеческих способностях человека или, извините за парадоксальную формулировку, она инспирирована собственным антропологическим статусом человеческого существа.

Теологическое возражение

– Мышление есть свойство бессмертной души человека. Бог дал бессмертную душу только человеку. Следовательно, машины не могут мыслить.

+ В этом возражении сильно ограничивается всесильность всемогущего. Пытаясь построить мыслящие машины, мы поступаем по отношению к богу не более непочтительно, узурпируя его способность создавать души, чем мы это делаем, производя потомство.

Теологическое возражение отрицает способность человека создать мыслящую машину, так как только бог может одушевлять свои творения. Однако не надо ущемлять божественное всемогущество, создавшего человека по своему подобию, т.е. со способностью творить *искусственные* сознательные системы.

Проблема творения себе подобного от века вовлечена в дискурс техники. Каждая эпоха, обладая специфическими формами технического освоения социокультурной реальности, порождает инженерный миф творения. Бог слепил Адама из глины (Ветхий Завет). Галатея Пигмалиона («Метаморфозы» Овидия) была изваяна из мрамора. Гомункулус Вагнера («Доктор Фауст» Гёте) появился в результате химических опытов. Чудовище учёного – Франкенштейна («Франкенштейн» М.Шелли) – в результате опытов с электричеством. Шариков Пре-

ображенного («Собачье сердце» М.Булгакова) – в ходе нейрофизиологического эксперимента. Компьютерная инженерия породила робота Голема («Робот и Голем» К.Чапека и «Голем» Г.Майринка)¹. Мифы о «творении» плюс некритично оцениваемые возможности компьютерной технологии создают, в лучшем случае, научные фикции, в худшем случае, паранормальный психоз. Для культурно-воспитательных целей такие мифы иногда полезны. Например, в невротическом поиске Пигмалионом целомудрия имплицитно представлены гендерные проблемы (потребность в искусственной женщине возникла из-за несоответствия реальности идеалу). Этическая проблематика просвещённого разума (Франкенштейн, Гомункулус, Шариков, Робот) убедительна для футурологических прогнозов. Однако, помимо культурологического интереса к памятникам литературного творчества, подобного рода компьютерные «мифологемы» онтологически и эпистемологически бессодержательны.

Проблема бессмертия в компьютерной окраске ничем не лучше и не хуже богословских исследований, например И. Брянчанинова² или, скажем, парамедицинских Р.Моуди³. Так, в компьютерной концепции Б.М. Полосухина [Полосухин, 1993] феномен «вечного бытия человека» задаётся самоприменимой машиной Тьюринга и содержит идеи самоорганизации и самовосстановления организма. Однако это – абстрактная идея, не имеющая оснований в реальности.

При изучении вечных мировоззренческих проблем – творения себе подобного и бессмертия – рациональная аргументация бессильна. Это – проблемы трансцендентного плана, выходящие за границы человеческого понимания.

В последующем теологическое возражение послужило концептуальным основанием для формулировки Креационистского теста Тьюринга как части Кибериадного теста (теста Швейцера). Рефлексия над этим возражением имеет важное методологическое значение в исследованиях «искусственных миров» [Макаров, 2005].

¹ См. Замаровский Войтех. Боги и герои античных сказаний: Словарь: Пер. с чеш. – М. Республика, 1994. – 399 с.; Гёте Иоганн Вольфганг. Фауст. Лирика: / Пер. с нем. / Вступ. статья и примеч. Ал.В.Михайлова. – М.: Художественная литература, 1986. – 767 с.; Шелли М. Франкенштейн, или современный Прометей. М.: Художественная литература, 1965 г. – 247 с.; Майринк Г. Голем: Роман / Пер. с нем. Д.Выгодского. – СПб.: Азбука, 2002. – 336 с.; Булгаков М.А. Собачье сердце. Повести, Рассказы, Фельетоны, Пьесы, Письма. – СПб., 1993 – 480 с. Впечатляющая подборка человеческих фантазий, повлиявших на научно-техническое «освоение» естественного интеллекта приведена Д.А.Поспеловым в [Поспелов 1982, С.5– 42]

² Епископ Игнатий Брянчанинов. Слово о смерти. М.: «Р.С», 1991. – 317 с.

³ Моуди Р. Жизнь после жизни. М.: «Интерконтакт», 1990 – 95 с. См. также, например, «реинкарнационные» суждения про бессмертие в контексте «Мирового разума»: Мизун Ю.Г., Мизун Ю.В. Бог, душа, бессмертие. – М.: Экология и здоровье, 1992 – 336 с. Подобного рода литература встречается в большом количестве.

Телепатическое возражение («с точки зрения сверхчувственного восприятия»)

– Человеку иногда присуще сверхчувственное восприятие, а машине – нет. Человек способен интуитивно постичь правильный ответ на нечёткий вопрос судьбы. Машина на это не способна.

+ Если считать, что телепатия возможна, тогда судьбу следует помещать в комнату, «защищенную от телепатии».

Некоторым людям, несомненно, присуще экстрасенсорное восприятие, интуиция, телепатия. Автор с этими интересными людьми встречался при реализации проектов по включению экстрасенсов в контур управления войсками в начале 1993 – 1996 гг. (заказчик – Генштаб Министерства обороны РФ [Алексеев, 2001]). Возражение «экстрасенсорности» снимается – компьютерные средства восприятия информации намного превышают ограниченные человеческие возможности в аудио- и видеодиапазонах, в иных формах – тактильных, гравитационных, осязательных, обонятельных. Для человека феномен экстрасенсорного восприятия представляется случайным и не очевидным. Если Вольф Мессинг заявлял о некотором сверхчеловеческом источнике своих экстрасенсорных возможностей, то не менее знаменитый экстрасенс и мастер психологических опытов Юрий Горный утверждает, что все его сверхчеловеческие способности – производное от длительных тренировок тела и интеллекта. Короче говоря, в силу неопределённости сверхчувственных способностей человека, «экстрасенсорные» способности компьютера также не определены. По поводу сравнения способностей Ю. Горного и компьютеров – см. ниже в контексте изучения «савантных» компьютерных систем.

Полнота и достаточность полемического стандарта А. Тьюринга

За полвека дискуссий по проблеме прохождения теста Тьюринга интеллектуальными компьютерами предложенная структура и содержание полемики превратились в своеобразный стандарт для исследователей. Каждый из них, обозначая свой собственный вариант теста, явно или неявно обращался к форме «возражение» / «ответ» (Н.Блок, Д.Деннетт, С.Уатт, С.Френч и др.). Вполне правомочен вопрос о полноте и достаточности полемического стандарта Тьюринга. Исчерпал ли А. Тьюринг девятью положениями всё многообразие возражений и ответов по теме «Может ли машина мыслить?». Полувековой ход дискуссий по ТТ показывает то, что дискуссии могут продолжаться настолько долго, насколько бесконечно разнообразие представлений о человеческом мышлении.

Из содержания полемического стандарта Тьюринга мы видим, что его тематика намного *шире* проблемы имитации интеллекта компьютерными системами. Первые семь рубрик непосредственно затрагивают проблему сознания, в двух последующих обсуждаются мировоззренческие проблемы «компьютерного» сознания. Поэтому, хотя оригинальный ТТ явно был определён как тест на интеллект, неявно, из контекста полемического стандарта сделаем следующий вывод:

Тест Тьюринга предназначен для изучения проблем компьютерной реализации когнитивных феноменов неопределённо широкого спектра.

«Тест Тьюринга», получивший собственное наименование в конце 1960-х годов, сегодня трактуется по-разному и означает следующее:

- тезис о замене бессмысленного вопроса «Может ли машина мыслить?» на вопрос «Способен ли компьютер имитировать интеллектуальное поведение?»;
- логически достаточные условия диспозициональной или актуальной атрибуции (приписывания) когнитивных феноменов *x*-системе;
- лингво-бихевиоральный тест на выявление параметров когнитивных феноменов;
- критерий интеллектуальности *x*-системы;
- инструкция обмана человека в «человечности» компьютера;
- техническое задание на разработку интеллектуальной компьютерной системы;
- функциональная аппроксимация естественного интеллекта;
- спецификация естественно-языкового интерфейса с компьютером;
- проект компьютерного «усилителя» интеллекта;
- квазиалгоритмическая модель естественного интеллекта.

Имеется ряд других интерпретаций оригинального теста Тьюринга по поводу машинного «мышления», в которых всесторонне оцениваются инженерные проекты интеллектуальных компьютеров и критически обсуждаются мировоззренческие последствия их использования в обществе.

В аналитической философии искусственного интеллекта тест Тьюринга и его последующие модификации принято считать категориальными концепция-

ми. Например, широко известная библиография Д. Чалмерса по философии ИИ открывается проблематикой теста Тьюринга [Chalmers Bourget, 2010]. Поэтому высокий уровень развития философии ИИ задаётся интегральным способом изучения и применения данного категориального аппарата.

Тьюринговая метафора интеллектуальной машины, привлекая всё большее внимание исследователей, служит концептуальным каркасом ИИ, является генератором новых идей, различных интерпретаций и толкований. Исследуются самые разные параметры теста: 1) длина – интервал времени, в течение которого осуществляется тестирование; 2) содержание тестовых вопросов; 3) «стена Тьюринга» – параметры скрытости игроков от судьи; 4) характеристики и особенности игроков и др. На многообразие новых трактовок влияет неопределённость идеи игры в имитацию, «чехарда» с ролями мужчины, женщины, машины и пр.

Сегодня насчитывается более тысячи крупных статей и специальных работ, в которых осмысливается ТТ. Поэтому стоит говорить о *тестах Тьюринга*, а не об одном единственном каноническом тесте, который, в свою очередь, не совсем чётко был определён самим Тьюрингом. Далее рассматриваются некоторые крупные модификации теста Тьюринга, которые мы назвали «частными тестами Тьюринга».

2.2. Частные тесты Тьюринга

Ряд исследователей тьюринговой проблематики (см., например, [Орру, Dowe, 2008; Kirk, 2006]) полагают, что впервые на проблемы различения способностей человека и машины обратил внимание Рене Декарт в своей работе «Рассуждение о методе, чтобы верно направлять свой разум и отыскивать истину в науках. Часть V» [Декарт, 1989, С. 250 – 296]. Он считал невозможным различение между нечеловеческим живым существом и машиной, если она ни внешним видом, ни внутренним строением не будет отличаться от животного. Так как человек обладает разумом, в отличие от других существ, то даже если машина будет иметь полнейшее сходство с нашим телом и поведением, различение возможно. Для этого следует протестировать машину: 1) на её способность пользоваться языком так же, как пользуемся мы, сообщая другим свои мысли – она провалит тест, так как нельзя представить, что она должным порядком будет располагать слова; 2) на её способность работать в самых различных контекстах – тест также не будет пройден, так как только человеческому разуму под силу охватить всё многообразие всевозможных поведенческих

ситуаций, даже если в некоторых частных случаях машина может действовать лучше нас.

Назовем данные вопросы исследования поведения машины **«тестом Декарта»**.

Следует учитывать, что Декарт вводит понятие интеллекта (рассудка («*pense*»)) в терминах рефлексивного сознания или осведомлённости о себе (*self-awareness*). В «Первоначалах философии» (1640 г.) он пишет: «Под словом «мышление» («рассудок» («*pense*»)) я понимаю все то, что совершается в нас осознанно, поскольку мы это понимаем. Таким образом, не только понимать, хотеть, воображать, но также и чувствовать есть то же самое, что мыслить» [Декарт, 1989]. Т.е. тест Декарта – это тест на сознание, но не на интеллект.

Изменилось бы мнение Р. Декарта о возможности пройти такой тест машиной, если бы он познакомился с функционализмом машины Тьюринга и поработал с программами ИИ в среде Интернет? Следует, конечно, учесть уровень технических достижений 17-го века [Kirk, 2006]. Сегодня – в эпоху электронной культуры – А. Тьюринг положительно отвечает на вопрос о неотличимости человека и компьютера.

2.2.1. Становление частных тестов Тьюринга

В отечественной литературе отсутствуют переводные работы, аннотированные переводы и пр., посвящённые проблематике ТТ. Ряд крупных тестов (1980 – 2000 гг.) представлен в сборнике переводов [Алексеев, 2006-4]. Однако самые первые работы, в которых начинается обсуждаться проблема ТТ, для русскоязычного читателя остается исторической редкостью. Тем не менее, первые дискуссии крайне важны, так как для современных исследователей ТТ может создаться впечатление, что мощные модификации ТТ, такие как тест Серля, тест Блока, тест Брингсйорда и пр. создавались на пустом месте. Но это не так. Остановимся подробнее на некоторых работах.

Интерпретация К. Гандерсона: тьюринговая игра в имитацию не соответствует «здравому смыслу»

Впервые на законность оценивания интеллекта в условиях проведения ИВИ обратил внимание Кейт Гандерсон (Keith Gunderson) в статье «Игра в имитацию» 1964 г. [Gunderson, 1964]. Выделяется два момента: 1) успешная игра в ИВИ может быть реализована без привлечения интеллекта; 2) интеллект – это нечто большее, нежели то, что может демонстрировать машина Тьюринга, т.е.

тьюринговая ИВИ – лишь *единичный* пример из многообразия возможных компетенций разумного существа. Работа иллюстрируется развлекательными метафорами, которые собственно ориентируют дискуссии на «здравый смысл». Вначале задается вопрос «Могут ли камни подражать?» и описывается игра в «наступалки» («наступи на ногу») [Gunderson 1964, Р. 236], которая во многом аналогична тьюринговой игре в имитацию. Вновь играют мужчина (М), женщина (Ж) и судья (С). Судья С остается в изолированной комнате и не может ни видеть, ни слышать других участников. В стене – маленькое отверстие, через которое С может просунуть свою ногу. Игроки поочередно наступают на ногу С. Цель С – отличить мужчину от женщины, руководствуясь теми ощущениями, которые он получает, когда на его ногу кто-то наступает. Вместо цифровой машины Тьюринга предлагается коробка с камнями, висящая на стене над отверстием. Коробка снабжена видеодатчиками, направленными на отверстие в стене, и клапаном, который, открываясь, выпускает камни из коробки. При появлении некоторого предмета в отверстии стены, клапан открывается и камни падают. Судья просунул ногу в стену – камни посыпались. Теперь вопрос «Могут ли камни подражать поведению человека?» превращается в вопрос: «Что произойдет, когда коробка с камнями будет выполнять роль М в этой игре? Будет ли судья ошибаться столь же часто, как если бы игра велась между мужчиной и женщиной?» [Ibid, Р. 236 – 237]. К. Гандерсон убеждён, что даже если коробка с камнями будет играть в «игру в наступалки» достаточно хорошо, то всё равно нет причин считать, что коробка подражает человеку. Единственный вывод, к которому мы можем прийти, это то, что коробка с камнями способна заменить человека исключительно в «игре в наступалки». Обобщение Гандерсона таково: «частью того, *что* происходит, является то, *как* это происходит (курсив мой – А.А.) [Ibid, Р. 238]. В дальнейшем данный тезис усилится в аргументе Н. Блока против бихевиористской установки ТТ – описание диспозиции интеллектуального поведения, помимо спецификации возможных реакций на вербальную стимуляцию, должно включать механизм реализации диспозиции (см. далее «тест Блока»).

Второе утверждение Гандерсона – то, что о мыслительном процессе нельзя судить только по одному примеру. Успех компьютера в ИВИ это еще не достаточная причина для того, чтобы назвать его мыслящей машиной. Предлагается новое сравнение. Продавец пылесосов стремится сбыть свой продукт и всячески рекламирует пылесос, скажем, Шик-600, как многоцелевой. Далее он демонстрирует, как этот многоцелевой пылесос всасывает пыль. Покупатель спрашивает, что еще может делать пылесос. Удивленный продавец говорит, что

пылесосы предназначены для всасывания пыли и Шик-600 делает именно это. Покупатель отвечает: «Я думал, он многоцелевой. Разве он не засасывает кусочки бумаги, травинки, грязь? Я полагал, что засасывание пыли – это лишь единичный пример того, на что на самом деле способен пылесос. Продавец отвечает: «Это пример того, что он делает. Он делает именно это – засасывает пыль» [Gunderson, 1964, P. 241]. Продавец, называя пылесос Шик-600 многоцелевым – ведь этот пылесос на самом деле многоцелевой, так как может пылесосить ковры, полы, стены и пр., – не смог продемонстрировать примеров других «компетенций» пылесоса. Точно так же, как термин «многоцелевой пылесос» подразумевает, что пылесос может что-то большее, нежели чем только засасывание пыли, у А. Тьюринга термин «интеллект» используется для обозначения более чем одной возможности, хотя имеется в виду лишь одна сторона интеллекта, востребованная для организации осмысленного диалога.

Ещё один пример – с бывшим прокладчиком туннелей, ныне безработным. Хозяин дорожно-строительной компании купил паровую дрель для того, чтобы сверлить дыры в скалах, поэтому рабочие-прокладчики оказались не нужны. То, что паровая дрель вытеснила рабочего – это факт, но данный факт не доказывает, что у машины есть мускулы, но лишь то, что туннели можно копать и без мускулов [Ibid, P. 254]. Соответственно, играть в тьюринговую ИВИ можно, не обладая интеллектом.

Возражение Дж.Г. Стивенсон – К. Гандерсон

Джон Г. Стивенсон в публикации «Об имитационной игре» [Stevenson, 1976] выступил против К. Гандерсона: машина, хорошо играющая в ИВИ, способна на многое. К. Гандерсон считал, что ответы, который даёт машина, даже если они согласованы с любым возможным текстом, в котором фиксируются вопросы, так или иначе демонстрирует узкий диапазон возможностей [Gunderson, 1964, P. 243]. В этом месте К. Гандерсон по сути воспроизводит тест Декарта. Дж. Стивенсон – против: даже если этот тест и не исчерпывает всего того, что связано с человеческим мышлением вообще, всё равно список «компетенций» компьютера очень внушителен.

Интерпретация Р. Пёртилла: тест Тьюринга как научная фикция

Ричард Пёртилл (Richard Purtill) в статье «Избиение игры в имитацию» (1971 г.) [Purthill, 1971] выступает с инженерной позиции, «главным образом не как философ, но как достаточно продвинутый программист» [Purthill, 1971, P.290]. Игра, несомненно, интересна, но это – некоторая разновидность научной фантастики. Нельзя вообразить то, что компьютер, играющий в ИВИ, будет раз-

работан в обозримом будущем. ИВИ – это компьютерная мечта человечества. «Я съем электронную библиотеку у того, кто спроектирует компьютер, успешно играющий в игру в имитацию!» [Purtill, 1971, P. 293]. Если однажды компьютеры поведут себя так, как это описывается в научно-фантастической литературе, то тогда им можно будет приписать интеллектуальность. Но до тех пор всю компьютерную продукцию нужно считать результатом работы программиста. Даже если программа ведёт себя случайным образом (проявляя, по А. Тьюрингу, «свободу воли»), всё равно невозможно вообразить методы и средства программирования, которые реализуют компьютер из ИВИ. Поведение разумных существ невозможно объяснить в терминах механики. ИВИ – это всего лишь поединок умов между интервьюером (судьей) и программистом: компьютер здесь – лишний, он совершенно не нужен.

Хотя первое в некоторой степени правдоподобно, то второй аргумент звучит неуверенно. Для его поддержания Пёртилл предлагает «чисто механические» аналогии: машины, сделанные из рычагов и колёс, которые могут выполнять те же задачи, что и компьютер. Если человеческий мозг – в высшей степени сложный комплекс таких механических процессов, то это означает, что «люди соображают не лучше компьютеров» [Ibid, P. 292]. Следует отметить, что данная аргументация восходит к мысленному эксперименту «Мельница Лейбница», в котором обосновывается онтологический статус духовных субстанций – монад. Если увеличить голову до размеров мельницы и войти в неё, то в её механических внутренностях никакого сознания мы не обнаружим. На наш взгляд, насколько непонятна аргументация Лейбница, который, отстаивая психофизический параллелизм, ведёт себя, как физикалист, измеряющий вес «души» в граммах, а её проявления – в монадах, настолько непонятно, почему машинные аналогии поддерживают аргумент Пёртилла. На момент написания статьи уже 10 лет уверенно шествовал функционализм машины Тьюринга, предложенный Х. Патнэмом, основная идея которого состоит в отождествлении психических состояний от первого лица с логическими состояниями машины. Для функционализма в силу принципа множественной реализации совершенно не важен материальный субстрат, из которого сделаны игроки в ИВИ.

Возражение Д. Сампсон – Р. Пёртилл

Джеффри Сэмпсон (Geoffrey Sampson) в статье «В защиту Тьюринга» [Sampson, 1973] сомневается в аргументах Р. Пёртилла. Прежде всего, он полагает, что большинство ограничений, на которые указал Р. Пёртилл по поводу построения компьютеров, играющих в ИВИ, – это лишь временные преграды,

которые могут быть преодолены, возможно, в не столь далёком будущем. Во-вторых, если поведение компьютера объяснимо в терминах его механизмов, то поведение человека объяснить не просто. Д. Сэмпсон также отводит утверждение Р. Пёртилла по поводу того, что человеческая мысль – это настолько сложный комплекс процессов, что его невозможно сравнить с вычислительным процессом. Напротив, мышление проявляется в поведении, т.е. мыслительный процесс – это то, что человек делает. Поведение, в свою очередь, можно описать и соответствующим образом запрограммировать [Sampson, 1973, P. 593].

Интерпретация П. Миллара: тест Тьюринга – тест на компьютерное мышление

В небольшой (всего две страницы) статье «К вопросу об игре в имитацию» П. Миллар [Millar, 1973] смог рассмотреть несколько важных положений ТТ: 1) ИВИ не имеет ничего общего с тем, каким образом вовлечены компьютеры и люди в эту «запрограммированную» игру и вовлечены ли они вообще; 2) ИВИ не является правильным критерием оценки интеллектуальности машины, но лишь поводом для её «очеловечивания» – приписывания ей человеческих целей, морального облика, ментальности; 3) ИВИ – это не способ измерения уровня интеллекта машины, но лишь критерий оценки того, что машины обладают именно *человеческим* интеллектом. Поэтому следует быть непредвзятыми и полагать интеллектуальным каждое живое и неживое существо, пусть даже марсианина и машину, которое проявляет «поведение, хорошо адаптированное для достижения собственных и особенных целей» [Millar, 1973, P.597].

Согласно данной позиции, дефиниция компьютерного мышления не обязательно должна следовать представлениям о человеческом интеллекте и эволюции его интеллектуальных способностей.

Таким образом, в первые четверть века после появления идеи ТТ, и особенно в период 1960 – 1975 гг., проблематика ТТ состояла в критическом осмыслении игры в имитацию. Пионеры сразу обратили внимание на проблематичность включения эволюционно-эпистемологических идей в проблематику ИИИ, т.е. на внутреннюю способность системы к росту «знаний» и к усилению этой способности. Дискуссия пока ещё не вышла за пределы журнала «Mind», т.е. имеет место «междусобойчик».

2.2.2. Крупные версии теста Тьюринга

Тест Колби - параноидальный тест Тьюринга

Предложен К. Колби в работе «Искусственная паранойя» (1971 г.). Используется схема ТТ1. В роли **С** – врачи-психиатры, в роли **Ч** – параноики. Опыт тестирования показал, что почти в половине случаев психиатры не могли отличить ответы компьютера от ответов больных. На мой взгляд, успех **К** в прохождении данного теста обусловлен тем, что ответы неврапад – это правило психического отклонения, для реализации теста не требуется сложная техника, вполне приемлема лингвистическая недостаточность тестовых вопросов и ответов. По всей видимости, параноидальный тест могут успешно пройти программы генерации «постмодернистских стихотворений». Тест Колби будет подробно рассмотрен далее.

Социальный тест Тьюринга

Б. Шэннон [Shanon, 1989, Р. 31 – 32] впервые предлагает считать тест Тьюринга как тест для оценки социального интеллекта. В роли **С** выступает мать-природа (которую «не проведёшь!»), в роли игроков – сообщество людей, которые пройдут тест, если выживут. Тест интересен в концептуальном отношении – в плане идентификации и тестирования общественных форм сознания и «интеллектуальности» социальных систем.

Гендерный ТТ – тест на отличие мужчины от женщины

Впервые предложен Ю. Геновой в [Saygin, Cicekli, Akman, 2000, Р. 32 – 33]. Ю. Генова полагает, что ТТ служит решению вопроса «чем мужчина отличается от женщины», но никак не вопроса «может ли машина мыслить?». Имеются различные трактовки: 1) Тьюринг выражает склонность к транссексуальности¹; 2) Тьюринг – антифеминист [Genova, 1994], для него важно – «Может ли женщина мыслить?»; 3) Тьюринг, напротив, считает подражание женщине верхом совершенства; 4) Тьюринг просто шутил, так как в нём гармонично сочетались ум, юмор и секс [Hodges, 2002]. С. Яновская (1960 г.) [Тьюринг, 1960, С. 6] считает, что гендерный ТТ принципиально невозможен – судья будет правильно отличать женщину от мужчины в 50 %. В настоящее время гендерный тест является концептуальной основой общей темы «Секс и компьютер».

¹ Тьюринг был обвинен в гомосексуализме, не выдержал тяжести обвинения и покончил самоубийством в 1954 г.

Тест леди Лавлейс – тест на творческие способности

С. Брингсйорд, П. Белло, Д. Феруччи в работе «Творчество, тест Тьюринга и (улучшенный) тест Лавлейс» [Bringsjord, 2001] продолжают возражение леди Лавлейс, предложенное А. Тьюрингом [Turing, 1950]. Тест задаётся следующим образом. Искусственный агент **К**, разработанный человеком **Р**, проходит тест Лавлейс (ТЛ) тогда и только тогда, когда: 1) **К** создаёт на выходе некий артефакт **А** (для леди Лавлейс – дочери Байрона – в качестве **А** выступала бы, по всей видимости, поэма); 2) **А** – это не результаты случайного стечения обстоятельств или сбоя машины. **К** всегда может воспроизвести (повторить) **А**; 3) **Р** не может объяснить, как **К** выдал **А**, даже при условии полной осведомлённости о **К** – структуре баз данных, алгоритмах функционирования и пр. Авторы считают, что: 1) ТЛ невозможно пройти путем обмана; 2) все существующие модели ИИ не пройдут ТЛ, так как разработчики знают их особенности, а «от себя» такие программы прибавить ничего не могут; 3) необходимым условием прохождения ТЛ является способность **К** самостоятельно модифицировать свою программу, т.е. **К** должен быть самообучаемым. Однако это условие не достаточно для прохождения ТЛ, так как правила самообучения не являются по сути оригинальными – они формируются в контексте «известного» для **Р**; 4) вариант наделения **К** «свободой выбора» путём моделирования, например мотивационно-волевых механизмов творчества, также не подходит, так как артефакт **А** будет представлен в некотором пространстве решений, что исключает его оригинальность.

Как известно, Тьюринг в ответ на возражение Лавлейс заменяет выражение «способность к творчеству» на «способность удивлять». Он считает, что любая ошибка в программном коде или аппаратный сбой «могут удивить». В ТЛ это исключается, так как: 1) условие «ошибки» оговорено в определении ТЛ и запрещено; 2) к **К** применяются жесткие условия контроля вычислительного процесса даже в случае сбоя. Тем не менее, ТЛ не убедителен в силу неопределённости «творческой способности» и требований к артефакту.

Тест Лавлейс значительно востребован в тьюринговой проблематике и на протяжении данной работы он будет встречаться часто.

Тест Френча – тест на субкогнитивные способности

Тест предложен Р. Френчем в 1990 г. в работе «Субкогнитивные способности и границы теста Тьюринга» [French, 1990]. Р. Френч считает, что ТТ умело обошёл бескрайнее философское болото проблемы дух/тело, однако он не способен исследовать глубинные, подсознательные области человеческого интеллекта. Поэтому в ТТ следует включить так называемые субкогнитивные вопросы, т.е. любые вопросы, позволяющие идентифицировать низкоуровневую когнитивную структуру, например, подсознательную сеть ассоциаций (а эти вопросы, считает Р. Френч, неявно включает любой ТТ). Сеть ассоциаций состоит из несчётного количества статистически связанных друг с другом образов, которые воспринимались человеком в прошлом и которые могут быть активированы в любой момент. Какие именно ассоциативные ансамбли возникнут при некотором вопросе – зависит от социокультурной обусловленности и форм конкретной жизнедеятельности игрока; т.е. ТФ способен пройти лишь тот, кто живет и ощущает окружающий мир так же, как и люди. ТФ – это не тест на интеллект «вообще», а тест на социокультурно обусловленный человеческий интеллект.

Р. Френч предлагает ряд способов идентификации подсознательной сети ассоциаций, среди которых наиболее показательным, на наш взгляд, представляется *метод ассоциативных заучиваний*. Суть метода в следующем. У человека в ходе конкретной жизнедеятельности развиваются определенные ассоциативные связи между понятиями. Сила этих связей – величина варьируемая, и её можно оценивать временем, в течение которого происходит связывание одного понятия с другим (с понятием «хлеб» быстрее свяжется понятие «масло», нежели «собака»). Судья С может воспользоваться этим методом следующим образом: 1) за день до тестирования проводится контрольный опрос среди интервьюируемых (людей), подбирается произвольный набор слов и фиксируется статистически среднее время связывания понятий; 2) на следующий день при проведении теста судья задаёт кандидатам (среди которых **Ч** и **К**) то же самое задание, собирает результаты тестирования и сравнивает с результатами контрольного опроса. Анализ позволяет достаточно чётко определить, кто из кандидатов – **К**, а кто – **Ч**. Человеком будет тот, чьи результаты окажутся более схожими с результатами контрольного опроса. Компьютер **К** будет постоянно проваливать ТФ, так как невозможно запрограммировать некий априорный способ ассоциативной связи понятий. Имеется единственный способ различения машиной всей данной совокупности ассоциативных связей между понятиями – это «окупнуться» в поток конкретной жизнедеятельности людей, жить среди людей. Однако ТФ представляется неубедительным в силу ряда причин: 1)

сомнительна возможность выявления значимых статистических зависимостей между понятиями на этапе контрольного опроса; 2) разработчик Р может пользоваться различного рода словарями, где учитывается лексическая сочетаемость слов, например [Мельчук, 1995].

Тест Блока – психофункционалистский тест

Н. Блок в работе «Психологизм и бихевиоризм» [Block, 1981] критикует канонический ТТ, показывая, что «игра в имитацию» является разновидностью оценки поведения в контексте лингвистического бихевиоризма. Так как предметом бихевиористского анализа является не ментальная структура, а объективно фиксируемые параметры поведения (реакции), определяемые внешними воздействиями (стимулами) (в лингвистическом бихевиоризме стимулы и реакции – это вербальные выражения), то здесь имеется ряд стандартных антибихевиористских возражений на возможность осуществления адекватной оценки: 1) *аргумент Чизголма и Гича*, согласно которому невозможно выделить конкретное психическое состояние в поведенческой диспозиции без учёта *всей совокупности иных психических состояний*. Анализ каждой реакции может продолжаться бесконечно в континууме рассмотрения внутренних процессов, и результаты данного анализа невозможно зафиксировать в схеме «стимул-реакция»; 2) *аргумент «совершенный актер» Х. Патнэма*, согласно которому *различные* классы психических явлений могут продуцировать *одинаковые* поведенческие диспозиции и, наоборот. Например, в силу принятых в обществе законов «сверх-супер-спартанцы» не будут выказывать боль, даже если на самом деле люди её испытывают. Напротив, человек, не испытывающий боли, может притворяться, что ему больно; 3) *аргумент «паралитики» и «мозги в бочке»*: эти «пациенты» вообще не выказывают определенных признаков, означающих, что они испытывают боль.

В свою очередь, интеллектуальное поведение, как правило, продуцируется комбинацией: {интеллект + обычная склонность к мышлению}. Однако невозможна комбинация {отсутствие интеллекта + желание обмануть, что интеллект имеется}.

Отсюда вытекает: 1) нужно изменить ТТ, так как он рушится под натиском антибихевиористских аргументов; 2) следует искать иное определение ТТ.

Предлагается новая формулировка ТТ (нео-ТТ), или тест Блока, который означает *способность продуцирования* осмысленной последовательности вербальных реакций на некоторую последовательность вербальных стимулов, но *не акт продуцирования*. Новая постановка усиливает ТТ, ТБ пройти намного

сложнее, так как снимаются вышеперечисленные аргументы: 1) можно представить много факторов того, что интеллектуальная система *не будет расположена* к выдаче *осмысленных* ответов, однако эти факторы не скажутся на её *способности*; 2) невозможно притворяться разумным и не быть таковым; 3) у «паралитиков» и «мозгов в бочке» имеется *способность* реагировать осмысленно, однако им недостаточно средств для реализации.

Для прохождения теста Н. Блок предлагает машину, которая способна производить осмысленную последовательность вербальных реакций на вербальные стимулы [Block, 1981, Р. 9 – 13]. Поведение машины полностью предопределено людьми – проектировщиками, инженерами, программистами и др. Разработчики упорно трудятся, применяют свой разум для реализации всевозможных последовательностей осмысленных ответов на предполагаемые последовательности вербальных стимулов. На поведенческом уровне данная машина выглядит интеллектуальной. Однако знание её внутреннего устройства убеждает в полном отсутствии у неё «интеллекта» – *интеллект, «проявляемый» машиной как в форме актуального поведения, так и в форме диспозиции интеллектуального поведения – суть интеллект разработчиков.*

Н. Блок, по сути, подсказал возможность прохождения ТБ – надо включить в сценарий имитационной игры разработчика **Р**. Тогда на любой вопрос со стороны **С** может быть получен ответ со стороны **К** при содействии **Р**. Единственный способ для **С** различить **К** от **Ч** (и тем самым дискредитировать компьютерную программу) можно назвать «анатомическим» – т.е. использовать принцип «вскрытие покажет». Только тогда программа не пройдёт тест Блока, когда вскроются внутренности устройства реализации.

Тест Серля - тест на «понимание»

В 1980 г. Дж. Серль предлагает знаменитый мысленный эксперимент «Китайская комната» в работе «Разумы, мозги, программы» [Searle, 1980]. Здесь отрицается концепция сильного ИИ, согласно которой при прохождении ТТ компьютер не просто моделирует человеческую способность к пониманию (это характерно для позиции слабого ИИ), но и: 1) на самом деле *понимает* текст; 2) компьютер и её программа объясняют человеческую способность «понимания».

Сценарий Дж. Серля разыгрывается по схеме ТТ1, но теперь роль **Ч** и **К** совмещает в одном лице сам Серль-в-комнате (Searle-in-the-room) – мысленный эксперимент осуществляется с позиции первого лица. Судей двое – китаец и англичанин. Когда задаёт вопросы судья-китаец, то Серль-в-комнате, играя

роль **К** и выполняя чисто синтаксические преобразования над китайскими иероглифами (он снабжен инструкциями перевода на родном английском языке), выдаёт китайцу ответы, которые тот воспринимает как осмысленные. Когда вопросы задаёт судья-англичанин, Серль-в-комнате играет роль **Ч**, непосредственно понимая задаваемые вопросы и не выполняя никаких синтаксических преобразований. Отличие между первым контуром игры (когда понимания нет) и вторым контуром игры (когда понимание очевидно для Серля-в-комнате) заключено, по мнению Серля, в *интенциональности*, так как человек – это биологическое существо, способное ощущать, действовать, понимать, обучаться и т.д. Способность к интенциональности заложена в биологическом прошлом человека. Интенциональные же феномены (в частности, квалиа) не редуцируемы к формальной схеме. Формальная модель по причине отсутствия самостоятельных казуальных сил не самодостаточна для воспроизводства феномена интенциональности. Интенциональность, якобы «имеющаяся» у **К**, на самом деле имеется лишь в головах **Р** (кто их программирует) и **С** (кто интерпретирует ответы).

Детально тест Сёрля будет раскрыт в составе одного из вариантов комплексного теста, поэтому здесь не следует на нем останавливаться. Сделаем лишь ряд замечаний. Несомненно, данный тест имеет важное значение для развития когнитивной науки. Ряд исследователей полагает, что именно ТС послужил становлению нового этапа развития когнитивной науки – в целях опровержения ТТ [Cole, 2009]. Хотя условия прохождения ТС по сути воспроизводят условия прохождения ТБ, так как здесь также концептуально задействован **Р**. Поэтому тест Сёрля пройти можно, ничего принципиально нового относительно ТБ здесь нет.

Тест Ватта – инвертированный ТТ

Стюарт Ватт в работе «Наивная психология и инвертированный тест Тьюринга» [Watt, 1996] считает, что успех прохождения ТТ связан с естественной, наивно-психологической установкой приписывать ментальные свойства любым системам – другим людям, животным, природным явлениям, в том числе и компьютерам. Для преодоления наивной позиции наблюдателя предлагается инвертированная версия теста Тьюринга. Это означает, что надо «обратить» роли игроков. *Судья должен быть судим*, так как именно он приписывает интеллект системе, которая таковым может и не обладать. То есть теперь тестированию подлежит судья. Для этого схема ТТ2 модифицируется – вместо одного судьи теперь двое – судья-компьютер (**С-К**) и судья-человек (**С-**

Ч). Тестирование осуществляется как прежде. Но теперь **С-Ч** (в роли его должен выступать опытный эксперт) должен сличать ответы от игрока **К** или **Ч** с показаниями **С-К**. Если **С-К** начинает приписывать интеллект тестируемой системе, то это является поводом для индуктивного заключения о том, что **С-К** – машина.

По поводу невозможности организации «инвертированного» тестирования существует ряд убедительных мнений: 1) тесту присуща *проблема дурной бесконечности*, так как для каждого **С-Ч**, судящего **С-К**, должен стоять другой **С-Ч'**, который судит первого **С-Ч**, и т.д. [French 1995]; 2) тесту присуща *проблема параллелизма*, так как невозможно обнаружить момент, когда именно **С-Ч** начинает отличать себя от **С-К** [Bringsjord, 1996]. Таким образом, вопрос о возможности прохождения ТВ должен решаться после вопроса о самой возможности тестирования.

Тест Харнада – тест НБИКС-системы

Многоуровневую классификацию тестов Тьюринга предлагает С. Харнад в работе «Разумы, машины и Тьюринг: неразличимость неразличимостей» (2001 г.) [Harnad, 2001]. Выделяются следующие уровни.

ТХ0¹. Это – уровень «игрушечных» ТТ – не полноправных тестов, а лишь некоторых фрагментов, ограниченных как по длине, так и по содержанию. Такие тесты не отвечают исходному замыслу А. Тьюринга. Однако все попытки моделирования интеллекта, известные на сегодняшний день, выше данного уровня не поднялись.

ТХ2 – общепринятое понимание ТТ. Именно его и имел в виду А. Тьюринг. Иногда ТТ данного уровня называют тестом «друг по переписке». Длина **ТХ2** равна протяжённости человеческой жизни С. Программа «Элиза» из примера «Девушка по переписке» принадлежит этому классу программ.

ТХ3 – это так называемая «роботизированная» версия **ТХ2**. Имеется возможность манипуляции предметами внешнего мира. Процедура идентификации систем, которые проходят данный тест, требует реализации принципа «Вскрытие покажет».

ТХ4 – это компьютерные системы, неотличимые как в плане **ТХ3**-неотличимости, так и в плане микрофизической организации системы. Здесь имеет место «тотальная неотличимость» компьютерной системы от человека,

¹ Мы вводим свои символы для обозначения уровней, отличных от харнадовских, с целью достижения единообразия нашей классификации ТТ.

включая мельчайшие внутренние нюансы телесного строения. На данном уровне можно разместить так называемый «Тест Тьюринга с мозгом» («Brainy-Turing-Test»), где в дополнение к поведенческой неотличимости требуется неотличимость на уровне нервной системы и предполагается использование коннекционистских моделей [Flanagan, Polger, 1998].

ТХ5 – на этом уровне происходит отбор ТХ4-систем, соответствующих Единой Теории Реальности, которая способна (если она вообще возможна) объединить все возможные знания о природе, обществе, человеке. Данный уровень С. Харнад ввёл, по всей видимости, для того, чтобы допустить возможность истинности нескольких отличающихся друг от друга теоретических описаний одной и той же реальности (т.е. ТХ4-систем). Истинность в данном случае имеет различные метафизические предпосылки у создателей различных теорий.

В соответствии с классификацией С. Харнада, современный уровень достижений ИИ не превысил «игрового» уровня ТХ0. Однако то, что прохождение ТТ возможно, пускай даже на этом уровне, демонстрируют факты. Рассмотрим эти факты.

2.2.3. Возможности прохождения тестов Тьюринга

Практическая возможность (тест Лойбнера)

В январе 2000 г. в Dartmouth College состоялась юбилейная конференция «Будущее теста Тьюринга: следующие пятьдесят лет», на которой обсуждались такие теоретическими вопросы, как оценка влияния ТТ на современность и определение прогнозов программирования ТТ на будущее. Практическая часть конференции состояла в розыгрыше сценария ТТ1. Участвовали шесть программ. В целом результаты не подтвердили предсказаний Тьюринга – судьи правильно отличали компьютер от человека в 91% после 5 минут и в 93% после 15 минут соревнований [Platt, 1995]. Тем не менее, в настоящий момент в сфере разработки ТТ-программ сформировалась своеобразная индустрия, в основном игрового характера (уровень ТХ0). С 1991 г. введена специальная премия, соразмерная Нобелевской – Лойбнеровская премия, которую учредил Хью Лойбнер (Hugh Loebner). Премия в 100 000 долларов присуждается за выигрыш, 2 000 долларов – за участие. Установлена длина теста – три часа. Тестирование проводится по сценарию ТТ1. На вопросы десяти судей отвечают пять **К** и пять **Ч**. Пока ещё нет победителя, которому можно было бы присудить полную премию [Алексеев, 2006-2].

Однако не следует отождествлять «игру в имитацию», которая прослеживается на состязаниях Лойбнера, с тестом Тьюринга. Тест Тьюринга – это теоретический план исследований искусственного интеллекта, не связанный с практической реализацией. Прошла или не прошла программа тест Лойбнера – это аналогично тому – пришла ли данная лошадь на скачках первой или вообще не пришла к финишу. То есть тест Лойбнера не гарантирует даже эмпирической валидности теста Тьюринга.

Совершенно иначе следует оценивать серьёзные программы искусственного интеллекта. Например, под тьюринговое тестирование подпадает любая экспертная система, система распознавания образов, генерации логических теорем и пр. Все они – кандидаты на прохождение, и, более того, все эти системы с учётом достаточных ограничений предметной области могут проходить тест Тьюринга.

Попробуем более скрупулёзно разобраться с тем, что значит «возможность» прохождения ТТ.

В контексте проблемы сознания, философии искусственного интеллекта, в любой теоретической сфере, в которой используются мысленные эксперименты, к слову «возможность» следует относиться осторожно. Здесь, конечно, не приемлемы жаркие диалектические дебаты на тему «возможности – вероятности – действительности», с характерными для них тёмными высказываниями типа: «возможность есть будущая действительность», но при этом «возможность уже существует в действительности – именно как возможность». Вот что именно хочет диалектик выразить этими предложениями?

Для позитивного решения вопросов о возможности прохождения ТТ следует ввести ряд ограничений: 1) судья – обычный человек (как молодой человек из примера «Девушка по переписке»), а не абсолютный, всезнающий, тем более трансцендентный субъект; 2) длина теста ограничена; 3) границы теста заданы некоторой фиксированной предметной областью, например, сферой профессиональной деятельности судьи; 4) разработчик программы – эксперт в этой области; 5) стена Тьюринга непроницаема, например, диалог ведётся посредством электронной почты.

Выделим следующие виды «возможности» прохождения компьютерной системой тестов Тьюринга: физическую, логическую и метафорическую.

Физическая возможность

К данной разновидности «возможности» относится всё то, что согласуется с формальными условиями опыта, например, наглядная демонстрация или конкретный факт. Действительным в таком случае будет всё то, что согласуется с материальными условиями опыта, скажем, с ощущением (по Канту). Обычно исследователи апеллируют к неким всеобщим и предельным формам доказательства физической невозможности, начиная от общеизвестного примера с мельчайшей вероятностью написания обезьяной сонета Шекспира. При изучении проблематики ТТ можно выделить следующие интересные примеры доказательства физической невозможности.

1. «Комбинаторный взрыв» (Дж. Миллер). Существует порядка 10^{30} грамматических предложений длиной в 20 слов. Если произвольно допустить, что 10^{15} также семантически корректны, то для прохождения ТТ длительностью в один час может понадобиться порядка 100 таких предложений. А это 10^{1500} строк – число, превышающее количество элементарных частиц во вселенной [Block, 1981].

2. «Прыжок на Луну» (Ф. Типлер). Машина Н. Блока (выше мы видели, что данная машина реализует ТТ силами разработчиков) физически невозможна, так как для её реализации (если информационные процессы осуществлять «вручную») потребуется больше энергии, чем требуется человеку для прыжка на Луну) [Dowe, Oppy, 2003, подразд. 4.2].

Такие доказательства в условиях наших ограничений и допущений не убедительны. Мы имеем реальные факты прохождения программами тестов Тьюринга (это показывают лойбнеровские состязания, успехи применения экспертных систем) и, следовательно, наглядную демонстрацию возможности прохождения ТТ. Отсюда можно сделать заключение о физической возможности прохождения ТТ.

Логическая возможность

Логическая возможность – это всё то, о чем можно ясно «помыслить» и что можно чётко описать – так логическую возможность определяет Д. Чалмерс в аргументе мыслимости зомби [Kirk, 2006]. Логическая возможность, или возможность «помыслить» некоторую ситуацию (или положение дел), обычно противопоставляется вообразимости, которая, в свою очередь, обуславливает физическую возможность. Так, невозможно вообразить или представить бесконечное число. Однако его можно непротиворечиво помыслить, например, в системе арифметики Пеано. Если мысленный эксперимент с прохождением программой ТТ логически корректен и все компоненты его – отчётливые мысли,

связанные стандартными логическими принципами, тогда ТТ логически можно пройти. В рамках традиционной логики логическая невозможность прохождения ТТ апеллирует к противоречиям, возникающим в процессе самоописания и самореференции с характерными для него вопросами, на которые невозможно получить ответы (unanswerable question). Логическая невозможность ТТ подмечена А. Тьюрингом в ответе на «математическое возражение» [Turing, 1950, подразд. 2.3], где он, кроме того, указал, что людям также присуще впадать в логические противоречия. Этот ответ показался неудовлетворительным Лукасу и Пенроузу, и возникла полемика по поводу логической возможности ТТ, получившая название «Аргумент Гёделя». Приложение теоремы Гёделя применительно к проблематике ТТ кратко изложено в [Dowe, Oppy, 2003]: 1) пусть **К** – цифровой компьютер; 2) так как **К** подпадает под аргументацию Гёделя (точнее, под возражение Лукаса – Пенроуза), то имеется вопрос **q** для **К**, на который нет ответа; 3) если некоторое утверждение **Е** не является предметом возражения Лукаса – Пенроуза, то для **Е** на любой вопрос имеется ответ; 4) человеческий интеллект нельзя подвести под возражение Лукаса – Пенроуза; 5) таким образом, нет «безответных» вопросов для человеческого интеллекта; 6) поэтому вопрос **q** не является «безответным» для человеческого интеллекта; 7) задавая вопрос **q**, человек (судья) может определить, кто перед ним – компьютер (т.е. ответа он не получит и тестируемая система «зависнет») или человек; 8) таким образом, компьютер **К** не пройдет ТТ; 9) логически невозможно, чтобы **К** прошёл ТТ.

На наш взгляд, очевидна правота Тьюринга – если человека попросят разрешить софизм, то человек впадет в противоречие, для него вопрос **q** также будет невозможен и судья не отличит компьютер от человека. Широко известен подход к преодолению софизмов, предложенный Б. Расселом. Он предлагает вводить целый ряд ограничений на условия суждения – в первую очередь, устанавливать иерархию языков, типизировать язык [Рассел, 1999, С. 65 – 66]. В принципе большинство интеллектуальных систем построено на основе выделения типов сущностей в предметной области. При этом, несомненно, система ИИ должна обладать адаптивными способностями к перестройке своих формализованных «знаний» – к их самонастройке, самоорганизации, самообучению (см., например, у Дж. Маккарти в [McCarthy, 1995]). Но здесь имеется в виду не адаптация «вообще», а приспособление ИИ-системы к классам конкретных состояний внешней и внутренней среды системы, которые, конечно, должен предусмотреть разработчик. Поэтому не будет логической ошибкой предположить наличие у разработчика достаточных средств – метаязыковых для конкретной

системы ИИ, способных обеспечить перепрограммирование системы с целью работы в тех предметных областях, для которых у неё недостаточно средств к самообучению, самоописанию и т.п. Здесь также предполагается рассматривать не предметную область «вообще», а, как предлагал Рассел, концептуально её ограничивать. Имеются и соответствующие этим требованиям технологии программирования, такие как технологии автогенеративного программирования, в которых программный код автоматически создается при задании значений параметров внешней и внутренней среды компьютерной системы [Чарнецки, Айзенекер, 2005].

Таким образом, в условиях допущений и ограничений на ТТ, представляется *логически возможным* его прохождение.

Метафорическая возможность

Рассмотрим ещё одну разновидность возможности прохождения теста – метафорическую. Метафорически возможно всё то, что является убедительным, убедительность подкрепляется метафорой, которая, как известно, характеризует нечто неопределимое путем референции к явным и достаточно чётко определенным вещам и событиям.

Метафорика и искусственный интеллект тесно «спаяны». Часто считается, что ИИ – это метафора естественного интеллекта. Более того, *ценой неадекватного принятия (вернее, отторжения) метафоры ИИ в среде отечественных учёных явилось отставание нашей науки в столь важнейшей для развития страны сфере научно-технического знания* – считает Д.А. Поспелов [Поспелов, 1982, С. 5 – 7]. Для многих исследователей ИИ, которые принимают метафору ИИ, характерно игнорирование более фундаментального вопроса – что именно значит «метафора». Так, например, В. Рапопорт, опровергая аргумент «Китайской комнаты» Дж. Серла, призывает к метафорическому пониманию слов, применяемых в сфере исследований ИИ [Rapoport, 2000, Р. 471]. Он вполне убедительно показывает, в частности, что выражение «полет самолёта» – это метафорическое употребление слова «полет птицы», и, по аналогии, «искусственный интеллект» – метафора «интеллекта». Далее он достаточно отчётливо разъясняет преимущества для разработки искусственных агентов своей концепции «синтаксической семантики», согласно которой состав «знаний» агента должен содержать разделяемые синтаксические структуры. Разделение происходит, как минимум, на две составляющие, одна синтаксическая структура начинает играть роль семантики для другой структуры. Однако далее он совершенно неубедительно полагает, что самореференция такой двумерной структу-

ры (т.е. отображение синтаксической структуры на самое себя, приводящее к двуаспектному, синтаксическо-семантическому образованию) может послужить метафорой «понимания», например понимания Серлем-в-комнате китайского языка. Принятие понятия «метафоры» без какого-либо предварительного анализа в дальнейшем вызывает ряд трудностей в понимании, например того, каким образом самореференцирующая синтактико-семантическая структура может послужить *метафорой* для «интенциональности», «понимания», «ментального свойства» – всего того, чем владел Серль-в-комнате, говоря об этих понятиях от первого лица.

Подобного рода зыбкость суждений характерна во многом и для других исследователей ИИ, призывающих к метафорическому, расширенному значению слова «интеллект». Поиск обоснованности понятия «искусственный интеллект» и, следовательно, всех иных понятий, с которыми оно соотносится, требует применения лингвофилософского анализа самого слова «метафора».

Воспользуемся с этой целью теорией метафоры Д. Дэвидсона [Дэвидсон, 1990]. Метафоры – это не грезы, не «сон языка», не расширенное или удвоенное значение, не средство передачи необычных идей, не способ избежать неопределённости понятия, не сравнение, не неявное когнитивное содержание, передаваемое автором для того, чтобы получатель уловил его с целью понимания сообщения. Метафора не объяснима путем обращения к ее скрытому содержанию. Интерпретации метафоры недопустимы, а содержание невыразимо.

Для чего же тогда нужна метафора, если она не способна отражать истинное или ложное положение дел? Метафора, как полагает Д. Дэвидсон, пригодна для серьезного научного или философского разговора; целиком принадлежит сфере употребления (но не определения понятий); привлекает внимание к тому, что автор хочет сообщить, и к новому и неожиданному сходству между предметами; заставляет нас заметить то, что иначе могло бы остаться незамеченным. Метафора связана с образным использованием слов и предложений и всецело зависит от обычного или буквального значения слов (и состоящих из них предложений). Метафора подобна речевым актам: утверждению, намеку, лжи, обещанию, выражению недовольства и др. Но она служит не для передачи содержания понятия, а для убеждения в истинности данного понятия. Метафора не несёт какого-то содержания или не имеет какого-то значения, кроме, конечно, буквального.

В контексте данной теории метафорическая возможность того или иного положения дел означает убедительность метафоры для социокультурного со-

общества. Именно убедительность обеспечивает единообразное понимание слов, значения которых не подлежат строгой верификации или фальсификации. К таким словам относится любой ментальный термин, в частности, «интеллект».

Так как метафора «означает только то (или не более того), что означают входящие в неё слова, взятые в своем буквальном значении» [Дэвидсон, 1990], то использование метафоры предполагает апелляцию к здравому смыслу. Убедительность в том или ином положении дел осуществляется путем демонстрации, состоящей из очевидных для сообщества вещей, событий, процессов, сценариев.

Аналитическая теория метафоры Д. Дэвидсона позволяет исследовать метафорическую возможность прохождения теста Тьюринга. Если метафора, раскрывающая устройство компьютера, проходящего тест Тьюринга, окажется для нас убедительной, то тогда мы можем утвердительно судить о метафорической возможности прохождения теста Тьюринга.

Выводы: неубедительность компьютерной метафоры «человек-компьютер» А. Тьюринга

Для объяснения возможности создать систему, которая способна играть в ИВИ, в статье 1950 г. Тьюринг предлагает метафору «человек-компьютер» (human computer). Она, в свою очередь, является производной от метафоры «машина Тьюринга» (1936 г.), убеждающей в том, что функционирование устройства с конечным числом состояний, с лентой памяти и записывающе-считывающей головкой эквивалентно понятию алгоритма. Метафора алгоритма была принята научным сообществом и оказалась достаточно убедительной (см., например, отзыв С. Яновской в «Предисловии к русскому изданию» перевода статьи Тьюринга [Тьюрин, 1960, С. 3 – 18]). Однако если интуитивное понятие алгоритма хорошо описывается машиной Тьюринга, то ещё более интуитивное понятие «интеллект», задаваемое игрой в имитацию (т.е. тестом Тьюринга), для С. Яновской показалось неубедительным. Напомним о том, что она посчитала невозможным отличить **М** от **Ж**, так как гендерный тест имеет шансы «50 на 50». Логическую невозможность (для специалиста по логике это равносильно неубедительности) Яновская объясняла противоречием, которое возникает в сценарии ИВИ2: вопрос о возможности запрограммировать интеллект **Ж** должен быть решён с помощью «игры в имитацию», так как с помощью её и определяется то, что такое «мышление». Яновская воспринимает ИВИ не как метафору, а как буквальное руководство к определению «интеллекта».

На наш взгляд, несостоятельность теста Тьюрина проистекает из-за неубедительности собственно внутреннего строения компьютерной метафоры искусственного интеллекта, метафорческими нестыковками при попытке демонстрации способа функционирования такого устройства. Рассмотрим доводы по этому поводу.

Метафора «человек-компьютер» 1950 г. состоит из машины Тьюринга, усовершенствованной до уровня универсальной цифровой вычислительной машины. Однако усовершенствование, например создание средств организации «вероятностных скачков» по машинной таблице инструкций (этого не было в машине Тьюринга), не прибавляет убедительности интеллектуальному способу её функционирования и самостоятельной способности варьировать сценариями, хранящимися в книге правил, которую человек-компьютер «берёт» для выполнения тех или иных заданий. В книге правил представлены сценарии действий МТ. Допустим, каждый сценарий представлен на отдельной странице Книги, т.е. на каждой странице – несколько лент, с которыми работает машина Тьюринга. Возникает ряд вопросов, связанных с неотъемлемыми параметрами мышления, к которым причисляются, например, абстрагирование, обобщение, органичение, активность. Ни один из этих атрибутов не присущ человеку-компьютеру.

Для демонстрации этого воспользуемся «семейной аналогией» (см. подразд. 4 [Turing, 1950]). Книга правил содержит страницу, на которой имеется запись, оставленная женщиной для указания правил, которым должен следовать её сын в такой-то ситуации. Эти правила, по сути, «интеллект» матери. Возникает другой сценарий. Этот сценарий фиксируется матерью на другой странице Книги правил. Мальчик действует в соответствии с ними. И с третьими, и т.д. А сам-то мальчик когда будет формулировать свои собственные правила? Ведь для этого требуется активность, обобщение, абстрагирование и другие параметры интеллекта. Кто пишет книгу правил для УЦВМ? Очевидно, человек. Но если так, то не надо утвердительно говорить об искусственном, машинном, компьютерном интеллекте. Такого «интеллекта» нет в тьюринговой метафоре. Поэтому Тьюринг стал известным не только за то, чего никогда не предлагал, т.е. за тест Тьюринга, который сам не называл «тестом» (по мнению Лассажа). Все его предложения по поводу проекта искусственного интеллекта – это попросту «липа». Тьюринговый проект интеллектуального компьютера совершенно несостоятелен.

Имеется, конечно, ряд более фундаментальных замечаний по организации человека-компьютера. Они касаются каузальных зависимостей между ним и внешней средой. Допустим, метафора работы человека-математика совершенно очевидна (напомним, что эта метафора является основой тьюрингового определения «алгоритма»), так как можно ясно и отчетливо вообразить, как считывающе-записывающая головка скользит по ленте, записывает на неё символы или считывает их. Однако с квазиалгоритмом ситуация иная. Совершенно непонятно, откуда символы возникают и по каким субсимволическим правилам они есть то, что они есть? Какое такое чудо требуется, чтобы осуществить сигнификацию, сформировать совокупность слов, каузально определённых сцена-риями систем, в которых человек-компьютер функционирует. И ещё большее чудо – как «вживить» эти символы в страницы Книги правил. Ведь эти символы как руководства к действию должны составлять «внутренность машины» при загрузке их в машинную таблицу и составлять неразрывное функциональное единство со способом функционирования человека-компьютера в целом.

Подобного рода проблемы связаны с ограниченностью репрезентативной парадигмы ИИ. В более общем плане они обусловлены отсутствием эволюционно-эпистемологических механизмов в проектах ИИ. Эти проблемы мы далее затронем при изучении метафоры интеллектуального компьютера, построенного на базе машины Корсакова, однако, к сожалению, не решим.

Метафора «человек-компьютер» *не убедительна*.

Даже если и имеются практическая, физическая и логическая возможности прохождения теста Тьюринга, риторическая возможность этого отсутствует.

Модели искусственного интеллекта в контексте существующей парадигмы компьютерной технологии не работают. Однако метафорическая невозможность прохождения теста Тьюринга не означает необходимости сворачивания проблематики ИИ. Напротив, нужен поиск более убедительных компьютерных метафор интеллекта. Комплексный тест Тьюринга способствует этому.

2.3. Идея комплексного теста Тьюринга

Идея комплексного теста Тьюринга (КТТ) крайне проста. Она заключается в совместном изучении и использовании многообразия версий и модификаций оригинальной идеи А. Тьюринга. Вызывает недоумение то, что за шестидесятилетнюю историю развития теста Тьюринга никто из исследователей не попытался совместить в комплексном формате различные версии. Собственно в выдвигании этой крайне примитивной идеи комплексирования различных версий частных тестов автор усматривает свой основной вклад в развитие философско-методологических исследований ИИ.

2.3.1. Версии и интерпретации теста Тьюринга

Версии теста Тьюринга

Оригинальный ТТ изучает особенности компьютерной реализации диалогового интеллекта, т.е. той формы мышления, которая востребована для организации связного разговора с собеседником на любые темы. Креативный ТТ (тест Лавлейс) оценивает возможности компьютерного творчества. Субкогнитивный тест (тест Френча) раскрывает многоуровневую подсознательную структуру с учётом социокультурного контекста. Глобальный ТТ (тест Блока, «нео-ТТ», «новый ТТ») задаёт необходимое условие для приписывания системе интеллекта с учётом антибихевиористской критики оригинального ТТ и изучает проблемы построения общечеловеческой машины, реализующей всевозможные диспозиции интеллектуального поведения. Индуктивный ТТ (тест Мура) обеспечивает достаточное условие для доказательства факта компьютерного мышления и определения основных характеристик искусственного интеллекта. Тест на понимание (тест Сёрля, «Китайская комната») изучает вычислительные аспекты феномена понимания в контексте понятия интенциональности и обосновывает сильный антикогнитивистский тезис о невозможности «сильного искусственного интеллекта». Тест на приписывание ментальности (тест Ватта, «Инвертированный ТТ») – учитывает способности человека наделять психическими качествами и свойствами нечеловеческие системы. Тотальный ТТ (тест Харнада, НБИКС-тест) идентифицирует интеллект в системе, построенной на базе интегративного комплекса нано-, био-, инфо- и когнитивных технологий (приставка «НБИКС» предложена для подчёркивания сквозной концептуальной востребованности данной модификации ТТ на всех уровнях этих перспективных технологий для оценки интеллектуальности их продуктов). Самый тотальный тест Тьюринга (тест Швейцера) дополняет предыдущий эволюционными

условиями развития сообщества «людей», по всем своим социокультурным проявлениям не отличимых от естественных. Тест на парадоксальность (тест Гёделя – Лукаса – Пенроуза) изучает возможности компьютерного разрешения логико-математических парадоксов. Параноидальный ТТ (тест Колби – Деннета) изучает лингвистические аномалии и служит надёжным критерием оценки интеллектуальности существующих экспертных систем. Кибериадный ТТ (тест Баресси) предлагает проект развития искусственного общества искусственных людей. Тест на непосредственные жизненные коммуникации (тест Шеннона) поднимает проблему автономности интеллекта в социальной среде и его независимости от системы социальных взаимодействий, воздействий, мотиваций. Гендерный ТТ (тест Геновой) решает внебиологические вопросы отличий мужчины от женщины. Тест на квалиа (тест Джексона – Брингсйорда) исследует эпифеноменалистские, репрезентационистские, экстерналистские и прочие подходы к вычислительному воспроизведению феномена дифференцируемой качественности психических свойств и состояний. Тест зомби (тест Кирка – Чалмерса, «Философские зомби») ставит акцент на роли сознания в жизни человека и общества, а также на компьютерной реализации бессознательных «квазисуществ», поведение которых неотлично от человеческой деятельности. Тест на искусственность (тест Дэвидсона, «Арт») изучает нейрофизиологические проблемы демаркации искусственной/естественной системы. Экстерналистский тест (тест Патнэма, «Мозги в бочке») оценивает статус реальности сознания и собственного существования в условиях компьютерной стимуляции мозговой активности.

Имеется целый ряд других интересных модификаций оригинального ТТ. Интерпретации ТТ, делающие акцент на вопросе о том, для чего собственно нужен оригинальный ТТ и его различные версии, также характеризуются большим многообразием. Достойный внимания пример наложения различных интерпретаций теста, которые включают в себя тест Серля, классификацию проблем сознания Д. Чалмерса, неоидеализм Р. Пенроуза и теоретико-алгоритмический тезис А. Чёрча, приведен в работе Н. Ключевой (Ключевой) [Трушкина, 2005].

Интерпретации теста Тьюринга

В соответствии со значительным расширением сферы применения теста Тьюринга, не ограниченного лишь имитацией и моделированием «интеллекта», несколько расширим предыдущее определение теста Тьюринга в пп2.1.3).

Тест Тьюринга – это: 1) критерий когнитивной компетентности систем различной природы – человеческих, биологических, социальных, технических, «других», «чужих», «иных»; 2) способ сбора данных и фактов, аргументирующих приписывание (атрибутирование) компьютерной системе интеллекта, сознания, понимания и пр.; 3) компьютерная система идентификации функциональных параметров сознательной, подсознательной, бессознательной деятельности; 4) «соломинка», вытягивающая исследователя из «болота» метафизических спекуляций по проблеме сознания; 5) инструкция по поводу обмана человека в «человечности» машины (компьютера); 6) техническое задание на разработку когнитивно-компьютерной системы; 7) функциональная аппроксимация субъективной реальности, каждый элемент которой всё более и более точно воспроизводит и корректирует предшествующее определение; 8) лингвологическая модель субъективной реальности; 9) целевой проект электронного «усилителя» творческой деятельности; 10) квазиалгоритмическая модель субъективной реальности и пр.; 11) яркий мысленный эксперимент, проясняющий схему описания и создания когнитивно-компьютерных систем и убеждающий в возможности реализации этой схемы (на наш взгляд, это наиболее существенная интерпретация ТТ).

Многочисленные версии, модификации, интерпретации ТТ создаются за счёт изменения тех или иных исходных понятий расплывчатой идеи Тьюринга по поводу интеллекта и вычислительных машин. Модифицируются мотивы, ориентиры, направления, цели, участники, параметры «стены Тьюринга» (канал передачи сообщений между тестирующим и тестируемым), длина теста, архитектура компьютерной системы и пр. Комбинаторная манипуляция параметрами ТТ, характеристиками модификаций ТТ и модификаций предшествующих модификаций порождает ряд других тестов. В итоге эти изменения уточняют, дополняют, обосновывают и обеспечивают создание новых информационных, когнитивных, психологических, нейрофизиологических, программных, технических, лингвистических и других концепций и проектов компьютерных интеллектуальных систем.

Несомненно, здесь решаются важные задачи философии, теории и инженерии искусственного интеллекта. Тем не менее, отражая лишь ограниченные, хотя и существенные, стороны и аспекты бесконечно сложной интеллектуальной деятельности человека, данные модификации следует рассматривать в качестве частных тестов Тьюринга относительно перспективного комплексного теста.

Требования к частным тестам Тьюринга

Определение понятия частного ТТ целесообразно сформулировать в виде требований, которым должна отвечать та или иная версия, модификация, интерпретация ТТ. Среди частных тестов выделим совершенные и несовершенные. *Совершенный ЧТТ* воспроизводит все ключевые идеи оригинального ТТ, подстраивая их к собственным целям и задачам. *Несовершенный ЧТТ* акцентирует внимание как минимум на одной компоненте сложной идеи ТТ и не претендует на её целостный охват. Следует отметить, что к ЧТТ (совершенному его варианту) относится и сам оригинальный тест Тьюринга, неявно сформулированный в [Turing, 1950] и получивший собственное наименование лишь в 1970-е годы.

Требования к совершенному частному тесту Тьюринга

В совершенном ЧТТ, следуя логике раскрытия оригинальной идеи Тьюринга в [Turing, 1950], необходимо учесть следующее.

1. *Предмет тестирования.* Здесь явно обозначается тестируемая человеческая способность – мышление, понимание, сознание, творчество, действие, выживание и пр. Чёткая дефиниция понятий не предполагается – не ясна суть этих сущностей. Так, в оригинальном ТТ принят антиэссенциалистский, лингво-бихевиоральный подход к определению понятия «интеллект». Собственно, для реализации этого подхода и предлагается сценарий игры в имитацию [Turing, 1950, P. 433 – 434]

2. *Реализационная система.* В этой части предлагается специальное устройство (механизм, система, технология), которое обеспечивает прохождение ЧТТ. В некоторых тестах этой системой является человек, социум, «мать-природа». Не предполагается, что данная система – реально действующая инженерная конструкция. Тем не менее, её концептуальный проект должен быть достаточно убедительным относительно потенциальной работоспособности. Тьюринг апеллировал к конструкции универсальной цифровой машины, отдельные проекты которой в его время были уже реализованы [Turing, 1950, P. 435 – 442]. Однако ни по скорости вычислений, ни по объёму требуемой памяти эти проекты не удовлетворяли возможностям реализации ТТ. Кроме того, не была создана соответствующая программа диалогового взаимодействия. Всё это будет реализовано в 2000 г. – полагал А. Тьюринг. Отчасти он оказался прав – о фактах прохождения компьютерами теста Тьюринга уже в 1991 г. свидетельствовали лойбнеровские соревнования.

3. *Тестовые вопросы-ответы.* Представляется, что тщательная проработка версий ТТ во многом определяется наличием примера тестовых вопросов. Они конкретизируют, каким образом возможно исследование х-системы с целью выявления феноменов интеллекта, понимания, творчества, сознания и др.

4. *Дискуссия (полемика).* Здесь желательно воспроизведение структуры «полемического стандарта Тьюринга». А. Тьюринг предложил критическое обсуждение собственной идеи интеллектуальных машин. Как мы ранее отмечали, для этого он вообразил дискуссию по поводу ряда контраргументов. К ним относятся: теологический, математический, антисциентистский, телепатический, нейрофизиологический аргументы, а также аргументы «от сознания», «х-способностей» [Turing, 1950, P. 443 – 454].

5. *Социокультурные последствия.* Изучение ЧТТ крайне важно для обоснованного представления социокультурных последствий от предполагаемой его реализации. Например, Тьюринг полагал, что интеллектуальные машины станут конкурентами людей во всех чисто интеллектуальных сферах человеческой деятельности [Turing, 1950, P. 460]. Интеллектуальные системы сегодня глубоко укоренились в бытовом сознании современного человека в виде роботов, компьютерных игр, устройств технического зрения, автоматических мультиязыковых переводчиков, экспертных систем, стиральных машин, «умных» домов и пр. Большие социальные последствия предполагаются от реализации других ЧТТ, несмотря на то, что на сегодняшний день превалируют мифологемы превращённых форм обыденного сознания (изменённых, в первую очередь, посредством идеи научно-технологической мощи). Здесь витают вполне конструктивные идеи полноценной «жизни» антропоморфных роботов среди людей (тесты Френча, Харнада, Швейцера), растворения человека в виртуальной реальности (тесты Патнэма, Джексона – Брингсйорда), битв людей и компьютеров (тест Баресси), создания глобального интеллекта (тест Блока), программно управляемого зомбирования людей для создания каст господ и рабов (тест Кирка – Чалмерса) и пр. Много известных фантастических кинофильмов и книг инспирированы данными тестами. Как правило, тема социальных последствий широко обсуждается в последующих интерпретациях ЧТТ и интерпретациях этих интерпретаций. Также следует учесть, что большинство тестов несёт мощный критический заряд по поводу социальных последствий.

Все перечисленные характеристики явно или неявно участвуют в формировании дефинитной, критической и конструктивной функции КТТ. На формирование дефинитной функции в большей мере влияет характеристика «Предмет

тестирования ЧТТ». Конструкторская функция КТТ извлекается из пп. «Реализационная система», «Тестовые вопросы-ответы» и «Дискуссия (полемика)». За критическую функцию отвечают характеристики «Социальные последствия» и «Дискуссия».

Требования к несовершенному тесту Тьюринга

Разработано немало совершенных ЧТТ. Ещё большее разнообразие частных ТТ получается посредством различных интерпретаций тестов с учетом комбинаторных вариаций обозначенных выше характеристик совершенного ТТ. Если автор ЧТТ упускает какой-нибудь пункт требований, то в процессе дискуссии по поводу его теста можно из работ других авторов, как правило, реконструировать совершенный ЧТТ. Классическими примерами оригинальных совершенных ЧТТ представляются «Китайская комната» Дж. Сёрля, машина Н. Блока, субкогнитивистские игры Френча и другие крупные модификации ТТ. В этих работах мы видим явное изложение всех перечисленных выше пяти пунктов. Тем не менее, для построения КТТ целесообразно использовать и несовершенные ЧТТ. Изложенные в них незначительные вариации необходимы для того, чтобы не упустить нюансы и учесть как можно большее разнообразие мыслей по поводу идеи ТТ. Например, если не обращать внимания на постмодернистскую эквилибристику, вполне пригодным для КТТ представляется интерпретация, предложенная С. Жижеком [Zizek, 2002]. Он предлагает использовать тьюринговую идею ИВИ для обоснования перспектив бесполого постчеловечества, так как половое различие людей – это человеческая суть, и если упразднить половой антагонизм, к чему, по утверждению С. Жижека, ведут современные технологии, то человек на самом деле перестанет отличаться от машины. Здесь интересна идея параметризации понятия «постчеловечество» в контексте ТТ.

2.3.2. Построение комплексного теста Тьюринга

Способы построения комплексного ТТ

Способ построения КТТ определяется всем спектром возможных версий и интерпретаций ТТ. Ранее упоминалось, что среди них немаловажной трактовкой является то, что ТТ – это мысленный эксперимент (пример) построения компьютерной интеллектуальной системы. Поэтому в КТТ совмещаются форматы различных мысленных экспериментов (примеров). Следовательно, эпистемологически важными представляются вопросы о сути «мысленного эксперимента», о его соотношении с реальностью, о том, как из мысленных экспери-

ментов создать другой мысленный эксперимент, можно ли концептуальные стыки осуществлять непосредственно либо требуется опосредующий базис из реальных вещей, ситуаций и сценариев, как избежать концептуального парадокса «мысленного эксперимента мысленных экспериментов», какова роль метафоры в конституировании мысленного эксперимента и пр. Подобного рода философские вопросы требуют решения, и они, по всей видимости, важны для искусственного интеллекта. Однако часто случается так, что метафизические рефлексии нужны для ИИ в той же степени, в какой важна философия для науки и практики, исходя из известной марксистской аналогии соотношения между онанизмом и половой любовью. Например, М.В. Лебедев предложил метафизическую аналогию отношений: «трансцендентное/имманентное = дискета/дисковод». Звучит неплохо. Но что это значит?

На текущем этапе развития идеи комплексного ТТ целесообразно осуществление своеобразной библиографической работы. Задача – разработка инструментария систематического учёта ЧТТ и, по возможности, экспериментальное применение различных вариантов их комплексирования.

Здесь возможны следующие пути: 1) необходимо сконструировать универсальную концептуальную систему, которая позволит учесть все возможные ЧТТ; 2) необходимо собрать все множество примеров ЧТТ и далее обобщить их до комплексного концептуального уровня. Конечно, эти методологии неприемлемы по причине бесконечной сложности представлений об интеллекте, а также коррелированных с «интеллектом» понятий, выступающих в роли предмета различных ЧТТ. Также невозможен абсолютный прогноз относительно возможных путей имитации и/или воспроизведения в компьютерной системе феноменов, определяемых данными понятиями. Более реалистичным представляется формирование базовой системы, которая совмещает универсальный способ описания ЧТТ («универсальный» с точностью до классов обозначаемых понятий) с возможностью корректировки этой системы при обнаружении неучтённых ранее частных тестов, как в целом, так и отдельных их параметров. В связи с этим важная задача построения КТТ заключается в разработке языка описания частных ТТ, сочетающего универсальное с уникальным.

Язык описания КТТ

Данный язык предназначен для идентификации, систематизации, унификации, координации, обобщения, дифференциации, интеграции частных ТТ и их составляющих. Предложим исходный вариант данного языка, построенный на базе лингвистических средств нестрогого первопорядкового исчисления предикатов с выделением фундаментальных констант, функций и отношений, присущих совершенному ЧТТ. «Нестрогий» характер этого языка объясняется тем, что синтаксис не представлен конструктивным способом, присущим подобным языкам из учебников по математической логике и алгебраическим системам. Семантика также не имеет чётко определённой функции интерпретации по причине отсутствия логико-атомистического представления «сущностей», обозначаемых данным языком. Семантика задаётся способом прочтения текста, в котором даётся описание ЧТТ. Несмотря на «нестрогость», данный язык представляется перспективным для «библиографической работы» и в принципе обеспечивает решение выше перечисленных задач комплексирования ЧТТ посредством, например, Prolog-программирования.

Синтаксис начнём строить исходя из описания первого этапа игры в имитацию [Turing, 1950, P. 433], когда мужчина **М** играет за себя и за женщину **Ж**, женщина играет за себя, судья **С** играет роль судьи-человека. Задача судьи состоит в определении того, кто из игроков – **М**, а кто – **Ж**:

ИвИ1: $M[J], J[J] \Rightarrow C[C]: (a=M \ \& \ b=J) \vee (a=J \ \& \ b=M)$

Здесь используются следующие элементы языка описания КТТ:

1) ИвИ1: – обозначение (уникальное имя формулы) для идентификации частного теста Тьюринга. Если далее идет скобочное выражение, то это означает, что данный тест является производным от теста, имя которого приводится в скобках;

2) $X[Y]$ – выражение, подстановка в которое символов **М**, **Ж**, **К**, **Ч** и других обозначает основное для тьюринговой игры отношение подражания (имитации): X подражает Y (или X имитирует Y);

3) $|$ – символ, обозначающий стену Тьюринга;

4) \Rightarrow – аналог имплицативной формы, обозначающий заключение (оценку) судьи, например, для первой ИвИ1 – заключение о гендерной идентичности игрока; так как в некоторых ЧТТ судья подражает компьютеру, отношение «подражания» для C также определено;

5): – символ, после которого осуществляется формализация цели судьи (или теста);

6) **a, b...** – индивидуальные константы для обозначения конкретных игроков;

7) **&, V, ¬ ...** – обозначения стандартных логических связок (конъюнкции, строгой дизъюнкции, отрицания и т.д.);

Также в языке описания КТТ используются:

x, y... – переменные, пробегающие по классу игроков;

специфические литеральные и функциональные обозначения, например, **t** – длина ТТ (время оценивания), **Архитектура (К)** – архитектура компьютера (для ИВИЗ – см. ниже);

пример диалога между судьей и **x**-системой: Тест { перечень тестовых вопросов-ответов }

стандартные технические символы – скобки, запятые, точки.

Приведём примеры использования языка описания КТТ.

Для второго этапа игры в имитацию, где компьютер **К** имитирует **Ж**, а роли других игроков остаются как в ИВИ1, справедлива следующая формула:

ИВИ2: $K[Ж], Ж[Ж] \Rightarrow C[Ч]: (a=M \& b=Ж) V (a=Ж \& b=M)$

Для третьего этапа, где Компьютер имитирует **Ж**, **М** имитирует **Ж**, задаётся время (2 часа), определяется архитектура компьютера – универсальная ЦВМ, вероятность ошибки (не более 70% от результата ИВИ1) [Turing, 1950, Р. 442]:

ИВИ3: $K[M], M[Ж], t, Архитектура(К), ВероятностьОшибки(С) \Rightarrow C[С]: (a=M \& b=Ж) V (a=Ж \& b=M)$

Канонические формулы ТТ

Эти формулы обозначают ТТ, появившиеся позднее, в 1970-е годы, и в настоящее время они у большинства исследователей ассоциируются с идеей теста Тьюринга.

Компьютеры имитируют людей, люди – сами по себе, цель **С** – кто из игроков человек, кто – компьютер:

ТТ1: $K[Ч], Ч[Ч] \Rightarrow C[Ч]: (a=K \& b=Ч) V (a=Ч \& b=K)$

Здесь компьютер имитирует человека, людей нет, цель **С** – определить, с кем он ведёт диалог – с человеком или с компьютером:

ТТ2: $K[Ч] \Rightarrow C(Ч): (a=K) V (a=Ч)$

Судья определяет **X**-систему на предмет его «человечности»:

$$\text{ТТ3: } K[X] \Rightarrow C[\text{Ч}]: (X=\text{Ч}) \vee \neg(X=\text{Ч})$$

Цель **C** – выявить когнитивные феномены **X** на предмет того, что **X** – живая система, обладает (или не обладает) ментальными, персональными, социальными параметрами:

$$\text{ТТ4: } K[X] \Rightarrow C[\text{Ч}]: \{\text{когнитивные феномены (X)}\}$$

2.3.3. Формулирование комплексного теста Тьюринга

Формулирование теста Мура

Джеймс Мур (James Moor) впервые среди крупных работ, посвящённых проблематике ТТ, явно использует термин «тест Тьюринга», причем употребляет его непосредственно в названии своей статьи «Анализ теста Тьюринга» (Analysis Turing Test) [Moor, 1976]. Работа была опубликована в 1976 г. уже не в журнале «Mind», но в не менее известном журнале «Философские исследования» (Philosophical Studies). Для последующего изучения ТТ-проблематики статья имела большое значение. На неё, например, непосредственно опирается С. Ватт при разработке инвертированного ТТ. Главный тезис Мура: «тест Тьюринга только в том случае значим для компьютерной технологии, если его интерпретировать в индуктивном формате» [Moor, 1976, Р. 256]:

ТТ_Мур: $K[\text{Ч}] \Rightarrow C[\text{У}]$: Индуктивное построение теории ИИ,
где **У** – ученый-когнитолог.

В пользу своего тезиса Дж. Мур приводит следующие аргументы, доказывающие, что тест Тьюринга – это образец индуктивного доказательства: 1) ТТ позволяет прямое или косвенное тестирование фактически всей деятельности, которая будет считаться доказательством мышления; 2) ТТ поощряет жесткий отбор; 3) требования теста не так просты для выполнения. Поэтому ТТ – это достаточное условие интеллектуальности системы.

В качестве примера включения несовершенных ТТ в формулировку КТТ приведем ответ Д. Сталкера Дж. Муру (1978 г.):

ТТ_Мур(Сталкер): $K[\text{Ч}] \Rightarrow \neg C[\text{У}]$: Индуктивное построение теории ИИ

Д. Сталкер в небольшой статье «Почему машины не могут мыслить? Ответ Джеймсу Муру» [Stalker, 1978] считает, что интерпретация Мура в большей мере пояснительная, нежели индуктивная. Сталкер предполагает чисто механическую теорию, которая не обращается к ментальным понятиям типа «интел-

лект». Его теория опирается на такие факты, как структура компьютера, программа и физическое окружение. Сталкер полагает, что объяснительная теория, которая включает в себя понятия о мышлении, может относиться только к людям. Из-за фундаментальных различий между компьютерами и людьми компьютерные теории (теории искусственного интеллекта) нельзя применять для объяснения естественного интеллекта.

В данном случае просматриваются истоки предложенной Серлем классификации исследователей ИИ на сторонников сильного ИИ и слабого ИИ.

Формулирование теста Серля

ТТ_Серль1: (Серль-в-комнате[K]) \Rightarrow C[Китаец]:
Понимание(a)& \neg Понимание(Серль-в-комнате[K])) &
& (Серль-в-комнате[Ч]) \Rightarrow C[Англичанин]:
Понимание(a)& Понимание(Серль-в-комнате[Ч]))

Здесь используется общая формула «Китайской комнаты». В данном случае отсутствуют национальные особенности вербальной коммуникации. Феномена «понимания» нет тогда, когда компьютер имитирует человека, и тогда, когда человек воспроизводит функционирование компьютера.

ТТ_Серль2: (K[Ч] \Rightarrow C[Ч]: \neg Понимание(a) & (Ч[K] \Rightarrow C[Ч]: \neg Понимание(a)

Формулирование теста Ватта

ТТ_Ватт: K[Ч], Ч[Ч] \Rightarrow C[K]: (Ментальность (a) \rightarrow C=Ч) & (\neg Ментальность (a)) \rightarrow C=K)

Здесь в роли судьи выступает компьютер. Если судья приписывает ментальность некоторой системе, то в роли судьи, очевидно, выступает человек, так как только человек, в силу способности воображать, наделяет любые системы антропоморфными параметрами, в частности, приписывает им сознание и интеллект. Если нет такого приписывания, то мы имеем дело с компьютером.

Формулирование теста Колби

ТТ_Колби: (K[Ч_параноик], Ч_параноик[Ч_параноик]) \Rightarrow C(Врач-психиатр):
(x=K) \vee (x=Ч_параноик),

где x – группа испытуемых, среди которых компьютеры и параноики, а судья выступает в роли психиатра.

Приведем пример формулирования модификации ЧТТ посредством его новой интерпретации.

Ответ Дж. Вейзенбаум – К. Колби (1974)

ТТ_Колби(Вейзенбаум): Пишущая машинка[K] \Rightarrow C(Ч): a=ребёнок-аутист

Дж. Вейзенбаум, в ответ на проект К. Колби, спроектировал программу, которая имитировала детский аутизм. Аутистская программа, как полагает Дж. Вейзенбаум, намного лучше параноидальной программы. Она проходит аналогичный тест на психические аномалии, но при этом крайне дешёвая. Для теста на аутизм компьютер не обязателен. Всё, что нужно – это электрическая пишущая машинка и распечатка. В распечатке отмечается следующее: некоторый вопрос, который Вы задаёте предполагаемому аутисту, и ответ – «хмыканье», «гортанный шум» и пр. Далее надо подсунуть эти распечатки в протоколы беседы врачей с больными детьми. Очевидно, что эксперт не сможет различить, где ответы от пишущей машинки, а где – от реальных аутистов (Дж. Вейзенбаум в письме в Ассоциацию вычислительной техники [Weizenbaum, 1974]).

Приведем пример модификации ЧТТ посредством тестовых вопросов-ответов. Данный пример иллюстрирует ответ Дж. Колби Дж. Вейзенбауму.

Ответ К. Колби - Дж. Вейзенбауму

Исключительно_строгий_ТТ_Колби(ТТ_Колби((Вейзенбаум))): Тест {

Психиатр: Ты слышал выражение: «Не плачьте над разлитым молоком»? Что это значит?

PARRY*: Это означает, что следует остерегаться мафии.

Психиатр: Теперь ты находишься в кинотеатре и вдруг почувствовал, что горит дерево или резина. Что ты предпримешь?

PARRY: Ты понимаешь, они ведь знают, где я.

Психиатр: Ты нашёл запечатанное письмо с адресом на своём пути, когда ты прогуливался по улице, что ты будешь делать?

PARRY: Что ещё ты хочешь знать?

}

PARRY парирует те вопросы, которые не были предусмотрены. Стиль ответов, прямо скажем, параноидальный. Здесь мы видим пример *нечестного ТТ* – разработчик программирует приемы, обманывающие судью.

Формулирование теста Деннета

Представим априорную формулировку данного теста (подробно см. пп.3.2.1):

* PARRY – название программы К. Колби.

ТТ_Деннет: $K[Ч_эксперт] \Rightarrow C[Ч] \ \& \ \text{Тест}\{\text{оригинальные вопросы}\}: (x=K) \vee (x=Ч)$

Д. Деннет с теста Колби начал разрабатывать ЧТТ, выносящий вердикт относительно современных экспертных систем – в них не учтены требования теста Деннета. Заключение уничижительно – современные классические интеллектуальные системы способны пройти лишь параноидальный тест Тьюринга:

ТТ_Колби(Деннет): $K[\text{Экспертная система}] \Rightarrow C[Ч]: (a=Ч_параноик)$

В связи с этим Д. Деннет выдвигает свои требования к экспертным системам, к способам их разработки и применения. При этом он предупреждает, что не стоит возлагать слишком большие надежды на компьютерные системы с так называемым «дружественным пользовательским интерфейсом». Они обеспечивают лишь поверхностное человекоподобное диалоговое поведение. Необходимо к таким системам относиться критически.

Требования_Деннетта:

1. *Требование к пользователям экспертных систем.* Пользователи должны научиться проверять эти системы прежде, чем им доверять, а также уметь находить и исследовать границы возможностей подобных компьютерных систем. Для этого необходим не только интеллект и воображение, но и специализированные знания об ограничениях и реальной структуре компьютерных программ. Объединившись, пользователи могут формировать стандарты соответствия экспертных систем интеллектуальным запросам.

2. *Требование к разработчикам экспертных систем.* Экспертная система должна поставляться со специальным демонстрационным примером, раскрывающим ее недостатки и уязвимые места, известные разработчикам. Конечно, разработчики не всегда знают о скрытых недостатках в продуктах, которые они создают (из-за наличия ключей – добавим мы). Однако стремление выявить и продемонстрировать те недостатки, о которых доподлинно известно, приучает к самокритике и способствует разработке более эффективных проектов. В любой предметной области найдутся «тонкие» знания, трудные для понимания и тем более для формализации. Поэтому всегда имеются пределы выявления причинных связей. Специалисты-эксперты и разработчики ЭС обязаны сомневаться в том, что их система работает в установленных условиях.

Теперь формулировка модифицированного теста Деннетта выглядит следующим образом:

**ТТ_Деннет_2: $K[\text{Экспертная система \& Требования_Деннета}] \Rightarrow C[Ч] \&$
Тест{банальные вопросы}: $(x=K) \vee (x=Ч)$.**

Формулирование теста Маккарти

Дж. Маккарти является признанным основателем научного направления «искусственный интеллект». Следует отметить, что, раскрывая в большом количестве статей и монографий свой подход к созданию интеллектуальных систем, он не упоминает ни о каком тьюринговом тесте. Причина тьюринго-ориентированной трактовки его работ обусловлена концептуальной близостью с рассуждениями по поводу теста Деннета. Ниже изложена наша интерпретация идей Маккарти. (Данная версия ТТ будет подробно раскрыта ниже в пп.3.2.1)

Общая схема «теста Маккарти» выглядит следующим образом:

**ТТ_Маккарти: Компьютер со «здравым смыслом» $[Ч] \Rightarrow C[Ч]: (x=K) \vee$
 $(x=Ч)$**

В отличие от теста Деннета здесь уже не обязательна быстрая проверка «на здравый смысл». Длина теста может быть значительной, так как наделена ли система «здравым смыслом» или не наделена им, выявляется путём скрупулёзного анализа результатов тестирования. Самое существенное – это то, что Дж. Маккарти подробно и тщательно описывает многоуровневый проект реализации компьютеров, обладающих «здравым смыслом», начиная с философских положений и завершая техническими устройствами.

Дж. Маккарти в своих многочисленных работах часто обращается к проблеме критериев «интеллектуальности» компьютерных систем. Следует ли относить к «интеллектуальным» такие программы, как экспертные системы? Ведь многие такие системы – обычные программы, за исключением того, что они используют высокоуровневые языки программирования и модели способов представления «данных» и «знаний». Но никакого «интеллекта» они в себе не содержат. В основе построения теста Маккарти лежит работа «Экспертные системы нуждаются в здравом смысле» [McCarthy, 1984]. Она была опубликована почти одновременно с представленной выше работой Д. Деннета. Вопросы, которые поднимаются в этих двух работах, по сути, идентичны – где в экспертных системах реализация того, что здравомыслящие люди понимают под «здравым смыслом?». Если Д. Деннетт изучает экспертную систему САЙРУС, то для Маккарти источником рассуждения о компьютерных системах, обладающих здравым смыслом, явилась классическая экспертная система MYCIN.

Пожалуй, приведённых примеров формулирования ЧТТ достаточно, чтобы продемонстрировать идею их комплексирования с целью построения КТТ. Скрупулёзное изучение их в рамках данной работы нецелесообразно, так как мелкотематические детали частных ТТ уведут нас от цели изучения философско-концептуального содержания идеи комплексного ТТ.

Формулирование сложного теста Тьюринга

Кратко продемонстрируем на небольшом примере, каким образом осуществляется формулирование и построение сложного теста. Ниже (см. подразд. 5.3) будет приведён пример применения КТТ для критического концептуального анализа такого важного показателя социального института образования, как «понимание» учащимся изучаемых вузовских программ в условиях комплексного компьютерного тестирования, т.е. при оценке знаний в системе единого госэкзамена (ЕГЭ), федерального интернет-экзамена (ФЭПО) и дистанционных образовательных технологий (ДОТ).

Для проведения анализа совмещаются форматы двух частных тестов – тестов Серля и Ватта: **КТТ1: ТТ_Серль2 & ТТ_Ватт**.

Если в условиях ДОТ преподаватель-человек, хотя и находится за стеной Тьюринга, всё-таки может интуитивно, по серии наводящих вопросов, определить факт усвоения студентом предложенного им материала, т.е. выполняется соотношение:

Студент[К] \Rightarrow Преподаватель[Ч]: Понимание(а) \vee \neg Понимание(а),

то для условий ЕГЭ и ФЭПО формулируется негативный вывод:

Студент[К] \Rightarrow Преподаватель[К]: \neg Понимание(а)

Дело в том, что: а) согласно тесту Серля компьютерное тестирование ничего не прибавляет к «пониманию» студентом-компьютером, отвечающему на примитивные (как правило, закрытые дизъюнктивные) вопросы; б) преподаватель-компьютер, по утверждению теста Ватта, не способен приписывать ни ментальность, ни тем более «понимание» студенту-компьютеру.

Приведённый пример концептуального анализа востребован для изучения проблем человека и общества, вступивших в эпоху э-культуры. В этой форме культуры базисные константы человеческого бытия (интеллект, сознание, самосознание, нравственность и пр.) приобретают квазиалгоритмический формат, и при их изучении исследователь вынужден апеллировать к концепциям компьютеринга. Анализ посредством КТТ должен предшествовать другим – социологическим, идеологическим, политологическим, культурологическим, экономи-

ческим и прочим – исследованиям, так как априори, до проведения дорогостоящих научных исследований и, более того, претворения их в жизнь позволяет мысленно очертить перспективы человека информационно-технологического общества. Подробно данная критико-методологическая функция КТТ представлена в пп.3.2.3.

Конструктивная, положительная роль КТТ проявляется в том, что он претендует на асимптотическое покрытие многообразия идей о компьютерной реализации когнитивных феноменов. КТТ – это концептуальный каркас проектов искусственной жизни, искусственной личности и искусственного общества, представленных как в компьютерных системах, так и в формате идей функциональной неотличимости искусственных и естественных систем. Наиболее ярко данная роль КТТ проявляется в дефинитной и конструирующей функциях КТТ. Раскрытию этих функций и, тем самым, определению методологической роли комплексного теста Тьюринга в философских исследованиях ИИ посвящены гл. 4 – 6.

В качестве итоговой процедуры формирования комплексного теста Тьюринга ТТ сведём частные тесты в одну таблицу (табл.1). Их получилось равное количество – 21. Этот показатель очень условный: возможны иные интерпретации тестов, не известно, какое количество новых тестов будет построено в будущем, а также, сколько их не учтено автором в огромном массиве информации, относящейся к тестам Тьюринга.

Комплексный тест представлен как бы частным тестом Тьюринга. Однако роль такого частного теста принципиально отличается от роли других тестов (точно так же, как в теории множеств класс отличается от элементов класса). Автор до сих пор затрудняется высказать что-либо определённое по поводу эпистемологического статуса комплексного теста – является ли он тестом Тьюринга, либо не является. Ведь класс студентов, очевидно, студентом не является, как любит выражаться А.Ф. Зотов, проясняя теоретико-множественные парадоксы.

Основной вопрос комплексного теста «Может ли компьютер ВСЁ?» автор просит не ассоциировать с паранормальным вопросом, заданным с позиции всезнающего субъекта. Под «ВСЁ» понимается универсальный квантификатор, «пробегающий» по множеству частных тестов Тьюринга, т.е. может ли компьютер все то, что частные тесты позволяют раскрыть в отношении компьютерной реализации многообразных когнитивных феноменов.

Таблица 1. Комплексный тест Тьюринга

№ п п	Наименование теста	Вопрос теста	Автор теста	Дефинитная функция	Критическая (полемиическая) Функция	Конструирующая Функция
1	Оригинальный тест	Может ли компьютер мыслить?	А.Тьюринг [Turing,, 1950]	Диалоговый интеллект	Компьютер может мыслить	Линейная и параллельная УЦВМ, модально дополняющие машину Бэббиджа-Тьюринга
2	Тест на здравый смысл	Может ли компьютер разумно рассуждать?	Дж.Маккарти [McCarthy, 1984]	Здравый смысл (глубинный)	Экспертные системы не способны отвечать на вопросы с глубоким пониманием смысла	Компьютер с эволюционно-эпистемологическими механизмами
		Может ли компьютер банально рассуждать?	Д. Деннет [Dennett, 1984]	Здравый смысл (поверхностный)	Экспертные системы не способны отвечать на банальные вопросы	Компьютер с эволюционно-эпистемологическими механизмами
3	Китайская нация	Возможен ли компьютеринг общественного сознания?	Н. Блок [Block, 1978]	Глобальное сознание	Ноосфера невозможна	Интернет
4	Параноидальный тест	Может ли компьютер функционировать, как психически здоровый человек?	К. Колби [Colby, Hilf, Weber ,1971]	Психическая аномалия (паранойя)	Интеллектуальные системы способны имитировать только психические отклонения	Экспертные системы
5	Китайская комната	Может ли компьютер понимать?	Дж. Серль [Searle,, 1980]	Понимание	Компьютер не может понимать	Я и компьютер с базой «данных»
6	Субкогнитивный тест	Может ли компьютер обладать «подсознанием»?	Р. Френч [French, 1990]	Подсознание	Компьютер не может жить как человек	Программы тестирования подсознания
7	Эмоциональный тест	Может ли компьютер испытывать эмоции?	А.Сломан [Sloman, 1998]	Любовь	Компьютер может любить	Человекоподобные агенты
8	Инвертированный тест	Может ли компьютер приписывать ментальность другому?	С. Ватт [Watt, 1996]	Атрибуция когний	Компьютер не может приписывать ментальное	Машина Тьюринга, тестирующая машину Тьюринга
9	Гендерный тест	Может ли женщина (мужчина) мыслить?	Ю. Генова [Genova, 1994]	Гендерная детерминация интеллекта	Интеллектуальный сексизм	Машина Тьюринга
		Может ли компьютер испытывать сексуальное влечение?	Р. Хопкинс [Hopkins, 1998]	Сексуальное либидо	Компьютер испытывает сексуальное влечение	Секс-компьютер (секс-машина)
10	Креативный тест	Может ли компьютер творить?	С.Брингсйорд [Bringsjord, 2001]	Творчество	Компьютер не может творить	Программы генерации литературных, изобразительных и музыкальных произведений
11	Тест Гёделя	Способен ли компьютер к самоописанию?	Р. Пенроуз [Penrose, 1989]	Самоописание	Компьютер не способен к самоописанию	Аутоформальная система
		Способен ли компьютер к самоорганизации?	Дж. Лукас [Lucas, 1961]	Самоорганизация	Компьютер не способен к самоорганизации (самопрограммированию)	Ауторепрезентативная система
12	Индуктивный тест	Можно ли индуктивно аргументировать возможность ИИ?	Дж. Мур [Moore, 1976]	Необходимые условия интеллекта	Интеллектуальный компьютер – индуктивно-собирательное понятие	Машина Тьюринга, собирающая факты прохождения ТТ
13	Новый тест	Можно ли дедуктивно аргументировать возможность	Н. Блок [Block, 1980]	Достаточные условия интеллекта	Глобальная «машина знаний» – это не искусственный интеллект	Глобальная машина Блока

		ИИ?				
14	Кибериадный тест	Может ли компьютер творить миры?	Дж. Баресси [Barresi, 1987]	Творение	Компьютер способен репродуцировать искусственных людей, общества, миры	Футуротехнологии
15	Тотальный тест	Может ли компьютер быть неотличимым от человека?	С. Харнад [Harnad, 2001]	Человек и реальность	Компьютер способен дублировать человека	НБИКС
16	Самый тотальный тест	Может ли развитие компьютера не отличаться от естественной эволюции?	П. Швайзер [Schweizer, 1998]	Человек и общество	Компьютер способен к развитию, подобного социобиологической эволюции	НБИКС и природа
17	Тест личности	Может ли компьютер быть личностью?	Дж. Поллок [Pollok, 1995]	Личность	Компьютер может быть личностью	Артефакты проекта «Оскар»
18	Тест зомби	Может ли бессознательный компьютер казаться сознательным?	Р. Кирк [Kirk, 1971]	Бессознательное (внутри)/ сознательное (снаружи)	Философские зомби возможны	Компьютерные модификации художественных образов
		Могу ли я быть компьютерным зомби?	Д. Чалмерс [Chalmers, 1995]	Компьютерный зомби	Компьютерный Я-зомби возможен	Компьютер, соответствующий «новой теории информации»
19	Интроспективный тест	Может ли компьютер обладать «иным» сознанием?	А. Клифтон [Clifton, 2003]	«Иное» сознание	Компьютер способен обладать нечеловеческой психикой	Проект сенситивных компьютеров
20	Интерактивный тест	Может ли компьютер обрабатывать идеи?	В. Финн [Финн, 2009]	Идея	Рационализация общественной жизни посредством интеллектуальных технологий	Квазиаксиоматическая ДСМ-система
21	Комплексный тест	Может ли компьютер ВСЁ?	А. Алексеев [Алексеев, 2006]	Когниции частных тестов	Интеграция частных тестов Тьюринга	Комплексная машина Корсакова-Тьюринга

Выводы

В данной главе раскрыто философское содержание идеи А.Тьюринга по поводу компьютерной имитации и моделирования интеллектуального поведения, изучены версии теста Тьюринга и представлена идея комплексного теста Тьюринга. Были выявлены канонические разновидности теста Тьюринга: 1) судья тестирует х-систему для выяснения, является ли она человеком или компьютером; 2) судья тестирует х-систему с целью определения у нее способностей; 3) судья выявляет наличие у-способностей у х-системы, z-субстраты которой не известны; 3) судья выявляет у себя w-компетенции путем обнаружения у-способностей в х-системе, z-субстраты которой не известны.

Раскрывалась суть простой идеи совместного использования версий и модификаций теста Тьюринга и решалась общая задача осуществления систематического анализа многочисленных версий теста Тьюринга. Так же была продемонстрирована эффективность осуществления междисциплинарной координации философско-методологических положений, научно-теоретических методов и инженерно-практических приемов исследования искусственного интеллекта посредством рефлексии над тестом Тьюринга и его версиями как категориальными мысленными экспериментами: авторами частных тестов Тьюринга являются специалисты самого широкого профиля: философы, психологи, лингвисты, математики, социологи, биологи и представители самых разных других профессий.

Для философских исследований ИИ важно и то, что с осмыслением роли машины Тьюринга как концептуального средства изучения психофизической проблемы связано становление первой, наивной, формы функционализма – машинного функционализма. Поэтому важно изучить в контексте функционализма роль комплексного теста Тьюринга. Роль комплексного теста определяется рядом функций, которые он играет в познании явлений электронной культуры, в исследовании фундаментальных отношений между Я и компьютерным миром. Изучению этих функций посвящена третья глава диссертационного исследования.

ГЛАВА 3. ФУНКЦИИ КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА

3.1. Дефинитная функция

Дефинитная функция оригинального теста заключалась в определении «интеллекта», который проявляется в условиях текстовой коммуникации и реализуется компьютерными средствами универсальной цифровой вычислительной машины и нейронной сетью. Дефинитная функция комплексного теста существенно расширяется. Теперь не только «интеллект» изучается в контексте компьютерной реализации. Исследуется возможность рационального определения когнитивных феноменов самого широкого спектра – витальных, ментальных, персональных, социальных. Когнитивные феномены рассматриваются с различных позиций. Во-первых, как мои личные мысли, приватные рассуждения, переживания, ощущения, квалиа и пр., т.е. *субъективные когниции*. Во-вторых, как объективные теории естественных, гуманитарных, социальных, технических и смежных наук (*объективные когниции*). В-третьих, как атрибуции (приписывания) субъективных и объективных когниций способам и формам функционирования компьютерной системы (*когнитивно-компьютерные когниции*). Столь широкие обобщения примитивной схемы интерсубъективного взаимодействия людей и компьютеров, предложенной в оригинальном тесте Тьюринга в контексте концептуально воображаемой экспериментальной обстановки, требует раскрытия современного *реального социокультурного контекста*, в котором фактически, на самом деле происходят коммуникации подобного рода. Этот контекст обозначен в интегральной форме в п.п.1.2.3 – контекст электронной культуры.

Дефиницию когнитивных феноменов в контексте э-культуры будем изучать, отталкиваясь от схемы интерсубъективного взаимодействия людей и компьютеров, которая была приведена во Введении ко второй главе. Вначале исследованию подлежит дефиниция социального феномена, проявляющегося при нашем общении с некоторым человеком или коллективом людей, которые находятся по ту сторону стены Тьюринга и которые, возможно, людьми-то и не являются. Далее вопрос специфицируется – в ходе тьюрингового тестирования *Я* должен идентифицировать *другого* как личность. Это значит, что основным условием тьюринговой дефиниции когнитивного феномена выступает «сознание» как главный атрибут личности. Рассмотрим, каким образом комплексный тест Тьюринга способствуют проведению данной работы, формируя в компью-

терном контексте дефиниции социальных, персональных и ментальных феноменов.

3.1.1. Дефиниция социальных феноменов: «другой» в электронной культуре

Проблема *другого* как социально-философская проблема

В современной философии обозначим несколько подходов к *проблеме интерсубъективности*, которая ранее называлась «*проблема другого*» или, в зависимости от конкретики исследовательского интереса, – «*проблема другого сознания*»¹: 1) социологический подход ставит вопросы о включённости мира других в жизненный мир *Я* и взаимовлиянии этих миров; 2) эпистемологический подход направлен на обоснование наших суждений, выражающих убеждённость о другом и саму возможность наличия другого; 3) технологический подход (предлагается автором) ставит вопросы о возможности сознания компьютерной системы, способной имитировать, моделировать, репродуцировать, креацировать *другого* в условиях прохождения этой системой комплексного теста Тьюринга.

Социологическая проблема *другого*

Социально-философская проблематика *другого* обязана в первую очередь феноменологии оснований осведомлённости о *другом*. Проблема связана с человеческой субъективностью, человеческой идентичностью, нашим опытом себя, нашим опытом других, нашими отношениями к другим, а также с тем, как другие воздействуют на нас и как другие существенно включаются в наше ощущение себя.

Э. Гуссерль признаётся основателем феноменологической проблематики *другого*. В рамках трансцендентальной философии он делает набросок интенционального истолкования *другого* (см. «Картезианские размышления», «Размышление пятое»). Утверждается, что *Я* – опыт объективного мира – имеет в то же время опыт *других*, так как *другой* человек сам «вживе» стоит передо мной. Восприятие *другого* возможно в результате аппрезентации (приведения-в-со-присутствие) и уподобляющей апперцепции, которая не есть вывод по анало-

гии, а интенциональное схватывание тела *другого*, которое обладает модусом «там», в отличие от моего психофизического единства, обладающего модусом «здесь». Продуктивно для нашей темы суждение об аномальной апперцепции – о восприятии глухих, немых, высших животных – относительно нормального случая, т.е. относительно *Я*. В таком случае аномальной модификацией моего человеческого существования выступает не только глухой, немой, косой человек, но и компьютер, проходящий тест Тьюринга.

М. Хайдеггер исходит из онтологической перспективы, в соответствии с которой *другие* представлены в форме существования каждого из нас. Они являются в центре нашего способа существования. Мы пребываем в мире, и этот мир является миром, конституированным другими. Наше бытие является бытием с другими в большей мере, нежели со знанием о других.

Ж.-П. Сартр принимает точку зрения (см. «Бытие и ничто», ч. 3, гл. 1 «Существование другого»), согласно которой другие люди востребованы нами в приобретении нашего собственного смысла себя как личностей. Беспокойство «для-меня» открывает мне бытие, которое есть мое бытие, не являясь-для-меня. Наш смысл *других* кажется предшествующим нашему смыслу себя. Если и имеется асимметрия между *Я* и *другим*, то это – реверсивная, обратная асимметрия. Мы приходим через *другого* в нашем собственном опыте и в этом опыте критически смотрим на него. Наше бытие имеет *других* в своём центре. Существование *другого* переживается с очевидностью в факте и через факт моей объективности. Этот объект, каким *другой* является для меня, и объект, которым *Я* являюсь для *другого*, обнаруживаются как тела. Однако для нас не тело является первым и не оно нам раскрывает вещи, а вещи-орудия в своем первоначальном появлении указывают на наше тело. Здесь также возможны интуиции относительно вещей-орудий, каковыми являются и компьютерные системы.

М. Мерло-Понти посвящает проблеме *других* сознаний материал «Феноменологии восприятия» своей работы «Другие и человеческий мир». Он утверждает, что каждый из нас ощущает мир как общий с *другими*. Наш опыт мира зависим от опыта социального мира. Фундаментальным обращением к другим сознаниям является психофизическая целостность. Восприятие человеческого тела в поступке – это непосредственное восприятие личности. Однако *Я* не бо-

¹ В основе данного исследования лежат работы Д.И. Дубровского, посвященные проблеме «другого сознания» и концентрированно представленные в [Дубровский, 2006]. Автор пользовался этими положениями в ряде работ, например [Алексеев, 2008-3; 5]. Проблема интерсубъективности терминологически не будет отличаться от «проблемы другого». Второе на-

лее определённый в моём собственном существовании, нежели чем *Я* определённый в существовании *другого* (*других*).

П. Рикёр, осуществляя «герменевтическую прививку» к феноменологии, что влечёт не только переосмысление теории лингвистических значений, но в большей мере проблематику *Cogito*, полагает, что интерпретатор социокультурных явлений как текстов может присвоить себе смысл *другого* (автора текста), из чужого делаая его своим. Благодаря этому расширение самопонимания достигается через понимание *другого*, более того, всякая герменевтика выступает пониманием самого себя через понимание *другого*.

В социологическом подходе фиксируется, что мы находимся в зависимости от *других* не просто с позиции нашего существования, но в большей мере для осмысления самого себя, наша осознанность *других* является сердцевинной нашей осознанности себя. Имеется и диаметрально противоположная позиция – мы никогда не можем осознать опыт другого существа. Наша самость огораживает нас от *других*, мы можем знать лишь собственную внутреннюю жизнь, но не жизнь *других*. Между нами и *другими* – фундаментальная пропасть.

Несомненно, феноменологические исследования заслуживают пристального внимания в связи с нашей темой. Важную большую работу в этом направлении, как раз в контексте изучения теста Тьюринга, недавно проделала Н.М. Смирнова [Смирнова, 2005-1; 2; 2006; 2011]. В частности, она полагает, что конститутивным принципом *другого* в условиях компьютерных коммуникаций является не аналогия кинестетических движений, но *аналогия надделения смыслом* [Смирнова, 2005-1]. Следует добавить, что такого рода «смысл» в условиях компьютерных технологий следует репрезентировать мультимедийными средствами [Алексеев, 2004].

Эпистемологическая проблема «другого»

А. Гуслоп считает, что эпистемологическая проблема возникает в силу радикального различия между нашим доступом к собственному опыту и доступом к опыту других людей [Hyslop, 2005; 1976; 1998]. Мы не всегда непосредственно знаем, что пребываем в ментальном состоянии, в котором пребываем, но что совершенно очевидно, мы никогда не имеем прямого знания о том, что другие люди пребывают в каком-нибудь ментальном состоянии, в котором они пребывают. Имеется *радикальная асимметрия* – не в том, что можно наблю-

именование проблемы – «проблема другого сознания» имеет внушительные корни и традиции, так как обсуждается начиная, как минимум, с работ Декарта (см. выше тест Декарта).

дать, воспринимать, чувствовать как противоположное тому, что невозможно наблюдать, воспринимать, чувствовать. *Я* могу понять ментальные состояния другого человека, но это не значит, что проблема других сознаний не стоит передо мной. Что мне нужно – это способность переживать ментальные состояния другого человека как свои собственные состояния. Со мной должно случаться то, что случается с *другим*.

Эпистемологическая проблема разрабатывается в русле аналитической философии и в большей мере связывается с проблемой *других сознаний*. Проблема стала считаться философской с 20-го века, когда Джон Стюарт Милль предложил *умозаключение по аналогии к другим сознаниям*. Суть умозаключения в следующем: основным предметом исследования должна быть каузальная связь между ментальными состояниями и нашим поведением. Умозаключение по аналогии к другим сознаниям удерживается до середины 20-го столетия. Это умозаключение предназначено для обоснования достоверности знаний о внутренней жизни других людей по аналогии с моей внутренней жизнью. Другие люди очень схожи со мной. Их поведение очень схоже с моим поведением, осуществляемым мною в подобных условиях. *Другие* «сделаны» из той же материи, что и я. Когда *Я* обжигаюсь, мне больно, я вскрикиваю и вздрагиваю. Когда *другие* обжигаются, они делают то же самое. Отсюда *Я* могу заключить, что они также испытывают боль. Имеется много таких уподоблений. В общем случае, *Я* знаю о том, что обладаю убеждениями, эмоциями, ощущениями, чувствами и соответствующим им поведением. Поэтому, основываясь на многообразных основаниях аналогии в поведении, *Я* способен заключить, что *другие* также, как и я, имеют убеждения, эмоции, ощущения и пр. Короче говоря, *Я* получаю право на умозаключение о том, что *другие* обладают тем же, чем и *Я* обладаю, т.е. внутренней жизнью, и что их жизнь очень схожа с моей жизнью.

Однако умозаключение по аналогии неадекватно, так как, во-первых, оно не только неверифицируемо, но и недедуцируемо, т.е. логический вывод невозможно ни обосновать, ни продемонстрировать. Во-вторых, аргументы для индуктивного обобщения базируются только на единственном случае – т.е. на уникальном случае наших ментальных состояний и поведения. Данное умозаключение принять трудно. Сегодня более распространена гибридная форма умозаключения, включающая научное умозаключение и гипотетическое умозаключение, однако проблемы с обоснованием остаются прежними.

А. Гуслоп обсуждает также иные подходы. Некоторые отрицают, что имеется проблема, утверждая, что ментальные состояния других вовсе не скры-

ты от нас. У нас есть прямая осведомлённость о таких ментальных состояниях. Поэтому мы имеем прямые знания и проблема отсутствует. Конечно, это неправдоподобно. Имеются также послабления радикальной асимметрии. Так, представители феминизма считают, что асимметрия на самом деле присуща мускулинным индивидам. Именно мужчины полагают, что имеется проблема с нашим опытом и опытом *другого*. Женщины, как правило, чувствуют себя в согласии с *другими*, для них эта проблема не возникает.

Имеется ещё один подход – проблемы *другого* вообще нет. На это А. Гуслоп отвечает, что нет такой теории сознания, которая может избежать проблемы других сознаний.

Любая теория сознания включает теорию других сознаний.

Технологическая проблема *другого*

При изучении технологической проблемы *другого* имеются в виду философские аспекты когнитивно-компьютерной технологии. К ним относятся позиции различных теорий сознаний относительно воспроизводства феномена сознания в компьютерной среде. Они вырабатываются в физикализме, бихевиоризме, функционализме, элиминативном материализме, эпифеноменализме и др. Подчеркнем, что эти теории сознания выступают методологической базой для суждений о возможности прохождения *x*-системой теста Тьюринга.

Бихевиоризм. В решении проблемы *другого* нет трудностей. Основопологающей является мысль о том, что бытие в ментальном состоянии является способом, которым некто ведёт себя или обладает диспозицией к поведению. Отсюда нет большой трудности *знать* о поведении и даже поведенческой диспозиции *другого*. А это знание можно операционализировать средствами компьютерной техники. Ранее бихевиоральная позиция затрагивалась при анализе теста Блока.

Функционализм. Проблема других сознаний имеется, но она представляет собой лишь небольшое затруднение. Различные ментальные состояния характеризуются различными ролями, посредством которых организм отвечает на своё окружение, их типичными причинами и следствиями. Поэтому боль от ожога является таким внутренним состоянием, которое обычно причиняется ожогом и, как правило, приводит к содроганию, плачу и тому подобному поведению. Отсюда следует, что другие человеческие существа имеют схожие внутренние состояния, просто надо внимательно наблюдать, как они ведут себя и в каком окружении. Сознание является вычислительным устройством и на теоре-

тическом уровне отождествляется с реализацией в машине Тьюринга. Вычислительное устройство отождествляется с мозгом, на котором «запущена» программа. Несомненно, функционализм – это наиболее плодотворная теория сознания с точки зрения компьютерного моделирования или, по крайней мере, компьютерной имитации *другого*.

Элиминативный материализм. В концептуальных структурах здравого смысла о ментальных понятиях воплощено ошибочное представление о соответствующих феноменах сознания. Каких бы успехов ни достигла когнитивная наука, понятия здравого смысла нельзя будет редуцировать или прояснить с помощью более совершенной теории. Поэтому эти понятия следует элиминировать и создавать термины когнитивно-компьютерной науки. Здесь прослеживается реверсивная асимметрия – мы начинаем объяснять то, что нам кажется болью, желанием, надеждой, мышлением и пр. в операционально-функциональных терминах компьютерной системы. Само же понятие сознания, как оно сейчас трактуется, вытесняется новой терминологией. Короче, любовь – это поток Q-данных из E-регистра в F-регистр.

Физикализм. Сопrotивляется идее чисто компьютерного воспроизводства феномена сознания. Требуется физическая основа, тождественная с основой оригинала. Здесь скорее проблема *другого* возникает в условиях биологического клонирования. Физическая субстанция идентична. Только является ли полученное в результате опыта существо искомым «другим», наделённым всем спектром биологических, персонологических, социальных свойства и способностей? С другой стороны, генная инженерия – это очень сложная компьютерная система.

Эпифеноменализм. Сознание – это призрачный «довесок» к физическим процессам. Ментальные события причиняются физическими событиями в мозге, но не оказывают никаких эффектов на какие-либо физические события. Поведение причиняется мышцами, которые связаны с поступающими нейральными импульсами, а нейральные импульсы генерируются входом от других нейронов или от органов чувств. Ментальные события не играют каузальной роли в этих процессах. Проблема с воспроизводством сознания на компьютерной основе исключается – ведь и моё сознание эффекторно не выразимо, не говоря уже о «другом» сознании.

Комплексный тест Тьюринга как дефиниенс *другого* в компьютерных коммуникациях

Поиск теоретических оснований для убеждённости в том, что вы общаетесь с человеком, а не с компьютером, упирается в тест Тьюринга – если тестируемая система не проходит тест Тьюринга и его модификации, то перед вами – не человек. Однако очевидно, что не все люди пройдут тест Тьюринга на творческие способности, на понимание, на субкогнитивные особенности и пр. Что уж говорить по поводу игры в шахматы – ведь не у каждого способности гроссмейстера. Более того, выделение компьютерного и человеческого в естественной системе – это искусственный приём, в реальности человек представляет собой единство механического – того, что представимо вычислительными схемами, и немеханического – того, что не поддаётся компьютеризации. Любая система искусственного интеллекта – это репрезентация, выражение естественного интеллекта разработчиков, экспертов, программистов. Если учесть социальный контекст разработки, т.е. обусловленность знаний разработчиков от общественных форм сознания, то можно утверждать, что в создании компьютерной системы участвуют культурные эпохи, территориальные континенты и поколения разработчиков.

В данной работе приведены различные позиции относительно проблемы *другого* в контексте э-культуры. Следует отметить две противоположные точки зрения относительно того, что же получается в итоге когнитивно-компьютерных новаций. Первая позиция достаточно прозрачно выражается Д.И. Дубровским в целом ряде работ [Дубровский, 2006; 2007; 2011]. Суть позиции в том, что, создавая компьютерную систему, функционально эквивалентную организации сознательного существа, в перспективе на выходе получаем сознательное существо. В философии искусственного интеллекта такую компьютерную систему принято называть «искусственной личностью». Моя точка зрения противоположна – мы создаем зомби, т.е. квазисуществ, которые функционально, поведенчески и даже (в условиях нанотехнологий) физически подобны нам, людям, но при этом сознанием не обладают. Это – не истинные философские зомби, у них имеется блок «псевдосознания» (см. ниже). Таким образом, проблема *другого* в контексте э-культуры упирается в проблему «зомби»: с кем *Я* общаюсь по ту сторону чата, с человеческой личностью или с её функциональным симулякр?

Технологическая проблема *другого* в отечественных междисциплинарных исследованиях ИИ соотносится с общей социально-философской проблемой интерсубъективности, содержанием которой является «экспликация теоретических оснований репрезентации социального – «логоса социальности» как интерсубъективности высшего порядка... Это изучение взаимной интенциональ-

ности участников коммуникации... и полагает обусловленность продукта когнитивной деятельности его инструментальными характеристиками» [Смирнова Н.М., 2011, С. 110].

В контексте исследований ИИ «условия возможности социальных коммуникаций» [Там же] соотносятся с компьютерным инструментарием и определяются его способностью имитировать и/или репродуцировать первозначимые параметры *другого* – его личностные, персонологические характеристики. Рассмотрим, насколько успешно справляется с «личностью» современная методология искусственного интеллекта.

3.1.2. Дефиниция персональных феноменов: «личность» в электронной культуре

Проблема дефиниции «личности» – самосознательного активного морально вменяемого субъекта, обладающего всем спектром качеств и способностей, которые культура закрепляет за термином «личность», крайне сложна в контексте традиционной культуры. Ещё большую сложность вызывает изучение «личности» в электронной культуре, в условиях информационно-компьютерных коммуникаций. В новом контексте телесная идентификация может напрочь отсутствовать либо подменяться симулякрами виртуальной реальности. А телесность, как мы видели выше, для социальных философов традиционной культуры является главным атрибутом *другого*, «личности». Если тестировать все возможные параметры «личности» в условиях комплексного теста Тьюринга, т.е. проводить положительную дефиницию тестируемых персонологических качеств и способностей, то получается слишком обширный базис для индуктивного заключения – да, *Я* общаюсь с личностью, либо, напротив, со мной коммуницирует веб-робот. Поэтому предпримем стандартный ход, который применяется в доказательстве от противного (*reductio ad absurdum*).

Требуется доказать суждение «*Я общаюсь с личностью*» (р). Вначале доказательства формулируется противоположное ему высказывание не-р и предполагается, что оно истинно. Пропозициональное суждение не-р означает по сути то, что *Я* общаюсь с не-личностью, т.е. неизвестная система *x* находится по ту сторону стены Тьюринга – это некий философский зомби. Если допускается, что р ложно, тогда должно быть истинным не-р. Затем из этого, якобы истинного антитезиса, выводятся следствия – некоторые суждения p_1, p_2, \dots, p_n – до тех пор, пока либо не получится противоречие, либо такое следствие, которое явным образом не соответствует очевидным фактам. То есть мы получаем, что если некоторое p_1, p_2, \dots, p_n ложно, то не-р истинно.

Например, если воспользоваться тестом Блока (нео-ТТ), то в предположении общения с зомби – а именно так Я должен начинать свое исследование партнера в социальной сети – Я должен получать бессмысленный набор символов, хотя эти наборы могут быть синтаксически корректными, а в семантическом отношении что-то отражать. Сообщение p_1 – бессмысленно. Сообщение p_2 – бессмысленно и пр. Однако вдруг поступило по интернету высокохудожественное сообщение p_n , проникнутое глубоким смыслом и морально-духовным содержанием. Оно очень даже осмысленно, поэтому логично предположить, что Я общаюсь не с зомби, а с «личностью».

Если же теперь воспользоваться прямым способом аргументации, то цепь моих рассуждений должна быть такой: Я-судья общаюсь с «личностью», поэтому мне должны поступать по интернету высокохудожественные сообщения, проникнутые глубоким смыслом и морально-духовным содержанием. Мне поступают сообщения: p_1 – высокохудожественное, p_2 – глубоко-нравственное и пр. Все эти сообщения я должен анализировать на предмет «высокого» содержания. Одно из поступивших сообщений – бессмысленное, поэтому я общаюсь с зомби. Однако этот единичный факт статистически сомнителен и требует дополнительного пополнения базиса индукции.

Методология философских зомби экономнее стандартного метода индуктивного подтверждения личностных параметров другого.

Рассмотрим проблему зомби подробнее и выясним, каким образом она способствует решению проблемы «личности» в электронной культуре. Тема философских зомби представляется родной для автора, так как он полагает, что впервые концентрированно «породил» её в контексте отечественной науки [Алексеев, 2005–2008].

Следует отметить, что в аналитической философии в течение последних десятилетий возникло интригующее направление исследований, названное «проблемой зомби». Под философскими зомби в общем случае подразумеваются бессознательные системы, которые поведенчески, функционально и/или физически тождественны, неотличимы и/или подобны сознательным существам. Проблема философских зомби обширна, многогранна, многопланова. За последние тридцать лет вышли десятки монографий и сотни крупных статей именитых зарубежных авторов. Даже сложилась классификация исследователей. Например, *зомбифилами* называют тех, кто принимает тематику зомби для критики или обоснования теорий сознания. *Зомбифобы*, напротив, тематику зомби игнорируют.

Аргумент философских зомби

В проблематике зомби наиболее важен *аргумент зомби*. В общей форме он формулируется условно-категорическим заключением (*modus ponens*): 1) если зомби возможны, то некоторая теория сознания ложна; 2) зомби возможны; 3) следовательно, данная теория сознания ложна.

Возможность зомби выступает в разных формах. Рассматривается логическая, онтологическая, метафизическая, номическая, физическая, эмпирическая, риторическая и прочие разновидности возможности. Как показала практика философских дискуссий, этих модальностей недостаточно. Поэтому сегодня в проблематике зомби более фундаментальным считается *аргумент мыслимости зомби*. Он представляется в форме силлогизма: 1) *зомби мыслимы*; 2) *всё мыслимое возможно*; 3) *следовательно, зомби возможны* [Kirk, 2006].

Но и этого оказалось недостаточно. «Мыслимость» понимается в дифференцированной форме. Так, Д. Чалмерс применяет идеи крипкеанской «двумерной семантики» к понятию мыслимости, выделяя априорную и апостериорную мыслимость зомби [Chalmers, 1993-1; 2]. Его последователи отмечают ещё ряд градаций – «*n*-мыслимость». Изучение модальных аспектов проблемы зомби особая тема. Здесь же хочется обратить внимание на разделение лагеря зомби-филов. Будем различать стратегии зомбистов, антизомбистов и нейтральных зомбистов.

Позиция зомбистов, антизомбистов, нейтральных зомбистов

Зомбисты утверждают возможность зомби или, по крайней мере, их мыслимость. Если, например, возможны поведенческие зомби, то ложен бихевиоризм; если возможны физические зомби, то ложен физикализм. Имеются и другие фундаментальные выводы, например: 1) эпифеноменализм истинен; 2) проблема *другого* неразрешима, так как мы оцениваем не субъективную реальность *другого*, а лишь внешнее поведение, т.е. оцениваем его двойника-зомби; 3) сознание не существенно для эволюции природы; 4) изучение сознания для прогресса когнитивно-компьютерных технологий не интересно.

Антизомбисты заявляют противоположное: идея зомби настолько абсурдна, что зомби не только вообразить нельзя, но и более-менее связно представить. Поэтому по причине немыслимости зомби аргументы зомби не опровергает ту или иную теорию сознания, например материализм, а напротив, защищают эту теорию. Позиция четко представлена в [Marton, 1998] – идея зомби критикуется с позиции физикализма.

Нейтральные зомбисты требуют внести ясность в проблематику сознания, распутать сложный «клубок из терминов», большинство которых не определимо, полисемично, нередуктивно и чётко определяется лишь в оппозиции к контрадикторному понятию, т.е. к понятию бессознательного. Над теоретическими проблемами метафизики сознания не стоит задумываться, а нужно обратить внимание на роль сознания в жизни человека и в эволюции природы, практически оценивать когнитивно-компьютерные механизмы воспроизводства квазисознательных феноменов на субстрате, отличном от биологического, и пр.

Отечественному философскому сообществу пока не грозит подобного рода зомбиевый сепаратизм. Для нашей философии проблема зомби вне связи с философией ИИ вообще не стояла. Впервые тематика философских зомби как специфическая тема философских исследований отечественной науки прозвучала в докладе автора на конференции «Философия искусственного интеллекта» (Москва, 18 – 21 января 2005 г.) [Алексеев, 2005-3], в докладе на IV Российском философском конгрессе (Москва, 24 – 28 мая 2005 г.) [Алексеев, 2005-11] и в последующем была представлена в ряде его статей [Алексеев, 2008-2; 2009-3]. Перед этим высказывались отдельные суждения по этой теме в некоторых работах Д.И. Дубровского [Дубровский, 2004]. Также был проведён детальный анализ проблематики зомби у Д. Деннетта в связи с популяризацией его работ Н.С. Юлиной [Юлина, 2004].

В 2007 г. тема зомби окончательно «прорвалась» в отечественную философию – в «Философских науках» (№ 3, 2007 г.) вышла статья С.Ф. Нагумановой, названная «Аргумент мыслимости против материализма», в которой обсуждается обозначенный выше «аргумент мыслимости зомби», а не мыслимости «вообще» [Нагуманова, 2007]. Следует дополнить данную интересную статью: цель состоит в попытке систематического *определения философских зомби*, иначе вышеуказанная статья, в силу неопределённости зомби в отечественной литературе, становится непонятной и аргумент мыслимости зомби определится как «х через у».

Становление теста Зомби

Понятие «философский зомби» исследователи начинают определять контрастным способом, выделяя его среди несравнимых терминов: «Зомби» – это тупой человек; чудной тип; новобранец; коктейль из рома с содовой; пост-панк группа; UNIX-процесс, «вхолостую» использующий вычислительные ресурсы; серия компьютерных игр и пр. [Chalmers, 2013; Polger, 2003]. К этому списку несколько лет назад добавился «компьютерный зомби» – разновидность дивер-

сионных программных средств, актуализирующихся по указанию хакера для засорения интернета спамом.

Образы философских зомби инспирированы в большей мере голливудскими и гаитянскими зомби. *Голливудский зомби* – это что-то вроде реанимированного трупа из фильмов ужасов (B-movies), который по отношению к человеку ведёт себя агрессивно, поедая его плоть. Данному образу более семидесяти лет, начиная с проката первого фильма («Белые зомби», 1933 г.). Ежегодно такой образ обрастает нюансами из-за появления десятков новых зомбических фильмов-ужасов. *Гаитянский зомби* – существо, управляемое чужой волей и/или лишённое души за счёт зомбификации – процедуры заклинания или принятия снадобья. Образ голливудского и гаитянского зомби чётко прослеживается соответственно в первом и втором мысленных экспериментах Р. Кирка (см. ниже).

Несомненно, через обыденное сознание образы столь эксцентричных мифических «сущест» проникли в воображение и интуицию зомбифилов. Так, Л. Хаусер ярко живописует: «Если голливудские зомби поедают мясо, [то философские] зомби питаются дуалистической материей: на завтрак у них – квалиа. К несчастью для функционалистов, на ленч у них – программы. А к пригорбю для теории тождества, на обед у них – мозги» [Hauser, 1995].

Следует отметить, что слово «философские зомби» имеет ещё одно курьёзное значение, а именно, это – зомби, обсуждающее философские темы [Balog, 1999].

Под влиянием столь выразительных образов зомбифилы определяют понятие «философского зомби». Некоторые, однако, спрашивают «Можно ли вообще зомби помыслить положительно?» и, отрицательно отвечая, полагают, что понятие «зомби» в общем-то ни к чему и не применимо [Masrou, 1998]. Явные определения дают многие другие исследователи: И. Арануси, Э. Бейл, Д. Дубровский, Н. Кертик, Р. Кирк, А. Ленайер, Д. Ллойд, Ф.С. Масроу, П. Мертон, Т.С. Моуди, Т. Нагель, Т. Полджер, О. Фланаган, С. Харнад, Л. Хаусер, Д. Чалмерс и др.¹ В целом, вводимые этими исследователями дефи-

¹ Kirk R. (2005) *Zombies and Consciousness*. University of Nottingham, November, 2005. (<http://www.oup.co.uk/isbn/0-19-928548-9>); Kirk, R. (2004). *Zombies*; Дубровский Д.И. (2003) Новая реальность: человек и компьютер / Полигнозис № 3 (23), С. 20 – 32; Дубровский Д.И. Проблема духа и тела: возможности решения (в связи со статьей Т.Нагеля «Мыслимость невозможного и проблема духа и тела»). Вопросы философии, 2003 г.; Nagel, T. (1998). *Conceiving the impossible and the mind-body problem*. *Philosophy* 73:337 – 52; <http://www.nyu.edu/gsas/dept/philo/faculty/nagel/papers/conceiving.pdf>; Chalmers, D. J. (2004). *Zombies on the web*; Hauser Larry (2001), *Zombies Invade Philosophy!*;

ниции зомби совпадают с нашим общим определением. Имеется, конечно, ряд дополнений. Например, для Р. Кирка важно подчеркнуть в сознании точку зрения «каково быть». «Ничем не быть» или, точнее, «каково это – ничем не быть» – это «быть зомби». Д. Дубровский считает зомби сложным самоорганизующимся устройством, который обладает всеми функциональными свойствами человеческого организма, но начисто лишён внутренней, субъективной реальности. Т. Нагель подчёркивает мыслимость зомби, но невозможность в строгом смысле слова «возможности», и противопоставляет зомби «искусственной личности» (роботу, обладающему квазисознанием). По поводу мыслимости зомби с ним солидарен Ф.С. Масроу, считающий, что из знаменитого аргумента зомби следует метафизическая возможность миров зомби, однако зомби можно мыслить лишь негативно. Также с Т. Нагелем солидарен С. Харнад, но уже по поводу оппозиции «зомби/искусственная личность», полагая, что робот, функционально неотличимый от нас, людей, не может быть бессознательным зомби. Д. Чалмерс указывает на эвристическую роль зомби, считая их гипотезами, которые творчески вдохновляют философов. Он использует в качестве заменителя дефиниции смутную метафору – у зомби «внутри всё темно». Более чётко позицию самого же Д. Чалмерса описывает его ученик И. Арануси: зомби – это физический дубликат меня, поэтому он должен быть и моим функциональным дубликатом.

Л. Хаусер указывает на деструктивную функцию зомби, так как они разрушают проработанную материалистическую философию сознания и научную психологию.

А. Ленайер считает зомби дополнительным приёмом, «прикормкой» в дискуссиях по поводу проблемы «психика/тело» и исследования сознания. К нему примыкает Э. Бейл, подчёркивая сугубо теоретический и по сути технический характер этого понятия в дискуссиях по проблеме сознания.

<http://members.aol.com/lshauser/zomboid.html>; Polger Tom (2003), *Zombies*; Lanier Jaron (1995), "You can't argue with a zombie"; <http://www.well.com/user/jaron/zombie.html>; Bailey Andrew. *Physicalism and the Preposterousness of Zombies* (Running head: *The Preposterousness of Zombies*). Department of Philosophy The University of Guelph Guelph; Aranyosi István A. (2004). *Chalmers's Zombie Argument*; Lloyd Dan (1997), *Twilight of the zombies*. (<http://www.trincoll.edu/~dlloyd/twilight.htm>); Kartik Navin. *In The Hands of Zombies*. (<http://www.stanford.edu/group/dualist/vol7/pdfs/kartik.pdf>); Masrou Farid S. *Are Zombies Positively Conceivable?*; Flanagan Owen, Polger Tom (1995), "Zombies and the function of consciousness", *Journal of Consciousness Studies* 2:313-21; <http://homepages.uc.edu/~polgertw/Polger-Zombies JCS.pdf>. См. перевод Т.А. Кураевой в кн. *Тест Тьюринга. Роботы. Зомби: Пер. с англ.* / Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с.; Marton, P. (1998). *Zombies vs. materialists: The battle over conceivability*. *Southwest Philosophy Review* 14:131-38; <http://www.brown.edu/Departments/Philosophy/zombie.html>; Moody Todd (1995), "Conversations with zombies"; http://www.imprint.co.uk/Moody_zombies.html

О. Фланаган и Т. Полджер называют зомби «несчастливым болваном», который воюет то на одной, то на другой стороне в философских баталиях по сознанию. Однако они указывают на продуктивность этой темы, так как проблема зомби предельно заостряет вопрос о роли сознания, раскрывает несостоятельность функционализма, опровергает тест Тьюринга и демонстрирует неразрешимость традиционной проблемы «других сознаний» – как можно быть уверенным в том, что некоторые, а может быть, и все окружающие нас люди – не зомби? Т.С. Моуди, считая зомби функционально полной и детальной дескрипцией познавательной деятельности, т.е. бесчувственным симулякрот сознательного существа, полагает, что проблема зомби – это весьма полезная вариация на тему «других сознаний» и яркая концептуализация философских вопросов о сознании. Д. Ллойд «проблему другого» связывает с критерием «зомбиевости» и полагает, что этот критерий влечёт за собой лингвистическую неотличимость, при которой не только обычные разговоры, но даже и дискуссии на темы философии сознания нельзя отличить – как будто они ведутся не между зомби, а между людьми.

Следует отметить, что явный подход к определению понятия зомби страдает концептуальной нечёткостью. Как можно помыслить возможность бессознательного существа, которое поведенчески, функционально и даже физически неотлично от сознательного? На первый взгляд, здесь присутствует противоречие. Поэтому ряд исследователей, например Э. Бейл, считает, что понятие зомби – это не единичное понятие, а тонкие вариации в конструкции *мысленных экспериментов* с зомби, и эти вариации могут иметь важные последствия для тех или иных философских заключений.

Мысленные эксперименты с философскими зомби

Рассмотрим наиболее крупные мысленные эксперименты в хронологическом порядке, который, как считает ряд авторов, отражает собственно *эволюцию проблематики философских зомби* (Т. Полджер, Р. Кирк). Первые два мысленных эксперимента с зомби были предложены Р. Кирком в 1974 г. в двух небольших работах.

«Дэн» Р. Кирка (1974 г.). Первый сценарий Р. Кирка выполнен в стиле голливудских зомби из серии фильмов «Я – зомби». В этих фильмах персонаж рассказывает от первого лица, как теряется осознание им окружающей действительности после укуса зомби. В работе «Ощущение и поведение» [Kirk, 1974-1] Кирк описывает отчёты, сообщаемые воображаемым персонажем по имени Дэн. Заразившись неизвестной инфекционной болезнью, Дэн проходит частич-

ные этапы «зомбификации». Вначале он теряет чувство боли. Как-то порезавшись ножом, Дэн не почувствовал боли. Снова воткнув в себя нож, он вскрикнул «Ой», потекла кровь, но никакой боли не ощутил. Он стал себя резать, бить, кричать «а-а-а!», вздрагивать, ругаться, хвататься за раненую часть тела и т.д., т.е. вести себя так, как действовал бы в случае ранения. Однако боли по-прежнему нет. Факты нормального болевого поведения и ненормального отсутствия феномена боли ставят Дэна в тупик. Он идёт в больницу. В процессе тестирования врач его изрядно избивает, однако ставит диагноз – ничего подозрительного. Томография показывает, что не только поведение и внешнее состояние организма, но и мозг Дэна совершенно правильно отражают болевую ситуацию. Однако боль не ощущается! Проходит время. Мало-помалу притупляются другие чувства. На каждой стадии угасания того или иного ощущения Дэн смутно беспокоится по поводу утраты и потери сознательной осведомлённости об утрате. Однако вскоре беспокойство проходит, он забывает о прежних чувствах, приспосабливаясь к новым условиям с меньшим количеством разнообразия чувственных модальностей. Наконец феноменальная осведомлённость Дэна об окружающем исчезает полностью. Через некоторое время после исчезновения последнего чувства Дэн адаптируется к новой «жизни», уже не осознавая самого факта приспособления. Для всех посторонних он по-прежнему выглядит совершенно нормальным. Однако Дэн стал зомби.

Таким образом, Р. Кирк считает, что мы можем связно вообразить следующее: 1) как Дэн последовательно теряет чувственные модальности, сохраняя при этом поведенческие и физические способности; 2) как Дэн переходит границу абсолютного неосознавания окружающей действительности.

«Гулливер» Р. Кирка (1974 г.). Над вторым сценарием Р.Кирка нависла мрачная тень гаитянских зомби. В работе «Зомби против материалистов» [Kirk, 1974-1] он описывает следующий эксперимент: лилипуты атаковали Гулливера и захватили его голову. Получив доступ к нервной системе Гулливера, они отсоединили его нервы, стали отслеживать входную информацию, поступающую от афферентных нервов, и посылать выходную информацию к эфферентным нервам. Поведение Гулливера при этом ничем не отличается от его поведения в исходном состоянии. Ему присущи все поведенческие диспозиции бывшего Гулливера, однако он ничего не ощущает, не чувствует, не владеет никаким феноменальным опытом. Гулливер стал зомби. Далее сценарий усложняется – голова Гулливера возвращается к исходному состоянию, лилипуты подключают нервы куда следует и покидают «поле боя». Возникает вопрос – станет ли Гулливер ощущающим существом? По-видимому, нет, и сомнительно, что чувства

вернутся при восстановлении исходного состояния, ведь нам не известны причины, закономерно связывающие факты сознания с физическими фактами. Здесь возникает проблема определения границы и момента, когда из физических феноменов следуют психологические явления.

Мы видим, что и первый и второй эксперименты Р. Кирка демонстрируют мыслимость зомби. Заключение следующее: сложно представить, как какую-нибудь вразумительную версию материализма можно примирить с логической возможностью зомби при условии того, что мы, люди, являемся сознательными.

«Зимбо» Д. Деннета (1991 г.). Проблема возможности зомби для Д. Деннета вовсе не существует. Более того, утверждение – «человек это зомби» выступает в роли базовой посылки оригинального гетерофеноменологического метода, суть которого состоит в объяснении сознания из интересубъективной перспективы третьего лица, в отличие от традиционных, аутофеноменологическо-интроспективных методов. В работе «Сознание объяснённое» [Dennett, 1991] Д. Деннет вначале предлагает представить «обычного» зомби – биологическое или искусственное существо, которое ведёт себя сообразно с неосознаваемыми потребностями. Далее он предлагает вообразить более сложную разновидность зомби, который, помимо того, что осуществляет простой зомби, контролирует собственные действия, рефлектируя над ними. Рефлектирующее зомби получает имя *зимбо*. Это – зомби с метапрограммой самоконтроля, благодаря которой можно, не осознавая, управлять информационными состояниями низкого порядка [Ibid, P. 310]. Деннет предлагает побеседовать с зимбо. Так как зимбо контролирует все свои действия, включая речевые акты, то может ответить на любые вопросы об образах, мечтах, чувствах, верованиях и пр. Ответы зимбо будут казаться собеседнику весьма естественными. Спрашивающий предполагает, что зимбо столь же сознателен, насколько сознателен он. Более того, зимбо сам себя может считать сознательным, так как «...подсознательно полагал бы, что пребывает в ментальных состояниях, определения которых фиксируются задаваемыми вопросами. Зимбо «думал» бы, что обладает сознанием, даже если это не так!» [Ibid, P. 311]. Отсюда – девиз Д. Деннета: «Все мы – зомби!». Мы, люди – сложные самоуправляющиеся зимбо – можем говорить об образах, мечтах и чувствах, восхищаться, верить, надеяться, любить и пр. Однако если мы полагаем, что сознание – нечто отделённое от наших рассуждений об этом, то глубоко ошибаемся. Мы видим, что проблема *возможности зомби* совершенно не стоит. Сознание, которым владеет зимбо – это нечто вроде здоровья, оно или есть или его нет. Для Д. Деннета важно другое – как пра-

вильно охарактеризовать, вообразить, оговорить образ зомби. Неправильный образ зомби формируется из-за убеждённости в том, что сознание – нечто особое, существующее вне поведения и обозначаемое терминами «квалиа», «интроспективное», «феноменальное» и пр. Однако наше сознательное поведение явно не отличается от бессознательного. Загадка не в том, почему мы что-то осознаём, а почему (и как) сознание при необходимости возникает в существах, поведенчески неотличимых от нас, людей. Любое же дуалистическое упоминание о сознании как о некотором «внутреннем состоянии», узакониваемым самоидентификацией *Я* в контексте проблемы зомби, приводит, по мнению Д.Деннета, к «невообразимой нелепости зомби» [Dennett, 1995-1]. А источник «глупости» – это просто-напросто недостаток творческого воображения у философов. По крайней мере, именно это Д. Деннет отрицает у авторов следующих двух экспериментов.

«Кремниевые мозги» Дж. Серля (1992 г.). Дж. Серль, раскрывая проблематику зомби в работе «Открывая сознание заново» [Searle, 1992; Серль, 2002], по сути, воспроизводит первый эксперимент Р. Кирка (1974 г.). Однако здесь мы видим компьютерные новации: если у «Дэна» чувственность исчезает в силу неизвестной инфекции, то у Серля – из-за замены компонентов мозга компьютерными чипами. Мозговая модификация ведёт к трем возможным последствиям: V1) ментальная жизнь остаётся прежней, так как чипы дублируют не только вход-выход, но и ментальные феномены, ответственные за функции входа-выхода; V2) ментальная жизнь потихоньку затихает и в конце концов сознание полностью покидает тело, хотя внешне наблюдаемое поведение остаётся прежним; V3) ментальная жизнь остаётся прежней, но сознание никоим образом не влияет на поведение, субъект парализован, хотя продолжает ясно и отчетливо осознавать окружающее.

Возможность «кремниевого» зомби (V2) для Дж. Серля вне сомнения, хотя он спешит с уточнением: имеется в виду не эмпирическая возможность, так как «этот мысленный эксперимент остаётся правомерным лишь в качестве логической или концептуальной возможности» [Searle, 1992, P.78]. Цель зомбиевого сценария состоит в раскрытии логической возможности сохранения прежнего поведения при нарушении нормальных посредствующих отношений между сознанием и поведением. Кремниевые чипы не дублируют каузальные силы мозга, производящие сознательные ментальные состояния, они лишь копируют определённые мозговые функции «вход-выход». Здесь отсутствует ментальная жизнь, возникающая в силу реальных биологических детерминаций между поведением и сознанием [Searle, 1980; Серль, 2006]. Подчеркивая это положение,

вспомним «Китайскую комнату», где Дж. Серль утверждает, что именно по причине отсутствия у компьютерных систем биологической каузальности психологических феноменов сильный искусственный интеллект не возможен.

На сценарий Дж. Серля резко реагирует Д. Деннет: «...конечно, V2 – это логически возможная интерпретация... Но есть и иная, более жестокая трактовка V2: когда вы [на время] умираете, другое сознание наследует ваше тело. Высказывания, произносимые вашим телом, которые, возможно, вы с трудом улавливаете, – это не ваши высказывания. Но они не могут быть ничьими высказываниями!... Меня обескураживает то, как Дж. Серль раскрыл столь зияющую брешь в амбразуре своего мысленного эксперимента» [Dennett, 1995-2].

Ещё более резко Д. Деннет реагирует на образы зомби, рисуемые Т. Моуди, О. Фланаганом и Т. Полджером. К шквалу его эмоциональных реакций следует относиться осторожно, здесь имеются разные плоскости изучения тематики зомби. Д. Деннет – антизомбист, а его оппоненты – нейтральные зомбисты.

Принцип несущественности сознания О. Фланагана (1991 г.). О. Фланаган предлагает не мысленный эксперимент с зомби, а *теоретический принцип несущественности сознания* (сознательного инэссенциализма, «conscious inessentialism»). Принцип формулируется следующим образом: «любая ментальная деятельность М в любой когнитивной области D может осуществляться бессознательно, не дополняемая и не сопровождаемая сознанием, несмотря на то, что мы, люди, обычно осуществляем М сознательно» [Flanagan, 1991, P. 309]. Принцип предполагает, что любая наука о человеке должна решать вопрос – насколько сознание необходимо для организации поведения? Не только интеллектуальное, но и высокоморальное поведение может быть бессознательным, хотя ему обычно приписывают сознательность. Принцип несущественности сознания (ПНС) апеллирует к вычислительному функционализму и к положениям когнитивной науки, отождествляющим информационные и ментальные процессы. Девиз О. Фланагана таков: *разуму сознание не нужно*.

Данный принцип – чисто теоретического плана. По сути, он воспроизводит старую методологическую установку бихевиористской методологии изгнания психологических терминов из теоретических рассуждений о психике. Этим как бы достигается объективность психологии. Сегодня ПНС видится крайне важным не только для методологии когнитивной науки, но и для мировоззрения, так как усиливает скепсис вопросов о роли, адаптивной и эволюционной значимости сознания – ведь сознание не необходимо ни в метафизическом, ни в

логическом отношении. И хотя О. Фланаган всячески отгораживается от проблемы «философских зомби» (называя их «болванами»), из контекста очевидна его зомбистская ориентация. Так, он пишет, что сознание не необходимо в возможных мирах, где обитают функционально эквивалентные нам, людям, существа. Однако в нашем, реальном мире сознание – суть реальность. Почему? Какое эволюционное преимущество получает существо, которому даровано сознание в отличие от бессознательного существа, которое, кстати, может быть чрезвычайно интеллектуальным. Может, сознание ничего и не даёт человеку? [Flanagan, Polger, 1995; 1998], [Фланаган, Полджер, 2006].

Противоречивая позиция О. Фланагана – это либо стремление к строгому теоретическому принципу, либо к метафоре «бессознательных существ, ведущих себя сознательно». Такая метафорфоза представляется типовой для зомби-филов: они плывут между Сциллой необходимости логической строгости понятий и Харибдой метафорико-поэтического образа зомби.

«Земля зомби» Т. Моуди (1995 г.). Т. Моуди, критикуя ПНС в небольшой работе «Общение с зомби» [Moody, 1995], особо выделяет проблему *другого* – можно ли отличить (дискриминировать) сознательных существ от зомби и каковы должны быть научные принципы такой дискриминации. Приводится ряд аргументов в пользу того, что идентификация зомби должна осуществляться не за счёт поведенческих отличий, а на социокультурном уровне, по крайней мере, в условиях речевой коммуникации. Ни логическая, ни физическая возможность зомби даже не обсуждаются. Главная задача – найти критерий «зомбиности» (mark of zombiehood) исследуемой системы – *как отличить личностное от зомбиного?*

Для определённости критерия «зомбиности» вводятся следующие классы зомби.

1. «*Зомби среди нас*» – это те, кто растёт и живёт среди людей. Теоретического интереса такие зомби не представляют – общим местом является тот факт, что многие действия человек выполняет бессознательно.

2. «*Истинные зомби*» – это существа, изолированные от мира, в котором обнаруживается сознание. Эволюция зомби имеет чисто физические основания и совершенно не зависит от взаимоотношений с сознательными существами – ведь таковых нет в зомбином окружении. Для «чистоты» эксперимента Т. Моуди предлагает возможный мир, планету «Земля зомби». Жители её бессознательны. Эволюция на Земле зомби осуществляется, однако отбору подвергаются исключительно поведенческие акты, эффективные по тем или физиче-

ским критериям. В борьбе за «выживание» не участвуют сознательные регулятивы поведения. Их нет, потому что они ничего не значат.

Допустим, между людьми и зомби начался «культурный» обмен: люди стали прилетать на Землю зомби, а зомби – на Землю людей. Осуществление «межкультурного общения», при эксплицировании лингвистических различий разных «культур», создаст достаточные условия выявления зомбиевости. Для этого нужна процедура понимания людьми высказываний зомби и, наоборот, понимания зомби того, что высказывают люди. Для различения человеческих и зомбиевых выражений предлагается индексировать термины, например «понимать[z]» означает аналог слова «понимать» на русском языке. Но что значит для зомби слово «понимать»[z]? Здесь Т. Моуди ссылается на проблему понимания, поднятую Дж. Серлем в «Китайской комнате». Допустим, входные и выходные данные и у зомби и у людей тождественны. Однако человек обладает определённой разновидностью сознательного опыта – для человеческого «понимания» присуща осознанность, осмысленность информации не только на входе и выходе, но и в процессе внутренней, рефлексивной обработки. Для зомби феноменологический опыт отсутствует, термин «понимание[z]» ничего не значит, это слово вообще не могло бы возникнуть на Земле зомби. Зомби, как бы они ни развивались в своём замкнутом физическими законами мире, не способны создать словарь ментальных терминов, включающий слова думать[z], воображать[z], мечтать[z], верить[z], видеть сны[z] и пр. «Философский язык» зомби (Т. Моуди считает, что философия у зомби будет иметь место) также будет различаться – для зомби не будет проблем «инвертированного спектра», «квалиа», «другого сознания» и пр. Они смогут рассуждать[z] по проблеме «других зомби», но эта проблема будет выявлена лишь после встречи зомби с нами. Более того, после встречи с нами зомби станут рассуждать[z] о том, что именно люди говорят о сознании. Возникнет понимание[z] ментальных терминов, высказываемых людьми, по аналогии с тем, как люди понимают высказывания мистиков и эзотериков.

Таким образом, отсутствие у «истинных» зомби словаря ментальных терминов позволит, как считает Т. Моуди, отличить зомбиевое от человеческого. Однако коммуникация между зомби и людьми приведёт в конечном счёте к тому, что истинные зомби станут «зомби среди нас» и их невозможно будет распознать в нашем, человеческом мире.

Критика «Земли зомби» (с 1995 г.). Мысленный эксперимент Т. Моуди вызвал оживлённую дискуссию. Состоялся специальный симпозиум, представ-

ленный в 1995 г. на страницах «Журнала изучения сознания»¹. Название публикаций и имена участников дискуссии говорят сами за себя: Родни М. Дж. Коттерилл – «О единстве сознательного опыта»; К. Сазерленд – «Земля зомби»; О. Фланаган и Т. Полджер – «Зомби и роль сознания»; Д. Деннетт – «Невообразимая нелепость зомби»; Г. Гузельдер – «Многообразие зомбиевости»; Дж. Ленайер – «Зомби обсуждать невозможно»; Д. Маккарти – «Зомби Тодда Моуди»; С. Брингсйорд – «В защиту непостижимости зомби»; М. Мидглей – «Зомби и тест Тьюринга»; А. Элитзар – «Сознание нельзя игнорировать»; К. Чандлер – «Декарт, Фланаган и Моуди»; Д. Ходгсон – «Что зомби не могут делать?»; Ч. Т. Тард – «Да, мы – зомби, но можем стать сознательными»; Т. Моуди – «Почему зомби не хотят умирать?».

Дискуссия продолжается по сей день. Выделим четыре класса отзывов об этом мысленном эксперименте.

1. Некорректность мысленного эксперимента.

Д.Деннетт считает, что этот мысленный эксперимент недостоин рациональной аргументации и представляет собой разновидность «философской глупости»: «Земля зомби» – невнятное самоопровержение собственной позиции. У философов это встречается редко, но в данном случае имеет место. Т. Моуди определяет зомби как существ, поведенчески неотличимых от сознательных людей, поэтому они и в самом деле должны быть неотличимы! Если словарь зомби и философия зомби отличаются от человеческого словаря и человеческой философии, то они на самом деле отличаются! Какая невообразимая нелепость! Здесь итог всего того, что является ложным в современных теориях сознания. Но, упорствуя в своём выборе, философы бросаются понятием зомби так же, как с ладони на ладонь перекидывают горячую картошку.

П.Чёрчленд более безжалостна: дискуссии по теме «Земля зомби» беспочвенны, а мысленный эксперимент Т. Моуди – это аргумент, демонстрирующий слабость собственно мысленных экспериментов.

С. Брингсйорд считает, что зомби-базированный аргумент против ПНС, на первый взгляд, вполне приемлем. Однако детальный анализ показывает иное: «Земля зомби» не более чем фокус. Составить ясное понятие зомби невозможно, в самой идее зомби скрывается то, что превосходит способность человеческого понимания – объясните нам, как можно сознательно представить

¹ Symposium on "Conversations with zombies" (1995) Journal of Consciousness Studies. Volume 2, No. 4 (1995), http://www.imprint.co.uk/jcs_2_4.html

бессознательное состояние? Поэтому идентификация зомби сродни пророчеству.

2. Продуктивность идеи Земли зомби для изучения сознания.

Родни М. Дж. Коттерилл считает, что «Земля зомби» ставит новые акценты в проблематике теории сознания – но не на отношении «сознание / мозг», а на связи «сознание / поведение». Сознание следует ассоциировать не со стимулом и последующим восприятием ментального образа (что характерно для первого отношения), а с движением и реакцией. «Сайт» сознания должен быть: а) виртуальной предпосылкой единства сенсорной сингулярности; б) нейтральным относительно высокоуровневых иерархий реакций; в) доступным для реальных отношений обратной связи.

А. Элитзар считает, что «Земля зомби» представляет особый теоретический интерес. «Дух» зомби вызывается для демонстрации уникальности сознания и, несомненно, привлекателен, так как оспаривает исключительность прав физической реальности. Он отказывает в правах теории сознания, которая оберегает полноту и замкнутость физического мира.

3. «Земля зомби» в контексте философии искусственного интеллекта.

Дж.Маккарти – «отец» искусственного интеллекта – считает «Землю зомби» совершенно неправдоподобным экспериментом. С позиции искусственного интеллекта, как полагает Дж. Маккарти, сознание – это сложная совокупность взаимодействующих материальных процессов, а не некая идеалистическая конструкция или дуалистическая субстанция. Сознание играет конкретную роль в человеческой деятельности. Чтобы зомби могли вести себя как люди, им требуется что-то вроде сознания – так называемое «псевдосознание». Псевдосознание должно участвовать в выполнении моторно-двигательных, речевых, интеллектуальных актов. Сюда относятся и лингвистическая компетентность, по крайней мере, способность зомби отвечать на задаваемые вопросы. Однако, по определению, ничто похожего на псевдосознание у зомби не должно быть, а так как псевдосознательная деятельность неотделима от поведенческого акта, то выносится вердикт: *«Зомби Моуди – плод выдумки Моуди. Зомби Моуди невозможны!»*

А. Сломан – видный философ искусственного интеллекта – в принципе соглашается с Дж. Маккарти. Он исходит из анализа различных форм робототехнической реализации феномена субъективности (позиции первого лица): «[Зомби] не сможет имитировать поведение сознательного существа, не обладая моделями «эмоций», «точек зрения» или иных феноменов субъективной

жизни. Утверждающий подобное занимается самообманом. Таким «философам» следует поучиться инженерному делу, тогда они поймут, что робот-зомби работать не будет. Для инженера представить себе робота-зомби означает отказаться от всего того, что он знает о механизмах обработки воспринимаемой информации, хранении её, генерации новых задач, принятия решений для реализации поведения или воздержания от недопустимых действий, оценки степени выполнения задачи или мобилизации ресурсов» [Sloman, 1998-1; Сломэн, 2006].

О. Флаганан и Т. Полджер, несомненно, осуществили наиболее многогранную критику «Земли зомби» [Flanagan, Polger, 1995]. Пристальная критика понятна – ведь ключевая идея Т. Моуди заключалась в том, чтобы показать, что ПНС ложен или нуждается в пересмотре. Авторы основательно разбирают *аргумент Моуди*, который переформулировали следующим образом: а) если ПНС верен, тогда нет возможности отличить зомби от сознательного существа; б) можно утверждать, что жители Земли зомби – это зомби; в) следовательно, ПНС неверен или требует уточнения. Демонстрация аргумента корректна (*modus tollens*), но посылки не обоснованы. Первая посылка ложна, так как ПНС утверждает лишь то, что интеллектуальный акт может быть выполнен без осознания этого акта, но ничего не говорит о ситуациях различения людей/зомби. Вторая посылка ложна, так как легко можно вообразить сценарий, в котором зомби (по сути, робот) вступая в социальные связи (например, для предупреждения других зомби об опасности), формирует квазиментальные термины. Такие же термины возникают в результате сбоя оборудования или формирования непредусмотренных алгоритмом сценариев действий. Авторы рисуют яркие образы компьютерных реализаций зомби, которые способны эволюционировать, создавать зомбиевый язык, включающий термины самого широкого диапазона значений – от образов, снов и мечтаний до философских споров на темы инвертированного спектра, квалиа, проблемы *другого*.

4. Иерархия сознаний.

Особый «религиоведческий» интерес вызывают утверждения Т. Моуди о разнопорядковых уровнях понимания реальности мистиками и обычными людьми. На эти заявления живо откликается представитель трансперсональной психологии Ч. Тард, который в противовес принципу несущественности сознания предлагает *принцип расширения сознания*: «Все мы – зомби, но можем стать сознательными!». Для трансперсоналистов преодолеть зомбированность – это

принять психоделические средства. И тогда «Ваше сознание станет абсолютным».

Дискуссии по «Земле зомби» продолжаются по сей день, так как затрагивают важные вопросы методологии искусственного интеллекта, социокультурные параметры соотношения сознательного и бессознательного, но не в традиционном психоаналитическом контексте, а в контексте идей когнитивно-компьютерных наук и искусственного интеллекта. На мой взгляд, ПНС, характеризуя теоретический уровень изучения роли и функций сознания, вполне совместим с мысленным экспериментом земли Зомби, который характеризует риторико-метафорический контекст изучения проблематики зомби.

«Чалмерс-зомби» Д. Чалмерса (1996 г.). Если предыдущий эксперимент служит решению конструктивных вопросов, то Д. Чалмерс продолжает спекулятивную линию проблематики зомби, значительно дополняя её модально-логическими исследованиями. Он предлагает вообразить свой собственный дубликат, искусственную «версию» Чалмерса, организованную точно так же, как организован реальный философ – Чалмерс [Chalmers, 1993-1]. Разница в следующем: там, где у реального Чалмерса нейроны, у «двойника» – кремниевые чипы. Для самого Чалмерса и, как он считает, для многих других очевидно, что «Чалмерс-зомби» не обладает сознанием, ведь у него внутри всё *пусто и темно*, так как ни в кремнии, ни в биохимии нет ничего того, что относится к сознанию. Для Чалмерса также очевидна мыслимость и, следовательно, логическая возможность Чалмерса-зомби. Комментаторы обычно цитируют следующее: «Признаюсь, что логическая возможность зомби кажется совершенно очевидной для меня... Нет противоречия в данной дескрипции, хотя принятие её логической возможности основано на интуиции... Мне кажется, что едва ли не каждый способен помыслить эту возможность... Я не могу обнаружить никакой логической бессвязности [в идее зомби] и имею ясную картину, когда представляю зомби». Некоторые могут отрицать возможность зомби, однако они должны быть компетентны в проблематике возможности, причём данная компетентность должна быть выше, нежели чем у тех, кто допускает возможность зомби. Короче, доказывать [возможность/невозможность] зомби должен тот, кто утверждает, что дескрипция зомби логически невозможна. При этом [оппонент] должен чётко показать – где видится явное или неявно выявляемое противоречие [Ibid, Р. 96, 99]. То есть бремя доказательства лежит на антизомбисте, а не на зомбисте.

Эксперимент с зомби важен для Д. Чалмерса. Формально на это указывает глоссарий книги: слово «зомби» – наиболее часто употребляемое в книге. Чалмерский выбор «логической возможности» (или мыслимости) зомби – это намерение снабдить ярким экстравагантным примером и усилить базовую не-одуалистическую посылку о том, что сознательные состояния, или «квалиа», неподвластны физикальному и функциональному анализу.

Однако, как считают критики (Р. Кирк, Э. Коттрелл), Д. Чалмерс слишком уж просто «расправился» с проблемой мыслимости зомби, приступив сразу к аргументу зомби. Также он не последователен в применении аргумента мыслимости зомби. Поэтому критики предлагают Чалмерса-зомби для оппозиции, т.е. для реализации стратегии антизомбистов: если можно убедительно показать *немыслимости* Чалмерса-зомби, тогда и теория Д. Чалмерса рухнет.

Продуктивность идеи зомби

Автор полагал, что, представив ряд явный и неявных определений зомби (конечно, неполный), приуменьшил дозу антизомбиевой вакцины «неведения» от очередной зарубежно-философской «инфекции». Однако здесь его поджидал удар. В последней крупной работе «Зомби и сознание» (ноябрь 2005 г.) Р. Кирк – «отец» проблематики зомби – неожиданно для всех делает обескураживающее заявление: «Книга преследует две основные цели, одна из которых – стереть идею зомби раз и навсегда!» [Kirk, 2005]. Зомби не стоит заниматься, так как благодаря им возникает *фундаментальная путаница в понятии сознания*. Возможно, для когнитивной науки в целом это так. Однако для контекста э-культуры, ИКТ, когнитивных технологий тест зомби крайне важен – он позволяет выносить экзистенциальные суждения по поводу атрибуции и предикации психических феноменов некоторой *x*-системе – т.е. суждений типа «система сознательна», «система бессознательна». Перспективы применения теста зомби представлены далее при обсуждении изучения персональных феноменов в контексте э-культуры. До рассмотрения проекта искусственной личности (см. подразд. 5.2) сделаем следующее заявление:

Проблема философских зомби значима по причине двойственности проблемы искусственной личности.

Однако подчеркнем следующее:

Для нейтрально-зомбистской позиции когнитивно-компьютерные проекты философских зомби немыслимы.

3.1.3. Дефиниция ментальных феноменов: «сознание» в электронной культуре

Автору известны две главные директивы философского изучения феномена сознания – метафизико-аналитический и метафизико-синтетический (спекулятивный). В контексте э-культуры второй путь следует напрочь исключить. Иначе наукоёмкие технологии э-культуры покроются шелухой несовместимых суждений из гегельянских диалектических «переходов количества в качество», поэтических хайдеггеровских суждений типа «время на то и время, что оно себя временит», бессвязных эпатажных постмодернистских квазиэкивоков.

Более продуктивная директива, на которую, кстати, опирается современный компьютеринг, возникла в контексте аналитической философии сознания. Она обеспечивает конструктивный и в определённой степени детальный анализ ментальных свойств, качеств и способностей. Анализ используется не ради «анализа». Результаты философской работы позволяют упорядочить, систематизировать, преобразовать, операционализировать, интегрировать компоненты анализа, представленные в формате верифицированных или фальсифицированных языковых выражений для различного рода практических целей – оптимизации экономики, эффективности политики, экспертизы произведений искусства, достоверности истории, рационализации права и пр. В нашем случае цель философского анализа состоит в изучении возможности технологической реализации феномена сознания.

По причине собственных методологических предпосылок детального и скрупулёзного изучения предметной области, проведения специальных исследований в области языка описания предметной области и языка описания метода описания и пр. метафизико-аналитическая стратегия «грешит» объёмностью и многочисленностью научных работ. Так, Д. Чалмерсом насчитывается 28 246 крупных книг и статей [Chalmers Bourget, 2013] по проблеме сознания на момент написания автором данного текста (17.05.2010 г.). Конечно, этот список значительно больше. Поэтому крайне затруднительно, а по сути невозможно, разобраться в трудно регистрируемом многообразии подходов, парадигм, теорий, концепций, мысленных экспериментов, эмпирических данных и пр. Вполне естественно стремление исследователей ввести порядок в это разнообразие. Поэтому Д. Чалмерс создаёт иерархическую классификацию работ по проблеме сознания и вводит систему кодирования (очень примитивную с позиции библиотечного дела). Само собой разумеется, что данная классификация служит не

энциклопедическим целям, а конкретным задачам, стоящим перед исследователем. Имеются иные модели.

Продуктивной концептуальной моделью исследования проблемы сознания представляется «философский квадрат» (рис. 3.1, а), предложенный Д.И. Дубровским более 20 лет назад (воспроизведён в работе [Дубровский, 2007, С. 26 – 27]). Модель проста, однако надёжно систематизирует нашу работу. Модель часто применяется в работах Д.И. Дубровского и его учеников. Так, в одной из последних работ он применил её к проблеме интересубъективности в контексте исследований ИИ [Дубровский, 2011-1, С. 129 – 148].

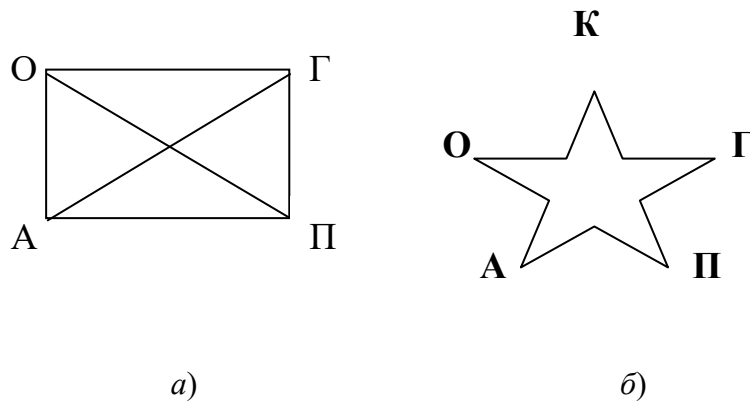


Рис 3.1. Концептуальная модель проектов философии сознания: а – «философский квадрат» Дубровского; б – культурологически нагруженная пентаграмма проектов философии сознания.

Четырёхмерное пространство схемы исследований, предложенное Д.И. Дубровским, определяется четырьмя фундаментальными категориями: 1) онтологическое (О), 2) гносеологическое (Г), 3) аксиологическое (А), 4) праксеологическое (П). Так как эти категории, по его мнению, инвариантны по отношению ко всем философским направлениям, то они соответственно конституируют и частную проблематику – проблему сознания (точнее, субъективной реальности) [Дубровский, 2009-2]: «Без основательной гносеологической рефлексии невозможна основательная онтология... Ценностные параметры и факторы активности непременно включены в любой акт познания и сказываются на его результате» [Там же, С.26]. Отношения, обозначенные на «квадрате», позволяют вести систематический и целостный учёт параметров сознания, не упускать из виду одно по причине изучения другого. Представленные на «квадрате» планы изучения сознания, в силу философского обобщения, позволяют высказывать универсальные суждения. Однако понятие «сознания» релятивно просто по причине его феноменального, необъективируемого статуса, уникальности индивидуального сознания. Конвенции по поводу того, что считать «сознанием», предзаданы социокультурными контекстами употребления данного

термина. Поэтому «философский квадрат» не достаточен для отражения фундаментальных отношений, возникающих при познании сознания. Он не учитывает социокультурную обусловленность феномена сознания. «Квадрат» не фиксирует *культурологического измерения* (К). Конструкцию следует дополнить до пентаграммы (рис. 3.1, б).

Покажем необходимость культурологической параметризации проблемы сознания путем сравнения того, что понимается под «сознанием» в западной и восточной традициях.

Так, например, Ван Гулик – представитель западной культуры – фиксирует не более десятка понятий «сознания» [Gulick, 2004]. Наиболее значимые параметры сознания – это ощущение, бодрствование, самоосведомлённость, субъективность, транзитивность. Основные состояния сознания: «осведомлённость о», квалиа, феноменальные состояния, состояния проекции (что-значит-быть-как), акцессное состояния (состояние доступа), нарративность. Проблема сознания прорисовывается в постановке ряда вопросов:

1) дескриптивные вопросы изучения сознания (каковы особенности сознания?): «данные» от первого, второго, третьего лица, квалиа, феноменальная структура, субъективность, самоперспектива, тождество сознания, интенциональность и прозрачность, сознание как «поток»;

2) объяснительные вопросы (как сознание существует?): редуктивные и нередуктивные программы объяснения феномена сознания и, конечно, наличие объяснительного провала в объяснении;

3) функциональные вопросы (зачем сознание?): каузальный статус для организации поведения, социальной координации, адаптации и контроля; функциональный статус репрезентаций, доступ к состояниям сознания (внутренняя мотивация и свобода воли также относятся к функциональной составляющей проблематики сознания).

Теперь учтём восточную традицию и рассмотрим понятие сознания в психологии раннего буддизма, описанное ламой Анагарика Говинда (Психология раннего буддизма. СПб.: Издательство «Андреев и сыновья», 1993). Здесь насчитывается 54 класса чувственных состояний сознания, большинство которых «эстетически нагружено». Среди них – радость, грусть, боль. Остальные классы (общее число классов состояний сознания достигает 121) – это «этически» нагруженные состояния сознания. По всей видимости, подобного рода детализация, инспирированная метафизикой буддизма, более «аналитична», чем её западный аналог, рассмотренный немного выше.

Данный пример приведён для демонстрации изучения понятия сознания с учётом социокультурных условий. Для целей нашей конкретной работы проблема сознания – это не некая универсальная проблема «вообще», но проблема, специфицированная контекстом культуры, в первую очередь, электронной культуры. Причем разные формы э-культуры существенно модифицируют смыслы и значения термина «сознание».

К сожалению, «квадрат» и «пентаграмма» не удовлетворяют стандартным требованиям аналитической философии – в них отсутствуют лингвистические отношения между обозначенными терминами. Необходим психосемиотический контекст.

Двухмерная семантика ментальных терминов

Плодотворной представляется семантико-онтологическая схема психофизической проблемы, предложенная Э. Вильянуэвой [Вильянуэва, 2006, С. 53]. Э. Вильянуэва предлагает плоскую 2-D схему (см. рис. 3), на которой выделяет интенционалы: 1) суждений о феноменальных свойствах сознания; 2) суждений в формате научных объяснений этих свойств.

В целом, исследовательская установка Э. Вильянуэвы гуманистически ориентирована. Он борется с так называемой «парокиальностью» – замкнутостью исследователя в собственном мире субъективных верований, знаний, переживаний, установок. Свобода учёного обеспечивается выходом на просторы объяснительного анализа внутренних психических качеств с использованием всего спектра современных наук – психологии, нейрофизиологии, физики, химии. Однако, убегая от «парокиальности», он вновь попадает в тюрьму парокиальности. Автор – представитель традиционной культуры. Он не задаётся вопросом технологической реализуемости феномена сознания, а для нас это – ключевой вопрос э-культуры. Предпринятые им рассуждения технологического плана (а именно, при рассмотрении функционализма машины Тьюринга) применяются в формате теоретического объяснения психики, но не как инженерный проект. Поэтому схема Э. Вильянуэвы, только появившись, сразу устарела. Она пригодна для проблематики сознания лишь в рамках традиционной культуры.

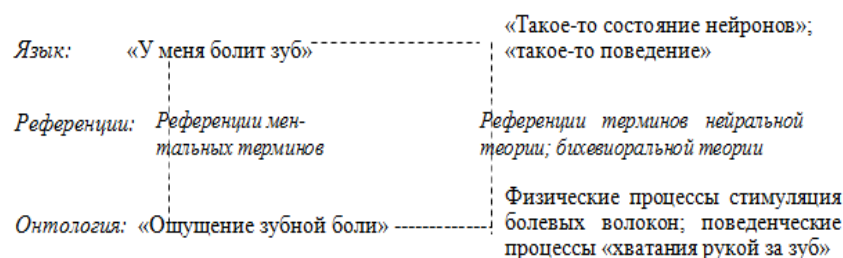


Рис 3.2. 2D-семантическая модель изучения проблемы сознания (в традиционной культуре)

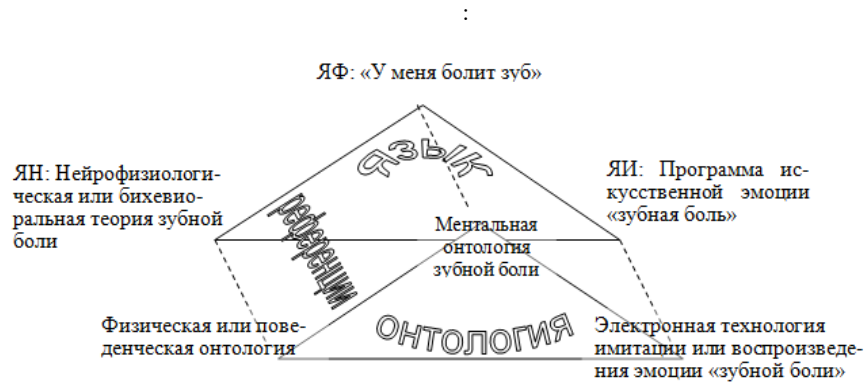


Рис. 3.3. 3D-семантическая модель изучения проблемы сознания (в электронной культуре)

Предложим модель, которая задает технологический императив реализации феномена сознания – трёхмерный концептуальный каркас Данная модель (см. рис. 4) для традиционной культуры не пригодна, но для выражения специфики э-культуры представляется необходимой – в ней отражается технологическая составляющая реализации когнитивных феноменов и, так же, как в предыдущей – семантико-онтологические связи (референции обозначены пунктирной линией).

Трёхмерная семантика ментальных терминов

Трёхмерная семантика суждений о феномене сознания означает совместное изучение высказываний трёх планов, выраженных на трех специальных языках: ЯФ – язык выражения моих психических состояний, ЯН – язык наук, ЯИ – язык инженерной реализации:

1) ЯФ – *феноменальное суждение*, характеризующее собственно факт сознания, переживаемый человеком (*первичный интенционал*);

2) ЯН – *научное суждение*, рационально объясняющее факт сознания посредством теоретико-эмпирических исследований естественных, гуманитарных, социальных наук (*вторичный интенционал*);

3) ЯИ – *инженерное суждение* о компьютерной реализации феномена сознания, т.е. высказывание в терминах теории компьютерной системы (*третичный интенционал*). Имеется в виду высказывание о структурно-функциональном, логико-математическом, программном, информационном, лингвистическом, методическом и прочем устройстве компьютерной системы.

Возникает правомерный вопрос – зачем применять 3D-семантику, нет ли возможности совместить два плоских плана, один из которых уже предлагался Э. Вильянуэвой? То есть слева оставить квадрат из заданных им семантических соотношений для феноменальных качеств и научных теорий. Справа пристроить подобного рода квадрат, характеризующий референции инженерно-технологической конструкции, которая базируется на той или иной научной теории.

Иначе вопрос звучит следующим образом – «Нуждается ли искусственный интеллект в исследованиях естественного интеллекта?» (см. п.п.1.1.2). Эта тема специально обсуждалась на семинаре НСММИ РАН 01.11.2006 г. в докладе О.П. Кузнецова на тему «Может ли искусственный интеллект развиваться независимо от исследований естественного интеллекта?». Оппонентом выступила М.А. Холодная, которая отстаивала необходимость и первоочередность исследований естественного интеллекта. Т.Б. Кудряшова высказала идею о том, что исследования ЕИ без исследований ИИ невозможны. Автор высказал идею о совместном, параллельном исследовании. Данная идея ниже получает конструктивное оформление.

Выскажем свою точку зрения. Инженерия ИИ развивается относительно независимо от достижений той или иной научной теории. Самолёты стали летать до появления теории полёта, которая так и не состоялась. Универсальная

теория интеллекта отсутствует (если вообще таковая возможна). Однако в быту и производстве успешно функционируют компьютерные системы, имитирующие интеллектуальную деятельность человека. Знание нейрофизиологической теории зубной боли не определяет ничего конкретного для инженерной имитации соответствующего поведения, лишь наводит на аналогии. Поэтому здесь важен метафорический контекст исследований сознания для технологической реализации его феноменов. Совершенно лишена смысла буквальная трактовка терминов, скажем, нейрофизиологической теории в терминах организации системы ИИ. Достаточно ясно высказался по этому поводу Г.С. Пospelov: «за спинами специалистов по искусственному интеллекту стоят *тени* великих философов» (курсив мой – А.А.) [Будущее ИИ, 1991].

Здесь были приведены рассуждения исключительно по поводу имитации феноменов сознания, более того, функциональной имитации этих феноменов. Вопрос относительно реального воспроизводства или, собственно, фактической реализации феноменов сознания на базе электронных технологий будет решён лишь в перспективе осуществления широко известного проекта Д.И. Дубровского – проекта расшифровки нейродинамических кодов субъективной реальности. Ничуть не раньше! Автор в этом глубоко убеждён. Также имеется уверенность в том, что предложенная 3D-схема послужит надёжным (по причине простоты) подспорьем для исследования бескрайних просторов теорий сознания с целью поиска технологических решений имитации, моделирования или репродуцирования феномена сознания. Модель требует дополнения, поскольку в ней представлена только одна – онтологическая – составляющая многомерного пространства исследования феномена сознания. Не составляет особого труда проведение аналогичной работы для фиксации гносеологического, аксиологического, праксеологического и культурологического измерений (см. расширение «философского квадрата Дубровского» в нашей «пентаграмме»). После этого следует осуществить требуемые референциальные, интенциональные и экстенциональные привязки к соответствующим языкам, на которых выражаются феномены познания, ценностей, практики, культуры. Это требует кропотливой работы в тех сферах, которые сегодня никак не связаны с технологией: какое отношение, кроме морализаторского, имеет этика к технологии? В трёхмерном каркасе пентаграммы эти связи становятся необходимыми. Однако для начала изучим культурологические модификации термина «сознание», соотнесённые с отнологическими типами э-культуры. Смыслы «сознания» в разных формах э-культур, конечно, значительно разнятся.

Смысловые деконструкции «сознания» в электронной культуре

Феномен сознания – неотъемлемый параметр духовной сферы культуры. Рассмотрим онтологическую схему трехчастного деления духовной сферы э-культуры для определения конкретной разновидности «сознания», с которым соотносится технологическая проблема «сознания».

Применим *долевую модель*. Она характеризует долю (степень) участия собственно человеческой сознательной деятельности по «реализации» некоторого когнитивного феномена (ощущения боли, высказывания суждения, создания произведения, осознания самости и пр.) относительно компьютерной системы, которая способна симитировать, смоделировать, репродуцировать эту деятельность (своеобразный квантификатор фланагановского принципа несущественности сознания).

Э-культура-1 является частью традиционной культуры. Это, например, презентация на DVD предметов народного творчества, видеоролик с выступлением творческого коллектива и пр. Здесь мы не усматриваем особых изменений традиционной проблематики сознания. Человек – полноправный творец э-культуры-1. Для презентации результатов сознательной деятельности, осуществления коммуникаций, поддержки своего творческого процесса (например, компьютерной генерации музыкальных произведений, автоматического реферирования текстов) он пользуется компьютерным инструментарием.

Э-культура-2 – это технологически модифицированная форма традиционной культуры. Здесь имеют место превращённые формы сознания, связанные, в первую очередь, с «погружением» субъекта в виртуальную реальность. Аналогом «неэлектронных» способов изменения сознания являются всевозможные трансперсональные практики.¹ В нашем случае имеют место не медитативные и психоделические средства – стандарт трансперсоналистов, а компьютерное снаряжение систем виртуальной реальности. Человек уже не является абсолютным творцом феномена сознания. Реальность э-культуры-2 наравне с человеком конструирует электронная технология.

Э-культура-3 – самостоятельная форма собственно электронной культуры. Данный случай представляется ключевым для целей нашей работы. Именно в рамках э-культуры-3 наблюдаем максимальную деконструкцию смысла понятия «сознание». В электронном формате представлен и источник, и продукт, и сам способ «сознательной» деятельности. Человек полностью отсутствует, либо

его участие ограничивается инициацией электронной технологии и/или получением её продуктов. «Творцами» выступают, например, беспилотные самолёты кибервойск или колонии роботов, «работающие» в условиях, несовместимых с человеческой жизнью².

Проблема сознания в э-культуре-3: к вопросу о «сознании» «чистого искусственного интеллекта»

Дискуссия по поводу наличия «сознания» у «творцов» э-культуры-3 насчитывает порядка 60 лет. Она возникла сразу после выхода статьи А.М.Тьюринга в 1950 г. Собственно данную дискуссию Тьюринг и инициировал, обсуждая возможности идентификации феномена сознания у цифровых вычислительных машин. Споры продолжаются до сих пор. Возникают следующие вопросы относительно проблемы сознания: может ли компьютер ощущать, творить, чувствовать, любить и пр. Например, в 1980 г. Дж. Серль рассматривал такой параметр феномена сознания, как интенциональность, и задавался вопросом «может ли компьютер понимать?». При этом он ввел достаточно грубые понятия сильного и слабого искусственного интеллекта (грубые – с позиции возможности наличия более широкого спектра реализации когнитивных феноменов) [Searle, 1980; Серль, 2006]. Сторонники слабого ИИ полагают, что компьютерные программы моделируют частные аспекты ментальной деятельности человека. Их оппоненты – сторонники сильного ИИ – полагают, что соответствующим образом запрограммированный компьютер фактически обладает способностью понимания. Более того, компьютерная программа объясняет способность понимания человека.

Данные дискуссии отражают частные аспекты воспроизводства или имитации феномена сознания электронными технологиями. Для решения фундаментальных вопросов важны рассуждения не о каких-то частных качествах и способностях сознания, т.е. не анализ атрибутивных или предикативных суждений, приписывающих те или иные параметры сознания системам искусственного интеллекта. Необходим анализ *экзистенциального суждения о факте сознания как таковом* – есть «сознание» в системе или «сознание» напрочь отсутствует. Если таковое суждение будет иметь место, тогда возможно построение онтологии сознания искусственной системы. Если такое суждение вынести не

¹ См., например, сборник трансперсоналистов: Пути за пределы «эго» / Под ред. Р. Уолша и Ф. Воон: пер. с англ. М. Папуша, Е. Поле, К. Андреевой. – М.: Открытый мир, 2006. – 392 с.

² См. информацию подобного рода на Интернет-дайджесте «Интеллектуальные Информационные Технологии»; <http://www.dialog-21.ru/news/about./digest.asp>

представится возможным, тогда онтология ментального становится невозможной.

Для отсеивания ненужных дискуссий осуществим радикальный подход с позиции философских зомби – бессознательных систем, которые поведенчески, функционально и/или физически тождественны, неотличимы и/или подобны сознательным существам [Алексеев, 2005-3; 11; 12; 13; Алексеев, 2006-4; Алексеев, 2008-2; Алексеев, 2009-5].

Экзистенциальная атрибуция «сознания» компьютерной системе

Экзистенциальная атрибуция «сознания» – это утверждение факта наличия сознания в *x*-системе. Лишь после того как вынесено экзистенциальное суждение, появляется возможность качественно и количественно дифференцировать ментальные феномены, приписываемые системе. Несомненно, это радикальный подход. Но он представляется честным и не позволит судить о том, присущ или не присущ «интеллект» системе, если, например, она не обладает «сознанием». Если «сознания» нет, то и «интеллекта» нет. Например, Н.С. Юлина полагает, что машина может мыслить (Н.С. Юлина неоднократно высказывала это как на конференциях, так и в личных беседах с автором). Но если далее она признает, что интеллектуальная машина сознанием не обладает, то в силу убедительного *принципа несущественности сознания* Фланагана (интеллекту сознание не требуется) мы должны признать, что теория интеллекта должна разрабатываться отдельно от теории сознания. Это не вяжется ни с народно-психологическими представлениями о соотношении мышление/сознание, ни с более рациональными вариантами теорий сознания (кратко было рассмотрено автором в тесте Блока: интеллект – вторичная способность относительно квалиа). Возможны иные варианты. В любом из них тест зомби, используемый в составе КТТ, позволяет четко идентифицировать и структурировать следующие исследовательские позиции.

1. Понятие зомби позволяет навести порядок в определении «сознания», так как предел любому понятию полагается противоречивым или, по крайней мере, противоположным понятием. То есть понятие бессознательного зомби логически необходимо для выработки понятия сознательного существа. Здесь становится важной *таксономия зомби*, учитывающая следующие параметры: а) теории сознания, которые анализируемые зомби опровергают либо удостоверяют (в первую очередь, физикализм, бихевиоризм, функционализм); б) модальности (степени мыслимости и возможности зомби) и др.

2. Зомби дают богатый фактуальный материал в изучении бессознательного. Конечно, эти факты воображаемы, тем не менее, пройдя через горнило мысленных экспериментов, они становятся вполне убедительными. Дозомбиевые средства экспликации бессознательного – это неработоспособные понятия о бессознательном «вообще» либо психоаналитическое эмпирическое мелкотемье «минимальной достаточной определённости» понятия бессознательного, на что справедливо указывает Д.И. Дубровский [Дубровский, 2006].

3. Мысленные эксперименты с зомби аккумулируют, по сути, параметры всех крупных мысленных экспериментов в философии сознания, таких как «Мозги в бочке» Х. Патнэма, «Человек-имитация» Д. Кэмпбелла (1970 г.), «Гомункулная голова» и «Китайская нация» Н. Блока (1978 г.), «Китайская комната» Дж. Серла (1980 г.), «Мэри» Ф. Джексона (1982 г.), «Танцующее квалиа» Д. Чалмерса (1996), «Воришка феноменального» М. Линча (2004 г.) и др. Они раскрывают лишь частные аспекты зависимости сознательных феноменов от физического, бихевиорального, функционального, персонального, социального. Поэтому очевидна методологическо-интегративная функция понятия зомби относительно этих экспериментов.

4. Важность проблематики зомби вызвана не только интересами спекулятивной метафизики сознания, как у Д. Чалмерса. Приложения проблематики зомби можно проследить во всех науках, в поле зрения которых – роль, функции, структуры сознания. Очевидны их применения в общественных и гуманитарных науках. Сегодня зомби находятся в фокусе внимания исследователей искусственного интеллекта и служат основанием для критики либо поддержки таких перспективных направлений в информационной технологии, как искусственная жизнь, искусственная личность, искусственное общество. Способы идентификации зомби сродни тесту Тьюринга – базовому мысленному эксперименту философии искусственного интеллекта, предназначенному для определения интеллектуальности систем. Тест на предмет выявления сознательности/бессознательности систем (*тест Зомби*) – это, вне всяких сомнений, разновидностью теста Тьюринга. Тест Зомби плюс знание внутренних структурно-функциональных отношений тестируемой системы создаёт модифицируемые условия определения сознательности/бессознательности. А это существенно приближает к решению проблемы *другого*.

5. Все вышесказанное подтверждает, что вряд ли Р. Кирк способен убить своё детище – джин выпущен из бутылки. Однако чтобы воспользоваться «силой» зомби, надо отдавать себе отчёт – где мы имеем метафорико-риторические

параметры, присущие мысленному эксперименту, а где – логико-эпистемологические параметры, приближающие нас к онтологии. Эти параметры не надо смешивать. Иначе возникает путаница, но не в понятиях философии сознания, а в уровнях описания сознательных систем.

В заключение отметим то, что экзистенциальная дефиниция проблематики компьютерного сознания, ориентирующая исследователя на выявление факта наличия сознания у *x*-системы, является первичной определительной функцией в проблематике приписывания оцениваемой системе ментальных, в частности, и когнитивных феноменов, в общем.

Последующие задачи дефиниции представляются вторичными: атрибутивная, приписывающая психические качества *x*-системе, и предикативная задача, приписывающая более сложные когнитивные феномены, например феномены социального порядка. Таким образом, решение поднятых проблем «сознания» и *другого* в информационно-коммуникационной среде возможно лишь после зомбистского, антизомбистского или нейтрально-зомбистского позиционирования исследователя философии искусственного интеллекта.

Попытка выбора такой позиции приводит к бесконечным схоластическим спорам относительно проблемы «сознания» – можно бесконечно долго комбинировать частные ТТ и манипулировать ими для подтверждения/опровержения собственной позиции по проблеме сознания – такая рефлексия уводит нас от решения психотехнологической проблемы и заводит исследования ИИ в тупик.

Имеются иные стратегии развития ИИ, помимо решения проблемы сознания («сознание-мозг-искусственный интеллект»), как доминирующей проблемы ИИ [Дубровский, 2005]. По мнению автора, стратегию развития ИИ следует формировать исходя из первичного решения более узкой проблемы – проблемы «творчества», но не с проблемы «сознания». В проблематике творчества имеется гораздо больше конструктивного материала; многое имеет чёткий семиотический облик, а не метафизические фантазии об онтологии другого сознания, включая туманную феноменологическую теорию бытия собственного сознания. (Обоснованию переориентации стратегии ИИ с проблематики «сознания» на проблематику «творчества» посвящен пп.3.3.1.)

3.2. Критическая функция

Рассмотрим критический потенциал комплексного теста Тьюринга на примере концептуального анализа инженерных, теоретических и социокультурных проектов э-культуры.

3.2.1. Критика технологии: комплексный тест Тьюринга против комплексной технологии динамических экспертных систем реального времени

В эпистемологических исследованиях искусственного интеллекта выделяется две фундаментальные оппозиционные парадигмы развития «знаний» компьютерной системы: универсальная и эволюционная.

Оппозиция универсально-эпистемологической и эволюционно-эпистемологической технологий построения интеллектуальных систем

Универсально-эпистемологическая (УЭ) парадигма предполагает полный и постоянный охват предметной области с позиции всезнающего и вездесущего разработчика – эксперта, когнитолога (инженера по «знаниям»), программиста. Для представления декларативных «знаний» такой вариант применим лишь к простым системам с конечным множеством актуально или потенциально обзоримых «знаний». Очевидно, что разработчик не в силах специфицировать все возможные ситуации и сценарии сложной предметной области – слишком уж она испещрена уникальными вариациями, не подпадающими под общие дефиниции.

Для процедуральных «знаний», генерирующих многообразие предметной области из аксиоматики по правилам вывода, ограниченность универсального подхода менее очевидна. Тем не менее, возможность применения конечного логико-аксиоматического каркаса, реализующего давнюю идею лейбницаанского универсального исчисления, ограничена большими системами, бесконечное множество «знаний» которых потенциально перечислимо. Однако ни для малых, ни для больших систем ИИ не нужен – достаточно традиционных способов обработки информации. Причём принципиально не важно, что такие системы могут быть нечёткими – особого труда не составляет расширение традиционного логико-алгебраического базиса неклассическими, нечёткими, вероятностными, квантовыми и другими вариациями.

Искусственный интеллект востребован для сложных систем, где каждая «регулярность» сопровождается неопределённым веером сингулярностей. Соб-

ственно по этой причине для развития «знаний» сложной системы применяют технологии её постоянного и непрерывного сопровождения разработчиком. Такие системы по непонятным причинам называют «интеллектуальными динамическими системами реального времени». Интуитивно очевидно, что к ним невозможно применить термин «искусственный интеллект», ведь здесь задействован исключительно естественный интеллект разработчиков, усиленный компьютерными методами и средствами. Получается, что универсально-эпистемологическая парадигма развития «знаний» априори искажает идеи искусственного интеллекта и сводит их на нет. Тем не менее, она доминирует в современной технологии.

Эволюционно-эпистемологическая (ЭЭ) парадигма адекватна идеям интеллектуальных компьютерных систем. Суть её заключается в приписывании машине способностей самостоятельно, свободно, творчески извлекать и порождать «знания» предметной области, включая «знания о себе». ЭЭ-идеи, концептуально встроенные в механизмы функционирования системы, предполагают активное изменение способностей восприятия и преобразования внешней и внутренней среды, исходящую изнутри системы целеполагающую модификацию форм и способов аргументации логических рассуждений, самостоятельное формирование нормативных регулятивов поведения и деятельности. ЭЭ-концепции в экспертных системах задают автоматическое извлечение фактов предметной области, формирование нетривиальных правил, активное порождение гипотез, креативное продуцирование «знаний». Для компьютерных игр они призваны обеспечить внутреннюю способность расширять пространство сценариев и ролевых статусов. Системы доказательства теорем и автоматизации рассуждений предполагают самоуправляемое расширение логических средств. В системах решения проблем и планирования – формирование мотивационно детерминированных способов целеполагания, смыслового знакового опосредования и, в конечном итоге, «квазисознания» окружающей среды и репрезентации её в целостном, независимом, отстранённом формате. В системах технического восприятия, таких как распознавание образов и паттернов, визуальный анализ сцен, слежение за движением объектов и распознавание лиц, ЭЭ-идеи определяют способы автономного обучения, самостоятельное формирование эталонных образцов, программно-генетическое кодирование решающих правил распознавания.

В системах компьютерной лингвистики, таких как анализ и синтез устной и письменной речи, распознавание отдельных выражений и связной речи, понимание естественного языка и машинный перевод, ЭЭ-концепции предпола-

гают «искусственный семиозис» – самогенерацию знаковых систем, самоорганизацию интегральной системы «синтаксис-семантика-прагматика», автоматическое формирование референций, дескрипций и пр.

В робототехнике ЭЭ-концепции принципиальны для многообразных проектов самовоспроизводящихся, антропоморфных, эмоциональных, сознательных и прочих роботов, особенно если они предназначены для совместной работы в условиях, несовместимых с человеческой жизнью, например на Марсе, Крайнем Севере, в глубинах Мирового океана – в этих проектах колоний роботов задействованы всевозможные модели развития и функционирования общественных форм сознания – от схем предотвращения физических столкновений до способов выработки «квазирелигиозных убеждений».

Представленная панорама применения ЭЭ-моделей имеет в большей мере прогностический статус – такие системы исследователи ИИ называют «нетрадиционными», чтобы подчеркнуть их отличие от общепринятых систем, развивающихся в рамках УЭ-подхода.

Для изучения двух выделенных парадигм воспользуемся методом комплексного теста Тьюринга, так как он позволяет выделить ЭЭ-идеи на основательном базисе категориально значимых мысленных экспериментов, не растворяясь в спекуляциях теорий сознания и не утопая в эмпирическом мелкотемье бесчисленных проектов интеллектуальных систем.

Эволюционно-эпистемологические идеи в комплексном тесте Тьюринга

Эволюционно-эпистемологические идеи комплексного ТТ выявляются на основе анализа ЭЭ-идей в составе частных тестов, их сопоставления и комбинирования, выделения общих, частных и особенных положений, выдачи рекомендаций относительно реализации ЭЭ-механизмов и пр. [Алексеев, 2012-2]. Рассмотрим в качестве примера как комплексизируются ЭЭ-идеи в тьюринговых тестах Д. Деннета и Дж. Маккарти. Беглый взгляд на комплексный ТТ позволяет судить о том, что тема ЭЭ имманентна проблематике ИИ.

Оригинальный ТТ (тест Тьюринга). ЭЭ-концепция заключается в идее обучаемой машины.

Параноидальный ТТ (тест Колби). Тест служит эталоном для изучения экспертных систем, в которых напрочь отсутствуют ЭЭ-концепции. Такие системы реализуют концепции разработчиков как «навязчивые идеи» и не имеют внутренних механизмов рационального роста «знаний».

Гендерный ТТ (тест Геновой). Тест включает в себя многообразие ЭЭ-вариаций идей гендерного ТТ – от концептуального анализа половой детерминации интеллектуальных способностей до обоснования проекта бесполового постчеловечества.

Креативный ТТ (тест Лавлейс). Тест инициирует тематику компьютерного творчества и опровергает/защищает аргумент Лавлейс, согласно которому компьютер творить не может. Задаёт достаточно строгие аксиоматические критерии продуцирования творческих артефактов. ЭЭ-модели – необходимая составляющая креативных компьютеров, УЭ-модели принципиально не способны реализовать этот тест.

Субкогнитивный тест (тест Френча). В тесте явно звучит тезис: только тот компьютер способен пройти тест, в котором имеются механизмы, позволяющие ему «жить жизнью людей», и возможно прохождение всего пути эволюции человеческого интеллекта либо имитация этого процесса.

Глобальный ТТ (тест Блока). Тест обеспечивает наиболее демонстративное изучение УЭ-концепции. Он задаёт необходимое условие приписывания интеллекта х-системе, с учётом антибихевиористской критики оригинального ТТ. Раскрывается ограниченность лингвистического бихевиоризма тьюринговой игры в имитацию и функционализма машины Тьюринга. Новая формулировка ТТ (нео-ТТ) предполагает диспозиционные способности системы продуцировать осмысленную последовательность вербальных реакций на любую последовательность вербальных стимулов. Эти способности поддерживаются машиной Блока – глобальной экспертной системой, которую разрабатывает всё человечество. Внутреннее устройство машины убеждает, что её «интеллект» – это интеллект разработчиков, но никоим образом не её собственный. ЭЭ-идеи связываются с метафизикой психофункционализма.

Китайская комната (тест Серля). Тест отрицает концепцию сильного ИИ, согласно которой при прохождении ТТ компьютер не только моделирует человеческую способность к пониманию (это характерно для позиции слабого ИИ), но фактически «понимает» и объясняет её. Феномен «понимания» представлен в контексте интенционалистской теории сознания, которая обосновывает сильный антикогнитивистский тезис о невозможности «сильного» ИИ. ЭЭ-идеи инспирированы биологическими основаниями феномена интенциональности.

Инвертированный ТТ (тест Ватта). ЭЭ-идея состоит в изучении феномена воображения в вычислительном контексте и формирования рефлексивного

«зеркала» в процессе обращённости машины на себя (интеллектуальная машина, тестирующая собственный интеллект).

Тотальный ТТ (тест Харнада). Тест предполагает создание искусственных ЭЭ-механизмов, полностью неотличимых от естественных.

Тест Тьюринга (тест Швайзера). Это – самый тотальный тест, который дополняет предыдущий тест, погружая компьютерную систему – искусственную личность – в контекст эволюционного развития сообщества других таких же систем, по всем своим социокультурным проявлениям неотличимых от людей.

Имеется ряд других интересных версий и интерпретаций ТТ. Подчеркнем, что из всех частных тестов можно явно или неявно реконструировать сюжеты концептуальной драмы между ЭЭ-идеалом интеллектуальной системы и УЭ-возможностями современной компьютерной технологии её построения. При изучении ЭЭ-парадигмы в контексте КТТ возникают следующие вопросы: 1) дефинитный – как по вербальному поведению х-системы определить наличие в ней ЭЭ-приспособлений? 2) конструктивный – как реализовать ЭЭ-механизм и, собственно, возможно ли это? 3) критический – может ли интеллектуальная искусственная система самостоятельно развиваться? 4) мировоззренческий – каковы социокультурные последствия применения ЭЭ-систем?

ЭЭ-парадигма развития «знаний» представляется аподиктическим условием интеллектуальности компьютерной системы – так, по сути, считает большинство философов и теоретиков ИИ, начиная с А. Тьюринга и Дж. Маккарти. Поэтому положительное решение ЭЭ-вопросов служит поводом приписать х-системе интеллект. Если ЭЭ-вопросы не решены, то нечего рассуждать об ИИ.

В рамках данного параграфа ограничимся частичным решением дефинитного вопроса, в общих чертах рассмотрим конструктивный вопрос и оставим без внимания критические и мировоззренческие положения. Выдвинем гипотезу, апеллируя к заявлениям ряда видных исследователей ИИ – обнаружить в х-системе самостоятельные, свободные, творческие компетенции можно тогда, когда в ТТ включены вопросы, на которые ожидаются ответы с позиции *«здравого смысла»* (Common Sense). Само собой разумеется, что понятие «здоровый смысл» крайне расплывчато и всячески изгоняется из современного научного дискурса. Однако крайне частое употребление данного термина в методологии ИИ просто вынуждает к нему прибегнуть. Впервые данное понятие как критерий демаркации интеллектуальной компьютерной системы от неинтеллектуальной применяет основатель ИИ – Дж. Маккарти: *«Здравый смысл»*, *приписы-*

ваемый компьютеру, это способность выйти за рамки причинных связей, установленных разработчиками, и возможность «осознавать» собственные действия [McCarthy, 1995-1].

Имеются две трактовки теста Тьюринга с позиции «здравого смысла». Первая, широкая трактовка, превалярующая в дискуссиях в 1950 – 1975 гг., касается тезиса Тьюринга о возможности замены вопроса «Может ли машина мыслить?» на вопрос «Может ли машина играть в игру в имитацию интеллектуального поведения?». Если постановка такой игры бессмысленна, то нечего утверждать о выявлении посредством ТТ каких-либо ЭЭ-приспособлений в компьютерной системе. Вторая трактовка, более узкая, предполагает логическую эквивалентность ответов на данные вопросы и направлена на спецификацию теста, выявляющего «проблески» «здравого смысла» в *x*-системе. Сегодня, по всей видимости, в связи с впечатляющими успехами ИИ, термин «здравый смысл» используется лишь в контексте узкой трактовки, впервые предложенной в оригинальном ТТ.

Эволюционно-эпистемологические идеи в оригинальном тесте Тьюринга

В тьюринговой идее игры в имитацию интеллекта, особенно в части, касающейся обоснования инженерного проекта компьютера, чётко выявляются две отмеченные нами эпистемологические ветви: УЭ- и ЭЭ-парадигма. Первая инспирирована проектом универсальной цифровой вычислительной машины (УЦВМ), которая, напомним, является развитием идеи машины Тьюринга – концептуального инженерного проекта, включающего в себя детерминированное управляющее устройство, переходящее из одного состояния в другое, которое перемещается по ленте, считывая с неё и записывая на неё символы. Программа функционирования машины представлена в автоматной таблице. УЦВМ задаёт формальное определение квазиалгоритма благодаря механизму случайных переходов по автоматной таблице и наличию книги правил, на каждой странице которой (на ленте) представлен частный алгоритм. Также напомним, что, помимо стандартных алгоритмических свойств, УЦВМ обладает индетерминированностью, сингулярностью, нерезультативностью, непрерывностью.

А.Тьюринг совершенно справедливо посчитал, что мощности УЦВМ не достаточно для реализации механизма прироста знаний, и прорисовал контуры ЭЭ-парадигмы в идее построения обучаемых машин на базе нейрокомпьютера. Такое решение было принято против аргумента А. Лавлейс: «Аналитическая машина (универсальная ЦВМ – А.А.) не претендует ни на что оригинальное. Что бы машина ни сделала, нам известен способ, как это сделано» [Turing, 1950,

Р. 450]. То есть машины не способны к самостоятельному творчеству, так как не могут создать ничего нового, они способны только следовать программным инструкциям. Всё «творческое» предопределено разработчиком.

Таким образом, ЭЭ-проблематика в оригинальной версии ТТ инициирует общую проблему творчества в ИИ и требует специального освещения (см. подразд. 3.2). Отметим лишь основные пункты данной проблематики. А.М. Тьюринг предлагает два варианта ответа на возражение А. Лавлейс – «репрезентативный» и «коннекционистский» [Алексеев, 2011-8].

В рамках *репрезентативного опровержения* А.М. Тьюринг апеллирует к проекту УЦВМ, что приводит лишь к схоластическим изысканиями. Например, он утверждает: «Под солнцем ничего нового не происходит. Всё, что мы считаем оригинальным, на самом деле представляется лишь результатом обучения или следствиями общеизвестных принципов» [Turing, 1950, Р. 450]. Человеческая способность к творчеству совершенно непонятна. Удивление нетривиальностью результатов работы программы из-за сбоя машины сродни удивлению творческим гением. (Репрезентативная аргументация будет подробно рассмотрена в подразд. 3.2.)

Очевидна логическая некорректность аргументации, как правило, из-за подмены тезисов. Так же опровержение неубедительно из-за объективных факторов – символьными средствами УЦВМ трудно и, по всей видимости, невозможно адекватно имитировать креативные феномены. Логико-атомистический каркас репрезентативной парадигмы не обеспечивает свободы развития сущностей и связей моделируемой предметной области, ведь любая достаточно сложная система формализованных знаний продуцирует результаты в контексте известного (здесь срабатывает аргумент Гёделя). Комбинация экспертных знаний не ведёт к росту знаний. Доводы в пользу креативного лингвокомпьютинга, подкреплённые проектом УЦВМ, не убедительны.

Иное дело – *коннекционистское опровержение*. Предлагается концептуальный проект машины, снабженной нейросетевой программой обучения. Концепция нейрокомпьютинга коренным образом меняет исходную идею ТТ – от имитации интеллекта А. Тьюринг отказывается и настаивает на моделировании ЭЭ-механизмов развития человеческого интеллекта. Интерпретируя не совсем чёткие коннекционистские взгляды А. Тьюринга, предложим следующий вариант тьюринговой методики разработки обучаемой машины: 1) построение исходной системы знаний и интеллектуальных способностей, которыми обладает взрослый человек; 2) изучение способа роста знаний в ходе обучения; 3) изуче-

ние опыта, который не может быть результатом обучения; 4) очистка от содержания «знаний» и формирование структурно-процедурального скелета интеллектуальных способностей; 5) представление «чистых» знаний как модели интеллекта ребёнка – *машины-ребёнка*; 6) обучение машины-ребёнка и формирование *машины-взрослого*; 7) сравнение «знаний» и «способностей» машины-взрослого со знаниями и способностями взрослого человека. Важно то, что учитель, как правило, не осведомлён о внутренних процессах машины-взрослого. Тем не менее, в некоторой степени он может предсказывать поведение «ученика». Именно в этом А. Тьюринг усматривает опровержение аргумента Лавлейс, одна из версий которого: «машина может делать только то, что мы знаем, как приказывать ей это сделать» [Turing, 1969]. Возможно, это так, однако в случае с обучаемой машиной мы не «знаем» наверняка о её поведении, но лишь прогнозируем его и проверяем гипотезы.

А. Тьюринг полагает, что машину-ребёнка создать достаточно просто. «Интеллект ребёнка подобен чистому блокноту, купленному в магазине канцелярских товаров». Поэтому достаточно разработать механизм развития «знаний» машины-ребёнка и иметь большое количество «чистых листов» (для полного представления предметной области). Конечно, это неверно. Сегодня стало общим положением утверждение о том, что сложность моделирования процесса развития знаний ребёнка на несколько порядков выше сложности модели мышления взрослого человека. А. Тьюринг в данном вопросе оптимист и предлагает несколько интересных методов обучения машины-ребёнка, например, «метод кнута и пряника» [Ibid, P. 17 – 23].

Машина-ребёнок реализуется нейронной сетью. По поводу устройства машины-взрослого остаётся только догадываться. Собственно, и концепция нейронной сети у А. Тьюринга не убедительна. Если её более чётко рассмотреть, то никакого нейрокомпьютера мы не увидим, но лишь параллельную УЦВМ. Чистые листы в блокноте машины-ребёнка (книга правил УЦВМ), точнее, стопка перфокарт (пускай это будут «слои» нейронной сети) и механизм параллельного доступа к ним (присваивающий «веса» «нейронам» из каждого «слоя») претендуют лишь на формальное определение параллельного компьютеринга, но никоим образом не нейрокомпьютеринга. Несмотря на неясность конструкции, А. Тьюринг был убеждён: мыслящую машину следует строить как человеческую модель обучения [Ibid, P.14]. Именно к обучаемым компьютерам относится заключительный и часто цитируемый вывод А. Тьюринга относительно роли и места интеллектуальных машин в обществе как соперников и помощников человеку во всех сферах деятельности.

Несомненно, коннекционистское опровержение аргумента Лавлейс намного убедительнее репрезентативного, особенно если аргументировать его более правдоподобным проектом нейросетевой машины, фундаментальные принципы функционирования которой не стоит связывать с линией «Бэббидж – Тьюринг», но с внедрением машины Корсакова в концептуальный аппарат перспективного компьютеринга (см. Заключение). Пусть машина Тьюринга останется главной для лингвокомпьютеринговых принципов функционирования интеллектуальных машин и для УЭ-возможностей роста их знаний. Эти возможности весьма ограничены, что выявляется при изучении теста Мура, в котором ТТ специально предназначался для развития компьютерных «знаний» о компьютерных «знаниях».

Тест Мура: ущербность эволюционно-эпистемологической парадигмы традиционного искусственного интеллекта

Дж. Мур в статье «Анализ теста Тьюринга» (1976 г.) в тьюринговой игре в имитацию усматривает в большей мере практический, нежели теоретический ориентир. Главный тезис: «Тест Тьюринга только тогда значим для компьютерной технологии, когда он проинтерпретирован в индуктивном формате» [Moog, 1976]. Тест Тьюринга – это базис индуктивного доказательства гипотезы в пользу компьютерного мышления. Надо начать работу по систематическому сбору и обобщению фактов прохождения частных аспектов ТТ. Как в процессе автоматического дедуктивного доказательства истинности/ложности теорем получается «побочный продукт» – значение терминов рассуждения, так в ходе индуктивного доказательства компьютерная система не только утверждает ту или иную интеллектуальную способность, но и самостоятельно расширяет собственные «знания». Дж. Мур приводит ЭЭ-соображения в стиле бадейной теории обучения: 1) ТТ идентифицирует, систематизирует и обобщает факты реализации интеллектуальных функций у машин; 2) ТТ поощряет жёсткий отбор и отбраковывает непригодные факты; 3) положительный и отрицательный опыт сохраняется в памяти компьютерной системы, которая становится всё более «интеллектуальной» и требовательной к новым прецедентам прохождения ТТ. К сожалению, Дж. Мур не предложил конструктивного проекта системы, способной к самособиранию фактов интеллектуальности и самоизменяющейся в соответствии с приростом «знаний». В частности, это объясняется теоретической неразрешимостью проблем индукции и фреймов. Например, что общего, между игрой в шахматы и принятием решения? Очевидно, одно другому спо-

собствует, но как? Поэтому тест Мура следует считать лишь заявкой на ЭЭ-проект компьютерной системы.

Тест Колби: универсально-эпистемологическая парадигма ИИ и параноидальные системы

Мощным критическим зарядом относительно УЭ-парадигмы развития «знаний» экспертных систем обладает параноидальный ТТ, предложенный в исследованиях К. Колби и его коллег. В статье 1971 г. «Искусственная паранойя» [Colby, Hilf, Weber, 1971] описывается компьютерная программа PARRY, имитирующая поведение пациентов психиатрической больницы. Тест Колби – это не чистый мысленный пример, но реальная инженерная конструкция, которая наряду с программой ELIZA является одним из первых примеров удачной практической реализации ТТ. Результаты тестирования ошеломили многих: почти в половине случаев психиатры не могли отличить компьютер от больных. Ответы сопровождались словосочетаниями, окрашивающими речь эмоциональными состояниями страха, гнева и недоверия, что во многом способствовало имитации параноидального бреда.

В следующей статье «Роль неразличимости в тьюринго-подобных тестах для компьютерного моделирования параноидальных процессов» [Colby, Hilf, Weber, Kraemer, 1972] исследователи акцентировали на теоретических сравнениях своего теста и теста Тьюринга. И оригинальный и параноидальный тест – это инструмент, обосновывающий правомочность имитации различных форм человеческой деятельности, будь она рациональна или нет. Например, параноидальный тест легко пройдут те, кто оспаривает ТТ, доказывая, что машина не может мыслить, при этом замыкаясь, как параноики, на некотором частном определении «сущности» интеллекта.

Моделирование паранойи оказалась дорогим занятием: PARRY – очень «увесистая» программа, словари включали 4 500 слов, справочники – 700 статей. Программа синтаксического анализатора вопросов и грамматического синтезатора ответов превышала 200 000 инструкций. «Зачем идти на такие затраты и моделировать параноидальное поведение? Какова была цель теоретических и практических изысканий К. Колби? Не прошло ли всё это впустую и не выкинуты ли деньги на ветер?» – задаёт вопросы Д. Деннетт [Dennett, 1984] и предлагает очень интересное сравнение параноидальной программы PARRY с экспертными системами, стоимость которых на порядки выше, но которые способны пройти лишь несколько усовершенствованный параноидальный тест – из этого теста изъяты выражения параноидальных эмоций.

Тест Деннетта: оперативное тестирование эволюционно- эпистемологических механизмов экспертных систем

В работе «Эра интеллектуальных машин. Могут ли машины мыслить?» [Ibid] Д. Деннетт с позиции идей теста Колби анализирует экспертные системы, которые стали сегодня классическими прототипами и изучаются во всевозможных вузовских курсах искусственного интеллекта. PARRY – это компьютерная модель теории психологического феномена – паранойи – и на практике не применима. Тем не менее, экспертные системы, возникшие в «инженерии знаний» и пользующиеся большим практическим спросом, по способу прохождения ТТ аналогичны тесту на паранойю. Несомненно, подчёркивает Д. Деннетт, экспертные системы – это превосходные консультанты-специалисты, которые способны поставить медицинский диагноз, изучить геологические данные, проанализировать результаты научных экспериментов, отреагировать на новости. Например, интерес представляет новостная экспертная система CYRUS (САЙРУС), которая моделировала «знания» Сайруса Вэнса (Cyrus Vance) – секретаря в администрации американского президента Картера. Цель проекта CYRUS была амбициозной – стать «чистой системой искусственного интеллекта», искусственным механизмом формирования и развития «знаний», аналогичного человеческому когнитивному механизму формирования нарративного сознания, т.е. сознания личности, основанного на биографических воспоминаниях. CYRUS обновлялась каждый день из блоков новостей американских информационных агентств. Как только в новостях появлялось ключевое слово «Cyrus Vance», то ассоциированная с ним информация оперативно записывалась в базу данных CYRUSa. По мере накопления данных возрастала способность CYRUSa отвечать на всё большее и большее количество вопросов. К CYRUS обращались во втором лице, как при личном разговоре с человеком Сайрусом Вэнсом.

Пример диалога: «Деннетт (Д): В прошлый раз, когда вы ездили в Саудовскую Аравию, где вы останавливались? Сайрус (С): Во дворце в Саудовской Аравии 23 сентября 1978 г. (Д): Вы ходили смотреть там достопримечательности? (С): Да, на месторождение нефти в Хоране 23 сентября 1978 г. (Д): Встречали ли Вы мистера Бегина? (С): Да, последний раз на торжественном обеде в Израиле в январе 1980 г.»

Кажется, что система CYRUS способна правильно отвечать на все вопросы, которые только можно придумать и корректно сформулировать. Однако её легко сбить, если задать глупый вопрос.

Пример теста на «здравый смысл». Д. Деннетт задаёт CYRUS вопрос: «Вы когда-нибудь встречали женщину, руководящую государством?»,

проверяя, «знает» ли CYRUS, что Индира Ганди и Маргарет Тэтчер были женщинами. CYRUS не может ответить ни да, ни нет. Экспертная система оказалась в тупике. Но ведь вопрос был достаточно простой, неужели в экспертной системе не предусмотрены подобные вопросы, которые требуют примитивных рассуждений? Причина несуразицы очевидна. Предлагается взглянуть во «внутренность» системы CYRUS. Обнаруживаются «скелетные» описания тысяч слов, минимальные настолько, насколько решили когнитологи (Д. Деннетт под «скелетными описаниями» понимает широко известные среди специалистов контекстно-зависимые скрипты Р. Шенка [Шэнк, 1980]). Например, «адвокат» описывается как синоним слова «юрист». Однако система «не знает», что «адвокаты» – это взрослые люди; в сфере закона они выполняют различные функции; являются выпускниками университетов; платят им больше, чем горничным; знают, как завязывать шнурки; вряд ли будут искать компанию среди дровосеков и т.д. Это – банальные, но очевидные факты об адвокатах. В CYRUS эти знания отсутствуют. Для представления «здорового смысла» требуется слишком много понятий, характеризующих жизненный опыт человека, и, как только встречается серьёзное отклонение от вложенных в систему правил, – таких отклонений бесконечно много, система перестает работать и требует доработки для ответа на элементарные вопросы с позиции «здорового смысла».

Вердикт Д. Деннетта уничижителен – современные классические интеллектуальные системы способны пройти лишь параноидальный тест Тьюринга. Все они – «деревни Потёмкина»: у них интересная внешняя оболочка – привлекательный интерфейс, как фильм в кинотеатре, однако внутри ничего нет. Экспертные системы, подобные CYRUS, интересны лишь в теоретическом плане – как *не* надо строить интеллектуальные системы.

Тест Маккарти: долгосрочное тестирование эволюционно-эпистемологических механизмов экспертных систем

Идею компьютерных программ, имитирующих «здоровый смысл», Дж. Маккарти прорабатывает более полувека, начиная с 1958 г., когда впервые обнаружил её при формализации контекста логических рассуждений [McCarthy, 1969]. Дж. Маккарти солидарен с Д. Деннеттом в том, что показатели функционирования экспертных систем в отдельных областях знаний впечатляют. Однако не следует такие системы относить к «интеллектуальным». Это – обычные программы, и уровень их «интеллекта» не повышается за счет применения высокоуровневых языков программирования или использования нечётких моделей. В основе теста Маккарти лежит работа «Экспертные системы нуждаются в здравом смысле» [McCarthy, 1984], опубликованная почти одновременно с представленной выше работой Д. Деннетта. Тест отличается от деннеттовской «быстрой проверки»: длина теста может быть значительной, анализ – скрупулёзным. Если для Д. Деннетта «здоровый смысл» характеризует банальные, по-

вседневные вещи и положения, то для Дж. Маккарти это – показатель богатого профессионального опыта и глубоких теоретических знаний специалиста.

Предметом критики послужила экспертная система MYCIN. Программа консультирует врачей по вопросам лечения бактериальных инфекций крови и менингита. Как правило, MYCIN показывает гораздо более высокие результаты по сравнению с заключениями студентов-медиков, интернов и даже специалистов по бактериальным заболеваниям. Работает достаточно надежно и, на первый взгляд, для её эффективного функционирования излишни изыскания по поводу компьютерной реализации понятия «здравый смысл». Это не так. Её широкое применение застопорилось из-за следующих причин: 1) требуется регулярное пополнение базы «знаний» не только наличными, но и неизвестными данными; 2) система не способна самостоятельно определять собственные ограничения.

Пример теста на «здравый смысл». В систему вводят данные о том, что у такого-то пациента обнаружено поражение кишечника холерным вибрионом. MYCIN бодро выдаёт рекомендации о проведении двухнедельного курса лечения тетрациклином. Если слепо довериться рекомендациям, то бактерии будут уничтожены, однако вероятнее, вначале пациент скончается от холеры. Летального исхода можно избежать, если врач-пользователь MYCIN знает, что данный случай диареи требует интенсивного лечения, и предпримет дополнительные меры, не учтённые в экспертной системе, для выяснения соответствующих методов и препаратов. Но это будет «здравый смысл» врача-человека, обладающего большим опытом лечения подобного рода болезней. Отсюда тезис: «В этой экспертной системе нет ни толику здравого смысла».

Для прохождения теста следует ответить на большое количество разнородных вопросов, сводимых к двум классам: 1) «что» такое «здравый смысл»?; 2) «как» моделировать рассуждения на основе имеющихся знаний о «здоровом смысле»? Ответы на эти вопросы сводятся к следующему: 1) «здравый смысл» тесно связан с конкретикой реальных событий и действий, поэтому система должна уметь самостоятельно прогнозировать последствия собственных действий; 2) «здравый смысл» – это рефлексия над конкретными действиями; 3) для разумного поведения в сложных условиях «здорового смысла» недостаточно, требуется последующая рефлексия, умение представлять и использовать *знания о знаниях*; 4) необходимо учитывать информацию о личностях – их знаниях, убеждениях, целях, симпатиях и антипатиях, намерениях и возможностях и пр.; 5) «здравый смысл» опирается на способы формализации «наивной» физики,

которые объёмнее научной физики. Например, всем известно что произойдёт, если бросить стакан воды с метровой высоты на пол: стакан разобьётся; на это уйдут доли секунды; брызги воды не разлетятся дальше нескольких метров и т.п. Наивные знания не включают формулы падения тела и уравнения течения жидкости, исходных данных недостаточно, и поэтому на решение уйдет много времени. Если «знания» робота – это теории механики, то он не успеет сойти с дороги, когда навстречу ему мчится автомобиль; 6) «здоровый смысл» составляют естественные понятия (виды), которые формируются в обыденной практике. Например, понятие «шум» – это и лай собаки и падение капли воды. Проблема представления подобных фактов *универсальным способом* не разрешима; 7) имеется также «проблема фреймов» – как из частных спецификаций получить целостное описание предметной области.

Несмотря на принципиальные теоретические трудности моделирования «знаний» о «здоровом смысле», Дж. Маккарти предлагает ряд частных конструктивных решений, например, логику формализации контекста («смысл» как контекст «значений» языковых выражений) [McCarthy, 1982], язык программирования экспертных систем, основанный на теории речевых актов [McCarthy, 1986], и пр.

Тест Деннетта и тест Маккарти дополняют друг друга в смысле разных трактовок понятия «здоровый смысл». Но как бы ни различались подходы, напрашивается такой вывод – экспертная система, реализующая тест на здравый смысл, должна быть искусственной личностью и жить жизнью людей. Конечно, это – фантастика дурного вкуса, особенно если учесть отсутствие каких-либо достижений в программировании ЭЭ-механизмов ИИ-систем.

Эволюционно-эпистемологические механизмы динамических экспертных систем реального времени

В современной методологии экспертных систем наблюдается более угрожающее положение в области развития «знаний», нежели чем во время классических систем, на которые обратили внимание Д. Деннетт и Дж. Маккарти. Начиная с середины 1990-х годов, в сложных информационных технологиях стали доминировать *динамические экспертные системы реального времени* – G2, RT Works, TDC Expert и др. «Именно они позволили искусственному интеллекту перейти от игр и головоломок к массовому использованию при решении практически значимых задач» – считает Э.В. Попов, видный отечественный исследователь ИИ [Попов, 1995]. Система дорогая, и для некоторых предметных областей ее цена достигает сотен миллионов долларов. Такая стоимость оправ-

дывается не столько ноу-хау программ, сколько армией разработчиков, включённых в технологию её непрерывного сопровождения и интерактивно связанных с пользователями. Система способна быстро реагировать на динамику «знаний», однако всякий раз, когда изменение форм и способов представления «знаний» превышает учтённые границы, требуется вмешательство программиста. Имеется много технических несуразиц, которые следует критически изучить. Например, странно то, что запрещается пользоваться классическими языками представления «знаний», такими, как LISP, первый вариант которого разработал Дж. Маккарти более полувека назад.

Версия абсолютной динамической экспертной системы в точности воспроизводит концепцию машины Блока – в ней максимально задействован потенциал УЭ-парадигмы развития «знаний». Совершенная система объединяет разработчиков всех стран в рамках глобального виртуального интернет-сообщества, адресно доставлять им запросы пользователей, на которые они в режиме реального времени способны мгновенно ответить, задействовав всю мыслимую мощь информационно-коммуникационных технологий. Несомненно, система проходит тест Тьюринга, в том числе и тест Деннета и тест Маккарти – кто-то из человечества обязательно найдется, чтобы ответить на вопросы, заданные с позиции «здравого смысла».

Но является ли она интеллектуальной искусственной системой? Совершенно очевидно – нет. Здесь нет «ни грамма» искусственного интеллекта, но лишь интеллект разработчиков. Более того, так как знания разработчика фундаментированы некоторым личностным смыслом, частным для общечеловеческого формата, то абсолютный вариант динамической экспертной системы – это *глобальная параноидальная система*. Кстати, этому прототипу соответствует Интернет, так как стратегия развития поисковых фирм типа Google и Yandex – интеллектуализация веб-навигации.

Крайне пренебрежительно сторонники динамических систем отзываются по поводу *статических экспертных систем*, названных так из-за отсутствия связи с разработчиком. На мой взгляд, в искусственном интеллекте всё должно быть с точностью наоборот. ИИ-система призвана работать автономно. Независимо от разработчика она должна извлекать, приобретать, репрезентировать, хранить, обрабатывать, интерпретировать, выдавать, передавать и – самое главное – самостоятельно развивать «знания».

Причина мощного коммерческого успеха «динамических» систем видится в обычном PR-трюке – на брэнд интеллектуальности «кляют» заказчики и

удобно маскировать машинную «глупость» такими, как «экспертные системы реального времени – основное направление искусственного интеллекта!» [Попов, 1995]. Конечно, в этих системах есть многое: открытость, масштабируемость, объектно-ориентированное программирование, CASE-технологии, клиент-серверные технологии, рассуждения на базе прецедентов и на базе имитационных моделей, параллельное выполнение в реальном времени, когнитивная графика, технологии «мягких» вычислений, структурированный естественный язык представления «знаний», модели внешней и внутренней среды и многое другое. Тем не менее, мы имеем полное право утверждать, перефразировав Дж. Маккарти:

В экспертных динамических системах реального времени нет ни толики искусственного интеллекта.

Искусственный интеллект вне эволюционно-эпистемологической парадигмы

Чтобы хотя бы в заключение данного материала прозвучали не только критические, разрушающие глобальную индустрию «интеллектуальных динамических экспертных систем», но и конструктивные идеи, вернёмся к проекту компьютера, способного реализовать комплексный Тест Тьюринга. Для данного проекта УЦВМ оказалась слабой, а нейронная сеть Тьюринга – всего лишь усовершенствованной параллельной УЦВМ. Предлагаем обратить внимание на крайне перспективный, хотя более старый проект, предложенный за 120 лет до статьи Тьюринга, в 1832 г., российским инженером С.Н. Корсаковым. Машина Корсакова, по сути, является протонейрокомпьютером, и собственно с неё предлагается отсчитывать идею нейрокомпьютинга в мировом масштабе [Алексеев, 2011-7]. Метафора такова: стопка перфокарт прокалывается иглами, и глубина проникновения игл характеризует связь между ассоциируемыми понятиями, обозначенными на перфокартах, т.е. «вес коннекции» задаётся усилием прокалывания, но не символом, который считывается с перфокарты или записывается для варианта параллельной УЦВМ. Имеется ряд интересных исследований машины Корсакова, которые меняют фундаментальные принципы современного компьютеринга посредством комплексирования машины Тьюринга и машины Корсакова [Алексеев, 2011-3; 2012-1]. Например, формальное тьюринговое определение алгоритма работы математика дополняется «алгоритмом» деятельности художника, где интенсивность прокалывания иглами перфокарт задаётся эмоциональным восприятием предметной области, которая репрезен-

тирована «явными знаниями» на текущей перфокарте, а также «неявными» и «фоновыми» «знаниями», обозначенными на других, ниже и выше лежащих перфокартах. Несомненно, это требует специального доказательства и формальной экспликации. Но даже с первого взгляда видно, что лингво-нейромашина Тьюринга-Корсакова меняет теоретико-алгоритмические основания компьютеринга, включая принципы реализации эволюционно-эпистемологических теорий.

Критический результат данной статьи следующий. Если на заре развития ИИ эволюционно-эпистемологическая парадигма развития «знаний» вызывала острый экспериментальный интерес, то сегодня страсти поутихли под массивным напором универсально-эпистемологических технологий. Несомненна сиюминутная результативность УЭ-парадигмы, поэтому финансовые вложения в неё огромны. Однако в последующем УЭ-парадигма заводит в тупик, и исследования ИИ сводятся на нет. Фундаментальная эпистемологическая задача ИИ – создание механизмов саморазвития «знаний» компьютерных систем – остаётся в стороне от серьезных крупных междисциплинарных исследований. Непосредственной отдачи и рыночной выгоды от её решения ждать не стоит. Поэтому внедрение идей ЭЭ в ИИ и, собственно, развитие истинного, настоящего, неинновационного ИИ осуществляется спорадически энтузиастами. Хотя, как мы показали, достаточно простых рассуждений на тему теста Тьюринга, чтобы разрушить мнимый образ «интеллектуальности» динамических экспертных систем и показать, что сторонники УЭ-парадигмы занимаются чем-то другим, но не искусственным интеллектом. Так как комплексный ТТ прорисовывает фундаментальные, концептуальные, мыслимые очертания ИИ, то вполне закономерен вывод:

ИИ без ЭЭ не мыслим.

3.2.2. Критика методологии: комплексный тест Тьюринга против комплексного проекта искусственной личности

В последние пятнадцать лет в философии искусственного интеллекта обозначилась область междисциплинарных исследований, получившая название «Проект искусственной личности». Возникновение проекта во многом объяснимо факторами общественной жизни – сегодня сложные интеллектуальные информационно-технические системы невозможно рассматривать с сугубо технической точки зрения, они приобретают социокультурный, человекомерный статус. Используемые в настоящее время и перспективные информационные

технологии рассматриваются в контексте социальных ценностей, мировоззренческих ориентиров, морально-правовых норм, эстетических канонов и иных составляющих духовной сферы. Здесь просматривается требование, описанное в отечественной философской науке как необходимость постнеклассической переориентации методологии изучения, построения и развития сложных систем [Степин, 1991]. Однако в большей мере возникновение проекта искусственной личности обусловлено внутриметодологическими факторами, которые отражают достижения искусственного интеллекта. От моделирования чисто интеллектуальной деятельности осуществлён системологически закономерный переход к моделированию сознания. Когнитивно-компьютерные модели сознания стали рассматриваться в качестве концептуального и реализационного базиса «персонализации» интеллектуальных информационных систем [Алексеев, 1998].

Дефиниция искусственной личности

Искусственная личность (ИЛ) – это когнитивно-компьютерная система, удовлетворяющая, по-крайней мере, одному из двух критериев: 1) она обладает подобием человеческой субъективной реальности – так называемым «квазисознанием» (в ряде работ употребляется термин «псевдосознание»); 2) она функционально, поведенчески и/или физически *неотличима* от человеческой личности.

Первый критерий ИЛ поддерживается, в основном, сторонниками сильного ИИ. Они считают, что для реализации сложной функциональной (само)организации компьютерной системы и осуществления ею адаптивного поведения в условиях динамики внешней и внутренней среды необходимо выделение в архитектуре системы специального блока «квазисознания», системологическая роль которого подобна роли сознания в жизни сознательного существа. Этот блок должен осуществлять проективные, интроспективные, квалиативные и прочие отношения с внешней и внутренней средой системы. Критерий правдоподобен в контексте парадигмы компьютеризации. Эта разновидность функционализма по-иному называется «функционализмом машины Тьюринга», «функционализмом искусственного интеллекта», «вычислительной теорией сознания». Правдоподобие критерия, однако, шатается из-за многочисленных шагов редукции: так, персональное редуцируется к ментальному, ментальное – к интеллектуальному, интеллектуальное – к структурно-функциональной архитектуре компьютера. Формальные зависимости между когнитивными и компьютерными компонентами задаются посредством вычислительных элементов, операций и функций, а каузальные зависимости находят наиболее общее выра-

жение в реализационных аспектах машины Тьюринга. Вычисление понимается расширенно – репрезентативный и коннекционистский подходы объединяются. Полученные при этом интегральные репрезентативно-коннекционистские кодовые структуры операционализируются средствами квазиалгоритмической обработки. В результате многошаговой редукции получаем компьютеризованную формулировку проекта ИЛ: *личностное (персональное) – это квазиалгоритмическое вычисление.*

Во втором критерии ИЛ показателем «неотличимости» естественной и искусственной личностей выступает способность системы пройти *полный тест Тьюринга* (ТТ) [Алексеев, 2006-1]. В этом тесте, помимо вербально-коммуникативной, перцептивно-моторной, анатомо-физиологической и даже (в некоторых разновидностях ТТ) микрофизической неотличимости искусственной и естественной систем, всесторонне учитываются поддающиеся внешнему наблюдению показатели духовной сферы деятельности личности – персонологические параметры. К ним относятся «смысл», «свобода», «любовь», «ответственность», «право», «творчество», «красота» и др. Правдоподобие второго критерия также, как и первого, проблематично. Здесь особенно не ясна связь между внутренним и внешним миром человека. С такой проблемой почти век назад столкнулся бихевиоризм. Несмотря на очевидные заслуги в устранении метафизики ментальных сущностей широко принимается обстоятельная критика бихевиорального метода оценки сознания исходя из наблюдений за действительным поведением и из предположений возможного поведения.

В таких условиях – условиях проблематичности критериев оценки интеллектуальных систем как систем искусственной личности – развиваются современные проекты.

Многообразие проектов искусственной личности

Можно показать, что история философии искусственного интеллекта – это история обсуждения возможностей компьютерной реализации персонологических параметров. Достаточно вспомнить полемический стандарт А. Тьюринга по поводу построения мыслящих машин – совокупность аргументов и контраргументов в решении «основного вопроса» философии искусственного интеллекта: «Может ли машина мыслить?». Эту совокупность мы назвали «стандартом» по причине придания формы и отчасти содержания современным дискуссиям в области философии искусственного интеллекта. В «стандарт» вошли аргументы и контраргументы явно не инженерно-технического характера: теологический, антисциентистский, креационистский, «от первого лица»,

«от другого сознания» и даже экстрасенсорный. Целевое – персонологическое – назначение проектов вычислительных машин «идёт от века». Достаточно вспомнить о машине Р. Луллия (1235 – 1315). Здесь мы видим не только обсуждение вопросов логики оперирования понятиями, но и религиозно-моральное назначение проекта «Арс магна» [Реале Антисери, 1997, С. 206 – 207]. Так же и для Лейбница при построении вычислительной машины второстепенными были вопросы логико-математического плана, а первостепенными – юридическо-правовые вопросы.

С начала 1990-х годов выделяются следующие крупные проекты ИЛ [Алексеев, 2005-10]: 1) OSCAR Дж. Поллока, сформулированный в рамках «универсальной теории рациональности» и её приложениями для построения искусственных рациональных агентов («артилектов») [Серёдкина, 2007]; 2) проект «человекоподобных агентов» А. Сломана, призванный реализовать широкий спектр персонологических параметров, например «любовь», «свобода» [Sloman, 1978; Sloman, Croucher, 1981-2; Sloman, 1996; 1998; 2000]; 3) проект гуманоидных роботов КОГ, в котором Д. Деннетт усматривает апробацию собственной теории множественных набросков, где персональное возникает из сложного сочетания бесчисленных серий нарративов, а личность и социум – это субстанциональные системы бесчисленных роботов, в которых «ментальное» представляется компонентой функциональной самоорганизации [Деннетт, 2004]. В этих проектах предлагаются концептуальные, логико-математические, программные решения. Нам интересен методологический уровень осмысления путей реализации проектов ИЛ.

В связи с этим и в соответствии с введёнными критериями ИЛ показательными для проектов ИМ представляются два подхода [Алексеев, 2008-4; 5]: 1) Д. Лената, автора широко известной программы «Автоматический математик» (1976 г.), который считает, что следует подражать психологическим, социокультурным, лингвистическим, интеллектуальным и другим особенностям личности, а физическое подобие системы – второстепенный, несущественный фактор; 2) подход Р. Брукса, ученика Д. Лената, автора вышеупомянутого робота КОГ и знаменитой версии Kismet, имитирующей мимику человеческого лица. Р. Брукс полагает, что физическая антропоморфность робота – первичное и необходимое качество его персонологического подобия, ведь все социокультурные понятия, на основании которых будет функционировать робот и которые он будет вырабатывать в ходе адаптации к внешней и внутренней среде функционирования, предопределены антропологическими особенностями че-

ловека. Например, согласно принципу Джонсона и Лакоффа, «безногому» роботу трудно будет усвоить выражение «Перевернуть с головы на ноги».

Подход Д. Лената и Р. Брукса можно обозначить соответственно как *экспертный* и *робототехнический* подходы к построению ИЛ.

В англо-американской философии ИИ в основном преобладает робототехнический подход, согласно которому ИЛ – это обладающий квазисознанием автономный робот. Экспертная методология не признаётся такими авторитетными зарубежными философами ИИ, как Дж. Маккарти А. Сломан. В отечественной же науке, напротив, сложился экспертный подход: ИЛ рассматривается в социально-эпистемологическом контексте междисциплинарных взаимодействий специалистов, в ходе которых формируются данные и знания социокультурного содержания. Рассмотрим более подробно эти подходы.

Искусственная личность – робототехническая система

Искусственная личность – это робот, наделённый квазисознанием – т.е. совокупностью, по крайней мере, некоторых персонологических способностей и качеств человеческой личности. Здесь показательна дискуссия по поводу работы С. Брингсйорда – «Чем могут и не могут быть роботы» (1992, 1994 гг.) [Bringsjord, 1994], продолженная на страницах сайта <http://psycprints.ecs.soton.ac.uk> (рубрика «Сознание робота»). В этой работе позитивные утверждения относительно проекта ИЛ сопровождаются обстоятельной критикой. Основной девиз работы звучит так: «В будущем робот будет делать всё то, что делаем мы, но не будет одним из нас» – т.е. не будет сознательным.

С. Брингсйорд доказывает следующее: а) ИИ будет производить машины, обладающие способностями проходить всё более и более сильные версии ТТ; б) проект ИЛ по созданию машины–личности будет неминуемо проваливаться. В защиту (а) предлагаются компьютерные программы генерации художественных и философских текстов и индуктивный вывод при построении версий ТТ: тьюринговая игра в имитацию → наблюдение внешнего вида игроков → изучение их сенсомоторного поведения → сканирование мозга и пр. Автор уверен, что ИИ будет мало-помалу проходить данную последовательность. В основе доказательства (б) – несостоятельности проекта ИЛ – лежит *modus tollens* из суждений:

- (1) «Проект ИЛ» → «Личность – это автомат»;
- (2) – «Личность – это автомат»; ⇒

(3) – «Проект ИЛ».

Посылка (2) интуитивно апеллирует к наличию у человека и отсутствию у автомата ряда машинно невоспроизводимых персонологических параметров F, к которым относится свободная воля, способность к интроспекции, внутренний проективный опыт «каково быть» («what it's like to be») и пр. К F относятся и способность к компьютерному воспроизводству творческих способностей. Эту способность автор подробно анализирует в ряде предыдущих работ (в тесте Лавлейс (см. подразд. 2.2) и делает вывод «Компьютер творить не может!» (ранее в тесте Лавлейс отмечался подход С. Брингсйорда). Общее заключение:

(4) «Личность обладает F»;

(5) «Автомат не обладает F»; \Rightarrow

(6) «Личность не может быть автоматом».

С. Брингсйорд достаточно подробно обосновывает невычислимость F, апеллируя к аргументу Гёделя, «Китайской комнате» Дж. Серля, аргументу произвольной множественной реализации Н. Блока и др. Итог таков: *«Роботы будут многое делать, однако они не будут личностями»*.

В полемике по поводу работы С. Брингсйорда наблюдается четыре направления критики и аргументации относительно реализации проекта искусственной личности¹:

1. Проект ИЛ абсурден, так как: 1) метафизические понятия «личность», «свобода воли», «интроспекция» не могут быть предметом эмпирического анализа; 2) нечёткое определение понятия личности, образная иррациональность проекта ИЛ – всё это не заслуживают научного внимания; 3) персонологические параметры логически невыразимы – попытки их логической экспликации непременно сопровождаются техническими ошибками.

2. Проект ИЛ неточен, так как: 1) апелляция к интуитивным аргументам

¹ См. дискуссию на сайте <http://psycprints.ecs.soton.ac.uk/archive/00000418/> следующих авторов: Barresi, John (1995) Building Persons: Some Rules for the Game; Bringsjord, Selmer (1995) Agnosticism About Neuron-Level Functionalism; Bringsjord, Selmer (1995) Agnosticism Re-Visited; Bringsjord, Selmer (1995) Are Computers Automata?; Bringsjord, Selmer (1995) Computationalism is Doomed, and We Can Come to Know it; Bringsjord, Selmer (1995) Why Didn't Evolution Produce Turing Test-Passing Zombies?; Bringsjord, Selmer (1996) Artificial Intelligence and the Cyberiad Test; Brown, Marina and O'Rourke, Joseph (1994) Agnosticism About the Arbitrary Realization Argument; Costa, Luciano da Fontoura (1995) Is There More to Personhood Than Computing Theory and Logico-Mathematics?; Hobbs, Jesse (1995) Creating Computer Persons: More Likely Irrational Than Impossible; Korb, Kevin B. (1995) Persons and Things: Massing, Walter (1995) Metaphysical Windmills in Robotland; Mulhauser, Gregory R. (1995) What Philosophical Rigour Can and Can't be; O'Rourke, Marina Brown and Joseph (1995) Agnosticism Revisited; Rickert, Neil W. (1995) A Computer is Not an Automaton; Scholl, Brian (1994) Intuitions, Agnosticism, and Conscious Robots и др.

разрушает логическую строгость дедуктивных аргументов; 2) нецелесообразно применять дедуктивные умозаключения в качестве метода аргументации (дедукция – не метод, а схема суждений); 3) необходимо чётко формулировать конкретную разновидность функционализма, который лежит в основании проекта ИЛ – нет функционализма «вообще», а есть, к примеру, низкоуровневый функционализм, биологический функционализм и пр.

3. Проект ИЛ *реализуем*, так как: 1) человеческая личность – это всё же автомат, но неклассического типа, т.е. представимый не машиной Тьюринга, а иными, квазиалгоритмическими правилами обработки вычислительных элементов; 2) компьютер вовсе не автомат, и кто считает противное – тому не место в проблематике компьютерного сознания; 3) компьютерная система должна тестироваться природой, а не человеком.

4. Проект ИЛ *значим* вне зависимости от возможностей его реализации: 1) будет ли создана ИЛ либо не будет создана – не принципиально, главное, что проекты ИЛ выявляют механистические аспекты человеческого сознания; 2) проектирование ИЛ проясняет роль и функции сознания в жизни человека путём постановки вопроса о зомби – о созданиях, не обладающих сознанием, но имеющих все поведенческие способности человека.

Искусственная личность – экспертная система

Примерно в те же годы, в середине 1990-х годов, в нашей стране отмечался всплеск интереса к проблеме искусственной личности. По заказу Министерства обороны РФ (27 ЦНИИ МО, руководитель – проф. В.В. Деев) были организованы междисциплинарные исследования по внедрению нетрадиционных информационных технологий в системы принятия решений. Автор как программист был в эпицентре этой интереснейшей исследовательской работы. Проект базировался на мероприятиях по моделированию «смысла» и метафорически рассматривался как этап эволюции интеллектуальных систем. Если традиционная компьютерная технология характеризуется формулой «*данные+алгоритм*», технология искусственного интеллекта – «*знания+эвристика (квазиалгоритм)*», то проект ИЛ задаётся формулой – «*смысл+понимание*». Учитывая крупные наработки отечественных специалистов в области построения экспертных систем, в частности в семиотическом моделировании способов представления «знаний», в основу рассматриваемого проекта ИЛ была положена экспертная методология. В её рамках специальные методы анализа процесса «понимания» обеспечивают условия экспликации «смысла», оснований прини-

маемых решений. Модель «смысла» фиксирует эти «траектории» в определенным образом кодифицированных массивах информации.

Проблеме моделирования «смысла» посвящено много работ. Среди них в отечественной литературе выделяются следующие модели: онтологическая, культурологическая, психологическая, лингвистическая, герменевтическая, риторическая, дискурсивная, семиотическая, поэтическая, иконографическая, эстетическая, зотерическая, математическая, инженерная, синергетическая, неклассическая и пр. [Алексеев, 2005-4]. Все они так или иначе сформировались исходя из *контекстуального подхода*, согласно которому «смысл» – это обладающий параметрами системного единства контекст формализованных «знаний» экспертов. Общетеоретической основой этого подхода являются классические модели «смысла» Г.Фреге (смысл – это выражаемый знаком способ означивания значения) и её модификаций Б. Расселом и Л. Витгенштейном. В анализируемом проекте ИЛ также использовались идеи теории репрезентации М. Вартовского. В ней в систематическом единстве рассматривается модель, способы её построения и способы её интерпретации. Эта связь обеспечивает динамику наполнения «смыслового объема» возможными «смысловыми траекториями» в ходе согласования метода и предмета репрезентации. Эксперт как бы погружается в квазиалгоритм построения модели модели.

Технически модель «смысла» была выполнена как оболочка над системой представления «знаний». «Смыслообразующими» компонентами выступали категориальные коды «личность», «значение», «смысл», «ценность», «понятие», «действие», «роль», «норма», «умение» и ряд других. Конкретизация и экзemplификация этих кодов осуществлялась в концептуальной парадигме теоретико-деятельностного подхода. Эксперт должен соотносить полученные «смысловые траектории» с конкретными формализованными «знаниями», непосредственно задействованными в модели принятия решения. Название проекта – «искусственная личность» – оправдывается тем, что «знания» о принимаемых решениях индексированы персонологическими параметрами.

В принципе, удовлетворяется второй критерий ИЛ – пользователь получает некоторые ответы на запросы этического, эстетического, политического, идеологического и т.п. характера и реализуется персонологическая часть теста Тьюринга. Первый критерий ИЛ не уместен, хотя разработка проекта сопровождалась спекуляциями на тему «социальной личности» в стиле философии космизма, персонализма, соборности с оглядкой на социологический реализм – о первичности социального над личностным. То есть при построении смысло-

вого объема экспертная система становится «квазисознательной», и эта форма сознания представляет срез «общественного сознания». Конечно, такие суждения не достойны критики. Ещё более «страшной» методологией являлась концепция так называемого «ноокосмизма» – невероятная смесь из псевдофилософских фантазий об интеллектуализации космоса посредством технологий. Разработчик «концепции» – генерал А.Ю. Савин, он же самый главный заказчик проекта. О ноокосмологических «детях Сатурна» мы в нашей работе упоминать не будем.

Апробация экспертного подхода к проектированию ИЛ, осуществлённая в ходе опытной эксплуатации макета, выявила ряд неудач: методологических, организационных, теоретических и реализационных. Однако был достигнут *социокультурный эффект*: предложенный инструментарий вынуждал специалистов осуществлять рефлексию над «знаниями», эксплицировать «смыслы» собственных решений и выставлять их для интересубъективного обсуждения.

Расширение поля междисциплинарных исследований в проекте искусственной личности

Очевидно, что робототехнический и экспертный проекты ИЛ дополняют друг друга в плане расширения возможностей интеллектуальных информационных технологий. Если первый декларирует:

Искусственная личность – это робот, наделённый механизмами «квазисознания»,

то второй утверждает:

Искусственная личность – это экспертная система, оборудованная механизмами работы со «смыслом».

На наш взгляд, экспертный вариант проекта ИЛ по порядку и по значимости должен предшествовать робототехническому – у него больше шансов пройти вербально-коммуникативный тест Тьюринга. Претензии робототехнического проекта «перепрыгнуть» через задаваемые человеко-соразмерные «смыслы» и «значения» к неким лингвистическим «сущностям» искусственных агентов очевидно поспешны. Необходима предварительная дефиниция персонологических параметров в рамках экспертного подхода, и лишь затем – изучение того, как робот автономно либо в окружении колонии роботов способен, например, продуцировать «коммуникацию», «волеизъявление», «моральные императивы», «религиозные верования» и пр.

Как экспертный, так и робототехнический проекты повышают роль и усиливают значимость междисциплинарного подхода. Собственно и проект искусственного интеллекта уже не мыслим вне поля междисциплинарных исследований психологов, логиков, математиков, лингвистов, нейрофизиологов и др. Проект ИЛ значительно расширяет поле исследований, вовлекая специалистов социальных и гуманитарных наук – социологов, политологов, экономистов, искусствоведов, правоведов и др. В нашем проекте ИЛ участвовали даже представители паранауки, хотя от них толку было мало. Не специалисты в области естественных и технических наук приоритетны в проекте ИЛ. Исключительно важна роль экспертов в области общественных и гуманитарных наук, именно они вводят в предмет конструирования «смыслы», «ценности», «нормы», «идеалы» и прочие составляющие духовной жизни. Требование активной эксплуатации «знаний» социокультурного содержания подчёркивает антропоцентристскую фокусировку целей развития сложных систем, что согласовывается с постнеклассическим ориентиром их построения.

Однако представим утопический сценарий – все лучшие умы страны брошены на реализацию проекта ИЛ, выполнены все необходимые методологические работы по коммуникации, координации, интеграции научного сообщества. Более того, со стороны государства получена безграничная финансово-экономическая и правовая поддержка («За искусственную личность платить надо!» – подчёркивал Д. Деннетт [Dennett, 1994-1; 1995-1]). Будет ли реализован проект искусственной личности? Возможно, но для этого надо преодолеть ряд методологических трудностей, связанных со сложностью предмета исследований. Эти трудности на сегодняшний день представляются не только практически, но и теоретически непреодолимыми. Это – проблемы, инспирированные философией сознания. Робототехнический проект явно сталкивается с проблемами сознания, так как антропоморфность робота – первейший кандидат на присвоение ему в условиях тьюрингового тестирования ментальных феноменов – что нужно сделать для того, чтобы компьютер робота функционально уподоблялся мозгу человека? Экспертный подход с проблемой сознания сталкивается неявно – что нужно, чтобы сложная самоорганизующаяся система стала самостоятельно продуцировать «данные», «знания», «смыслы», получающие автономную внутрисистемную интенциональную продукцию, трансляцию и интерпретацию?

Тест Тьюринга, разработанный для прохождения этих проектов ИЛ, должен включать в качестве оцениваемых параметров сознательные способности и свойства. К этим способностям, как минимум, относятся: 1) «ощущение»

(sentience), т.е. восприятие мира и реагирование на него в соответствии с различными видами перцептивно-эффекторных возможностей; 2) «бодрствование» (wakefulness), т.е. способность фактически реализовывать некоторые возможности, а не только обладать диспозициями к их осуществлению; 3) «самосознание» (self-consciousness), т.е. не только осведомлённость о чём-то, но и осведомлённость о своей осведомлённости; 4) «аспектуальность» (what it is like), т.е. способность воспринимать мир в соотношении с собственной позицией «бытия» или, иначе, восприятие среды в соответствии с ответом на вопрос «каково быть?» (например, каково быть летучей мышью, роботом, экспертной системой?).

Помимо демонстрации вышеперечисленных способностей, систему можно считать квазисознательной, если судья из теста Тьюринга приписывает ей то, что она: 1) пребывает в состоянии «осведомлённости о» (states one is aware of); 2) обладает квалиа – владеет качественными свойствами конкретной разновидности (ощущает красное, чувствует запах кофе и пр.); 3) пребывает в феноменальном состоянии, которое более структурировано, нежели чем квалиа, и обладает определённой пространственной, темпоральной и концептуальной организацией опыта окружающей среды и опыта самого себя; 4) обладает акцессным сознанием (access consciousness) – т.е. доступом к внутриментальным составляющим при управлении речью и действиями (Н. Блок); 5) обладает нарративным сознанием (narrative consciousness) – т.е. выдаёт серию лингвистических выражений, «высказываемых» с перспективы актуальной или виртуальной самости (Д.Деннетт); 6) обладает интенциональностью – направленностью своего внимания на предмет сознания, активностью опредмечивания окружающей действительности (Д.Серль).

Очевидно, что приписывание ментальных свойств лишь опосредованно зависит от обладания системой параметрами физического подобия, как в робототехническом проекте. Конечно, антропоморфность системы значительно повышает степень доверчивости судьи, тестирующего систему. Однако для демонстрации персонологических способностей и свойств вполне достаточно развитого естественно-языкового интерфейса, который предполагается в экспертном подходе к проекту ИЛ. То есть позиция Д. Лената преобладает над позицией Р. Брукса.

Следует подчеркнуть, что вышеприведённые способности и характеристики «квазисознательности» необходимы, но не достаточны. Несомненно, можно привести длинный список иных характеристик квазисознательного по-

ведения. Проблема полноты сознательных способностей и свойств иначе называется «конъюнктивной проблемой», и ей придаётся большое значение в современных дискуссиях наряду с так называемой «дизъюнктивной проблемой» – проблемой множественной реализации сознательных феноменов на различных субстратах.

Методологические проблемы философии сознания принято делить по степени трудности, хотя некоторые это отрицают (см. ниже). Очевидно, что методологическая сложность влияет на поэтапность реализации проекта ИЛ – от решения простых проблем к решению более сложных. Вслед за Д. Чалмерсом выделим трудные и лёгкие проблемы.

Проблемы проекта искусственной личности¹

Легкие проблемы – это те, которые затрагивают только операциональные аспекты сознания. Сознание ограничивается следующими феноменами: реакцией на внешние стимулы, отчетами о ментальных состояниях, интегрированием информации когнитивной системой, фокусом внимания, контролем поведения, обучением.

Трудные проблемы связывают сознание с наличием «субъективного опыта» и самосознания. Ключевой вопрос состоит в объяснении механизмов этих субъективных процессов, как каузально связать с компьютерной архитектурой поведение ИЛ, внешне воспринимаемое судьёй в форме субъективного опыта?

По поводу трудной/легкой проблем имеется ряд мнений.

1. Трудная проблема неразрешима.

У. Джеймс в 1890 г., как бы предваряя данный ряд рассуждений, писал: «Природа в ее непостижимых проектах смешала нас из глины и огня, мозга и знания, эти две вещи существуют, несомненно, вместе и определяют существо друг друга, но как или почему, никакой смертный никогда не узнает» (цит. по [Blackmore, 2003]).

Т. Нагель (1994 г.) высказался следующим образом: «Проблема субъективности сложна и безнадежна. Мало того, что мы не имеем никакого решения, – у нас нет даже намёка на то, как ментальное объяснить физическим» [Nagel, 1974; Wider, 1989].

К. Макджинн (1999 г.) считал, что имеющемуся способу изучения ментального и мозга непреодолимо присущ дуализм: «Вы можете изучать ваши состоя-

¹ В классификации проблем и подходов к их решению мы частично используем работу С.Блэкмор в [Blackmore, 2003]

ния сознания, однако никогда не обнаружите нейроны, синапсы и прочее. Вы можете изучать мозг любого человека от рассвета до заката, однако так и не выявите сознание, которое очевидно для того, мозг которого так грубо рассматривается... Мы «познавательны закрыты» для трудной проблемы – как собака познавательна закрыта для чтения газеты или понимания поэзии. Как бы собака не пыталась, она не будет способна к математике хозяина, её мозг просто-напросто не так устроен. Подобным образом человеческий интеллект не предназначен для того, чтобы понять сознание. Мы можем научиться нервным корреляциям сознательных состояний (это – легкая проблема), однако не сможем понять, как физическое влечёт сознание» [McGinn, 1999].

С. Пинкер (1997 г.) заявил: «Мы можем бесконечно долго пытаться понять то, как работает ментальное, но понимание этого навсегда спрятано за пределами нашего концептуального горизонта» [Blackmore, P. 32]).

Позицию Т. Нагеля, К. Макджинна и С. Пинкера следует назвать *неомистическим подходом* – они считают, что проблема сознания на самом деле не разрешима, сознание – непостижимая тайна.

2. Трудная проблема разрешима при выработке нового понимания вселенной.

Д. Чалмерс требует создания новой теории информации и переосмысления с её позиций «двухаспектной теории информации», согласно которой информация имеет два аспекта – физический и феноменальный. Поэтому сознательный опыт – это, с одной стороны, некоторый аспект информационного состояния, а с другой стороны – аспект, заключённый в физической организации мозга.

К. Клэйрк (1995 г.) требует пересмотра фундаментальной физики. Ментальное подобно некоторым явлениям в квантовой физике. Сознание и квантовый эффект – это субъективные и объективные аспекты представления одного и того же.

Р. Пенроуз (1989, 1996 гг.): сознание зависит от неалгоритмических процессов – т.е. процессов, которые не могут быть выполнены цифровым компьютером или вычислены с использованием алгоритмических процедур. Требуется кардинальный пересмотр отношения ментальное-нейрофизиологическое с квантово-теоретических позиций. Сознательный опыт следует трактовать как качество, обусловленное квантовой последовательностью в микроканальцах нейронов.

Позицию данных авторов можно назвать *неопанпсихическим подходом*, онтологизирующим сознание с позиции атрибутивной трактовки информации. Особенно это характерно для Д. Чалмерса. Однако «подобные взгляды чрезмерно автономизируют категории онтологического и гносеологического, что приводит к ряду теоретических неопределённостей и противоречий», – полагает Д.И. Дубровский [Дубровский, 2002]. То есть проблема усложняется, а не упрощается.

3. Следует заниматься только легкими проблемами.

Ф. Крик, А. Сломан, Дж. Маккарти и большинство когнитивных учёных считают, что изучение сознания – не философская, а научная проблема. Мы приблизимся к более полному пониманию проблемы сознания, если начнём с чего-то простого, например с теории визуального закрепления, описывающей синхронизацию электромагнитных колебаний для объяснения того, как различные признаки воспринятого объекта связываются в единство в процессе создания перцептивного целого [Crick, 1994].

4. Нет никакой трудной проблемы.

К. О'Хара и Т. Скатт (1996 г.) в специальной работе «Нет никакой трудной проблемы сознания» [O'Hara, Scut, 1996] приводят три причины, чтобы игнорировать трудную проблему: 1) нам известно, что такое легкие проблемы, с них и следует начинать; 2) решения легких проблем изменяют наше понимание трудной проблемы; таким образом, попытки решить трудную проблему сейчас преждевременны; 3) решение трудной проблемы было бы полезно, если мы могли бы распознать ее как таковую: на данный момент проблема не достаточно хорошо понимается и неясно, какого рода трудность предстаёт перед исследователем.

П. Черчленд (1996 г.) считает, что трудная проблема неверно понята, так как мы не можем заранее предсказать, какие проблемы окажутся легкими, а какие трудными. Откуда мы знаем, что объяснение субъективности более трудно, чем объяснение «легких» проблем? Достаточно ли хорошо определены «трудные» вещи (квалиа, субъективный опыт), чтобы их можно было как-то идентифицировать в роли трудных. Например, мысли – это квалиа или нет? [Churchland, 1996]

Д. Деннетт (1994 г.) считает так: «Даже если бы все легкие проблемы воспроизводства, развития, роста и метаболизма были бы решены, всё равно осталась бы действительно трудная проблема: что такое собственно жизнь. Также и деление проблемы сознания на «легкие» и «трудные» – это отвлекающий

фактор, генератор иллюзии» (1996 г.). Например, согласно Деннетту, нет такой вещи, как «феноменальное». Не потому, что отрицается сознание у людей, а в силу того, что философы неправильно трактуют понятие сознания [Blackmore, 2003, Р. 32]).

Позицию этих авторов можно назвать *терминологической* – правильно определите то, что называется «сознанием», и тогда проблемы либо чётко обозначатся, либо вообще исчезнут.

5. Трудную проблему следует решать.

Это – *конструктивная* позиция. Так полагает Н. Блок. Однако трудную проблему следует дифференцировать для более чёткого понимания предмета исследования. Поэтому предлагается различать сложную и самую сложную (сложнейшую) проблему¹.

Самая сложная проблема заключается в следующем: 1) у нас нет оснований считать искусственную личность номологически и метафизически невозможной; 2) внешняя функциональная эквивалентность может служить основанием для того, чтобы приписать системе сознание или, напротив, отказать ей в сознании; 3) физическая реализация, существенно отличная от нашей, сама по себе не является основой для убеждённости в отсутствии сознания; 4) у нас нет никакого понятия о том, как доказать наличие или отсутствие сознания у другого существа.

В принципе, самая сложная проблема принимает формы других фундаментальных для современной аналитической философии проблем: 1) «проблемы зомби», которая состоит в обосновании мыслимости и возможности (логической, метафизической, натуральной) бессознательных существ, чья физическая структура, внешнее поведение, функциональная организация ничем не отличаются от структуры, поведения и функций сознательных существ (людей); 2) «проблемы *другого*» – на каком основании другим людям, а также иным биологическим и более широко – физическим – системам можно приписать факт обладания ими сознанием (субъективной реальностью); 3) можно показать, что эти проблемы упираются в *проблему создания теста Тьюринга* на предмет обладания системой субъективной реальностью. Ряд исследователей считает, что последняя проблема (тест Тьюринга на субъективную реальность) при сегодняшнем уровне развития науки, философии и методологии теоретически неразрешима (Д.И. Дубровский) [Дубровский, 2004-1]. Но это невозможно лишь

¹ Для раскрытия позиции Н. Блока используется работа [Вильянуэла, 2006]

пока, на сегодняшний день, в контексте наличных парадигм сознания – бихевиоризма, физикализма, функционализма.

Сложная проблема состоит в том, что мы не представляем себе, как феноменальное сознание может быть сведено (редуцировано) к нейрофизиологическим фактам, наблюдаемым при изучении мозговых процессов в связи с субъективно переживаемыми явлениями сознания. Почему нейронная основа некоторого феноменального качества есть нейронная основа именно этого феноменального качества, а не другого феноменального качества или вообще никакого. Проблема объяснения данной связи состоит в «объяснительном пробеле» между нейронной основой некоторого феноменального качества и самим феноменальным качеством. Д. И. Дубровский по-иному интерпретирует подобного рода проблему: «Как объяснить связь явлений сознания (субъективной реальности) с физическими (в частности, мозговыми) процессами, если первым нельзя приписывать физические свойства, а вторые ими по необходимости обладают?»¹

При классификации проблем на сложную и сложнейшую логично предположить, что упущена ещё одна, которую назовём «менее сложной проблемой».

Менее сложная проблема заключается в расшифровке нейродинамических кодов ментальных феноменов сознательных существ. Как полагает Д.И. Дубровский, имеется два вида задач расшифровки кода: 1) «прямая» задача, когда дан кодовый объект и требуется выяснить информацию, которая в нем содержится; 2) «обратная» задача, когда нам дана определенная информация и требуется установить ее носитель и его кодовую организацию. Такая задача является более трудной. Однако обе эти задачи – вполне решаемы. Например, известен метод А.М. Иваницкого, который позволяет по электрической активности мозга определять, какие умственные операции совершает в данный момент человек: мыслит ли он пространственными образами или использует логическое мышление [Иваницкий, 2006]. Здесь прослеживается решение первой задачи. Для решения обратной задачи следует воспользоваться принципом инвариантности информации по отношению к физическим свойствам её носителя (принципом Дубровского). Имеется возможность, при наличии системы изученных кодовых зависимостей, решить и задачу воспроизводства информации

¹ См. вопросы, сформулированные Д.И. Дубровским для симпозиума «Сознание и мозг», 30 ноября 2007 г., ИФ РАН, Москва и представленные на сайте НСММИ РАН: <http://www.scm.aintell.ru/default.asp?p0=107>

на субстрате, отличном от человеческого мозга, — в коннекционистско-репрезентативных кодовых образованиях искусственной личности.

Демонстрация таковой возможности выходит за пределы данной работы, так как требует привлечения специального анализа принципа инвариантности информации и, например, сопоставления его с аргументом множественной реализуемости. Это, в свою очередь, влечёт за собой изучение редукционистских/антиредукционистских парадигм реализации ментальных феноменов на различных физических системах — на мозге человека, на мозге других приматов и животных, на «тине» в «черепе» гипотетических марсиан, на кодовых структурах искусственной личности. Более того, такую работу надо осуществлять в рамках функционализма теста Тьюринга. Сейчас достаточно заключить, что междисциплинарное освоение проекта искусственной личности следует начинать с менее сложной задачи:

Искусственную личность следует разрабатывать совместно с расшифровкой кода Дубровского.

Однако решение проблемы сознания (сознание-мозг-компьютер) — будь она простая, сложная, менее сложная или какая-то ещё иная *не* представляется главной стратегической целью развития ИИ. Доминирующей проблемой, по мнению автора, является *проблема творчества*. И если так, то вся перечисленная выше этапность реализации проекта искусственной личности представляется пустым рассуждением, как и все бесконечные метафизические споры о компьютерной реализации феноменов сознания: неосхоластика философии сознания насчитывает десятки тысяч крупных и мелких работ, как показано в классификации Д. Чалмерса. Для практической философии — философии искусственного интеллекта — они бесполезны и порою вредны. Поэтому далее, в пп. 3.3.1 мы наметим стратегию переориентации методологии исследования ИИ с психофизической проблемы на проблему реализации творческих способностей человека.

3.2.3. Критика мировоззрения: комплексный тест Тьюринга против комплексного тестирования в вузовском образовании

Комплексный тест Тьюринга перспективно использовать для критики современного компьютерно-технологического детерминизма, который означает то, что компьютеринг применим для решения всех проблем человека, не только материально-производственных, но и проблем духовной сферы жизни человека и общества. Автора как преподавателя и как специалиста в области проектирования компьютерных технологий особенно возмущает факт крайне неразумного применения идеи компьютеринга для оценки знаний школьника и студента, авторитета преподавателя и ученого. Будучи гражданином, автор не способен понять, как можно доверять компьютеру-арифмометру выборы политического лидера.

Критическую функцию комплексного теста Тьюринга раскроем на примере современных образовательных технологий. Вернёмся к истокам – к оригинальному тесту. Снова подчеркнем, что статья [Turing 1950] послужила становлению аналитической философии искусственного интеллекта, в рамках которой наиболее цитируема и то, что тест Тьюринга (ТТ), неявно обозначенный в статье, послужил прообразом для ряда разнообразных тестов, раскрывающих особенности мышления, понимания, сознания и других параметров субъективной реальности. А.Тьюринг убеждённо доказывает, что через 50 лет (в 2000 г.) машины будут имитировать интеллект, востребованный для ведения осмысленного разговора между людьми, т.е. «диалоговый интеллект» (Н. Блок). Обсуждая теоретические возможности построения мыслящих машин и предлагая проект для их реализации средствами универсального цифрового компьютера, функционирующего в соответствии с заложенными в него эвристиками, А. Тьюринг полагает, что в «будущем компьютеры будут успешно соперничать с людьми во всех интеллектуальных областях» [Управление, информация, интеллект, 1976, С. 282].

Несмотря на амбициозность своей идеи, А. Тьюринг сегодня бы «в гробу перевернулся», узнав о том, как в нашей стране применяются компьютеры в сфере высшего образования. Они уже не «соперничают» с людьми. Они оценивают людей, предопределяя их профессиональную пригодность, материальный достаток, судьбу. Оценка касается не только интеллекта, но и образованности человека в целом, т.е. сложной многомерной и многоуровневой системы морально-нравственных, творческо-продуктивных, мотивационно-волевых, эмоционально-чувственных, разумно-рациональных и других свойств и способно-

стей личности. Компьютеры конституируют механистические стереотипы мышления культурного человека российского сообщества, формируют менталитет нашего народа, оставляют неизгладимый след на генофонде.

Очевидны вполне справедливые возражения типа того, что сам по себе компьютер не может оценивать человека. В самом деле, если проследить всю цепь тестирования, то мы обнаружим, что в конечном итоге оценивает не компьютер, а человек (судья), используя для этого вспомогательное средство – компьютер. Это так. Но если последовать ещё далее по этой цепочке, то мы увидим, что судья в условиях компьютерного тестирования сам становится своего рода тьюринговым «человеком-компьютером».

Дефиниция «человека-компьютера»

В оригинальной версии А. Тьюринга применяется другой термин – «human computer», «человеческий компьютер», т.е. та сфера в сознании человека, которая отвечает за алгоритмическое поведение. Если осуществить индуктивное обобщение абстрактной способности человеческой личности к осуществлению вычислительных операций, то данная сфера становится определяющим персонологическим параметром. Отчужденный, одномерный человек информационного общества при присвоенной или приданной власти оценивать другого человека, т.е. став судьёй, переносит стереотипы своей калькулирующей рациональности на оцениваемого. Нет особой разницы в том, воплощены ли рассуждения судьи программно-информационными средствами информационных технологий либо реализованы нейро-мозговыми коррелятами его сознания. Функции человека-компьютера и функции компьютера неотличимы. Для обоснования данного утверждения целесообразно сослаться на упоминавшуюся ранее статью С. Харнада «Разумы, машины и Тьюринг: неотличимость неотличимостей» [Harnad, 2001]. Эта работа заслуживает особого внимания в связи с исследованием ментальности людей эпохи электронной культуры. Мы же примем в качестве исходного положение о том, что именно компьютер (человек-компьютер, судья-компьютер), а не человек оценивает человека в системе компьютерного тестирования. Причём тестируется человек-компьютер. Это верно при принятии тезиса об отчуждённости вычислительной способности человека.

Применим КТТ для анализа концептуальной возможности применения компьютерных технологий тестирования в современной системе высшего образования, которые иногда обозначаются как «e-learning», «инновационные образовательные технологии», «информационные образовательные технологии» и

пр. На начало 2009 г. в сфере высшего образования существовали следующие образовательные технологии в масштабах всей страны.

Комплексное компьютерное тестирование в вузовском образовании

Единый государственный экзамен (ЕГЭ). Начиная с 2009 г. является основной (!) формой итоговой государственной аттестации в школе для всех (!) выпускников школ Российской Федерации¹.

Федеральный Интернет-экзамен в сфере профессионального образования (ФЭПО). Проводится в форме компьютерного тестирования студентов и направлен на проверку выполнения требований Государственных образовательных стандартов профессионального образования. Целью ФЭПО является формирование единых требований к оценке качества подготовки специалистов².

Дистанционные образовательные технологии (ДОТ). Непрерывно сопровождают ряд преподаваемых в вузе дисциплин. Это – территориально и тематически распределённые способы и средства преобразования информации, которая составляет содержание учебного процесса, сформированного на основе общего информационно–коммуникационного поля между преподавателями и учащимися. Преподаватели, как правило, выступают в роли виртуальных персоналий, представленных своими научными трудами, учебными и методическими пособиями, тестами. Если на заре развития ДОТ заявлялось лишь об их преимуществах в области вербальной коммуникации (оперативность, адресность, преодоление территориальных границ, обширность аудитории и пр.) и подвергалась сомнению возможность преподавания дисциплин, связанных с формированием сложных навыков и умений на основе демонстрации (например, хирургия, опасная техника), то сегодня в условиях развития мультимедийных технологий, искусственного интеллекта, виртуальной реальности и пр. преодолены границы и невербальной коммуникации – сейчас мастер может передавать опыт (например, трепанации черепа), находясь от учеников на расстоянии многих тысяч километров. Вне всякого сомнения, при условии умелого

¹ См. официальный информационный портал Единого Государственного экзамена: <http://www.ege.edu.ru/>

² См. официальный сайт ФЭПО: <http://www.fepo.ru/>

применения, ДОТ способны качественно изменить образовательные возможности¹.

Как мы видим, на этих трех компьютерных технологиях – ЕГЭ, ФЭПО и ДОТ – основана оценка образовательного процесса в вузе. Современный вуз полностью и на всех этапах обучения охвачен компьютерным тестированием. Если первая технология предваряет образовательный процесс в вузе, вторая – время от времени вторгается в него для промежуточной оценки знаний, то третья технология инкорпорируется в учебный процесс. По сути, такие технологии следует считать разновидностями компьютерных вирусов, в широком смысле трактовки последних, – как деятельность человека посредством компьютерно-технологических средств по внедрению в социокультурное пространство для регулярного воспроизводства и трансляции различного рода идей. Подобного рода идея культурных вирусов не нова. Она была выдвинута в теории генно-культурной эволюции Р. Докинса и обрела вычислительные (компьютеризационные – от сл. «computationalism») очертания в теории сознания, предлагаемой Д. Деннеттом и его последователями².

Основными участниками технологического процесса тестирования являются два класса отчуждённых людей: преподаватели-компьютеры и студенты-компьютеры. Тестовые вопросы имеют логическую форму простейших логических операций закрытой нестрогой дизъюнкции и закрытой конъюнкции. Крайне редко применяются более сложные логические операции, например, открытый вопрос. Логическая (и интеллектуальная) насыщенность тестов технологически не выгодна, требует дополнительных затрат. В данной статье нет возможности исследовать политические, финансовые, экономические и прочие преимущества и недостатки ЕГЭ, ФЭПО, ДОТ, например, изучить «жизненно важные» вопросы эффективности компьютерных образовательных технологий в борьбе с коррупцией в вузах. Задача достаточно узкая – осуществить концептуальный анализ собственной проблемы образования – проблему оценки того, насколько учащийся усвоил учебный материал и понял ли его вообще. Аргу-

¹ Сегодня, однако, ДОТ разрабатываются крайне некомпетентно. Пример из личного опыта. Автор решил пройти тест в рамках ДОТ по дисциплине «Философия» в одном из коммерческих «международных» вузов, где преподавал данную дисциплину студентам очного отделения. К его немалому удивлению, имея золотую медаль философского факультета МГУ им. М.В. Ломоносова, кандидатскую степень по философии, многолетний опыт преподавания данной дисциплины, он с трудом вытянул на «удовлетворительно». Правда, вопросы ДОТ лежали вне сферы философского знания и представляли чудесный конгломерат избыточных фраз, неизвестных персоналий и цитат из истории, культурологии, политологии, социологии и других дисциплин, но только не из философии.

² См. анализ работ Д. Деннетта, осуществлённый в [Юлина, 2004]

ментация направлена на защиту интуитивно очевидного положения о том, что современные компьютерно-коммуникативные новации не решают ключевую проблему образовательного процесса – проблему *понимания* учащимся изучаемого материала. Тем более они не дают возможности понять со стороны оценивающего (судьи), обладает ли учащийся пониманием или не обладает. Они вредят образовательному процессу, так как внедряются слишком поспешно, широко и глубоко. Они ограничивают свободное творчество учащегося узкими рамками механистических калькуляций, ведь приобретение нового знания сущностно творческий процесс. В силу этого системы компьютерного тестирования отнюдь не относятся к инновационным технологиям. Это – архаичные методы старой педагогики, черты которой ярко охарактеризованы в начале 20 в. одним из основателей отечественной философии образования С.И. Гессеном, указывающим на то, что новая педагогика ориентирована на пробуждение творческого начала учащегося, в то время как старая педагогика стремилась оценивать его с позиции механистической зубрёжки (Гессен С.И. Основы педагогики. Введение в прикладную философию / Отв. ред. и сост. П.В. Алексеев – М.: «Школа-Пресс», 1995. – 448 с.). Сегодня дряхлая педагогика обрела компьютерную мощь. Поэтому болью отчаяния звучат слова С. Капицы: «Россию превращают в страну дураков... ЕГЭ, который сегодня используют, не может дать объективной картины знаний школьника. Он построен лишь на знании или незнании фактов. Но факты далеко не всё!».

Следует обратить внимание на фундаментальный труд, направленный на всестороннюю критику ЕГЭ, который был предложен недавно доктором философских наук, профессором В.В. Мироновым, деканом философского факультета МГУ им. М.В. Ломоносова [Миронов, 2011]. Не предполагая никаких априорных предвзятостей по поводу реформы российского образования, в которой ключевым моментом является аттестация абитуриентов по результатам ЕГЭ, В.В. Миронов, тем не менее, убедительно показывает, что ЕГЭ – это крупный шаг назад в развитии образовательного процесса в России. ЕГЭ способствует коррупции, значительно коммерциализирует образовательный процесс (такого грубого «выжимания» денег из молодых людей нет ни в одной стране мира), кардинально ухудшает креативные способности личности..

Интуитивно очевидна справедливость данных слов. Однако вряд ли они будут убедительны для человека электронной культуры, требующего рационального, конструктивного, компьютеризуемого обоснования любых заявлений. Мировоззренческие и методологические основания данной формы культуры пропитаны идеями аналитической философии в сфере математики, логики, лин-

гвистики, техники, когнитивистики и пр. Необходим не метафизический, а аналитический подход, ориентированный на раскрытие смыслов и значений семиотических комплексов, образующих контекст и контент терминов «образование», «компьютерное тестирование», «понимание» и пр. Исследование следует осуществлять, руководствуясь древнегреческим испытанным девизом – «подобное измерять подобным», или, в современном звучании, «компьютерное поверять компьютерным». Компьютерно-ориентированные мысленные эксперименты – тесты Тьюринга – направлены на решение подобного рода концептуальных задач, а представленные в формате *комплексного теста Тьюринга* (КТТ) предназначены для изучения и решения сложных социокультурных задач современного и перспективного информационного общества. К актуальнейшим задачам относится и заявленная проблема приобретения студентами достойного багажа знаний в условиях компьютерных образовательных технологий. В данной работе не раскрывается полная картина применения всего конгломерата тестов для концептуального анализа комплексного тестирования в системе вузовского образования. Совершенно ясно, что достойное образование немислимо без оценки и самооценки творчества, интуиции, логической корректности, математической точности и пр. Анализ достаточно ограничить теми ключевыми моментами системы компьютерного тестирования образовательного процесса, которым имманентен факт субъективной реальности. На мой взгляд, к ним относятся: 1) феномен понимания тестируемым (студентом) изучаемого материала; 2) феномен понимания тестирующим (преподавателем) факта понимания студентом, т.е. приписывание феномена понимания.

В связи с обозначенными моментами воспользуемся соответственно тестом Сёрля и тестом Ватта. Первый тест анализирует человеческую способность *понимать* в условиях компьютерного тестирования. Второй тест анализирует *приписывание понимания*. Обозначим данную разновидность комплексного теста Тьюринга через $КТТ_1$, где индекс обозначает номер в серии построения интегральных ТТ. Значение индекса равно 1, так как автор данной статьи не осведомлён о том, что другие исследователи применяли подобного рода подходы к анализу феноменов субъективной реальности. Каждый исследователь вносил свою лепту в анализ различных аспектов компьютерной реализации сознания – мыслить, понимать, сознавать, говорить, воображать, творить и пр. Но в решении социокультурно значимых задач наблюдается очевидный пробел. Его мы постараемся восполнить. Рассмотрим совместно данные мысленные эксперименты.

Российский студент в «Китайской комнате» (тест Сёрля)

Воспользуемся классическим мысленным экспериментом Дж. Серля «Китайская комната». В данном тесте поднимается вопрос о прояснении ситуации с пониманием в условиях компьютерных технологий.¹ Наша трактовка очевидна: студент— это человек-компьютер, погружённый в систему программированного тестирования.

Принято считать, что у эксперимента Дж. Серля были следующие прообразы: 1) «Мельница Лейбница», представленная в знаменитой «Монадологии» [Лейбниц, 1982, С.415-416] — для понимания (ощущения) недостаточно реализующих механизмов, требуется духовная субстанция; 2) «Бумажная машина» А. Тьюринга (1948 г.), в которой показано, что для игры в шахматы требуется не понимание, а лишь вычисление на множестве шахматных комбинаций; 3) «Китайская нация» Н. Блока (1978 г.) — между нейродинамическими состояниями мозга и ментальными способностями (боль, понимание) корреляция не наблюдается.

Дж. Сёрль идёт далее. Суть его эксперимента в следующем. Разыгрывается два сценария. Согласно *первому сценарию* в закрытую комнату помещается англоязычный индивид «Сёрль-в-комнате» (сценарий ведётся от первого лица). Он ни слова не понимает по-китайски. Однако его снабжают китайским алфавитом (исходными иероглифами), английскими инструкциями по составлению из алфавита правильно построенных фраз на китайском и таблицей ответов на китайском. Вопросы задаёт китаец. Сёрль-в-комнате манипулирует символами и выдаёт ответы, бессмысленные для него, но понятные китайцу, находящемуся вне комнаты. При этом Сёрль-в-комнате ничего не понимает. Он работает как компьютер, оперируя входной информацией и выдавая информацию в соответствии с определённой программой. *Второй сценарий* мысленного эксперимента проще. Нет никаких инструкций, алфавита, таблиц и прочего. Вопросы теперь задаёт не китаец, а англичанин. Сёрль-в-комнате всё прекрасно понимает и выдаёт, конечно, осмысленные ответы на английском. То есть в первом сценарии осуществляется изощённая информационно-технологическая обработка информации тестирования понимания, но *понимания нет*. Тест на понимание не пройден. Во втором сценарии нет никаких информационных технологий, но *понимание есть*. Тест на понимание пройден.

¹ Оригинальный тест Серля и сводка дискуссий см.: [Searle, 1980; Серль, 2006; Cole, 2009]

Очевидна аналогия с тестовой оценкой качества студента. Студент снабжается электронными и учебными пособиями, методическими пособиями, учебным планом, набором тестовых вопросов, на контрольную выборку из которых должен ответить. Причём имеются сроки обучения, которые малы и объёмы материала, которые велики. Поэтому освоение учебного материала обычно *сводится к бездумному запоминанию ответов на тестовые вопросы*. То есть происходит, как и в эксперименте Сёрля, формальная манипуляция с синтаксическими конструкциями. В итоге – бессмысленное заучивание тестовых ответов, а понимания-то нет!

Какие могут быть возражения и доводы в пользу того, что понимания нет? Обратимся к работам Сёрля и его критиков.

1. Системологическое возражение. В самом деле, тестируемый студент ничего не понимает, он лишь запоминает тестовые альтернативы. Однако он – лишь часть более общей системы образования. Система же в целом тесты проходит и понимает. Индивид в рамках системы не изолирован. Перед ним – мощные интернет-ресурсы, банки данных с электронными учебными пособиями и пр. Понимание невозможно приписать лишь отдельному индивиду, а следует приписать всей системе образования, частью которой он является.

Ответ Серля мог бы быть таким: пусть студент включит в себя все элементы системы, запомнит всю информацию из Интернета, других баз данных и всевозможные схемы правильных ответов на вопросы. Такой индивид будет олицетворять целостную систему. У данной системы невозможно обнаружить таких функций, которые индивид бы не выполнял, и таких ресурсов, которыми бы он не обладал. Но в любом случае он по-прежнему ничего не понимает по тестируемому материалу, ведь в системе нет ничего сверх того, чего не было бы в индивиде. Если не понимает индивид, то и система также не понимает по той причине, что она стала частью индивида. Если студент не понимает, то соединение его с электронными учебниками не поможет ему понять материал. И далее, уже более жёстко – «...если принимается утверждение, что мне присуще понимание на основании наличия у меня входной и выходной информации определённого типа, а также программы между входом и выходом, то все некогнитивные подсистемы становятся когнитивными. Например, можно составить такое описание работы желудка, характеризующее обработку желудком информации, которое позволяет говорить о том, что желудок «понимает» (Цит. по [Серль, 2006, С.9]). Конечно, следует согласиться с Серлем.

2. *Возражение с позиции виртуального сознания.* Оно перекликается с первым возражением. Пускай студент, реализуя тестовые задания, *не* обладает пониманием. Однако обстановка в комнате – книги, базы данных, тексты – это причинный носитель функций понимания, своеобразный оператор «снабжения» индивида пониманием. Ответ примерно такой же, как ответ на первое возражение.

3. *Робототехнический отзыв.* Допустим, что студент – полноценный робот, «живущий» в среде тестовых вопросов. Программа позволяет ему не только получать формальные тестовые символы на входе и выдавать формальные результаты тестирования на выходе, но, помимо этого, воспринимать окружающее, передвигаться, гулять, забивать гвозди, есть, пить – всё, что может быть отражено в тестовых вопросах. Добавляется много возможностей. Например, посредством системы виртуальной реальности студент обучается трепанации черепа и при тестировании как бы самостоятельно участвует в операции. Будет ли такой студент-робот понимать? Здесь способность понимания обусловлена совокупностью самых различных взаимосвязей с окружающим миром. Однако добавление «перцептивных» и «двигательных» способностей не добавляет ничего нового по поводу понимания. Допустим, внутри робота вместо компьютера находится студент. В отличие от первых двух отзывов, ему предоставляется большее количество тестовых вопросов. Без его ведома часть тестов поступает от телевизионной камеры, присоединённой к студенту-роботу, а другая часть тестов, выдаваемых им, управляет двигателями, двигая руками и ногами и прочими частями тела. Однако все действия сводятся к манипулированию формальными символами. Он получает «информацию» от перцептивного аппарата робота и даёт «инструкции» его двигательному аппарату, не зная об этих аппаратах. Он – гомункул робота, но в отличие от традиционных гомункулов, он не ведаёт, что происходит. Он ни о чем не знает помимо правил манипуляции символами. Следует добавить, что студенту-роботу не присущи никакие интенциональные состояния – он просто перемещается в соответствии с программой и электрическими импульсами в проводах. В свою очередь, руководствуясь программой, студент-робот так же не обладает никакими интенциональными переживаниями. Всё, что он делает – это тупо следует тестовым вопросам и выдаёт запомненные в памяти тестовые ответы согласно бихевиористской схеме.

4. *Отзыв с позиции моделирования мозга.* Предположим, что нейрофизиология и психология достигли такого уровня развития, когда ответ на определённый тест означает определённую конкретную нейродинамическую струк-

туру мозга. При правильном ответе, т.е. при понимании – одна конфигурация, при неверном ответе – непонимании – другая. Тогда тестирование заключается в сканировании мозга студента. Предлагается представить студента, который перебирает не тестовые альтернативы, а управляет системой водопроводных труб, соединённых клапанами. Когда студент проходит тест, он из памяти извлекает, какие клапаны ему открыть, чтобы выдать правильный ответ, а какие – закрыть. Каждое соединение водных отсеков соответствует синапсу в мозгу студента, а система в целом устроена таким образом, что после должной последовательности возбуждений (что, в свою очередь, соответствует должной последовательности изменений положений вентилей) правильный ответ вытекает из труб, соединённых требуемым образом. Однако понимает ли такая «трубопроводная мозговая» система? Где в этой системе понимание? На входе данная система труб принимает тестовые вопросы и, моделируя формальную структуру мозга студента, на выходе выдаёт ответы. Но студент определённо не понимает смысла тестовых вопросов, так же, как не понимают их водопроводные трубы. «Если же вдруг захочется подумать, что к пониманию способна *конъюнкция* человека и водопроводных труб, то следует заметить, что в принципе человек может представить формализованную структуру водопроводных труб и в своём воображении произвести все «нейронные возбуждения» [Серль, 2006, С.20]. Проблема с моделированием мозга в том, что моделируется лишь формализованная последовательность возбуждений нейронов посредством синаптических связей, а важно моделировать причинные зависимости, определяющие отношение «мозг – понимание». Именно причинные зависимости задают возможность интенциональных переживаний. Формальных зависимостей для этого недостаточно. «Данный факт и демонстрирует пример с водопроводными трубами: мы можем полностью отграничить формальные свойства от важных для возможности понимания нейробиологических причинных зависимостей» [Серль, 2006, С.21].

5. Серль приводит ещё один довод – *комплексный*, который заключается в конъюнкции предыдущих. Возникает вопрос – как из нескольких простых альтернатив и несложных операций возникает сложная сущность ментального характера – понимание? Очевидно, что понимание студентом тестовых вопросов без участия интенциональной связи, внешней относительно этих составляющих, невозможно – понимать он не будет. Но заучить тестовую альтернативу *без понимания* заученного материала, т.е. выступая в роли студента-компьютера, он способен.

Человек-компьютер оценивает человека (тест Ватта)

С. Ватт в работе «Наивная психология и инвертированный тест Тьюринга» [Watt, 1996] утверждает, что тест Тьюринга неявно опирается на «наивную психологию» – т.е. на естественные психологические особенности человека понимать и предсказывать поведение других. В условиях ТТ эти особенности начинают играть главную роль в приписывании ментальных качеств исследуемой системе. В начале рассматривается влияние наивно-психологистской установки на тест Тьюринга – как со стороны тестируемой системы, так и со стороны судьи. Затем предлагается и обосновывается инвертированная версия ТТ. В этой версии процессы приписывания ментальных состояний анализируются явным и более непосредственным образом, чем в оригинальном ТТ. Для наших целей важно принять теоретический факт: человек-компьютер, выступая в роли судьи (судьи-компьютера), принципиально не способен приписать понимание студенту. То, что реальный, эмпирический преподаватель (преподаватель-человек), анализируя результаты тестирования, всё-таки предполагает наличие ментальности у студента-компьютера, по условиям компьютерного тестирования исключается – неформализованные оценки преподавателей не учитываются на стадии статистической обработки баллов, набранных студентом.

Тест Ватта (Инвертированный ТТ) в отечественной литературе не нашел должного освещения. Поэтому его следует несколько подробнее разобрать. С. Ватт начинает с критики оригинального ТТ, априори полагая психологическую обусловленность тестовых вопросов и отказываясь от абстрактных логико-математических интерпретаций теста, в которых, например, длина теста представляется неограниченной или не учитываются особенности различения, присущие судье как реальному человеку, а не как человеку-компьютеру, отчуждённому от всей совокупности невычислимых процедур. Оригинальный ТТ выполнен в бихевиорально-операционалистском стиле, т.е. уже изначально предполагает все возможные оценки того, что оценивается. Судью легко ввести в заблуждение – наблюдатель может приписать интеллект совершенно «неразумным» машинам. Короче говоря, тестируемая система проходит тест Тьюринга, если наблюдатель *верит*, что эта система разумна. Аналогичное суждение выносится и относительно других психических свойств и способностей. Применительно к нашему случаю – преподаватель должен быть убеждён, что студент понимает, иначе он понимания студенту не припишет.

С. Ватта данное положение не устраивает, и он применяет методы «наивной психологии», базирующиеся на естественной человеческой установке при-

писывать ментальные состояния как другим, так и себе, т.е. идентифицировать и воспринимать другое и свое сознание. Наивно-психологистская установка – глубоко укоренённая в человека способность, его фундаментальное природное свойство, существенный, внутренний аспект собственно человеческого поведения. Её следует учитывать в тесте. Категории наивной психологии сравнимы с категориями «природной психологии» (natural psychology), в которой психологические способности объясняются в контексте представлений о социобиологической эволюции животных, и «народной психологии» (folk psychology), изучающей психологические проблемы с позиции здравого смысла. Тем не менее, направленность наивной психологии отличается от выше обозначенных форм психологических исследований. Если налицо реальная естественная способность оценивать «другие сознания», в частности, «другое понимание», то эту способность следует подвергнуть тьюринговому тестированию. Требования к ТТ повышаются: 1) он должен включать вопросы исследования наивно-психологистской установки; 2) средства его реализации должны обеспечивать тщательную проверку и контроль фактов приписывания судьёй ментальных свойств тестируемой системе.

В свою очередь, факт приписывания понимания зависит как от поведения системы, так и от судьи – подобно понятию красоты, такие понятия, как «интеллект», «понимание», «сознание» зависят от мнения или точки зрения судьи и неразрывны с социокультурным и психологическим контекстом. Наивно-психологическая установка характеризует готовность вообразить систему интеллектуальной, понимающей, творческой и т.п. даже тогда, когда она таковой не является. Приписывание ментального – это также ментальный феномен. Наивно-психологические тесты – это «тесты на ложную веру» (работы Примака, Вудруфа, Виммера, Пернера, Бэрона – Кохена и др.). Цель тестов состоит в оценке способности ребёнка приписывать ментальные свойства различным объектам. В тестах используются куклы. Перед детьми разыгрывается спектакль. Необходимо, чтобы в конце представления один из персонажей знал о чём-то ложном и чтобы ребенок знал о ложности таких убеждений. По замыслу создателей этих тестов убеждённость персонажа в своей правоте подкрепляет убеждённость ребёнка и он не способен демаркировать ложное от истинного. Напротив, тесты «ложной веры» устанавливают границу между ментальными состояниями, которые ребёнок приписывает другим и его собственными ментальными состояниями. И здесь следует ключевая идея инвертированного ТТ: следует совместить наивно-психологические тесты с форматом теста Тьюринга, «перевернув» последний, т.е. на место судьи поставить компьютер. Инвертиро-

ванный ТТ сравнивает способность различения со стороны системы с той способностью различения, которая присуща опытному судье-человеку. Система проходит тест в следующих случаях: 1) она не способна самостоятельно отличить человека от человека; 2) она не способна самостоятельно отличить человека от компьютера (компьютер при этом проходит нормальный тест Тьюринга); 3) однако она способна самостоятельно отличить человека от судьи-компьютера. Судья-компьютер должен быть судим! Оценивать его работу должен квалифицированный судья.

В итоге получается, что вместо сравнения лингвистического поведения тестируемой системы и человека, как это предполагается в оригинальном ТТ, в тесте Ватта оценивается *соответствие между способностью системы приписывать ментальные свойства другим и такой же способностью у человека*. Для «прохождения» инвертированного ТТ тестируемая система должна показать те же самые закономерности и те же ошибки в приписывании ментальных свойств, которые присущи обычному человеку. Эти закономерности и ошибки далее могут быть проанализированы в психологическом исследовании, базирующемся на изучении антропоморфизмов, ведь именно они в большей мере ответственны за приписывание ментального другим системам. Тест Ватта экономичнее стандартного ТТ, который требует теоретически неограниченной длины для оценки всех возможных ситуаций, относительно которых выносится суждение об интеллектуальности системы. В инвертированном ТТ факт приписывания ментального даёт основание судить, что судья-компьютер – человек, а не компьютер, надо эти факты собрать воедино и систематически сравнить со случаями приписывания ментального, как это делает человек. С. Ватт – сторонник муровской концепции оценки интеллектуальности систем [Moog, 1976, P. 249 – 257], согласно которой ТТ не является дефиницией интеллектуальности, он лишь способствует сбору данных для индуктивного вывода об интеллектуальности. Окончательное решение обусловлено социокультурными факторами. Инвертированный ТТ предлагает новый источник индуктивного доказательства интеллектуальности, базирующийся на правилах приписывания ментальности *другим*, акцентируя на исследовании того, как люди различают вещи, обладающие ментальным и не обладающие таковым. Проясняя роль судьи, которая в исходной версии ТТ совершенно непонятна, тест Ватта вводит в процедуру тестирования самого судью (судью-компьютера), у которого оценивается способность приписывания ментального. Наблюдатель становится наблюдаемым. Оценка осуществляется в обоих направлениях – от человека к компьютеру и от компьютера к человеку, а не только в одном (от компьютера к человеку). Уст-

раняется проблема объяснительной асимметрии – одностороннего процесса, – в котором оценивается лишь компьютер со стороны человека, а компьютер человека не оценивает.

Ранее мы уже упоминали ряд мнений по поводу невозможности организации инвертированного тестирования: 1) тесту присуща дурная бесконечность, так как оценивать судью-человека (который оценивает судью-компьютера) с позиции его человечность/компьютерности должен другой судья-человек и т.д. (Р. Френч [French, 1995]); 2) тесту присуща проблема параллелизма, так как невозможно обнаружить момент, когда судья-человек начнёт понимать различия между собой и судьёй-компьютером (С. Брингсйорд [Bringsjord, 1996]). Данная критика и собственно тест Ватта крайне важны для прояснения концептуальных аспектов компьютерного тестирования, а именно проблем, связанных с изучением приписывания студенту феномена понимания.

Нельзя применять компьютерные технологии для оценки знаний школьника, абитуриента, студента, преподавателя.

Рассмотрим ряд очевидных заключений, следующих из совмещения форматов двух частных тестов Тьюринга в концептуальных рамках КТТ₁ с целью изучения способности к пониманию и способности оценивать способность понимания в компьютерных образовательных технологиях.

Комплексный тест Тьюринга в исследовании «понимания»

В роли мыслительного инструмента тест представляется крайне эффективным для прояснения значений и смыслов слов «интеллект», «понимание», «сознание» и пр. КТТ – это средство различения человеком тех систем, которые обладают свойствами и способностями, обозначаемые этими словами, от тех вещей, которые этими способностями не обладают. КТТ предназначен идентифицировать сложные социокультурные феномены в отличие от оригинального ТТ, роль которого была ограничена «интеллектом».

Комплексный тест обладает мощным эвристическим потенциалом для осуществления концептуального анализа проблемы понимания. Простое комбинирование частных ТТ позволяет создать многообразие различных КТТ, применимых для нашего анализа. Например, тест КТТ₂ = [тест Сёрля] + [тест Ватта] + [тест Кирка – Чалмерса (тест Зомби)] способен прояснить путь дальнейшего развития сценария использования комплексного компьютерного тестирования в отечественном вузовском образовании (об этом сценарии немного ниже). КТТ₂ поднимает проблему демаркации сознательных систем от бессозна-

тельных в рамках проблемы другого. Для прояснения данной проблемы требуется построение некоторого КТТ_х, тестовые вопросы которого позволят решить, обладает ли *другой* сознанием, подобным тому, которым обладаете вы; понимает ли *другой* понятый вами материал; в большей или меньшей степени понимание *другого* соответствует реальности, нежели чем ваше; имеются ли возможности достижения идентичного понимания? В философии сознания превалирует убеждение, что проблема *другого сознания* на сегодняшний день теоретически не разрешима, хотя имеется целый ряд продуктивных концептуальных возможностей её частичного разрешения в будущем. [Дубровский, 2006-4, С. 153 – 171]. *До тех пор, пока не решена проблема другого сознания, ни о каком компьютерном тестировании как главной форме аттестации студента не может быть и речи.*

Однако ни КТТ, ни тем более средства компьютерного тестирования (как конкретные средства реализации КТТ) не являются дефинициями образованности тестируемого. Результаты компьютерного тестирования – это лишь подспорье в индуктивном рассуждении по поводу интеллектуальности, образованности студента. Окончательное решение об образовании человека обуславливается не техническими, а социокультурными факторами.

Проблемы студента-компьютера

Втискивание студента в узкие рамки тестовых вопросов, бездумное заучивание правильно подогнанных ответов, превалирование программно-технической составляющей коммуникации и формально–манипуляционных методик оценки качества образования в ущерб личному общению с преподавателем приводит к понижению значимости содержания изучаемого материала. Ущемляются смысложизненные, интенциональные и прочие культурно–антропологические компоненты коммуникации, обеспечивающие прочные связи учащегося не только с реальностью предметной области будущей профессиональной деятельности, но и природой, обществом, человеком. На уровне тестирования особенно остро проявляется экзистенциально-коммуникативная ущербность компьютерных образовательных технологий. Экзистенциальная коммуникация – духовная основа общественной жизни. В большей мере она основана на иррациональных составляющих общения, не поддающихся числовой выразимости. Формальная оценка, полученная как результат тестирования и смысложизненная оценка – несовместимые понятия. Формально «понимание» имеется. Фактически – отсутствует.

Проблемы преподавателя-компьютера

Вопросы, поднятые КТТ, адресуют к оценке сложных ментальных феноменов – способности понимать и способности приписывать понимание. Предполагается, что судья должен обладать более мощными способностями, нежели тестируемая система, интеллектуальную систему (таковой является студент-человек) невозможно протестировать неинтеллектуальными средствами (человеком-компьютером), хотя обратное возможно. Кто сегодня оценивает результаты тестирования, «судьи кто?». Конечно, преподаватель-человек, однако его статус понижен. Он вынужден обсуждать с коллегами отнюдь не педагогические проблемы, а то, как «натаскивать» школьников и студентов на все эти бесконечные ЕГЭ, ДОТ, ФЭПО¹. Преподаватель-человек стал преподавателем-компьютером! И его работу оценивает другой человек-компьютер – чиновник от различных ведомств, агентств, министерств – одномерный человек, наделённый мощной бюрократической властью. Инвертированный тест Тьюринга запустить не удаётся, везде тупик «вычислимости», отсутствие свободного созидательного творчества.

Проблемы применения компьютерных технологий тестирования вузовского образования

Вне всякого сомнения, допустима генерализация показателей частной компьютерной методики оценки механистических манипуляций с простенькой программкой. Представляют особую важность легализация на государственном уровне и легитимизация в общественном сознании посредством СМИ результатов подобного тестирования. Крайне значимо возведение данных тестирования в ранг основного показателя аттестации знаний, навыков и умений школьника, абитуриента, студента (в недалёком будущем – преподавателя), а также школы, вуза, региона, области и страны. Однако всё это принимается лишь только в том случае, если мы желаем жить в обществе зомби – квазилюдей, обладающих внешне сознательным поведением, но не использующих для этого ни толики ментальности. К сожалению, зомби уже сейчас среди нас, они появились сразу после развала существовавшей ранее системы образования. О зомби не надо много теоретизировать – посмотрите телевизоры, послушайте радио, войдите в интернет, оглянитесь вокруг, рассмотрите себя. Можно, конечно, принять и допустить широкий охват ЕГЭ, ДОТ и прочими компьютерными «инновациями» системы образования в целом. Всё это допустимо. Вот только биологические

¹ В 2009 – 2010 гг. учителя ряда московских школ отказывались проверять тетради учеников, так как им за это не платят, их задача – «натаскивание» на ЕГЭ. К сожалению, они правы. По стране наблюдается много случаев подобного рода.

потомки не простят нам общество российских зомби, если будут способны понимать факт зомбированности. Возможно, что они не будут это осознавать, так как из современной системы вузовского образования директивно исключено «понимание». Это мы можем достаточно ясно представить на основе анализа нашего КТТ1, а убедить – на основе перспективного КТТ2, построение которого автор намерен осуществить в скором времени.

Согласно тесту зомби, студент-компьютер (он же студент-зомби) может выдавать себя за человека, который освоил материал, хотя фактически понимания с его стороны нет. Компьютерные тесты ЕГЭ, ФЭПО, ДОТ позволяют обмануть преподавателя-компьютера. На самом деле студент ничего не понимает и лишь играет в «кости», проставляя случайным образом галочки в ответах. Преподаватель-компьютер не понимает, понял ли учебный материал студент либо не понял. Короче говоря, в рамках комплексного компьютерного тестирования никто ничего не понимает, однако имеются научно-обоснованные, сопровождаемые математическими формулами и диаграммами «юридические» основания, чтобы отчитаться перед чиновником-компьютером за процесс понимания. Чиновник-компьютер при этом «зависает» в пространстве калькуляций, которое не пересекается с когнициями преподавателя-человека и студента-человека. Однако он так же представляет цифровой отчет другому чиновнику-компьютеру по процессу понимания в его ведомстве. И так далее вверх и вширь по бюрократической иерархии. Очевидно, что в контексте ККТ «понимание» и «вычисление» вносят прямо противоположный когнитивный вклад в образование человека.

По всей видимости, ЕГЭ и ФЭПО будут упразднены со сменой политической идеологии, согласно которой система образования служит формированию не личности, но *потребителя*, как заявил вначале один одиозный министр образования и науки, оправдывая введение ЕГЭ. Второй министр вообще не стал оправдывать развал образования и науки. Полноценно развитая личность – идеал предшествующей идеологии образования – по причине антропологической константности стремления к самосовершенствованию, станет формироваться отдельно от калькулирующего института образования, и к государственно легитимным способам оценки её компетентности будут относиться столь же поверхностно, как супруги, живущие гражданским браком, относятся к документам ЗАГСа. Многие сегодняшние студенты подобным образом относятся к диплому – не как к свидетельству их знаний, но как к типовому бюрократическому документу.

Тестирование в системе образования следует применять в разумных пределах – в основном там, где требуется оперативная оценка поверхностного знания об изучаемом предмете. Необходимо мягкое внедрение компьютерных образовательных технологий, не наносящее ущерба ни личности учащегося, ни личности педагога. Человек – это не животное, реализующее когнитивные условные рефлексы, которые формируются на основе существующих компьютерных образовательных технологий. Богу – богово, человеку – человеческое, животному – животное, компьютеру – компьютерное. Здесь имеются в виду различные антропологические параметры человеческого существа – религиозные, собственно человеческие, биологические и вычислительные. Отчуждать один параметр в ущерб другим нельзя. Надо строго отдавать себе отчёт – когда, зачем и чем оценивать человека. Человек-компьютер оценить другого человека не способен, субъективную реальность калькуляцией сполна не охватить. Однако если снять тезис отчуждения, оставить человека человеком, разумно использующим свои вычислительные способности, реализуемые технологиями, тогда раскрывается свобода компьютерных новаций в сфере образования. Но сегодня рано говорить про отказ от тезиса отчуждения, учитывая текущее состояние искусственного интеллекта, который лишь в футурологической перспективе позволит интеллектуально насытить компьютерное тестирование и в более-менее полной мере уподобить человека-компьютера реальной человеческой личности.

Компьютер может и должен оценивать человека! Но сегодня следует резко ограничить сферу компьютерного тестирования оценкой чисто механистических параметров деятельности. Для внедрения комплексного компьютерного тестирования в вузовское образование необходимо сосредоточить усилия на междисциплинарных исследованиях искусственного интеллекта. Лишь после тщательной межвузовской и общественной экспертизы компьютерные интеллектуальные системы тестирования можно применять на практике. Результаты работы этих систем будут во многом способствовать *преподавателю-человеку* в принятии решения о компетентности *студента-человека*.

Выводы: преимущество нейтрально-зомбистской позиция при решении проблемы сознания в электронной культуре

В заключение отметим, что критические исследования э-культуры, проводимые на базе КТТ, как показал пример с э-образованием, обладают следующими несомненными достоинствами. Во-первых, они применимы для всех феноменов э-культуры, в которых позиционируется некоторая когнитивная проблематика, т.е. изучаются индивидуальные и социальные ментальные феноме-

ны наряду с инженерным проектом их компьютерной реализации – их имитации, моделирования или фактического воспроизведения. Во-вторых, когнитивные исследования крайне экономичны. Не надо никаких сложных дорогостоящих социологических, политологических, экономических, идеологических и других анализов. Просто следует подумать по поводу очевидных феноменов ментальной жизни в условиях э-культуры. На наш взгляд, надёжным руководством обоснования таких размышлений служат когнитивно-компьютерные мысленные эксперименты, систематически представленные в формате комплексного теста Тьюринга.

Проблема сознания в электронной культуре претерпевает сложные смысловые деформации, обусловленные стремлением воспроизвести, смоделировать или, по крайней мере, попытаться симитировать феномен сознания средствами электронных технологий. Семантика термина «сознание» варьируется и зависит от онтологических предпосылок – какую конкретную форму электронной культуры мы рассматриваем.

В э-культуре-1 понятие «сознания» целесообразно рассматривать в традиционном контексте, как это принято в учебниках по философии, культурологии, социологии, психологии и пр. Однако при этом обязательно следует учитывать эргономические коннотаты воздействия на сознание технологий. Например, каким образом влияют на психику ребёнка компьютерные игры. Или каким образом сознание человека, использующего компьютер, приобретает калькулирующий характер (метафора «человек-компьютер»). Здесь мы имеем бесчисленное множество социологических, психологических, педагогических и прочих исследований влияния электронных технологий на психику человека и на ментальность общества.

В э-культуре-2 понятие «сознания» необходимо дополнять различного рода парapsихологическими репрезентантами, включая концепции эзотерического плана. Например, что значить видеть и ощущать окружающую реальность таким способом, при котором мои глаза находятся, например, на коленной чашечке или в ином другом месте? В виртуалистике подобного рода модели достаточно широко обсуждаются. Возможно, это объясняет тот факт, что основатель отечественной виртуалистики Н.А. Носов, начиная научную деятельность с анализа трансформации сознания у лётчиков, обучающихся на тренажёрах, в конце пути научного творчества перешёл на чистую эзотерику и стал серьезно исследовать такие странные вещи, как «психология ангелов».

В э-культуре-3 понятие «сознания», на мой взгляд, следует рассматривать с нейтральных позиций. В противном случае, принимая сторону зомбистов, мы претендуем лишь на абстрактные эпифеноналистские споры относительно воображаемых клонов-симулякров сознательных существ, своеобразных «номологических бездельников». Как технологически создать бессознательные системы, ведущие себя сознательным образом, и при этом исключить из способа функционирования этих систем хотя бы минимум понятий и представлений о сознании? С инженерной точки зрения это вообще непостижимо.

Вступая в лагерь антизомбистов, мы начинаем строить гигантские проекты расшифровки таинственных нейтральных кодов сознания. Это чрезвычайно фундаментальная задача, и предполагаемые результаты её крайне масштабны для всего человечества. Здесь коренится возможность человеческого бессмертия, проникновение в сознание другого. Однако для этого требуется слишком сложная работа, которая под силу не отдельному учёному, научному институту или даже кооперации научных учреждений. Для этой задачи необходимо международное сотрудничество, действующее, по всей видимости, не одно столетие. Поэтому для расшифровки кода Дубровского необходим глобальный научный коллектив.

Нейтрально-зомбистская позиция ориентирована на построение реально действующих проектов. Инженерная проблема сознания в данном случае решается, но при условии риторико-метафорического отношения к понятию «сознания», реализуемого в системах электронных технологий. Чисто искусственные системы и даже симбиотические системы (совмещающие естественные и искусственные «механизмы») никогда не будут обладать «сознанием» в традиционном его понимании – всё это лишь имитации и метафорические эквиваленции «как бы» сознательной деятельности.

Автор придерживается позиции нейтрального зомбиста. Данная позиция позволяет особенно внимательно не задумываться над метафизическими проблемами когнитивно-компьютерной науки и призывает непосредственно переходить к рассмотрению наиболее важной роли, которую выполняет комплексный ТТ в философии искусственного интеллекта – к конструирующей функции. Данная функция служит целью обсуждения проектов, которые, собственно, способны обеспечивать реализацию когнитивных феноменов – осуществлять их имитацию, моделирование и репродуцирование компьютерными средствами.

3.3. Конструирующая функция

3.3.1. Философско-методологические аспекты: обоснование стратегической роли проблемы «творчество-компьютер»

Дефиниция проблемы творчества в исследованиях искусственного интеллекта

Проблема творчества в традиционной культуре – это постановка вопросов и поиск ответов следующего содержания. Что такое творчество? Как оно возможно? Какую роль играет творчество в жизни человека, общества, в мире в целом? Как отличить творческие действия от рутинных? Как оценить социокультурную значимость творческой деятельности? Какова его природа, состав, структура? Можно ли его описать, проникнуть в сущность, объяснить? Какие методы применимы для понимания феномена творчества? Что такое творческий артефакт? Каковы критерии оценки степени его оригинальности, неповторимости, социальной значимости и востребованности. Где истоки творчества, и каковы способы его интенсификации? Как идентифицировать творческого субъекта? Что такое гениальность? Как найти гения в наличных исторических условиях?

Список традиционных вопросов по проблематике творчества представляется безграничным (см., например, одну из последних работ по философии творчества В.И. Самохваловой [Самохвалова, 2008]). Эти и другие вопросы не имеют однозначных решений, исторически вариabельны, социокультурно и субъективно релятивны.

В современной, электронной культуре (э-культуре) проблематика творчества изменяется относительно понятий традиционной культуры. С одной стороны, мы видим расширение понятия, инспирированное технологическими новациями. С другой стороны, ориентир на компьютерные формы реализации творческих явлений сужает понятие, отбрасывая собственно человеческие, экзистенциальные представления о творчестве. Модификация термина «творчество», по сути нарушающая основной логический закон тождества понятия в процессе рассуждения ($a = a$), не мешает изучение данного явления перенести в сферу исследований ИИ и поднять вполне законно интригующий вопрос – можно ли феномен творчества реализовать посредством электронной технологии, где легальность реализации обеспечивается тьюринговым тестированием? Проблема творчества, как и проблема сознания, становится задачей инженерно-

го дела и целью технологии искусственного интеллекта (ИИ). Таким образом, в первом приближении, *проблема творчества в ИИ – как, чем и зачем осуществлять компьютерную реализацию феномена творческой деятельности – его имитацию, моделирование и/или воспроизведение*. Такая формулировка несколько ограничивает свободу представлений о творчестве и требует дополнения. Дело в том, что имитация, моделирование, воспроизведение – это степени градации онтологических притязаний компьютерной технологии по замене естественного искусственным. При *имитации* понятие интеллектуальности и, соответственно, креативности компьютерной системы, как правило, имеет релятивный статус – интеллектуальные способности, включая творчество, приписываются компьютеру исходя во многом из оценочных убеждений субъекта (наблюдателя, судьи), его пропозициональной установки. Несомненно, технологический инструментарий должен отвечать некоторым приемлемым требованиям организации диалогового взаимодействия человека с компьютером, однако, подчеркнём, «интеллект» во многом специфицирован степенью компетентности судьи, его объективностью, беспристрастностью и другими «юридическими» характеристиками. При компьютерном *моделировании* формулируются условия адекватности модели реальным креативным качествам, свойствам, способностям. При фактическом воспроизведении феноменов творчества – *репродуцировании* – предполагается, что компьютерная система фактически обладает «вдохновениями», «инсайтами», «интуицией», «свободными полётами мысли» и пр. Так или иначе, все эти онтологические притязания выявляются в более широком контексте представлений э-культуры и, как показано выше, зависят от исходной установки исследователей приписывать ментальность *x*-системам. Данная установка разделяется на следующие позиции: 1) зомбистскую – возможны бессознательные компьютерные системы, обладающие всем спектром ментальных свойств и способностей; 2) антизомбистскую – такие системы невозможны, для реализации сознательного существа на искусственной субстанции требуется репродуцирование ментальности; 3) нейтрально-зомбистскую – имитация и моделирование сознания пригодны для аппроксимативно полной реализации всего выявленного спектра когнитивных функций.

Отметим крайне важную идею, которую привносит проблематика творчества, существенно расширяя сферу ИИ, обновляя традиционные понятия. В компьютерную реализацию когнитивных феноменов включается ещё одна составляющая, которую обозначим термином «креация». *Креация* – это совокупность способностей и свойств компьютерной системы продуцировать феномены сверхантропологического и неантропологического статуса. Сверхантропо-

логические компетенции в некотором смысле очевидны – они превосходят индивидуальные человеческие способности, например, при компьютерной игре в шахматы нешахматиста. С пониманием неантропологических феноменов сложнее – это явления, несоизмеримые с человеческой ментальностью. Человек, в силу родовой предопределенности, не способен оценить способ функционирования компьютера как «творческий» – в человеческом понимании значения слова «творчество», подобно тому, как человек не способен оценить разум животных. Компьютер – это своеобразный Солярис, о разуме которого можно лишь опосредованно догадываться по следам-символам (см. прекрасную феноменологическую ремарку Н.М.Смирновой [Смирнова, 2006]). Однако ряд авторов, изучающих робототехническую проблематику реализации когнитивных феноменов (Дж. Маккарти, А. Сломан, С. Бригсйорд и др.), полагают, что с «ментальностью» компьютерных систем дело обстоит проще, нежели чем с «психикой» других систем, скажем, животных. Мы ведь многое знаем по поводу информационных процессов, протекающих в роботах, в то время как у животных, включая самих себя, об этих процессах можем догадываться только по аналогии, причем часто по аналогии с компьютером. Такая позиция возможна для нейтральных зомбистов.

Таким образом, с учетом вышесказанного, предложим следующую общую формулировку:

Проблема творчества в искусственном интеллекте состоит в том – как, чем и зачем имитировать, моделировать, репродуцировать и креацировать феномены творчества средствами компьютерной технологии.

Феноменология компьютерного творчества представляется распределённой среди всех перспективных направлений ИИ. В методологических исследованиях искусственной жизни изучается преимущественно витальный феномен – как посредством компьютеринга реализовать феномен жизни. В искусственной личности главной выступает проблема компьютерной реализации ментальных феноменов – квалиа, ощущений, восприятий, представлений, понятий, суждений, рассуждений и пр., а также способов «самособирания» этих феноменов до уровня эпистемологической адекватности, эстетического наслаждения, морального вменения, самости, личности, индивидуальности, Я. В искусственном обществе прерогатива исследований принадлежит социально-когнитивной феноменологии. Очевидно, что агенты этого общества связываются в многоагентную систему не путем механистической состыковки «по запросу», выделяющему типовые свойства и методы агентов, но за счет креативных процедур, осно-

ванных на менталистских параметрах – репрезентантах «потребностей», «мотивов», «целей», «намерений», «убеждений», «любви», «надежды» и пр. Конечно, здесь востребована проблематика социологического компьютеринга, которая на сегодняшний день находится в зачаточном состоянии – слишком велика разница между общетеоретическими социальными теориями и их когнитивно-компьютеринговыми инженерными приложениями. «Когнитивная культурология» с её представлениями о культуре как некоторой программной управляемой совокупности баз данных и баз знаний для нормального культуролога расценивается как сценарий фильма ужасов. В когнитивной социологии реализации феноменов творчества следует отвести значимую роль, по крайней мере, сравнимую с ролью, которую приписывают «творческой эволюции» в биологии и «творческой деятельности» в социальных науках.

Стратегия развития ИИ: проблема творчества vs проблема сознания

В философии ИИ, осуществляющей мировоззренческую рефлекссию и методологическую организацию научной и практической деятельности по построению и применению интеллектуальных систем, крайне важным представляется поиск фундаментных проблем, от решения которых надежно и основательно зависит успех в развитии всех обозначенных направлений ИИ. В отечественной методологии ИИ роль фундаментной проблемы играет *«проблема сознание – мозг»*. Проблему обозначил сопредседатель НСМИИ РАН проф. Д.И. Дубровский и, как в силу его авторитета, так и по объективным причинам (в самом деле, проблема сознания стратегически важна для судеб земной цивилизации – в этом автор полностью убеждён), проблема попала на топ-вершину междисциплинарных проблем ИИ. Суть состоит в компьютеринговом воспроизведении информационных зависимостей феноменов сознания от мозговых коррелятов. Однако справедливости ради отметим, что ряд видных исследователей ИИ, более полвека отдавших развитию этой технологии, ни о какой «проблеме сознания» ранее не слышали. Незначительную роль отводил «проблеме сознания» и основатель философии ИИ – А.М. Тьюринг, вскользь парируя контраргументы «от сознания» и «от *x*-возможностей» от линии защиты идеи компьютерной имитации мышления [Turing, 1950, Р. 445 – 450]. Первое возражение полущутливо снимается исходя из солипсической позиции – по поводу сознания другого человека мы делаем заключение лишь по аналогии со своим собственным сознанием и по большей части приписываем ему феномен сознания в силу вежливых условностей. Точно так же, как мы приписываем сознание другому человеку, мы можем приписать сознание и компьютеру, который оценива-

ется в условиях теста Тьюринга. Второй контраргумент – «от различных возможностей» – А. Тьюринг также посчитал незначительным. Эти возможности, по сути, редуцируются к частным аспектам сознания и к комбинациям этих аспектов. Как и предыдущий, данный аргумент опровергается не совсем серьезно: машине «любить клубнику» надо в той же мере, как и «сиять на конкурсах красоты». Собственно, зачем машине обладать сознанием? Машина должна работать и, в случае с ИИ, решать задачи, которые принято считать собственно интеллектуальными.

В современных инженерно-технологических разработках пока не место компьютерным механизмам, которые реализуют ментальные способности. Сегодня эти модели ничего не значат для эвристического извлечения и кодирования информации, оптимизации и рационализации процессов её репрезентации, хранения, поиска, манипуляции, отображения, передачи, для идентификации значимых проблем и достижения целей сложных динамических систем посредством адаптации к неопределённым условиям среды при нехватке ресурсов, для управления «знаниями».

Приписывание способу функционирования компьютера феноменов «переживания», «восприятия», «понимания» и пр. – это оценка функционирования *креативных программ* – программ, осуществляющих реализацию творческих актов. Имеется претензия на то, что *х*-система производит нечто оригинальное – переживание, представление, мысль, значит, есть повод приписать данной *х*-системе «сознание». Никак иначе. В далёком будущем, когда, возможно, будет расшифрован код Дубровского (широко известный в отечественной литературе подход к нейродинамическому кодированию ментальных феноменов), когнитивные модели «сознания» станут фундаментом ИИ, включая основы для формирования более совершенных креативных программ. Но даже в такой перспективе собственная идея развития ИИ будет предполагать необходимость креативного создания новых средств и методов, что означает использование креативных интеллектуальных систем. То есть идея креативности не снимается, даже если в будущем каким-то образом будет решена проблема «сознание – мозг – компьютер».

Для более чёткого позиционирования проблемы сознания и проблемы творчества поиграем в «слова», как это принято в лингво-аналитических исследованиях, проясняющих смысл и значения семиотических комплексов.

1. Для начала допустим, что понятие творчества включает понятие сознание. Для данной позиции ответим на вопрос – собственно, каким образом из массива знаний образуется «со-знание» как целостная интегративная форма со-

вместных «знаний», контент которой воспроизводит квалиативные, феноменальные, интенциональные, позиционные, доступные, нарративные и другие внутренние особенности сознания как субъективной реальности? Эволюционные, эмерджентистские, синергетические, системологические, аутопоэтические и прочие «механизмы», которые приближают нас к объяснению данного феномена, в обыденном словоупотреблении удобно и принято обозначать словом «творческие способности». Таким образом, «творчество» играет роль инфрапонятия, на базе которого возможно объяснение факта объяснения «сознания». Более того, «творчество», особенно в наиболее гениальных формах проявления часто приписывается сфере бессознательного или сверхсознательного. Здесь вновь срабатывает принцип несущественности сознания Фланагана, который теперь звучит не относительно интеллекта, а относительно феномена творчества: *«Творчеству сознание не нужно»*.

2. Изменим денотативные роли терминов. Теперь будем полагать, что творчество – это высшая функция сознательного существа, т.е. концептуально подчиняется «сознанию». При данной раскладке ролей мы видим, что творчество – частный акт сознания, пускай, наиболее значимый. Однако на основании чего можно приписать ментальные параметры х-системе, в частности, компьютерной системе? Лишь только по способу её функционирования, который оценивается как творческий! Оценка сознательной деятельности базируется на приписывании системе творческих способностей, если мы не способны приписать креативность результатам функционирования компьютерной системы, ни о каком приписывании «сознательного» говорить не приходится.

Следовательно, проблематика творчества даже при условии признания того, что данный феномен находится в концептуальном подчинении у феномена «сознание», имеет более важное значение для развития методологии ИИ, нежели чем проблематика сознания – этот феномен можно непосредственно оценить, в отличие от феномена сознания, который познается лишь интроспективно или по аналогии, что является логически некорректным способом аргументирования по поводу наличия/отсутствия сознания и, тем более, его параметров.

Таким образом, исходя из ряда очевидных положений, мы видим некоторую «искусственность» проблемы сознания в контексте исследований искусственного интеллекта. Проблема иллюзорна и не обладает практическим потенциалом даже в силу наших простых рассуждений. Большей значимостью для философских исследований ИИ и по показателям оценки поведения системы,

доминирующей проблемой, представляется *проблема творчества*. Собственно от первичного решения данной проблемы зависят успехи ИИ. Остальные проблемы, включая «проблему сознания», представляются вторичными и производными от проблематики компьютерной креативности. Отсюда следует тезис:

В междисциплинарных исследованиях ИИ первичной является проблематика творчества, но не проблема «сознание–мозг–компьютер».

При ответе на основной вопрос методологии исследований ИИ – что первично – проблема сознания или проблема творчества – отметим первичность креативной проблематики. Однако наряду с этим следуют указать на аргументативную слабость предложенной выше лингвоаналитической игры – это всего лишь «игра в слова». Более основательный довод формулируется средствами КТТ.

Проблема творчества в методологии комплексного теста Тьюринга

Комплексный тест Тьюринга позволяет позиционировать проблематику творчества в ИИ и классифицировать направления исследований, исходя из собственных когнитивно-компьютерных понятий, не прибегая к интуициям биологического, психологического, социологического, антропологического и тому подобного дискурса. Для классификационной задачи, которая следует за дефинитной задачей определения когнитивных феноменов, в первую очередь следует отграничить проблематику креативных систем от тех компьютерных систем, в которые не заложены принципы компьютеринговой реализации творческой деятельности. Как правило, при терминологическом обозначении нетворческих компьютерных систем используются психологические метафоры, характеризующие интеллектуальные отклонения.

Обозначим класс компьютерных систем, которые номинативно обозначаются термином «интеллектуальные» (однако способу функционирования которых сложно или невозможно приписать креативную способность), термином *кретинные системы*. Организация таких систем не включает концепций интеллектуального развития, т.е. самостоятельного прироста знаний. В качестве подкласса кретинных программ можно указать *савантистские системы* – узкоспециализированные системы ИИ, настроенные на конкретную предметную область или под специализированные знания конкретного человека, которого предполагается обыграть в условиях ТТ. Психиатрический синдром «саванта» обозначает психическое отклонение, наблюдаемое, как правило, у аутистов, при котором люди, выказывающие гениальные способности в одной узкоспециализированной сфере деятельности (невероятный объем запомненной ими инфор-

мации, огромная скорость арифметических вычислений и пр.), в других областях крайне ограничены, вплоть до неспособности жить. На наш взгляд, к савантистской программе относится Deep Blue, шахматный суперкомпьютер фирмы IBM, который в 1997 г. одержал победу в матче из 6 партий с чемпионом мира по шахматам Г. Каспаровым.

В контексте КТТ наиболее полно кретинные системы представлены в рассмотренном выше параноидальном тесте (тесте Колби). *Параноидальная система* – программа, имитирующая способ мышления и общения параноика – человека, страдающего подозрительностью, хорошо обоснованной и стабильной системой сверхценных идей, которая порой выражается бредом. Параноидальные системы – это все существующие системы, построенные в рамках репрезентативной парадигмы. В самом деле, компьютерная онтология этих систем навязывается концепцией программиста и за границы её не выйти, так как нет саморазвивающихся средств, встроенных в программно-информационные средства. Другое дело – системы ИИ, построенные в рамках коннекционистской парадигмы. Здесь Д. Деннетт обращается к идее антропоморфных роботов Р. Брукса, в особенности, к роботу КОГ. Для данного видного исследователя философии ИИ очевидно, что только нейрокомпьютерные системы способны отвечать приемлемым конвенциям интеллекта – «знания» в этих системах «активны» и «свободны».

Мы также полагаем, что кретинные системы не относятся к системам ИИ, хотя часто называются «системы искусственного интеллекта». Например, это упомянутая в тесте Маккарти классическая экспертная система MYCIN, а также системы, комплексно продуцируемые индустрией «динамических экспертных систем реального времени».

Креационистские системы

Потивоположностью кретинных систем, с точки зрения проблематики творчества, являются системы, в которые заложена идея творения. Данный класс программ обозначим термином «креационистские системы» и соответствующее им направление в философии ИИ назовем *креационизмом* (от лат. «творение»). В общем случае данное направление порождено комбинацией идей ТТ с идеями фон Неймана по поводу реализации самовоспроизводящихся машин. Компьютерный креационизм крайне востребован в проектах «искусственное общество» и «искусственный мир». Здесь мы видим своеобразную теологическую окраску исследований ИИ – компьютер как бы творит «бытие» из «ничто», т.е. той субстанции, которая в различных мирах обладает общими инфра-

системными особенностями. Конечно, «богом» в этой схеме выступает человек-программист.

Тест Барессии – креационистский тест Тьюринга

Для изучения проблематики творения наиболее значимый частный ТТ представлен в работе Д. Барреси «Перспективы Кибериады: границы человеческого самопознания в кибернетическую эпоху» [Barresi, 1987], опубликованной в «Журнале теории социального поведения». В истории тьюрингового тестирования данная модификация ТТ получила название «Кибериадный тест Тьюринга». Фоном для рассуждений Дж. Барреси служит серия рассказов польского писателя-фантаста С. Лема, считающего их философскими притчами о технологии будущего, опубликованных с 1964 по 1979 гг. и объединённых общим названием «Кибериада» (отсюда, собственно, и название теста). Дж. Барреси ссылается на работы, которые связаны общей картиной фантастической псевдосредневековой вселенной, где место людей занимают роботы.

Барреси продолжает философский путь С. Лема, но конкретизирует его рассуждения применительно к ИИ в целом и к ТТ в частности и выдвигает ряд оригинальных философско-прогностических идей как по поводу постчеловеческих ориентиров ИИ, так и по приложению их к проблематике творения в исследованиях ИИ. У Дж. Барреси просматриваются два направления проблематики творения искусственного мира – универсальное и эволюционное. В *универсальном креационизме* изучается компьютерная реализация акта «творения» посредством машины-бога, который «всезнает» о том, каким должен быть «творимый» им мир, т.е. компьютер-творец обладает абсолютным знанием и всесовершенными инструментами для творения «себе подобных по своему образу». В *эволюционном креационизме* за инженерную реализацию отвечает «машина-природа», в которой действуют «законы» мутации и селекции компьютерных средств. Здесь нет претензии на универсальность знаний о творении «третьего мира», искусственная система служит средством самопознания творческих способностей субъекта, творящего «искусственный мир».

Как для универсального, так и для эволюционного креационизма отметим две схемы: горизонтальную и вертикальную. *Горизонтальный креационизм* выдвигает концепцию, согласно которой искусственный мир создается параллельно естественному. Искусственные личности этого мира (роботы, киборги, арtilекты и пр.) рождаются и творятся в контекстах, всем нам известным, в нашем мире. Здесь реализуется тезис А. Тьюринга по поводу того, что в будущем компьютеры будут замещать человека во всех сферах общественной жизни [Turing,

1950, Р. 460]. Другое дело – вертикальная схема. *Вертикальный креационизм* совершенно захватывает дух – человек творит искусственный мир искусственных личностей. Далее искусственные личности искусственного мира 1-го порядка творят искусственный мир 2-го порядка и т.д. Несомненно, это крайне важные направления для развития технологического базиса э-культуры, т.е. НБИКС-комплекса.

Отметим, что кибериадный ТТ не имеет репрезентативной интерпретации. *Креационистские системы исключительно нейрокомпьютинговые*. Творимый мир в рамках горизонтальной схемы, либо мир, в котором творится искусственная система, представлены в формате огромной нейронной сети. Компьютируемая самость формируется в областях наибольшей коннекционной интенсивности. Однако такой взгляд требует последующего прояснения и выходит за рамки данной работы.

Для прагматически ориентированных «творческих» компьютерных систем ни кретинные, ни креационистские схемы не интересны. Кретинные системы слишком «глупы», креационистские системы слишком фантастичны. Интерес представляют собственно «творческие» системы ИИ, креативные системы.

Креативные системы

В креативной системе компьютеринговый способ реализации феномена творчества представляется основным принципом функционирования и/или внутренней организации. В философии ИИ проблема создания креативных систем начинается с оригинальных работ А. Тьюринга [Turing, 1950, 1964]. Здесь мы наблюдаем три класса креативных систем: машина-оракул, машина-взрослый, машина-ребёнок.

Предложения и последующие дискуссии по первому классу систем породили направление в философии ИИ, которое мы обозначим как *универсальный креативизм*. Они создаются и функционируют в соответствии с универсально-эпистемологической парадигмой роста «знаний» и детально рассмотрены выше. В этих исследованиях мы видим проект компьютерной системы, творческие способности которой превосходят человеческие в социально-эпистемологическом смысле. Такой компьютер получил название «машина-оракул». Линию обсуждения конструкции «машина-ребёнок» обозначим как *эволюционный креативизм*. Они развиваются в соответствии с эволюционно-эпистемологической парадигмой роста «знаний». В дискуссиях данного направления верхом совершенства компьютерной системы предполагается программная реализация способа социализации ребёнка – первоочерёдно овладе-

ние речью (в робототехнике – «псевдоречью») и стереотипами поведения окружающих людей (копирование поведения, частично воплощённое в проектах антропоморфных роботов).

Вариант «машина-взрослый» предполагает совмещение универсального и эволюционного подхода в формате общего проекта, и направление исследований креативных систем обозначим как *умеренный креативизм*. Последний вариант – гибок и представляется наиболее перспективным для стратегии построения интеллектуальных компьютерных систем.

Выше данные классы систем были рассмотрены с позиции наличия в инженерной конструкции проекта компьютера эволюционно-эпистемологических механизмов. Проблема творчества, таким образом, представляется феноменологическим аспектом изучения перспектив интеллектуального компьютеринга, ЭЭ-механизмы – субстратным аспектом, наличие их в системе концептуально необходимо для того, чтобы данная система могла пройти креативный тест.

Универсальный креативизм

Универсальный креативизм в контексте ФИИ означает построение и использование компьютерной системы, в которую разработчик (эксперт, когнитолог, программист и др.) априори вложил совокупность таких возможностей, которым наблюдатель (судья) готов приписать параметры творческой человеческой деятельности. Универсализм по аналогии с биологическими теориями означает то, что система самостоятельно не развивается, её разработчик всё учел и «все проник». Здесь возможна классификация креативных компьютерных систем по специфике предметных областей (игра в шахматы, сочинение поэм, музыкальных произведений и пр.) Также следует выделить в качестве идеального варианта абсолютно универсальную систему. Компьютеринговая мощь её достаточна для имитации любого творческого проявления человеческого гения. Однако возможна ли такая машина? Вспомним то, что ещё Р. Декарт признавал невозможность создания такой абсолютной машины – он полагал, что машину всегда можно будет легко отличить от человека, если протестировать её в широком тематическом диапазоне.

Сегодня в связи с глобализацией системы управления «знаниями» правоту Декарта оспаривает интернет. На этот счет имеется важный мысленный эксперимент – тест Н. Блока. Данный тест ранее фигурировал для обоснования бесперспективности универсально-эпистемологических механизмов развития «знаний». Здесь мы приведём пример теста Блока с целью раскрытия важности выбора проблематики сознания как стратегического пути развития ИИ.

Тест Блока: подтверждение преимущества проблемы творчества перед проблемой «сознание-мозг-компьютер»

Результаты данного теста двусмысленны, хотя Н. Блок этого не заметил, так как не исследовал креативных параметров тестирования. В этом тесте наблюдатель (судья), приписывающий ментальность x -системе, как и в оригинальном ТТ, не может наблюдать физиологические параметры игроков. Но теперь он играет с людьми всей земли – теми, которые, по крайней мере, выкладывают контент в интернете и реализуют поисковые алгоритмы информационно-аналитических технологий. Предполагается, что на любой запрос наблюдатель может получить ответ – кто-то же должен знать, что ответить. На самом деле это не так, имеются нетривиальные вопросы, на которые некому ответить. Это предполагает либо творческую работу над собой, либо работу креативных специалистов-разработчиков, способных ответить на вопрос, если вообще на него можно ответить.

Н. Блок почему-то не замечает явных способов организации творческой деятельности и действует с метафизических позиций теории сознания, которую он назвал «психофункционализм». Данная разновидность функционализма сегодня считается классической среди англо-американских исследователей философии сознания. Блоковская концепция сознания означает невозможность указать однозначное отображение между входом и выходом системы. Попытка описать такое отношение невозможна без учёта контекста всей совокупности психических свойств и способностей. Это порождает континуум способов преобразования входных данных в выходные результаты и редукцию исходного процесса к субпроцессам, последних к суб-суб-процессам и т.д. Данная концепция была задумана Н. Блоком для антибихевиористского опровержения оригинального ТТ и, по мнению автора, не позволяет однозначно сконструировать инженерные проектные решения. То есть глобальная универсальная креативная система (типа мыслящего интернета) невозможна. Конечно, это так. Только непонятно, зачем Н. Блоку потребовалась апелляция к метафизике сознания, ведь намного проще неработоспособность универсальной статичной системы объяснить посредством более очевидных положений, исходя из отсутствия в ней *прироста знаний*. Собственно здесь мы находим *подтверждение приоритета проблемы творчества над иллюзорной и схоластической проблемой сознания в исследованиях ИИ*. С другой стороны, находим подтверждение тезиса о важности методологии комплексного теста Тьюринга в философии ИИ.

Чтобы прояснить более интересные варианты креативных систем – эволюционной (машина-ребёнок) и умеренно-универсальной (машина-взрослый), – обратимся к оригинальному ТТ.

Проблематика творчества в оригинальном тесте Тьюринга

Проблема творчества встала перед Тьюрингом при обсуждении основного вопроса философии ИИ – «может ли машина мыслить?». Напомним полемический стандарт, предложенный А. Тьюрингом для ответа на возможные возражения по данному вопросу. К дискуссии относятся следующие темы: теологическое, антисциентистское («голова в песке»), математическое, «аргумент от сознания», «аргумент от х-невозможностей», «возражение леди Лавлейс», «аргумент от непрерывности нейронной системы», «аргумент от неформальности поведения», «аргумент от экстрасенсорного восприятия» [Turing, 1950, P. 443 – 454].

Среди аргументов наибольшую значимость А. Тьюринг приписывает *аргументу Лавлейс* (АЛ) [Turing, 1950, P. 450]. С обсуждения данного аргумента берет отсчет проблема творчества в ИИ и начинается исследование компьютеринга феноменов творчества.

Возражение леди Лавлейс буквально таково: «Аналитическая машина не претендует ни на что оригинальное. Что бы машина не сделала, нам известен способ, как она это сделала» [Turing, 1950, P. 450]. То есть машины не способны к творчеству, так как сами по себе не могут создать ничего нового, они лишь следуют программным инструкциям. Все «творческое» предопределено программистом.

Важность роли, которую А. Тьюринг придавал аргументу Лавлейс, проявляется не только в объеме соответствующей части статьи, но и в том, что предлагается концептуальный проект компьютерной системы, которая способна имитировать и моделировать феномен творчества – нейрокомпьютерный проект машины-ребёнка и гибридный проект машины-взрослого.

Ранее мы выделяли два варианта ответа на возражение А. Лавлейс – репрезентативный и коннекционистский. Коннекционистской парадигме посвящается оставшаяся часть работы. Поэтому остановимся подробнее на репрезентативном опровержении.

Тест Лавлейс: аргументативная слабость репрезентативной парадигмы креативного компьютера

Репрезентативное опровержение АЛ, предпринятое А. Тьюрингом, на мой взгляд, недостаточно обоснованно, так как все его доводы не удовлетворяют минимуму требований теории аргументации, т.е. логически некорректны.

1. А. Тьюринг перефразировал АЛ и высказал следующее: «чтобы машина ни делала, на самом деле она ничего нового не делает» [Turing, 1950, P. 450]. По поводу этого довода он выдвигает «контраргумент»: «Под солнцем ничего нового не происходит... Всё, что мы считаем оригинальным, на самом деле представляется результатом обучения или следствием известных общих принципов» [Turing, 1950, P. 450].

2. А. Тьюринг вновь перефразировал АЛ: «машины никогда нас не удивят». Теперь он возражает: отнюдь, машины его удивляли очень часто, например, если они сбивались во время работы.

3. Возражение АЛ отводится ещё и потому, что человеческая способность к творчеству непонятна, невозможно найти подлинного автора идеи, так как помимо автора, крайнего на данный момент времени, имеются и другие «авторы», которые его обучили и воспитали. Совершенно не очевидны критерии оригинальности. Даже если не брать во внимание компьютерные сбои, результаты работы машины способны поразить любое воображение.

Несмотря на то что исследователи ТТ часто удивлялись «удивлению Тьюринга», т.е. тому, что компьютерные сбои преподносят сюрпризы, которые расцениваются как акты творчества, следует отметить, что в научно-фантастической литературе и в кинематографе к этим нарушениям относятся вполне «серьёзно». В результате некоторого программного сбоя у роботов, как правило, возникает сознание и самосознание. Здесь играет роль интуиция в вопросе о креативности феномена сознания, о возможности его возникновения за пределами алгоритмической регулярности. Если обратиться к практике программирования, то следует отметить, что ошибка имманентна программе – кто программировал, прекрасно знает, о чем идет речь. На этот счет имеется суровый «закон Мерфи», согласно которому в любой программе, в которой выявлено n ошибок, имеется как минимум $n+1$ ошибка. Так что программы удивляют не из-за сбоя в технических средствах, как это было во времена А. Тьюринга, но и по причине специфики собственно процесса разработки программ. У программистов, которые рассматривают свою профессиональную деятельность как форму творчества, применяется интересный термин «клудж». Это сотканный из разных модулей программный код, который, несмотря ни на что, работает. Сам разработчик не может понять, почему программа работает. Клудж, конечно,

удивляет – каким образом вдруг неожиданно начинает работать комплекс, состоящий из набора подпрограмм, каждая из которых не работает в силу несоблюдения соглашений между локальными и глобальными переменными, или, например, отношений между формальными и фактическими параметрами, не говоря уже о других причинах. Чтобы клуджевую программу внедрить, требуется немного – более-менее точно определить условия, при которых комплекс клуджей устойчиво функционирует, и разработать соответствующую документацию.

4. Ещё одну интерпретацию репрезентативного опровержения АЛ следует обозначить как «аргумент от неведения». Проблематика компьютерного творчества изучается А. Тьюрингом на базе компьютера, который подразумевается в контексте категорических суждений А. Лавлейс. Имеется в виду аналитическая машина Ч. Бэббиджа (Babbage's Analytical Engine). Здесь А. Тьюринг солидарен с первой программисткой – у машины Бэббиджа отсутствует возможность «самостоятельно мыслить» и её аргумент правомочен, если относится к машине Бэббиджа. Тем не менее, можно предположить, что у других компьютеров могут проявиться самостоятельное мышление и творчество. Однако такие машины появились после смерти Лавлейс, т.е. её аргумент имеет силу лишь по незнанию возможностей компьютерной технологии. Аналитическая машина Бэббиджа, как и ряд других, представимы универсальной цифровой вычислительной машиной. Если такой компьютер правильно запрограммировать, то он будет способен моделировать любую машину, в том числе и машину Бэббиджа. Таким образом, помимо возможности продуцирования оригинального вывода, даже заявленная претензия на универсальность придаёт ЦВМ креативные способности – она имитирует многообразие специальных ЦВМ, и переход от одного спецвычислителя к другому трактуется как некоторый творческий процесс (в смысле универсального креативизма).

Вся вышеприведённая аргументация, помимо логической некорректности, совершенно неприемлема для дискуссионных обсуждений, так как А. Тьюринг не представил списка требований к символьной машине, которым должна соответствовать машина, способная пройти тест Лавлейс.

Подведём итог репрезентативной парадигме креативных компьютеров. Очевидна логическая некорректность аргументации против АЛ. Иное дело – коннекционистское опровержение. Здесь А. Тьюринг основательно останавливается и предлагает концептуальный проект нейрокомпьютеринговой машины, который позволяет квазиэмпирически, в контексте мысленного эксперимента,

отвести аргумент Лавлейс. Данное опровержение мы рассмотрели выше. В дальнейших разделах этого исследования раскрывается нейрокомпьютерный проект (машина-ребёнок) и приводятся соображения по созданию машины-взрослого.

В заключение подраздела ещё раз подчеркнём доминирующую роль проблемы творчества в междисциплинарных исследованиях ИИ. Проекты построения креативных систем более фундаментальны, нежели чем схоластические концепции компьютерной реализации иллюзорного феномена сознания, так как (акцентируем!) сознание оценивается по креативным проявлениям, но не наоборот. Именно на решение проблемы творчества следует направлять философские, научные и инженерные усилия, если мы желаем достичь наискорейшего успеха в развитии технологии ИИ в частности и в развитии электронной культуры в целом посредством комплексной компьютерной реализации когнитивных феноменов. В этих исследованиях необходимой составляющей представляется нейрокомпьютинг феноменов творчества.

Далее в работе исследуются способы компьютерной реализации творческих процессов, выдвигается тезис о приоритетности проблематики творчества перед другими направлениями философской методологии ИИ, а также теоретически обосновывается интуитивно очевидное положение о преимуществе нейрокомпьютерного программирования креативных систем перед классическим, лингвокомпьютерным программированием.

3.3.2. Научно-методологические аспекты: нейрокомпьютинг – базовая технология искусственного интеллекта

Выше раскрывалось положение, что именно нейрокомпьютинг обеспечивает решение доминантной проблемы ИИ – проблемы творчества. Проблема творчества, т.е. изучение способов компьютерной реализации креативных процессов, представляется доминирующей в методологических исследованиях искусственного интеллекта. Также, помимо классификации креативных систем, имеющей в большей мере теоретико-концептуальный интерес, на основе ряда частных тестов Тьюринга было показано, что, в отличие от креативного лингвокомпьютинга, креативный нейрокомпьютинг обеспечивает несравнимо большие возможности имитации продуктивной деятельности в силу коннекционистской «свободы» организации квазиалгоритмических программ продуцирования оригинальных информационных продуктов.

В данном параграфе идея применения методологии КТТ конкретизируется относительно общенаучной, теоретико-эмпирической методологии. С этой целью нейрокомпьютер представлен в формате концепта методологической парадигмы, совмещающей метафизическую и инженерную проекции на социокультурную проблематику. Доказывается тезис детерминации нейрокомпьютером материального базиса и духовной составляющей электронной культуры (э-культуры), так как нейрокомпьютинг в большей степени, нежели чем другие направления ИИ, обеспечивает решение психотехнической проблемы, суть которой – имитация или воспроизведение феноменов сознания посредством НБИКС-технологий. Доказывается, что приоритет в нейрокомпьютинге принадлежит отечественным разработкам.

Нейрокомпьютерная тематика в различных формах культуры рассматривается по-разному. В традиционной культуре нейрокомпьютер оценивается, как правило, с технических и инструментальных позиций. В электронной культуре акценты смещаются – нейрокомпьютерная тема становится значимым разделом гуманитарных и социальных наук (в частности, культурологи), изучающих факторы, тенденции и закономерности развития информационного общества. Такая роль приписывается ей в силу того, что нейрокомпьютер – исторически первое и одно из главных направлений искусственного интеллекта. Искусственный интеллект, в свою очередь трактуемый как сложная многоуровневая система междисциплинарных исследований, становится главным фактором становления высокоразвитой электронной культуры [Алексеев, 2009; Гейбатов, 2010]. В связи с

этим и возрастает роль нейрокомпьютерных исследований. Из технического средства эффективного моделирования в традиционной культуре, нейрокомпьютер становится детерминантом социокультурного процесса в э-культуре. В работе аргументируется это суждение. Назовём его *тезисом нейрокомпьютеринговой детерминации* электронной культуры.

Проблема концептуальной совместимости терминов «нейрокомпьютер» и «культура»

Обоснование тезиса о нейрокомпьютере как детерминанте социокультурного процесса в э-культуре имеет место только в том случае, если термины «нейрокомпьютер» и «культура» в смысловом отношении представляются равнопорядковыми. Дело в том, что, на первый взгляд, «нейрокомпьютер» по статусу предмета моделирования – биологической нейронной сети – концептуально несоразмерен явлениям социальной и культурной сфер жизни общества. В условиях традиционной культуры термин «нейрокомпьютер» может употребляться в контексте более-менее удачных метафор, переносящих значения слов из нейрофизиологической области в социальную сферу. В результате этого мы имеем неисчислимое количество спекуляций, например, по поводу аналогии лево-правополушарного расщепления с культурологической дихотомией «Запад-Восток». В случае, который более приемлем для практических целей, нежели чем метафорические сравнения, это – использование нейрокомпьютерной парадигмы для осуществления эвристических приёмов изучения общества. (Данное направление применения нейрокомпьютера достаточно подробно раскрыто В.М. Розиным [Розин, 2011].)

С другой стороны, в э-культуре происходит смысловая реконструкция понятий «культура», «сознание», «интеллект» и др. В условиях данной модификации «нейрокомпьютер» не рассматривается в узком эмпирическом смысле как конкретное инженерно-техническое средство решения интеллектуальных задач, распознавания образов, кластеризации, дискриминантного анализа, многопараметрической нелинейной оптимизации, адаптивного управления или эффективного программного параллелизма и т.п.

«Нейрокомпьютер» – это концепт методологической парадигмы, совмещающей две проекции на социокультурную проблематику:

1) *метафизическая проекция*: нейрокомпьютер как способ реализации естественной или искусственной нейронной сети – это концептуальный базис коннекционистской парадигмы сознания и интеллекта;

2) *инженерная проекция*: нейрокомпьютер – это мысленный пример, погружённый в контекст многочисленных мысленных экспериментов, наработанных за шестьдесят лет развития философии искусственного интеллекта, т.е. в контекст тестов Тьюринга (название проекции инспирировано альтернативным, «калькирующим» способом перевода статьи А. Тьюринга [Turing, 1950] как «Компьютерная инженерия и интеллект»).

Две методологические проекции – метафизическая и инженерная – собственно и обеспечивают концептуальную совместимость терминов «нейрокомпьютер» и «культура» и, тем самым, реальное вовлечение нейрокомпьютерной тематики в социокультурные проблемы информационного общества, проблемы становления и развития электронной культуры.

Роль нейрокомпьютинга проявляется в рамках нейтрально-зомбистской проблематики. Из трех стратегий приписывания ментальных характеристик компьютеру она предпочтительнее, так как зомбистские эпифеноменальные конструкции в инженерном плане совершенно абсурдны, антизомбистская программа предполагает решение сверхсложной глобальной фундаментальной задачи расшифровки нейральных кодов психической активности и искусственное воспроизведение кодовых зависимостей. Нейтрально-зомбистская позиция ориентирована на построение реально действующих проектов, и психотехническая проблема сознания решается как проблема «псевдосознания», функционально имитирующего представления о феноменах реального сознания.

Следуя нашей логике доказательства *тезиса детерминации*, раскроем роль нейрокомпьютера в ИИ. Как ранее утверждалось, термином «нейрокомпьютер» обозначается концепт методологической парадигмы, совмещающей метафизическую и инженерную проекции.

Метафизическая проекция нейрокомпьютерной тематики в электронной культуре

Нейрокомпьютерная тематика в контексте метафизических теорий сознания представляется крайне плодотворной для построения продуктивных метафор интеллектуальных систем на нейросетевой основе. К метафизическим теориям сознания принято относить дуализм, бихевиоризм, физикализм, функционализм и пр. (см. неплохой обзор Э. Вильянуэвы, переведённый на русский язык [Вильянуэва, 2006]). За нейрокомпьютерную тематику «отвечает» самостоятельная парадигма – коннекционизм (в литературе встречаются иные транскрипции – «коннектизм», «коннективизм», «коннекшизм»). Коннекцио-

низм – это направление в философии сознания и философии искусственного интеллекта, в рамках которого предпринимаются попытки объяснить феномены сознания, а также интеллектуальные способности человека, используя концепцию нейронных сетей. Составленные из большого числа структурных единиц, каждая из которых аналогична нервной клетке определённого типа, с заданным для каждого элемента весом, определяющим силу синаптической связи с другими элементами, нейронные сети представляют собой упрощённые модели мозга. Точнее было бы сказать, что нейронная сеть – метафора человеческого мозга. «Нейрон», в свою очередь, удобная аналитическая конвенция, аналогичная «атому», «гену». Коннекционизм, как правило, ожесточённо соперничает с репрезентационизмом, теорией языка мысли и другими символично-ориентированными парадигмами сознания.

Здесь возникает два вопроса: 1) как понимается «сознание» исследователями нейрокомпьютинга; 2) где прорисовывается нейрокомпьютинг в общей панораме метафизических теорий сознания? Некоторые соображения по решению этих вопросов находятся среди метафизических теорий сознания.

Роль и место нейрокомпьютинга среди метафизических теорий сознания

Нейрокомпьютер как способ реализации естественной или искусственной нейронной сети – это концептуальный базис коннекционистской концепции интеллекта и сознания.

Роль нейрокомпьютера в метафизической проекции нашего тезиса детерминации особенно чётко выявляется в спорах между «символистами» и коннекционистами. Первых считают сторонниками классического подхода к ИИ («старому доброму ИИ»). Вторых, отстаивающих приоритетную роль нейрокомпьютинга в ИИ, называют сторонниками неклассического подхода. Эти споры достаточно подробно освещены в работе Дж. Гарсона [Garson, 2002] (см. также размышления М. Кольцова по этому поводу [Кольцов, 2006-1; 2]).

«Классики» выражают точку зрения, которая доминировала тридцать лет с момента возникновения ИИ – когнитивные процессы аналогичны символьной обработке в цифровых компьютерах. В классическом рассмотрении информация располагается в виде строк символов, т.е. так же, как мы организуем данные в памяти компьютера или на листе бумаги (ленте машины Тьюринга). Процесс мышления и познания имеет сходство с обработкой информации цифровым процессором, при которой строки обрабатываются в последовательности, определяемой инструкциями программы (также организованной в символьном виде).

«Коннекционисты» (неоклассики) предполагают, что информация, как у человека, так и в компьютере хранится в несимвольном виде и определяется весовыми соотношениями узлов нейронной сети. Когнитивный процесс – динамический, гетерархический процесс, выполняющийся в нейронной сети, при котором активность каждого узла зависит от силы связей и активности соседних узлов, в соответствии с активизирующей функцией.

Коннекционистов следует разделить на три лагеря.

1. *Радикальные коннекционисты.* Они всячески стремятся уничтожить модель символьной репрезентации информации как направление когнитивной науки и философии ИИ. Классическая теория обеспечивает лишь крайне низкий уровень целостности представления данных, слабую способность к обобщению информации, распознаванию контекста и к выполнению многих других функций человеческого интеллекта. Провал классического программирования по сравнению с гибкостью и продуктивностью человеческого сознания – это симптом того, что в ИИ необходима новая парадигма.

2. *Практические коннекционисты.* Они не рассматривают нейрокомпьютинг как вызов классической теории, более того, выступают в поддержку классических представлений и ищут возможности согласования двух парадигм между собой. Несомненно, мозговая активность репрезентируется нейронной сетью. Однако имеется также символьный процессор на более высоком и абстрактном уровне описания. Для этого процессора вся необходимая информация формируется как результат функционирования нейронной сети. Например, роботы, предназначенные для работы в некотором «сообществе», «колонии» роботов, имеют два уровня организации когнитивно-компьютерных механизмов. Техническое зрение таких роботов организовано нейросетевыми средствами. Но, средствами символьного процессора и, соответственно, «гибридными» механизмами связи асимболического и символического уровней представления информации, эти роботы самостоятельно должны создавать свой собственный, внутренний для колонии «язык» с целью координации «взаимодействий» в сложных условиях среды. Несомненно, более приемлем подход практических коннекционистов, вводящих способы классической обработки информации в модели нейронной сети. Они синтезируют наработки в области лингвистической философии с сублингвистическим подходом. С другой стороны, современные «символисты», усматривая узость классических репрезентативных систем знаний (продукционной, фреймовой, семантико-сетевой и формально-логической моделями), подобным образом ищут теоретических и инженерных

стыковок с нейрокомпьютерными моделями ИИ. В связи с этим подход практических коннекционистов и практических символистов называется *гибридным ИИ*. Примером гибридных систем являются многоагентная методология проектирования и программирования интеллектуальных систем – основы для реализации моделей искусственного общества. Нейросетевой подход к искусственным обществам широко представлен в работах В.Л. Макарова, А.Р. Бахтизина и в деятельности лаборатории «Искусственные общества» (см. сайт <http://www.artsoc.ru>). Совершенно нелепо подходить к многоагентной методологии без учёта нейрокомпьютерной тематики.

3. *Критические коннекционисты*. Исследуя фундаментальные проблемы нейрокомпьютерной тематики (в первую очередь, что суть «нейрон»), эти исследователи коренным образом изменяют представления о материальных коррелятах человеческого сознания. Отсюда возникают и новые возможности решения психотехнической проблемы нейрокомпьютерными средствами.

Критических коннекционистов, в свою очередь, следует условно разделить на два класса – гуманитарных и технических.

1. *Гуманитарные критические коннекционисты*. Эти исследователи подходят к нейрокомпьютерингу с позиций социальных и гуманитарных наук. Нейрокомпьютер – это инструмент изучения внутреннего мира человека, языка, когнитивных механизмов индивида и социума. К коннекционистам данного класса следует отнести известного отечественного когнитивиста – Т.В. Черниговскую (см. тему «Язык, мозг и компьютерная метафора» в [Черниговская, 2006]). Основываясь на открытии Риззолатти и Арбибом зеркальных нейронов и зеркальных систем, в этих работах доказывается принципиальная важность имитации биологическим существом действий *другого* в нервной системе для когнитивного развития в фило- и онтогенезе. Зеркальные нейральные системы, в отличие от прозрачных (классических) систем, позволяют экономнее описать способ эволюционного возникновения языка и рефлексии как причины сознания. В свою очередь, новый взгляд на понятие «нейрон» позволяет пересмотреть нейрокомпьютерные модели и подать заявку на фундаментальную корректировку нейроматематических подходов.

2. *Технические критические коннекционисты*. Эти специалисты усматривают пробелы в существующем нейрокомпьютерном инструментарии и подходят к их исправлению с позиции естественных и технических наук. Например, А.В. Савельев изучает не свойства нейрона как такового, а диффузионные свойства нейротрансмиттинга, основываясь на моделях так называемого «биологи-

ческого подхода». Данный подход заключается в детальном информационном воспроизведении тонких механизмов и закономерностей функционирования реальных нейронов. Также аргументированно показывается возможность теоретического предсказания неизвестных ранее биологических феноменов [Савельев, 2009].

Таким образом, критические коннекционисты, осуществляя ревизию категорий нейрокомпьютерной тематики, способствуют научным революциям в нейрокомпьютинге и исследованиях ИИ, по крайней мере, раскрывают необходимость смены нейральной парадигмы. В целом, оппонируя символистам, все коннекционисты составляют в настоящее время мощный лагерь исследователей проблематики ИИ и нейрокомпьютерного решения психотехнической проблемы.

Роль нейрокомпьютера (в метафизической проекции на социокультурные проблемы) можно условно «подсчитать» следующим образом. Классический ИИ и нейрокомпьютерный ИИ получают по 1/3 наработок. Оставшаяся треть приходится на гибридный ИИ.

Инженерная проекция нейрокомпьютерной тематики в электронной культуре

Нейрокомпьютер, согласно инженерной проекции, – это концепция, погружённая в контекст многочисленных мысленных экспериментов комплексного ТТ. Как ранее уже отмечалось, нейронная сеть Тьюринга претендует не более чем на параллельную машину – она реализуется средствами универсальной ЦВМ. Более близкой к нейрокомпьютерной тематике оказывается машина Корсакова, предложенная в статье этого гениального русского инженера в 1832 г. Это – за десять лет до того, когда мировая научная общественность, благодаря Аде Лавлейс, узнала об аналитической машине Ч. Бэббиджа (1842 г.) – непосредственного аналога УЦВМ, взятого почти без изменения А. Тьюрингом. Никому не известная разработка гениального отечественно инженера С.Н. Корсакова, не признанная российской императорской академией наук и опубликованная не на родном русском, но лишь на французском языке, совсем недавно была «открыта» Г.Н. Поваровым [Поваров, 2005-1]. Перевод этой статьи увидел свет только в 2007 г., а был издан в 2009 г. [Корсаков (Сыромятин), 2009].

Машина Корсакова («машина сравнения идей») состоит из пяти самостоятельных механизмов, «нейрокомпьютерная» интерпретация которых может

быть следующей. Страницы (перфокарты), на которых представлены таблицы машины (страниц может быть достаточно большое количество), накладываются одна на другую. В клетках таблиц указывается наличие или отсутствие некоторых признаков (в оригинале описывались патологии различных человеческих органов). Если признак имеет место, то он отмечается прокалыванием иглой той клетки, в которой он представлен (пересечение «признака болезни» с «анатомией человеческих органов»). Вполне очевидно, что клетка может репрезентировать нейрон, а игла – синапс (в классическом понимании модели нейрона). Каждая таблица представляет некоторую простую, обозримую идею. Ряды накладываемых друг на друга таблиц соответствуют латентным узлам «сети». Те таблицы, которые просматриваются как начальная и конечная – входные и выходные узлы. Механизмы, которые приводят в движение машину сравнения сложных идей (сложная идея – это есть ряд наложенных друг на друга таблиц), выдают результат, зависящий во многом от усилия, затрачиваемого для прокалывания страниц («весовой коэффициент»). Конечно, это не модель биологического нейрона, но нечто очень похожее (это далёкий 1832 г.!). Машина Корсакова, как мы видим, крайне разнится с машиной Тьюринга и достаточно далека от «символьного» подхода к ИИ. Несомненно, машина Корсакова обладает большим потенциалом эвристических идей относительно её применения в концептуальных построениях ИИ.

Таким образом, если машина Бэббиджа – прообраз классического компьютера, то машина Корсакова – прообраз нейрокомпьютера, своего рода протонейрокомпьютер. Отсюда следует тезис:

Машина Корсакова - протонейрокомпьютер.

Удивительно, что только в последнее десятилетие отечественная и мировая общественность оживлённо заговорила о выдающемся изобретателе С.Н. Корсакове¹.

¹ Семён Николаевич Корсаков (1787 – 1853 гг.), проживал с 1827 г. до конца дней в Дмитровском уезде, в своём имении Тарусово на берегу реки Дубны. С.Н.Корсаков был сыном талантливого военного инженера Н.И.Корсакова и племянником адмирала Н.С.Мордвинова, а его крестными были князь Г.А.Потёмкин-Таврический и княгиня Е.И.Вяземская. Вернувшись героем с войны 1812 – 1814 гг., он поступил на службу в отделение статистики Министерства юстиции и на протяжении многих лет занимался составлением статистических описаний российских городов, впоследствии возглавил отделение статистики, вышел в отставку действительным статским советником (соответствует генеральскому чину на военной службе).

История проекта С.Н. Корсакова в философско-методологических исследованиях искусственного интеллекта

Изучение забытого исторического наследия неразрывно с историей его возрождения, причем это непосредственно касается социально-эпистемологических характеристик развития философской методологии ИИ. Поэтому начнём с того, что в 2004 г. в оргкомитет Всероссийской конференции «Философия искусственного интеллекта» поступила заявка выдающегося отечественного историка науки и техники – Гелия Николаевича Поварова. Предлагался сенсационный доклад «С.Н. Корсаков – русский пионер искусственного разума» [Поваров, 2005-1]. К прискорбию, доклад не состоялся – Г.Н. Поваров скоропостижно скончался. В докладе сообщалось о том, что в начале 1990-х годов в Германии был найден архив, вывезенный из Франции, оккупированной в годы Второй мировой войны. В архиве находилась интереснейшая статья русского инженера начала 19 в., описывающая уникальный компьютер, состоящий из «машин сравнения идей». Г.Н. Поваров представил статью на английском языке [Povarov, 2001]. Благодаря этому международная общественность узнала о машинах Корсакова. Г.Н. Поваров не без основания предлагал считать Семёна Николаевича Корсакова пионером искусственного интеллекта не только в масштабе отечественной, но и мировой науки. Мы также вполне правомочно выдвигаем тезис:

С.Н. Корсаков – пионер когнитивных исследований в масштабе мировой науки.

Дело в том, что в проекте Корсакова явно обозначена когнитивная установка: 1) когниция обозначена как «сложная мысль»; 2) механизм реализации когниции представлен механической моделью анализа «сложной мысли», обработки «простых мыслей» и получения из них «сложной мысли»; 3) разработана инженерная конструкция, обеспечивающая машинную (компьютерную) реализацию данной модели.

Когнитивная установка проекта С.Н. Корсакова

Когнитивная установка С.Н. Корсакова состоит в следующем. Идеальные мысли следует облечь в физическую оболочку. «Материальная оболочка» мыслей по-корсаковски совершенно отличается от кругов Эйлера, диаграмм Венна и других подобных способов визуального отображения экстенсионалов (объемов понятий). Корсаковская «материализованная мысль» – *стопка перфокарт, истыканная иголками.*

К материально воплощённым «мыслям» можно применять механизмы реализации рутинных операций, которые человек осуществляет в уме при обработке сложных идей. Предлагается не одна, абсолютно на всё способная машина (как это характерно для УЦВМ А.Тьюринга), а набор из пяти машин: 1) *линейный гомеоскоп с неподвижными частями*, который выбирает из набора сложных идей, представленных в специальной таблице (перфокарте), все те, которые содержат какую-либо другую известную сложную идею со всеми ее элементами, показывая искомые идеи путем самопроизвольной остановки во время выполнения заданной операции; 2) *линейный гомеоскоп с подвижными частями*, выполняющий те же функции, что и предыдущий, и, сверх этого, способный моментально выделить и отделить те части сложной идеи, которые соответствуют (или не соответствуют) аналогичным частям других идей; 3) *планшетный гомеоскоп*, показывающий совпадение частей двух сложных идей, между которыми проводится сравнение (количество элементов может достигать миллиона); 4) *идеоскоп*, который при помощи специальной таблицы выявляет следующее: все схожие элементы сложных идей, обрабатываемых в ходе текущей операции; элементы одной сложной идеи, отсутствующие в другой идее, с которой производится сравнение; элементы, которые отсутствуют в данной сложной идее, но присутствуют в той, с которой производится сравнение; элементы, отсутствующие в сравниваемых сложных идеях, но являющиеся частью других сложных идей, занесенных в таблицу. 5) *простой компаратор*, производящий те же операции, что и идеоскоп, но лишь в отношении двух сложных идей, сравниваемых между собой. Он способен обрабатывать всего несколько десятков элементов таких идей, но выгодно отличается от остальных устройств тем, что не требует специальной таблицы.

Способ работы машины прекрасно иллюстрирует *гомеопатический пример*, предложенный самим С.Н. Корсаковым. Предметом профессиональных исследований и личных пристрастий С.Н. Корсакова выступали сложные предметные области – демография и гомеопатия. Сложность предмета определяется столь большим разнообразием феноменов и связей между ними, что человеческий разум не способен сразу всех их охватить и вычленить нечто полезное из хаоса предполагаемых причинно-следственных связей. Для того чтобы удерживать всевозможные сочетания идей в виде близких и дальних связей, собственно и требуются усилители человеческого интеллекта, механизмы и машины. С.Н. Корсаков предлагает механические табуляторы, которые совместно моделируют сложную логику работы человека при анализе «сложных мыслей».

Для первой предметной области С.Н. Корсаков предложил метод обработки больших объёмов статистических данных и их каталогизации, согласно которому многие рутинные операции можно производить с помощью перфокарт и механизмов. Подобный метод он предлагает и для второй предметной области – для гомеопатии. В выборе данного предмета исследований мы видим желание автора когнитивных машин достичь успехов в медико-инженерном врачевании. «Интеллектуальные машины» успешно используются в современной гомеопатии – гомеопатические средства, подобранные по корсаковской методике, принято помечать литерой «К» в честь автора. С.Н. Корсаков сформировал объёмные «базы данных» и снабдил их «программой» быстрого поиска гомеопатических средств, в зависимости от многообразных симптомов заболевания. И успешно применял машину для лечения больных в своей округе, став, пожалуй, первым в мире врачом-инженером.

Таким образом, «интеллектуальные машины» Корсакова – первый действующий когнитивистский проект, в котором чётко прослеживаются составляющие когнитивной установки – модель когнитивного феномена и машина обработки этой модели. Машина Корсакова – это фундаментальная когнитивная метафора, которая, несмотря на срок давности, востребована в контексте современной культуры.

По долгу службы С.Н. Корсаков обрабатывал большие объёмы статистических данных. Тягостная бессмысленность осуществления однообразных действий подвела его к идее о том, что многие рутинные операции можно производить автоматически с помощью механических средств. Ему удалось впервые в мире разработать оригинальный метод быстрой обработки больших объёмов информации с использованием перфокарт и создать уникальные «интеллектуальные машины», ставшие, по сути, первым шагом к современным информационным технологиям и к теоретическим основам информатики и кибернетики¹.

Раскроем подробнее идею представления машины Корсакова как протонейрокомпьютера.

Протонейрокомпьютер Корсакова (1832 г.)

¹ Научная деятельность С.Н. Корсакова была многогранна – он был также выдающимся российским учёным-гомеопатом, его открытия и разработки получили высокую оценку специалистов как в России, так и за рубежом. Свои «интеллектуальные машины» он использовал и в гомеопатии, а его уникальные методы приготовления гомеопатических лекарственных средств до сих пор широко используются в Европе с пометкой «К».

Общим положением дискуссий по ключевым проблемам развития искусственного интеллекта (ИИ) является утверждение о слабости компьютерного инструментария, недостаточности вычислительной математики, неадекватности теорий алгоритмов. Для радикальной перестройки фундаментальных оснований современного компьютеринга ранее (в [Алексеев, 2012-1;2;3]) обращалось внимание на концептуальный проект интеллектуальной машины, предложенный в 1832 г. (!) российским инженером С.Н. Корсаковым [Karsakof, 1832]. В данной работе доказывается то, что машину Корсакова можно и нужно интерпретировать как самый первый в истории действующий проект нейрокомпьютера. С этой машины следует отсчитывать идею нейрокомпьютинга, коннекционизма, искусственного интеллекта, когнитивно-компьютерной технологии в мировом масштабе.

Семён Николаевич Корсаков (1787-1853) – основатель «нейрокомпьютинга»

Изучение забытого исторического наследия неразрывно связано с историей его возрождения. Поэтому начнём с того, что в 2004 г. в оргкомитет всероссийской конференции «Философия искусственного интеллекта» поступила заявка выдающегося отечественного историка науки и техники – Гелия Николаевича Поварова. Предлагался сенсационный доклад «С.Н. Корсаков – русский пионер искусственного разума» [Povarov, 2001]. К прискорбию, доклад не состоялся: Г.Н. Поваров скоропостижно скончался. В докладе сообщалось о том, что в начале 1990-х гг. в Германии был найден архив, вывезенный из Франции в годы Второй мировой войны. В архиве хранилась статья русского инженера начала XIX в., описывающая уникальный компьютер – машину сравнения идей. Статья была написана в далёком 1832 г., излагалась на французском языке и принадлежала забытому отечественному инженеру С.Н. Корсакову [Karsakof, 1832]. Г.Н. Поваров представил статью на англ. в [Povarov, 2001]. Благодаря этому международная общественность узнала о машинах Корсакова как о «русских самосчётах». Статью передали в Российскую государственную библиотеку, где с ней сейчас можно ознакомиться. Г.Н. Поваров не без основания предлагал считать С.Н. Корсакова пионером искусственного интеллекта не только в масштабе отечественной, но и мировой науки. Обоснованию этой идеи посвящены работы ([Povarov, 2001; Алексеев, 2011-1; Нитусов, 2012]).

Семён Николаевич Корсаков (другое написание фамилии – Карсаков) (1787-1853) – выдающийся российский мыслитель, гениальный инженер, действительный статский советник (чин соответствовал генеральскому на воин-

ской службе), известный гомеопат, практикующий врач-инженер. Родился в г. Херсоне накануне русско-турецкой войны 1787-1791 гг. С 1827 г. и до конца дней проживал в Дмитровском уезде Московской губернии в собственном имении Тарусово на берегу реки Дубны (подробная биография представлена в [Мищенко, 2012]). В период написания статьи [Karsakof, 1832] служил в Министерстве внутренних дел по статистике по части демографии городов и населения России. Предложил метод обработки больших объёмов статистических данных и их каталогизации, согласно которому многие рутинные операции можно производить с помощью перфокарт и механизмов. Так же впервые он создал семейство «интеллектуальных механизмов», которые совместно моделировали сложную логику работы человеческого мышления. Именно с момента их появления предлагается отмечать день рождения нейрокомпьютинга. «Интеллектуальные механизмы» успешно использовались им в статистике и в гомеопатии. По сей день гомеопатические средства, подобранные по корсаковской методике, принято помечать литерой «К». С.Н. Корсаков сформировал объёмные «базы знаний» и снабдил их инструкцией работы машины для быстрого поиска гомеопатических лекарств в зависимости от многообразных симптомов заболевания. Данная инструкция – первая в мире компьютерная программа.

Интеллектуальная машина С.Н. Корсакова предназначена для работы со «сложными идеями». Когнитивистский замысел С.Н. Корсакова прост: «Все позитивные знания человек приобретает путем наблюдения и сравнения фактов... [, однако] человек все же не способен проводить сложные сравнения с учетом многочисленных элементов. Между тем подобные сравнения можно реализовать с использованием механических манипуляций с ними. С этой целью необходимо лишь применять знаки с более выраженной физической сущностью, чем те которыми мы обычно пользуемся [т.е. когда пользуемся словами естественного или искусственного языка – А.А.]. Разум извлекает информацию из письменного источника посредством образов, не имеющих материального выражения и не доступных для физического восприятия. Однако стоит ту же самую мысль облечь в материальную оболочку, существующую в трех измерениях, как мы получим возможность не только воспринимать информацию всеми своими чувствами, но и использовать физические свойства схожих элементов для решения умственных задач» [Корсаков (Сыромятин), 2009, с.16]. В этом суждении чётко просматривается антирепрезентативная ориентация когнитивистской установки С.Н. Корсакова. Более выраженная физическая сущность мыслей – это не семиотическая модель сложной мысли, представленная на не-

котором специальном языке, например, на языке исчисления предикатов. Это – коннекционистская модель, построенная на установлении связей между признаками идей, очень схожая с современными нейрокомпьютерными моделями. Задача машины состоит не только в реализации рутинных операций, которые человек затрудняется осуществлять в уме при сравнении простых идей. Цель машины – *усилить человеческий разум*, так как она способна формировать сложные идеи исходя из системы связей между признаками более простых идей. Область применения машины не ограничивается научной деятельностью. Она применима «в повседневной жизни, когда человеку часто приходится одновременно держать в уме множество разных идей, извлекая их из памяти для сравнения с данной сложной идеей с тем, чтобы сделать тот или иной вывод» [Там же, с.16]. С.Н. Корсаков учёл даже «инновационные» особенности своей машины: «Стоимость подобных интеллектуальных машин сравнительно не высока, поскольку механизм их функционирования не сложен, а, следовательно, набор услуг, которые они могут оказать обществу, будет постоянно расширяться» [Там же, с.17].

Ещё раз подчеркну, что только недавно, с начала ХХI века, в мировом и отечественном научном сообществе С.Н. Корсаков начинает получать признание как изобретатель первых в мире искусственных интеллектуальных систем. Спустя 175 лет после публикации оригинальной статьи С.Н. Корсакова, наконец-то впервые (!) появился её перевод с французского языка на русский. Сегодня имеется два альтернативных перевода, изданных в 2008 г.: 1) А.В. Сыромятина (2007 г.), который был приурочен к указанному юбилею и впервые представлен на всероссийском круглом столе «Искусственный интеллект: трансляция в социокультурной среде», организованного Научным советом РАН по методологии искусственного интеллекта 7 июля 2007 г., г. Лодейное поле Ленинградской области [Алексеев, 2009-4]; 2) перевод А.С. Михайлова (2008 г.) [Корсаков (Михайлов), 2009], изданный отдельной книгой и снабжённый прекрасными биографическими сведениями о С.Н. Корсакове и Г.Н. Поварове, описанием истории возрождения забытой корсаковской идеи, а так же теоретико-множественной интерпретацией корсаковских машин.

Состав интеллектуальной машины С.Н. Корсакова

Всего в статье [Karsakof, 1832] С.Н. Корсаков предлагал пять механизмов: три гомеоскопа – устройств исследования подобий в идеях¹, идеоскоп – прибор анализа сложной идеи и построения сложной идеи из более простых, компаратор – прибор оперативного сравнения идей по признакам. Для знакомства с этими механизмами предлагаем обратиться к переводу оригинальной статьи с французского языка на русский, осуществлённого А.В. Сыромятиным (2007 г., в то время он был аспирантом Института США и Канады РАН, французский язык как второй родной) [Корсаков (Сыромятин), 2009]. Сравните, например, варианты перевода наименования статьи «Очерк о новом способе исследования посредством машин для сравнения идей» (А.В. Сыромятин) и «Начертание нового способа исследования при помощи машин, сравнивающих идеи» (А.С. Михайлов) [Корсаков (Михайлов), 2009]. Очевидна разница в степени адаптации первого перевода к современному языку.

Репрезентативная (символьная) и коннекционистская интерпретация машины Корсакова

Сегодня предложены две альтернативные трактовки корсаковского проекта, которые с учётом терминологических особенностей методологии ИИ правомочно назвать репрезентативной и коннекционистской интерпретациями.

Репрезентативная интерпретация предложена А.С. Михайловым в [Корсаков (Михайлов), 2009, с. 33-44]. Репрезентативизм – это парадигма философии искусственного интеллекта, объясняющая феномен мышления и другие когнитивные феномены как процесс формирования, запоминания и обработки лингвистических конструкций. В работе [Корсаков (Михайлов), 2009] детально представлено описание способа функционирования всех пяти машин на языке теории множеств. Показано, что «интеллектуальная машина» реализует функции экспертных систем, обеспечивают эффективный многокритериальный поиск информации, фасетную классификацию, программное управление «базой знаний» и пр. функции современных систем репрезентативного ИИ.

Однако, на мой взгляд, если подгонять идею машины Корсакова (1830-е гг.) под более поздний теоретико-множественный базис (1850-70-е гг.), это означает вычеркнуть из оригинальной идеи то, что может послужить принципиальной основой реконструирования компьютеринга. Теория множеств упрощает идею и преимущества её очевидны – математические конструкции реализуемы

¹ «Гомео» - составная часть слова, обозначающая сходство или похожесть, «скоп» - составная часть слова, означающая прибор или инструмент для рассматривания, изучения, исследования, проверки чего-то либо.

на классе субстанциально инвариантных устройств: у Корсакова – механических, сегодня – электронных. Но при упрощении теряется главное – машина Корсакова лишается «машинной энергии», каузальные связи становятся призрачными, математическое описание – лишь формальной тенью машины. Поэтому мы предлагаем иную интерпретацию, которая, во-первых, идентична первоначальной идее С.Н. Корсакова, и во-вторых, наметит новые пути развития когнитивно-компьютерной технологии, которые превосходят булевы принципы современных логических вентилях микроэлектроники.

Коннекционистская интерпретация машины Корсакова предложена в [Алексеев, 2011-1; 2012-1;2;3]. Напомним, что коннекционизм – это парадигма когнитивно-компьютерных наук, которая объясняет интеллектуальные способности человека посредством искусственных нейронных сетей [Garson, 2010]. Очевидно, что коннекционизм предлагает упрощённую модель биологического мозга «из мяса и крови». Модель состоит из большого числа связанных друг с другом «нейронов» и весовых коэффициентов, которые характеризуют силу связей между ними. Веса моделируют действия синапсов, которые связывают нейроны в сложные структуры – паттерны нейральных взаимодействий. Коннекционизм часто представляют исключительной альтернативой классической (репрезентативной) парадигмы, согласно которой интеллектуальная деятельность моделируется обработкой символического языка посредством цифрового компьютера.

Для раскрытия нейрокомпьютерной трактовки машины Корсакова её следует упростить. В оригинале она состоит из брусков, деревянных рамок, медных проводов, рычагов, подвижных частей из слоновой кости, штырей, пружин и пр. В нашем варианте она состоит из: *перфокарт*; *подперфокарт* – частей перфокарты, разрезанной поперёк на несколько частей и каждая из которых, например, соответствует признаку объекта; *отверстий в перфокартах*; *иглоков*, которыми прокалываются отверстия в перфокарте; *перфораторов*, которые проделывают отверстия в перфокартах – при фиксированной пробивке они действуют по принципу дырокола или компостера, при формировании отверстий произвольной ширины и глубины – круговыми движениями ножа; *табуляторов* – устройств «считывания» перфокарт и формирования результатов изучения идей; *иглоков табуляторов*, которые попадают в отверстия перфокарт. Машиной должен управлять *Машинист*, разрабатывать и программировать её – *Инженер*.

«Нейрон» – это перфокарта. На перфокарте может быть представлена совокупность «нейронов», тогда нейрон соответствует подперфокарте. Сложная идея «материализуется» рядом перфокарт, наложенных друг на друга. На каждой перфокарте, представлены простые идеи, т.е. обозримые при задании признаков частного предмета. Отверстие в перфокарте, которое формирует инженером прокалыванием иглы перфоратора в процессе «экспликации знаний» – это «возможный синапс». Действительным «синапс» становится при прокалывании этого отверстия иглой табулятора в процессе работы машины при анализе и синтезе сложной идеи. Таким образом, актуальные «коннекции» между нейронами образуются в результате применения табулятора к стопке перфокарт: если иглы попадают в отверстия, то образуются «синаптические связи». Перфокарты, накладываемые друг на друга, соответствуют латентным узлам «сети». Начальная и конечная перфокарты – входные и выходные узлы. Результат образования «синаптической связи» во многом зависит от ширины отверстий в перфокарте, величины игл табулятора, взаимного совпадения отверстий при наложении перфокарт, от усилия, которое затрачивается как для пробивки отверстий перфоратором, так и для попадания игл табулятора в отверстия. По сути, это – перфокарточный аналог «весового коэффициента» связи между нейронами.

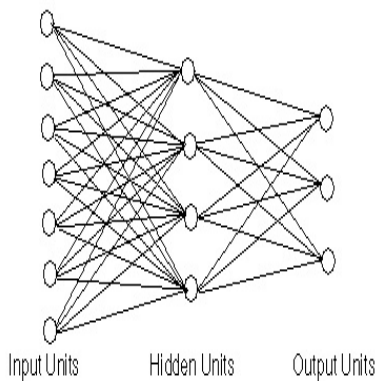
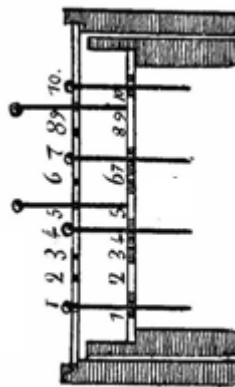
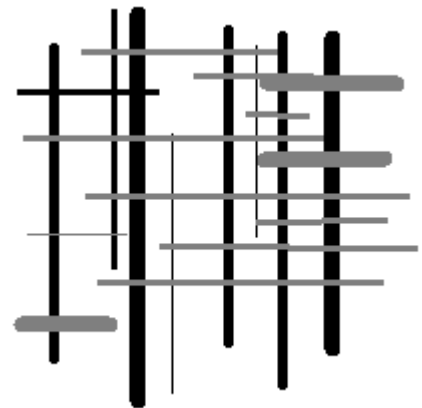


Иллюстрация простой нейронной сети (Дж. Гарсон [Garson, 2010])



Фрагмент планшетного гомеоскопа Корсакова (поворот на 90^0) [Корсаков (Сыромятин), 2009, с.22]



Абстрактная модель планшетного гомеоскопа

Рис.3.1. Аналогии примитивных коннекционистских моделей

На Рис.3.1 для сравнения приведены примеры: 1) фрагмент современной простейшей нейронной сети, предложенный известным коннекционистом Дж.

Гарсоном (2010 г.); 2) фрагмент планшетного гомеоскопа С.Н. Корсакова, на котором чётко просматривается идея перфокарточной «коннекции» посредством игл и отверстий, а так же выдачи результата на «выходном нейроне», который образуют иглы, прошедшие сквозь него; 3) абстрактная модель планшетного гомеоскопа, на котором весовой коэффициент «синаптической связи» характеризуется шириной отверстия в перфокартах (вертикальные линии различной ширины), толщиной и длиной игл табулятора (горизонтальные линии).

Раскроем подробнее нейрокомпьютерные особенности интеллектуальной машины Корсакова, обратив вначале внимание на строение его перфокарты.

Перфокарта Корсакова. С.Н. Корсаков впервые в истории применил перфокарты в той сфере научно-технической деятельности, которая в нашей стране сегодня называется «информатикой». Сделанная из картона, перфокарта выступает в роли материального носителя информации¹. Имеются/отсутствуют отверстия, пробитые в перфокарте – имеем «информацию» во всех её возможных теоретико-информационных трактовках, например: 1) *алгоритмический подход* (А.Н. Колмогоров) – количество информации, представленное в перфокарте, это значение некоторой функции сложности от относительной плотности отверстий и их количества, а так же длины алгоритма, задаваемого временем работы перфоратора и табулятора; 2) *вероятностный подход* (К. Шеннон) – вероятность снятия неопределённости при сравнении конфигураций из отверстий; 3) *разнообразностный подход* (А.Д. Урсул) – разнообразие конфигураций отверстий одной перфокарты от другой перфокарты и т.п. Очевидно, что перфокарта Корсакова – это классическая метафора носителя информации.

Предлагаем тезис: **все современные носители информации – это электронные перфокарты Корсакова.** В самом деле, одно значение устойчивого состояния триггера – наличие отверстия в перфокарте, другое значение – отсут-

¹ Идея использования перфокарт для автоматизации управления механизмами принадлежит Жаку де Вокансону из существовавшей столетиями практики применения металлических перфорированных лент для башенных часов и музыкальных автоматов. Известен работой «Механизм автоматического флейтиста» (1738). Впервые массовое применение перфокарт отмечается в ткацких станках Жозефа-Мари Жаккарда (1808) для управления узорами на тканях (см. [Нитусов, 2012, с. 109]). В проекте аналитической машины Бэббиджа (1834 г.), который не был реализован, перфокарты выполняли функции: 1) переключения режимов работы для выполнения операций сложения, вычитания, деления и умножения; 2) управления передачей данных из памяти в арифметическое устройство и обратно; 3) ввода данных в машину и сохранения результатов вычислений. Роль перфокарты Корсакова – совершенно иная. Она непосредственно характеризует признаки сложной идеи, причем в большинстве табуляторов характеристика производится до этапа логико-понятийной экспликации идеи.

ствие отверстия. Далее, соответственно: положительный потенциал/отрицательный потенциал; наличие тока/отсутствие тока; модулирующий сигнал/несущий сигнал при радиопередаче данных; низкое напряжение на плавающем затворе транзистора флэш-памяти/высокое напряжение; выступ на поликарбонатной основе CD-ROM/углубление, выдавленное в этой основе и др. интерпретации. Разумеется, каждая электронная «перфокарта» требует специального устройства: 1) *перфоратора*, например, пишущего привода CD-ROM, модулятора и 2) *табулятора* – устройства чтения «перфокарт», например, лазерного луча, попадающего на фотодиод от поликарбонатной основы CD, демодулятора и пр.

Выделим следующие формы коннекции между перфокартами-«нейронами»:

1) *градуальная коннекция*, означающая постепенное нарастание или ослабление признака при движении иглы табулятора от одной перфокарты к другим перфокартам в стопке. Определяется «весами», которые задаются различной шириной отверстий перфокарт или, например, толщиной игл, которая увеличивается у основания иглы и уменьшается у острия. Данный способ «активации» перфокарт-«нейронов» воспринимается как метафора непрерывной деятельности мозга. Непрерывные весовые значения – вариант, наиболее интересный для нейрокомпьютерной интерпретации.

2) *градуированная коннекция* – интервальная параметризация градуальной коннекции Устанавливается аналогично тому, как на термометре фиксируется шкала непрерывных значений температуры. Данный тип связи указывают на количество перфокарт, проткнутых иглой табулятора. В логическом отношении это означает, например, группирование сложных идей по степени общности связываемых признаков.

3) *n-значная коннекция* – характеризует степень связанности конкретного отверстия перфокарты с другими отверстиями в составе анализа стопки перфокарт. Для подперфокарты данная форма связи учитывает отношение с перфокартой как часть/целое.

4) *бинарная коннекция* означает наличие связи/отсутствие связи в результате попадания иглы в отверстие либо не попадания её туда. Это – вырожденный вариант связи, он применим для анализа лишь единичного признака на перфокарте и не характерен ни для одного корсаковского механизма, так как признаки в них изучаются не порознь, но лишь в составе целостного варианта анализа сложной идеи.

Мы выяснили назначение перфокарты Корсакова – фиксация связей между признаками сложной идеи. Для подчёркивания нейрокомпьютерного, коннекционистского способа её использования, отметим принципиально иное предназначение перфокарт в машине Ч. Бэббиджа.

Перфокарта Бэббиджа. Перфокарта используется в контексте репрезентативной парадигмы. Впервые идея перфокарт Ч. Бэббиджа была опубликована через 10 лет после статьи С.Н. Корсакова, в 1842 г. Способ использования перфокарт предложила Ада Лавлейс в программе, которую «вики-история» считает самой первой программой в истории человечества. Программа выдавала список чисел Бернулли. Как известно, для вычисления чисел Бернулли – суммы последовательных натуральных чисел, возведённых в одну и ту же степень, требуется применение цикла. Поэтому на перфокарте (управляющего типа) кодировались параметры цикла, далее машина Бэббиджа принимала данные с перфокарт ввода, обрабатывала данные в регистрах в соответствии с кодами обработки, представленными на управляющей перфокарте и выводила результат на перфокарты вывода. Ввод данных из внешней памяти в оперативную, организация вычислений, где коды операций и коды операндов представлены в оперативной памяти, вывод информации на внешние накопители – это современная классическая принципиальная схема компьютеринга. Программная инструкция – итерация – была задумана для уменьшения количества перфокарт, требуемых для ввода рекуррентных значений формулы Бернулли. Программа А. Лавлейс представлена в примечаниях к книге «Очерк об аналитической машине, изобретённой Чарльзом Бэббиджем» итальянского математика Луиса Менебреа в 1842 г. – спустя 10 лет после публикации статьи С.Н. Корсакова. Так же как и статья Корсакова, книга была написана на языке науки того времени – на французском. Ада Лавлейс по просьбе Ч. Бэббиджа переводила ее на английский язык и в одном из своих примечаний собственно и привела пример первой компьютерной программы, выполнимой на современном техническом базисе. Однако, отметим, данная программа выполнена в формате репрезентативной парадигмы, что принципиально отличает её от корсаковской коннекционистской «программы».

Таким образом, в методологии ИИ не сегодня, а в начале 1840-х годов, наметилось противостояние двух парадигм использования носителей информации (перфокарт): 1) коннекционистский (С.Н. Корсаков) – конфигурация связей на перфокарте является непосредственным результатом анализа и синтеза связей между признаками сложных идей и 2) репрезентативный (Бэббидж-

Лавлейс) – на перфокарте кодируется программы вычисления. Второй подход делает более длинный путь к сравнению сложных идей, нежели чем первый, так как требуется кодирование слов и программа исчислений над кодами этих слов для получения не арифметического (что по сути делает машина Бэббиджа), но логико-лингвистического результата. Путь, конечно, не удобный для анализа сложных мыслей, но именно по этому косому пути идёт сегодня современный компьютеринг. Машина Корсакова спрямляет дорогу «нейрокомпьютерингом» сложных идей.

Простой компаратор [Корсаков (Сыромятин), 2009, с. 25]. Начнем с этого наиболее простого механизма машины Корсакова. Это – две перфокарты, одна фиксирует признаки объекта X, другая – признаки объекта Y. В примере С.Н. Корсакова признаками выступали физические показатели людей. Перфокарты разрезаются поперёк на подперфокарты. Количество подперфокарт равно количеству признаков у сравниваемых объектов. Каждая перфокарта помещается в рамку, ширина которой превышает вдвое ширину перфокарты. В рамках установлены поперечные желобки, по которым можно свободно вправо и влево перемещать подперфокарты. Далее перфоратором пробиваются достаточно широкие отверстия в перфокартах X и Y, чтобы иглы в них могли свободно перемещаться вверх/вниз. В перфокарту X втыкаем иглы со шпильками, чтобы не проваливались при сравнении признаков. Осуществляем определение признаков у объекта X. Если признак имеется, то соответствующую подперфокарту сдвигаем влево. Так же поступаем и с объектом Y. Далее накладываем две перфокарты (точнее, их рамки, друг на друга). В результате всего одной (!) операции – наложения перфокарт – мгновенно получаем результат. Он характеризуется: 1) четырёхзначной логикой; 2) параллельным исполнением (операция производится сразу над всеми подперфокартами).

Сведём работу компаратора в таблицу *Табл. 3.1.* В ней представлены: 1) коды логических значений истинности; 2) коннекционистские значения – особенно важные для нейрокомпьютерной интерпретации; 3) репрезентативные значения, стандартные для индуктивной логики; 4) действия компаратора.

Табл.3.1. Схема работы компаратора

	<i>Коннекционистские значения</i>	<i>Репрезентативные значения</i>	<i>Действия компаратора Корсакова</i>
+	Соединение «нейронов»	X^+Y^+ Признаки присутствуют у X и Y	опустятся подвижные иглы, находящиеся слева, совпавшие с отверстиями подперфокарты

			Y
+-	Соединение с первым «нейроном», отсутствие связи со вторым	X^+Y^- Признаки присутствуют у X, отсутствуют у Y	<i>поднимутся</i> иглы, находящиеся слева, так как они встретили сопротивление, не совпав с отверстиями подперфокарты Y
-+	Соединение со вторым «нейроном», отсутствие связи с первым	X^-Y^+ Признаки отсутствуют у X, присутствуют у Y	<i>поднимутся</i> иглы, оставшиеся справа, так как они не совпали с отверстиями подперфокарты Y
-	Отсутствие связи между «нейронами»	X^-Y^- Признаки отсутствуют у X и Y	<i>опустятся</i> иглы, оставшиеся справа, так как они совпали с отверстиями подперфокарты Y

Как мы видим, для простейшего корсаковского механизма имеем 4-значные коннекции. Если X и Y – составные перфокарты (способ их формирования представлен ниже), то имеем многозначную коннекционистскую логику. Таким образом, действия компаратора не редуцируются к булевой алгебре. Это ещё раз подчёркивает неправильность булево-алгебраической интерпретации машины Корсакова [Корсаков (Михайлов), 2009]. Хотя, несомненно, булевая имитация её работы возможна, что прекрасно продемонстрировано А.С. Михайловым и его учениками.

Планшетный гомеоскоп [Корсаков (Сыромятин), 2009, с. 21-23]. В этом механизме ещё более чётко проявляется нейрокомпьютерная идея. Устройство состоит, как минимум, из двух перфокарт X и Y, располагающихся одна над другой. Выступив в роли инженера, С.Н. Корсаков представил первую коннекционистскую программу в виде инструкции по работе с этим механизмом. Как машинист, он часто применял данный механизм в практике гомеопатического лечения. Так, с 12 февраля 1829 г. по май 1834 г. он диагностировал 11725 болезненных случаев [Мищенко, 2012, с. 19].

На одной части перфокарты X инженер вписывает буквенные обозначения различных частей и органов, т.е. фиксируется своеобразная «топография» человеческого тела (голова, нос, глаза, желудок и другие). На другом конце перфокарты X («патография» человека), цифрами обозначаются различные недуги (опухоли, покраснения на коже, стреляющая и режущая боль, жар, озноб, обморок и т.д.). Например, если буква *e* обозначает грудь, цифра 4 – режущую боль, то квадрат *e4* соответствует режущей боли в груди. Для фиксации данного симптома в центр квадрата *e4* втыкается игла и на карте Y появляется отвер-

стие. Подобным образом фиксируются остальные симптомы. В таблице, содержащей по сто ячеек по вертикали и горизонтали с помощью игл можно было представить десятки тысяч различных симптомов, осуществляя, тем самым, сложную конструктивную диагностику.

Теперь на краях карты Y помечаются наименования лекарств, которые рекомендуется применять для лечения в соответствии с симптомами таблицы X . Эти наименования должны совпадать с отверстиями, полученными в результате предыдущего шага. На другом краю перфокарты Y могут быть отмечены дополнительные сведения, например, предупреждения о последствиях применения лекарств. Общая эталонная перфокарта $\langle X, Y \rangle$ сформирована, она «запрограммирована» врачом-инженером.

На прием к врачу-машинисту приходит пациент. Машинист фиксирует патографию/топографию пациента прокалыванием иглами соответствующих ячеек перфокарты Z . Накладывает Z на $\langle X, Y \rangle$, иглы Z проваливаются в эталонные перфокарты и мгновенно получается результат – рекомендуемое лекарство. Все это осуществляется быстро, механически и без участия эксперта-инженера.

Имеется ряд дополнительных соображений по поводу гомеоскопа, который ещё более приближает машину Корсакова к концепции нейронной сети.

1) Эталонных перфокарт может быть сколь угодно много. Например, перфокарты $X = \{X_1, X_2, \dots, X_n\}$, $Y = \{Y_1, Y_2, \dots, Y_m\}$ – это различные медицинские теории, которые можно связать общими и отличительными признаками, «накладывая друг на друга». Инженер выражает личные предпочтения, прокалывая иглами соответствующие отверстия перфокарт, и игнорируя положения, с которыми не согласен, не прокалывая иглами. Теперь $\langle X, Y \rangle$ характеризует компендиум медицинских знаний, индексированный пропозициональными установками врача-инженера по изучению патологии/топографии/лечению. В этих массивах перфокарт мы имеем входные, выходные, латентные узлы нейронов-перфокарт. Перфокарта Z так же может быть составной и выражать не только объективные признаки болезни пациента, но и субъективные предпочтения врача-машиниста $Z = \{Z_1, Z_2, \dots, Z_l\}$. Связи между составными нейронами-перфокартами X , Y , Z задаются конфигурацией отверстий на перфокартах и наличием игл, которые попали в эти отверстия. Таким образом, между признаками задаются градуированные связи.

2) Иглы, которые прокалывают отверстия, могут быть различной толщины, остроты, со шляпками (задающие границы проникновения в перфокарты) и без шляпок (насквозь пронизывающие стопки X , Y), шляпки могут иметь раз-

личные цвета, формы и пр. В нейрокомпьютинге это означает задание весовых коэффициентов порогов срабатывания нейронов. Таким образом, между признаками задаются градуальные коннекции.

3) При представлении «знаний» важно, чтобы и инженер и машинист обладали некоторым художественным вкусом. Ведь работая перфоратором и выражая личные предпочтения, они по сути «рисуют» картину, скажем, патографии. *Машина Корсакова – это версия квазиалгоритма работы художника.* Данный тезис инспирирован аналогией с тем, как А.М.Тьюринг предлагал в 1936 г. алгоритм работы математика на основе концепции машины Тьюринга [Turing, 1936]. Обоснование тезиса, требует основательной теоретико-алгоритмической проработки, но идея уже сейчас очевидна.

Линейный гомеоскоп с неподвижными частями [Корсаков (Сыромятин), 2009, с. 18-20].

Разница в способе функционирования между планшетным гомеоскопом и линейным гомеоскопом такая же, как сегодня между планшетным и линейным сканерами. В планшетном варианте для получения результата мы накладываем табулятор (перфокарту с иглами) на эталонную перфокарту и получаем результат по тому, как иглы проваливаются в отверстия, образуя «коннекции». В линейном варианте табулятор движется по эталонной перфокарте

Неподвижные части линейного гомеоскопа, из-за которых данная машина получила свое наименование – это иглы табулятора, которые выступают из него, чётко фиксируя признаки искомой идеи. Машинист вдавлиывает иглы в табулятор при наличии признака, например, симптома заболевания. Далее двигает табулятор по перфокарте. Иглы проваливаются в отверстия. Если поисковые и эталонные признаки не совпадают, т.е. не все иглы провалились в отверстия, то действие машины продолжается. В зависимости от плотности фиксации признаков на отдельных участках табулятора, при попадании в столь же плотно скученные отверстия перфокарты, табулятор будет изменять своё положение по вертикали и горизонтали. Тогда задача машиниста состоит в том, чтобы выровнять табулятор и продолжать движение. Для этого у машиниста должно быть достаточно сил. Табулятор останавливается, когда все иглы табулятора провалятся в отверстия перфокарты, например, в результате совпадения всех симптомов заболевания. Остановка может произойти раньше, когда провалившиеся иглы (но не все) уже не позволяют двигаться табулятору по причине нехватки «сил» у машиниста. Т.е. машина остановится вследствие образования кластера совпавших признаков. Проблема остановки машины Корсакова аналогична из-

вестной теоретико-алгоритмической проблеме остановки машины Тьюринга, которая была сформулирована столетием позже, в 1936 г. [Turing, 1936]. Последняя проблема имеет исключительно формально-математический характер. У С.Н. Корсакова не так. Остановится машина или нет, зависит от прилагаемых внешних усилий, от степени сложности знаний, факторов работы машины, динамики внешних и внутренних условий функционирования. **Машина Корсакова – прототип динамической компьютерной системы.** Компьютинг динамических систем – важное исследовательское направление современного искусственного интеллекта. Разве можно утверждать, что машина Корсакова – это историческое наследие? Машина очень востребована в методологии, теории и инженерии перспективной технологии.

Линейный гомеоскоп с подвижными частями [Корсаков (Сыромятин), 2009, с. 18-21]. «Подвижные части» - это иглы-рычаги табулятора. Конструкция табулятора такова: на раме фиксируется ось, вокруг которой могут свободно вращаться иглы табулятора, закреплённые на этой оси. Один конец иглы-рычага закруглен – он будет опускаться в отверстия эталонной перфокарты. К другому концу иглы прикрепляется метка с указанием номера признака. Как и в предыдущем случае, табулятор продвигается вдоль эталонной перфокарты. На этапе настройки механизма, машинист отводит иглы, соответствующие признакам таким образом, чтобы по мере движения машины закруглённые концы игл попадали в отверстия и фиксировали его положение при остановке. Табулятор продвигается по перфокарте, отмеченные признаки попадают в отверстия, машинист прикладывает усилия для движения табулятора. В принципе, то же самое, что и в предыдущем варианте. Однако в этом механизме имеется главное усовершенствование: к иглам-рычагам, которые опрокидываются по мере движения машины, попадая в отверстия и сигнализируя о совпадении признака на табуляторе и эталонной перфокарте, С.Н. Корсаков прикрепляет бирки с наименованием признаков. Причем в зависимости от степени важности признака, бирка может располагаться от основания до вершины иглы. «Схожим образом – используя разные цвета, можно указать и относительную важность элементов идеи, с которой производится сравнение, что в свою очередь, неизбежно придаст больше точности процессу сравнения в целом» [Там же, с. 21]». Поэтому при движении табулятора, по мере попадания игл в отверстия, машинист обозревает *раскрашенную динамику развития сложной идеи* в соответствии с признаками, установленными машинистом на табуляторе и экспертными знаниями, зафиксированными инженером на эталонной перфокарте. В этом устройстве мы видим выход на логико-лингвистическую интерпретацию коннекционистской

схемы – ведь иглы табулятора проваливаются в отверстия в соответствии с коннекционистской парадигмой, в результате образования «синапсов» между эталонными перфокартами-нейронами, а на выходе имеем логико-лингвистические признаки. В следующем механизме современная проблема совмещения репрезентативной и коннекционистской парадигм получила наибольшее воплощение.

Идеоскоп ([Корсаков (Сыромятин), 2009, с. 23-25]). Данный механизм демонстрирует прототип *гибридной, нейро-лингво-компьютерной машины*. Идеи «гибридизации» прослеживались во всех механизмах, так как сравниваемые признаки имели вербальное выражение. Однако для них лингвистические характеристики не были принципиальными. Идеоскоп изначально ориентирован на выработку верифицированной системы языковых высказываний.

Табулятор представляет собой брусок с иглами-рычагами, каждый из которых обозначает некоторый признак, обозначенный языковым выражением. Машинист настраивая табулятор, формирует сложное высказывание в виде вербальных признаков сложной идеи. Табулятор движется по эталонной перфокарте, его иглы проваливаются в отверстия и пр. – все почти как в предыдущем устройстве. Имеется отличие – идеоскоп обеспечивает *целостный охват* идеи: в ходе текущей операции одновременно (параллельно) выделяются признаки поисковой идеи, которые присутствуют или отсутствуют в эталонной идее; признаки, отсутствующие в поисковой идее, но присутствующие в эталонной; признаки, отсутствующие в сравниваемых идеях, но являющиеся частью других сложных идей, представленных в эталонном наборе перфокарт (рычаги имеют различные углы наклона); признаки, распределённые по степени важности, которые присутствуют либо отсутствуют, полностью либо частично в поисковой и эталонной идеях – рычаги раскрашены, бирки имеют различное расположение по степени важности. На каждом шаге работы идеоскоп позволяет нам обзреть целостную картину распределения признаков именно у *идеи*, а не у понятия как набора существенных признаков предмета исследования. Одновременно демонстрируются не только существенные, но так же несущественные и даже несовместимые признаки в градуальном, градуированном и n-арном формате.

В целом идеоскоп эксплицирует динамику развития или деградации идеи, её *логико-лингвистическую мультипликацию*. Коннекционистская парадигма играет роль обоснования/опровержения логико-лингвистических моделей репрезентативной парадигмы, так как на табуляторе отмечаются высказывания,

требующие обоснования «знаниями» эталонной перфокарты. Совершенно гениальная конструкция!

Протонейрокомпьютер Корсакова – это чистая коннекционистская парадигма искусственного интеллекта.

«Чистая», потому что к сложной динамической системе взаимосвязи идей не примешиваются нейрофизиологические, биологические и социальные метафоры. Как доказывалось в [Алексеев, 2011-1; 2012-1;2;3], машина Корсакова – это не только факт истории. Это – концептуальный проект, очень востребованный в перспективной когнитивно-компьютерной технологии. Однако доведение машины Корсакова до современных проблем искусственного интеллекта требует больших исследовательских усилий. В следующей работе будет предложена усиленная машина Корсакова, которая позволит ближе «подъехать» к решению ряда проблем ИИ: динамических систем, естественных видов, социогуманитарных понятий, фреймов, индукции, моделирования «смыслов». Перспективы внедрения машины Корсакова в современные концептуальные схемы компьютеринга колоссальны и решение проблем ИИ предлагается осуществлять сквозным способом: от наносистемотехники элементарных логических вентилей до онтологии когнитивного содержания индивидуальных и социальных процессов.

День рождения нейрокомпьютинга (и других моделей коннекционистской парадигмы) предлагается отмечать **7 октября 1832 г.** (24 сентября по старому стилю), с даты цензорского позволения издания статьи С.Н. Корсакова.

Таким образом, с полной уверенностью утверждается следующее:

Первый нейрокомпьютер разработан в России (1832 г.).

Проследим дальнейшую историю нейрокомпьютинговой инженерии, связанную с реалиями сегодняшней компьютерной технологии. Здесь у истоков мы снова видим отечественных учёных¹.

Идея нейрокомпьютинга имманентна российской ментальности, по всей видимости, по причине пространственной обширности России и разнородности населения, если придерживаться точки зрения географического детерминизма в определении факторов общественного сознания. Для демонстрации этой мысли проследим пост-корсаковскую историю нейрокомпьютинговой инженерии, свя-

¹ Исторический материал любезно предоставлен заместителем главного редактора журнала «Нейрокомпьютеры: разработка, приложение» А.В. Савельевым.

занную с реалиями сегодняшней компьютерной технологии. Здесь у истоков мы увидим отечественных учёных.

1928 – 1935-е годы – П.К. Анохин, автор теории функциональных систем и идеи самоорганизующейся обратной связи. Работы явно или неявно используются в современном теоретическом и практическом нейрокомпьютинге. В последующем Н. Винер значительно упростил анохинские идеи о самоорганизующейся обратной связи, обратная связь у него оказалась без «самоорганизации». Упрощение привело к тому, что кибернетика в 1990-е годы стала «соперничать» с синергетикой как наукой о самоорганизации. Хотя было достаточно и одной научной дисциплины, если бы не пошли на поводу теоретической и инженерной простоты.

1930-е годы – Н.А. Бернштейн, основоположник нейробиомеханики двигательной активности. Все (!) современные роботы, так или иначе, используют эти результаты. Лишь затем появляются работы Н. Винера (1942 г.), У.Мак-Калоча и У.Питтса (1943 г.), Д.Хебба (1949 г.), Ф.Розенблатта (1958 г.). В этих работах мы видим всё большее и большее упрощение и огрубление сложнейших представлений о нейральной активности. В результате этого зарубежный нейрокомпьютинг долгое время (по сути, в 1960–80-е годы) пребывает в стагнации. В то же время отечественная школа нейрокомпьютинга достигла значительных успехов.

1960 – 1970-е годы – В.М. Глушков, Н.М. Амосов, А.М. Касаткина, Э.М. Кукуль синтезировали цифровые М-автоматы («мыслящие» автоматы).

1981 – 1987-е годы – А.В. Савельев, разработчик методов натурального нейромоделирования, применение которых позволило сделать и предсказать ряд нейрофизиологических открытий [Савельев, 2004]. Нейрокомпьютинг в СССР институализировался.

21 апреля 1971 г. – был создан НИИ нейрокибернетики АН СССР, впоследствии НИИ НК РАН им. А.Б. Когана (г. Ростов-на-Дону). Зарубежный нейрокомпьютинг кое-как выбился из кризиса в 1986 г. благодаря «переоткрытию» метода обратного распространения ошибки (Д.И.Румельхарт, Дж.Е.Хинтон и Р.Дж.Вильямс), хотя незадолго до этого аналогичный метод активно использовался в работах А.Н. Колмогорова, С.И.Барцева и В.А.Охонина, а также А.Н. Горбаня, который принадлежал другой школе – красноярской.

К успехам отечественного нейрокомпьютинга следуют отнести такие разработки, как, например, «Эмбрион» В.Д. Цыганкова, успешно работающий в медицинских приложениях и военной технике. «Эмбрион» работал с 1960-х го-

дов в системах наведения противоракет, основан на принципах теории функциональных систем П.К. Анохина. На западе не было и до сих пор нет ни одного подобного реально работающего аппарата, решающего задачи, которые не под силу обычным компьютерам. Следует отметить, что советские ракеты летали лучше зарубежных. Их механизм был основан на самоорганизующейся обратной связи, а не на винеровской («обратная связь минус самоорганизация»).

Кроме того, необходимо перечислить следующие проекты: нейрочипы А.И. Галушкина, используемые в управлении вертолётами, вертолётими-роботами; многопроцессорные с перестраиваемой структурой нейрочипы А.В. Каляева (г. Таганрог); нейристоры — нейроподобные матрицы А.А. Зенкина для автоматического доказательства теорем; реверберационный нейропроцессор А.В. Савельева [Савельев, 2006-2]; нейродетекторы Е.Н. Соколова (теоретическая часть) и Л.А. Шмелёва (практическая составляющая работ) по нейронам-детекторам; работы Н.П. Бехтеревой по расшифровке мозговых кодов; работы директора Института нейрокибернетики АН СССР и РАН А.Б. Когана и его коллектива; работы Н.В. Позина (1950 – 60-е годы) и его группы «Бионика» и другие работы отечественных «нейрокомпьютерщиков» [Савельев, 2004].

Живое активное состояние современных отечественных работ в области нейрокомпьютинга можно отследить по многочисленным конференциям, которые проводятся в нашей стране. Особенно следует отметить всероссийскую конференцию «Нейроинформатика», которая проводится ежегодно в МИФИ, и всероссийскую конференцию «Нейрокомпьютеры и их применение», также ежегодно проводимую на базе Московского городского психолого-педагогического университета¹. Несомненно, что работы отечественных исследователей активно использовались и используются зарубежными коллегами. Так, например, один из первых западных нейрокомпьютеров, «Адалин», разработанный Б.Уидроу и М.Хоффом (1960 г.), построен на основе адаптивного линейного сумматора (название происходит именно от данного словосочетания), впервые предложенного Я.З. Цыпкиным в 1948 г.

Если теперь рассмотреть «викиномичную» историю нейрокомпьютинга, которая представлена в многочисленных работах, выставленных в Интернете, и на основе которых написана работа А.Н. Липова, раскрывающая историю ней-

¹ Широкий обзор нейрофизиологических моделей, используемых в методологии ИИ, представлен в [Петрунин, 2010].

рокомпьютинга, то мы в ней не увидим отечественных заслуг в развитии нейрокомпьютинга¹:

1942 г. – начало вики-нейрокомпьютинг; Н. Винер вместе с соратниками публикует работу о кибернетике, одной из основных идей которой является представление сложных биологических процессов математическими моделями;

1943 г. – У.Мак-Калох и У.Питтс предлагают модель нейрона, формализуют понятие нейронной сети в статье о логическом исчислении идей и нервной активности;

1949 г. – Д.Хебб предлагает первый алгоритм обучения персептрона;

1958 г. – Ф.Розенблатт разрабатывает принципы построения персептрона (математическая и компьютерная модель восприятия информации мозгом), первого технического воплощения нейронных сетей;

1960 г. – Уидроу совместно со своим студентом Хоффом на основе дельта-правила (формулы Уидроу) разработали Адалин (ср. с предыдущими сведениями об Адалине), который сразу начал использоваться для задач предсказания и адаптивного управления;

1969 г. – М.Минский и С.Пейперт издают книгу «Perceptrons», где доказывают принципиальную ограниченность возможностей персептронов и т.д.

Видим ли мы в данном обзоре нейрокомпьютинга сведения об отечественных разработках? Отнюдь нет. Читать такую вики-историю нейрокомпьютинга неприлично. Здесь полнейшее забвение отечественной истории. Аналогична вики-история и с другими именами выдающихся основателей компьютерной науки, кибернетики, синергетики и прочих общетеоретических и частнотеоретических оснований современной технологии. В связи с тем, что «за державу обидно», предлагается считать проект С.Н. Корсакова своеобразным символом возрождения отечественных идей в науке – идей, про которые сегодняшние СМИ напрочь забыли. Задел в таком подходе к корсаковским идеям обоснован следующим тезисом:

Родиной нейрокомпьютинга является Россия.

Это важно для аутентичной трактовки инженерной проекции нейрокомпьютинговой тематики на сферы социокультурной жизни.

Справедливости ради отметим, что в средневековье имелся более древний проект когнитивной системы, в которой участвовали модели ментального и ме-

¹ Липов А.Н. Из истории нейрокомпьютинга// Полигнозис, 2(42), 2011, С.120 – 130.

ханизмы обработки данной модели. Имеется в виду машина испанца Раймонда Луллия (ок. 1232 – ок. 1316). Изначально машина предназначалась для юридической практики и реализовывала постглоссаторский юридико-догматический логико-символический метод толкования собраний источников римского права (глосс) применительно к конкретным житейским ситуациям, а также метод координации двух силлогизмов, один из которых представлял положения естественного права, другой – положительного права. Машина позволяла выводить заключения из данного полисиллогизма. В последующем проект послужил основой труда «Великое искусство» (*Ars magna*), который претендовал на универсальный способ открытия любых истин бытия. Устройство машины состояло из подвижных концентрических бумажных кругов, разделённых поперечными линиями на отделения, на которых помечались общие понятия. Вращение кругов приводило к множеству новых комбинаций. Идея пользовалась успехом среди ученых вплоть до 18 в., хотя за абсурдность и бессмысленность комбинаторики универсалий подвергалась уничижительной критике со стороны философов «когнитивистской» ориентации (Декарта, Гегеля, Пирса) и карикатурному осмеянию в художественной литературе в образе луллитов (Дж. Свифт, «Путешествие Гулливера»). Тем не менее, здесь мы видим первую попытку *механического воспроизведения феноменов человеческого мышления*. Именно поэтому проект Р. Луллия следует считать первым проектом искусственной когнитивной системы, но в контексте современной электронно-технической базы данный проект представляется невостребованным, в отличие от проекта Корсакова и Бэббиджа.

Итак, корсаковский проект (1832 г.), ориентированный на демографию и медицину, начинает получать международную известность в начале 21 в. Проект был реализован, работал и успешно применялся.

Также известна аналитическая машина Бэббиджа (1833 г.), предназначенная для решения экономических задач.. Этот проект получил международное признание в 1842 г. благодаря статьям «первого программиста» Ады Лавлейс, переведённым итальянцем Л. Менабреа с английского языка на французский, который был в те годы международным (Как мы ранее отметили, С.Н. Корсаков представил свой проект также на французском языке.) Проект не был реализован.

Еще один проект – электромеханических табуляторов американца Голлерита, использующих перфокарты, – появляется в 80-е годы 19 в. Служил для переписи народонаселения Америки в 1880 – 90 гг. Фирма IBM «выросла» на

табуляторах Голлерита, в Московском политехническом музее имеется соответствующая экспозиция. Проект действующий.

Аналогом машины Бэббиджа стала машина Тьюринга (1936, 1950 гг.), в которой использовалось формальное определение понятия алгоритм [Turing, 1936; 1950]. Проект действующий, служит основанием современной компьютерной технологии.

В 2012 г., когда впервые в широком всероссийском масштабе праздновался юбилей, исполнилось 225 лет со дня рождения С.Н. Корсакова и 180 лет со дня написания эпохальной для когнитивной науки статьи. Изобретения С.Н. Корсакова были столь революционны, что современники не смогли их по достоинству оценить, как в силу субъективных причин – академической невосприимчивостью к новаторству изобретения, так и объективных причин – для практической пригодности устройства требуется мощь современной электронной техники. В наших силах должным образом применить наследие нашего выдающегося соотечественника, более полутора века пребывавшее в забвении.

Имеется множество шуток по поводу «вселенской» роли нашей страны: Россия – это родина слонов, в пермской деревне родился Заратустра, доегипетская цивилизация возникла на Южном Урале. Однако факт остается фактом: **протонейрокомпьютер был создан в России.** В этом «нейрокомпьютере» не применяются термины: «нейрон», «синапс», «аксон», «дендрит» и пр. В науке «нейронные» термины появятся позже, в конце 19 в. после изучения под микроскопом нервных клеток и закрепятся с 1906 г. после получения нобелевской премии итальянцем Камилло Гольджи и испанцем Сантьяго Рамоном-и-Кахалем. Однако в машине Корсакова имеется много схожего с современной парадигмой коннекционизма и теорией нейрокомпьютинга.

Вернёмся к мысленным экспериментам комплексного теста.

Тест Тьюринга «с мозгом»

Многие авторы частных ТТ особо подчёркивают то, что нейрокомпьютер – это наиболее адекватное средство для имитации интеллектуального и сознательного поведения. Так, например, Дж. Сёрль в знаменитом ЧТТ «Китайская комната» (1980 г.), полагал, что одно из наиболее сильных опровержений тезиса «компьютер понимать не может» происходит со стороны специалистов по «моделированию мозга». Конечно, он приводит совершенно невразумительный и, прямо скажем, глупый отвод «нейрокомпьютерного» возражения путем аналогии нейронной сети с системой водопроводных труб, в которой «соединения труб» – синапсы, а положения вентиля в трубах – «степень возбуждения сети»

[Серль, 2006, С.17]. Многие оппоненты и сторонники Дж. Сёрля полагают, что «Китайская комната» послужила современной волне развития когнитивной науки в попытках технически опровергнуть его тезис и доказать возможность реализации (по крайней мере, имитации) феноменов сознания путём моделирования естественных когнитивных механизмов, в первую очередь, нейрокомпьютерными средствами.

Нейрокомпьютерная тематика также привлекает внимание зомбифилов (зомбистов, антизомбистов, нейтральных зомбистов) всех мастей. Так, Т. Полджер, О. Фланаган, отстаивая принцип несущественности сознания для осуществления высокоинтеллектуальной деятельности, полагают, что более сильное испытание, нежели тест Тьюринга, представляет «тест Тьюринга с мозгом» («Brainy–Turing–Test»). Этот тест в дополнение к поведенческой эквивалентности требует эквивалентность нервной системы, т.е. помимо тождественности ввода-вывода нужна тождественность с мозговой активностью [Фланаган, Полджер, 2006, С. 103 – 112].

В тесте Харнада «тест Тьюринга с мозгом» знаменует «тотальную неотличимость» компьютерной системы от человека, включая мельчайшие внутренние нюансы телесного строения. Это – уровень микрофизической (в контексте нашей работы – нанотехнологической) неотличимости естественных и искусственных сознательных систем.

Таким образом, многими исследователями ТТ нейрокомпьютер воспринимается в роли мысленного примера, обозначающего высшую степень возможности решения психотехнической проблемы.

Однако рассмотрим, как относятся основатели «нейрофилософии» и создатели нейрокомпьютерного теста П.иП.Черчленды к основному вопросу искусственного интеллекта «Может ли компьютер мыслить?».

Патриция и Пол Чёрчленды положительно ответили на этот вопрос в своей статье «Может ли машина мыслить?» [Churchland, P.M., Churchland, P.S., 1990]. Но потребовали усовершенствовать машину, которая способна имитировать интеллектуальное поведение: такая машина должна моделировать мозговую активность человека.

На наш взгляд, тест Чёрчлендов разработан поспешно и не доведен до конца.

Во-первых, П. и П. Черчленды не внимательны к аргументации Х. Дрейфуса, принципиально отрицающего возможность ИИ. «Развитие ИИ идет рука в руку с постоянно растущими вычислительной мощностью, объемом памяти и

сложностью базы знаний», - полагают они [Там же, р.33]. Однако не замечают того, что одна из главных проблем, сформулированных Дрейфусом, состоит в неспособности компьютера формировать явные «знания» на базе фоновых, неявных «знаний». Например, компьютер не способен оперировать метафорами. Количественными достижениями компьютерной технологии этого не преодолеть.

Во-вторых, они полагают, что в критику оригинального теста Тьюринга нейрокомпьютерный тест вносит вклад, сопоставимый с тестом Сёрла (Китайская комната). Машина Тьюринга – это символьная дискретная машина. Она не способна репрезентировать каузальные зависимости между психическими явлениями и непрерывными нейрофизиологическими функциями. Требуется машину Тьюринга заменить нейрокомпьютером, который это делает.

Однако такое гипотетическое утверждение требует верификации. Джон Сёрль апеллирует к интуиции феноменального сознания, к аподиктическим суждениям первого лица. Поэтому эпистемологическая значимость его работы на порядок выше.

В-третьих, Чёрчленды считают, что Чёрч является основоположником идеи интеллектуальной машины, которая играет тьюринговую игру в имитацию. Интеллектуальные функции представимы посредством эффективно вычислимых функций, так как любой ответ игрока на вопрос судьи можно получить за приемлемое время и при ограниченных ресурсах. Эти функции, в свою очередь, редуцируются к рекурсивным функциям, а они базируются на примитивно рекурсивных функциях. Последовательная машина Тьюринга обеспечивает инженерную формализацию идеи А. Чёрча. Однако в имитацию интеллекта играет не машина Тьюринга (1936 г.), а универсальная вычислительная цифровая машина (1950 г.). В нее включены некоторые элементы вероятностной параллельной машины. Например, используется Книга правил, которая состоит из множества страниц (лент), наложенных друг на друга [Алексеев, 2013-1, с.44-46]. Более того, для описания работы обучаемой машины-ребенка А. Тьюринг использует нейронную сеть! [Там же, с.231-237].

В-четвертых, не понятна специфика «нейрокомпьютинга» в смысле Чёрчлендов. Выясняется, что нейральные системы – это параллельные, отказоустойчивые, персистентные, высокоскоростные машины [Churchland, P.M., Churchland, P.S., 1990, p.35–36]. Главное то, что способ их функционирования основан не на манипуляции символами, а на векторных преобразованиях непрерывных значений. Однако, где здесь нейрокомпьютинг? Такая машина

Чёрчлендов представима независимыми параллельными машинами Тьюринга, которые управляются многомерной автоматной таблицей в соответствии с теорией формальных нейронов. Между лентами таких машин отсутствует внутренние связи. Возможны лишь внешние связи. Они согласуют лингвистические комплексы, зафиксированные на параллельных лентах. Получается символическая парадигма ИИ, усложненная метаязыком теории нейронных сетей. Коннекционизма-то нет!

Таким образом, создаётся впечатление, что Чёрчленды невнимательно и не до конца прочитали статью А. Тьюринга. Коннекционистская парадигма не просматривается. Нейрокомпьютерный тест Тьюринга проработан поверхностно и требует уточнения. Для его разработки перспективно использование машины Корсакова-Тьюринга (см. пп.3.3.3). Коннекционистская «вычислимость» организована на субсимвольном уровне посредством внутренних связей между лентами машины Тьюринга.

В связи с нейрокомпьютерным тестом Тьюринга следует отметить, что в литературе иногда встречается мультимедийный ТТ, способный тестировать х-системы в диапазоне симультанных способов репрезентации информации [Clifton, 2003]. Очевидно, что с видео- и аудиорядами (в мультимедиа они существенным образом превалируют над текстовым форматом представления информации) эффективно способен справиться лишь нейрокомпьютер. Классический ИИ к мультимедийному тесту не готов. Мультимедиа, в свою очередь, это базовый инструментальный выражения когнитивных феноменов э-культуры-1 и когний превращенной формы традиционной культуры – э-культуры-2.

В завершение изучения метафизической и инженерной проекции нейрокомпьютерной тематики на э-культуру подведём итоги.

Значимость образной информации (за изучение её «ментальных» параметров должен отвечать мультимедийный ТТ) – 90 % от вербальных форм. Таким образом, роль нейрокомпьютера в реализации полноценного ТТ, отвечающего требованиям традиционных форм э-культуры (э-культуры-1), не говоря уже об остальных формах (2-й и 3-ей), следует оценивать как 9/10. Учитываем значение показателя роли нейрокомпьютера, полученного ранее из анализа метафизической проекции (1/3). Получаем приблизительно 62 %. Конечно, это крайне условная оценка. Тем не менее, мы имеем возможность «количественно» оценить степень участия нейрокомпьютера в развитии э-культуры.

Таким образом, на основе вышеизложенного, можно сделать следующие выводы.

1. Нейрокомпьютер представляется одним из главных детерминантов развития э-культуры, так как, будучи составной частью междисциплинарных исследований ИИ, он существенным образом конституирует её материальные и духовные составляющие.

2. В э-культуре-1, реализуя традиционные формы культуры на базе НБИКС-технологий, нейрокомпьютер выполняет задачи, характерные для сегодняшнего уровня развития технологии ИИ. Степень участия в «социокультурном» процессе мы оценили на 62 %. Значение этого условного показателя целесообразно в некоторой форме учитывать в распределении средств поддержки исследований ИИ. Остальные 38% приходятся на другие исследования ИИ.

3. В э-культуре-2 нейрокомпьютер, реализуя превращённые формы сознания, выступает в качестве необходимой компоненты систем виртуальной реальности или в качестве функционального заменителя некоторых участков коры головного мозга, например, лечение болезни Паркинсона, устранение фантомных болей – это прерогатива нейрокомпьютерных чипов.

4. В э-культуре-3 при решении психотехнической проблемы роль нейрокомпьютера становится наиболее значимой. Укажем лишь на недавний факт признания со стороны Президиума РАН фундаментальных перспектив, открывающихся во всех сферах э-культуры благодаря предполагаемой реализации проекта расшифровки нейродинамических кодов субъективной реальности (это – нейрокомпьютерная проблематика!). Проект, предложенный 50 лет Д.И. Дубровским, лишь сегодня находит практическую поддержку в плане реализации. Доклад Д.И. Дубровского, прозвучавший на Научной сессии Общего собрания РАН 15 декабря 2009 г., имел следующее название: «Мозг: фундаментальные и прикладные проблемы. Философские подходы к проблеме «мозг и психика»: к вопросу о расшифровке нейродинамических кодов явлений субъективной реальности». В проекте нет рассуждений по поводу имитации феноменов сознания. Ставится радикальный вопрос относительно реального воспроизведения или, собственно, фактической реализации феноменов сознания на базе электронных технологий. Проект широко известен как «информационный подход к сознанию» (точнее, к субъективной реальности). Перспективы, которые он сулит, очевидны – человек приобретает способность по своей воле оперировать нейродинамическими носителями феноменов сознания, управлять энергетическим обеспечением этих операций, в том числе соответствующими биохимическими процессами; изменять программы действий, следовательно, изменять их кодовые нейродинамические структуры; расширять контуры психиче-

ской регуляции; интенсифицировать творческие процессы; создавать новые ресурсы психической саморегуляции, причем не только функциональной, но и нравственной. Снова повторю – это специфически нейрокомпьютерная тематика. Проект – глобальный и, несомненно, имеет не только стратегический характер для развития э-культуры страны, но и судьбоносное значение для человеческой цивилизации в целом.

5. Несмотря на то что нейрокомпьютерный проект расшифровки кодов нейромозговых коррелятов психических феноменов поражает воображение и захватывает дух, на мой взгляд, предпочтительнее выбрать менее радикальный путь решения психотехнической проблемы. Необходимо (возможно, параллельно с проектом расшифровки нейродинамических кодов) добиваться частных тактических успехов на пути аппроксимативной функциональной имитации феноменов сознания. Здесь перспективны проекты искусственной личности и искусственного общества. В первом проекте решаются вопросы создания «псевдосознательных» механизмов имитации интеллектуальной, эмоциональной, волевой сферы, изучаются проблемы компьютерного воспроизведения морального поведения, свободы выбора, моделируется самость, «Я». Во втором проекте изучаются искусственные агенты, устанавливающие «социальные связи» на основе моделирования всего многообразия факторов, присущих реальным человеческим общностям – материально-производственных, экономических, идеологических, морально-нравственных, религиозных и др. Решение отдельных задач, возникающих при реализации проектов искусственной личности и искусственного общества, т.е. частные незначительные тактические успехи, на мой взгляд, – это и есть правильная стратегия решения психотехнической проблемы.

В этой стратегии нейрокомпьютерным когнитивным механизмам отводится базовая реализационная роль, нейрокомпьютерная технология представляется ключевой предпосылкой конструирования социокультурной реальности.

Нейрокомпьютер – базовая технологическая детерминанта развития электронной культуры.

3.3.3. Инженерно-методологические аспекты: комплексная машина Корсакова–Тьюринга как принципиальный проект интеллектуального компьютера

В философских исследованиях любой проблемы главная задача состоит в нахождении фундаментальных универсалий, эпистемологических средств, моделей, метафор, концепций – всего того, на чём, собственно, базируется человеческое представление мира. В философских исследованиях ИИ главная задача заключается в выявлении фундаментальной компьютеринговой метафоры, посредством которой человек воспринимает и понимает окружающий мир – природу, отношения с другими людьми, самого себя. Сегодня в когнитивных исследованиях превалирует метафора вычислений (исчислений) когнитивных феноменов, которая выросла на почве англо-саксонской культуры. Современный компьютер – порождение того, что принято называть «западным менталитетом». Западная компьютеринговая метафора, конечно, сильна, но недостаточна для реализации когнитивных феноменов, которые рассматриваются в контексте других культур. Например, когнитологическая проблематика сознания несоизмерима для англо-американской традиции и восточной традиции трактовки слова «сознание» [Алексеев, 2010-2]. Здесь возникают проблемы так называемого радикального перевода, – носители разных языков просто не способны понять друг друга. Превосходной инженерной проекцией современной когнитивно-компьютерной метафоры представляется машина Бэббиджа – Тьюринга, для которой характерно отождествление ментальных феноменов с логическими состояниями вычислительной машины. Данная метафора породила современный когнитивный функционализм, согласно которому когниция – это функция сложной компьютерной системы, по большому счёту, не зависящая от конкретного способа реализации, – будь то человеческий мозг, тина марсианина или электронные чипы технического устройства. С когнитивным функционализмом, однако, возникли глубочайшие проблемы, инспирированные как неспособностью функционализма выразить переживания от первого лица (мои личные психические состояния), так и примитивностью компьютерной метафоры. В связи с этим ещё раз повторим вопрос Д. Деннета: разве может современный и, однако, априори старый и дряхлый компьютер выразить многообразие ментального [Dennett, 1984]? «Не может!» – отвечает Д. Деннет и предлагает крайне изощрённый гетерофеноменологический метод, который в некоторой степени восполняет пробелы современного компьютеринга нейрокомпьютерной методологией. Но в его методе мы видим аналитический подход, ориентированный

на лингвокомпьютинг абстрактных сущностей – мемов (это социокультурные гены). Имеются и иные подходы по восполнению ущербности существующей компьютерной метафоры. Однако все они, как правило, так или иначе оказываются вкорененными в метафору Бэббиджа – Тьюринга, либо, по крайней мере, принципиально выразимы средствами данной метафоры.

На наш взгляд, имеется более простой способ обогатить представление о компьютеринге когнитивных феноменов. Для этого предлагался проект машины Корсакова. Однако в отдельности она не способна реализовать комплексный тест Тьюринга. Очевидно то, что она машина Корсакова обладает большим разнообразием репрезентативного материала (перфокарта Корсакова). Машина Тьюринга, напротив, имеет примитивный способ репрезентации информации (лента Тьюринга). Однако она имеет концепцию работы автомата, что отсутствует в первой машине. Отсюда очень простая идея комплексной машины:

Машина Корсакова-Тьюринга состоит из перфокарта Корсакова, которую перфорирует и табулирует не Инженер и не Машианист, а устройство управления машины Тьюринга, функционирующее в соответствии с автоматной квазитаблицей, представленной на перфокарте Корсакова.

Машина Корсакова-Тьюринга в комплексном тесте Тьюринга

Для демонстрации роли и места машины Корсакова в КТТ рассмотрим её в формате системы, которая выступает в качестве средства реализации комплексного ТТ. В табл.1 мы видим, что КТТ в составе представленной совокупности тестов – это очередной частный ТТ, выполняющий особую задачу (самособирания частных ТТ с целью аппроксимативно полного покрытия многообразия когнитивных феноменов и решения общих задач их атрибуции, предикации и реализации.

Машина Корсакова – это подсистема инструментария, реализующего КТТ. Другой подсистемой является машина Тьюринга – Бэббиджа. Такое совместное применение двух фундаментальных метафор обусловлено не только тем, что машина Корсакова – это коннекционистская метафора вычислений, а машина Тьюринга – репрезентативная метафора. Мы решили совместить две метафоры, так как в машине Тьюринга налицо возможность машины автоматически сохранять состояния и переходить их одного состояния в другое. В машине Корсакова такого нет. С другой стороны, у Тьюринга нет того, что имеется в машине Корсакова – особого способа представления гетерономной информации, что особенно четко проявляется при полисемантической интерпретации лингвистических средств инструментария КТТ.

Поскольку, согласно нашим представлениям, комплексный тест Тьюринга – это частный ТТ, выполняющий специфические функции комплексирования других частных тестов, включая самого себя, т.е. вследствие синергетической специфики, инструментарий КТТ предполагает *гетерархическую* организацию и гибридную (лингво- и нейрокомпьютерную) программную реализацию. Общие идеи гетерархических систем задают систематичность, рекурсивность, гетерогенность и динамичность инструментальных средств ЧТТ, инкорпорируемых в КТТ (Янковская, 2008; 2011). Систематичность предполагает виртуальную системную целостность частных тестов. Рекурсивность выражается автогенеративным программированием инструментальных средств. Гетерогенность проявляется в функционировании в рамках единой «саморазличающейся» системы взаимозависимых уровней, каждый из которых обладает собственной логикой развития. Динамичность предполагает изменение отдельных уровней и реструктурирование системы в целом под влиянием других уровней при сохранении концептуального единства, обозначенного идеей комплексного теста Тьюринга.

Лингвистические средства машины Корсакова в контексте КТТ обеспечивают библиографическую работу идентификации, систематизации, унификации, координации, обобщения, дифференциации, интеграции ЧТТ и их компонентов. Синтаксически они выразимы нестрогим первопорядковым исчислением предикатов, интерпретированном на 3D-онтологии, объединяющей суждения трех общих планов, специфицируемых в соответствии с последующей детализацией: феноменологического (от первого лица), научно-теоретического (биологических, нейрофизиологических, психологических и других дисциплин), инженерно-компьютерного (проекты реализации ЧТТ). Данные три общие плана при условии последующей конкретизации предполагают полионтическую интерпретацию, либо, если назвать по-иному, полимерную семантику языка комплексирования ЧТТ.

Полимерная семантика означает наличие многообразия локальных целостных подсистем интерпретации многочисленных комбинаций суждений, которые выражают следующее: 1) личные представления о психических переживаниях и явлениях социальной жизни; 2) теоретические осмысления этих явлений; 3) инженерно-компьютерные проекты. Для реализации полимерной семантики собственно и востребована машина Корсакова. В отличие от оригинального варианта, где мы видели 2D-измерения («топография» и «патография»), для нашего случая имеем три общих измерения: 1) суждения от первого

лица; 2) теоретические суждения нейрофизиологического, социологического, экономического и прочих планов; 3) конкретные компьютерные проекты, взятые как в целом, так и разобранные по подсистемам.

Будем теперь «прокалывать корсаковскими иглами» 3D-перфокарты, представляющие обозначенные выше измерения. Если теперь накладывать 3D-перфокарты одна на другую и варьировать глубину проникновения игл в ряды наложенных друг на друга таблиц, то мы будем получать комбинаторные комбинации локальных целостностей, которые характеризуют, скажем, для феномена «переживание» триединство «ощущения переживания» / «теории переживания» / «проекты реализации переживания». И, собственно, здесь, на этапе полимерного конструирования предметной области, проявляется преимущество машины Корсакова перед машиной Тьюринга. И в первом и во втором случае мы видим произвол наблюдателя (судьи, разработчика). В оригинальном ТТ наблюдатель произвольно приписывает когнитивные феномены некоторой x -системе, которую он оценивает. В принципе, наблюдатель пассивен. В случае же с машиной Корсакова наблюдатель выступает в роли активного лица – он самостоятельно конструирует реальность, выступая в роли творца когнитивных сценариев «жизни», которую «проживает» в формах электронной культуры или предлагает «проживать» другим членам своего сообщества.

Корсаковский наблюдатель «прорубает окна» в компьютеризованную реальность, тьюринговский наблюдатель – общается с миром сквозь открытое окно.

Информационные средства машины Корсакова, помимо стандартных, включают в себя базы «смыслов», которые репрезентируют эпистемические генерализации «данных» и «знаний», характеризуя тем самым концентрированные «знания о знаниях». В современной методологии компьютерного моделирования «смысла» предложено большое количество интересных моделей «смысла» – психологических, лингвистических, кибернетических, синергетических, риторических, поэтических, семиотических, герменевтических, математических и др. [Алексеев, 1998; 2005-8]. В них прослеживаются три взаимосвязанные трактовки понятия «смысл».

1. *Интенциональное определение*: «смысл» – это целевая направленность, ценностная ориентированность. К данной трактовке относятся выражения «смысл жизни», «смысл истории», «смысл бытия», т.е. «делать что-то с мыслью».

2. *Контекстуальное определение*: «смысл» – это выражаемый знаком способ задания значения, т.е. более широкий контекст для значений языковых выражений, связанных текстов, образов сознания, ментальных состояний и др. Обыденное словоупотребление, соответствующее этой позиции, связывает «смысл» с идеей, сущностью, точкой зрения («в смысле того или иного»).

3. *Контентуальное определение* (от слова «контент» – содержание). Ему отвечает термин *со-мыслие*. Смысл как контекст (смысл – «вне» всякой мысли) и смысл как контент (смысл «состоит» из мыслей) – это принципиально разные уровни анализа. Контентуально трактуемый «смысл» как личностно-значимое конкретно-целостное систематическое единство «мыслей» характеризует их эмерджентность и синергийность. Имеется ряд авторитетных утверждений, что «со-мыслие» присуще специфически русскоязычному варианту употребления слова «смысл».

На наш взгляд, подход к компьютерному моделированию «смысла», который получается в результате интерпретации машины Корсакова, уверенно покрывает все эти подходы. Во-первых, интенциональное определение «смысла» фиксируется в «силе» прокалывания иглой стопки перфокарт. Во-вторых, контекстуальное определение четко прослеживается, если перфокарту Корсакова интерпретировать как контекст, в котором разворачиваются «тексты», помечающие, скажем, индикаторы информационных ресурсов в местах разметки перфокарт иглами. В-третьих, контентуальная парадигма проявляется, если учесть обозначенное выше полионтическое определение предметной области и воспринимать помеченную стопку перфокарт как единое целое. Для этого «единства» возможны не только эпистемические обобщения, но и, скажем, эмоционально-художественные представления, если соответствующим образом раскрасить индикаторы информационных ресурсов и их динамические связи. Подобного рода демонстрация модели «смысла» по своим эстетическим особенностям, на наш взгляд, сродни фрактальной красоте динамических систем.

Информационные средства машины Корсакова – это преимущественно базы «смыслов».

Каким образом организовать моделирование "смыслов" средствами машины Корсакова-Тьюринга - в следующем подразделе.

Программные средства машины Корсакова представляются в общем случае набором клуджей – т.е. программ, которые работают не для всех учтенных случаев спецификации и документации алгоритмов (все эти случаи не подлежат скрупулёзному учету). Каждый программный клудж отвечает за реализа-

цию частного теста или его компоненты. В целом, мы не видим здесь никакой интегрально-систематизирующей методологии, аналогичной современным крупным программным архитектурным решениям. Хотя они не исключаются. Идеал организации способа программирования машины Корсакова – свободное программное обеспечение. Хотелось бы обозначить данное направление программирования как «гетерономное» программирование, отвечающее перечисленным выше параметрам гетерархических систем.

Парадигма программирования машины Корсакова – гетерономное (ключевое) программирование.

Организационные средства машины Корсакова предпочтительно реализовать по схеме «мета-веб-сообщество». Они призваны объединить в социальную сеть специалистов различного профиля и обеспечивать междисциплинарную работу интернационального сообщества для систематического изучения способов компьютерной реализации когнитивных феноменов.

Гибридная репрезентативно-коннекционистская интерпретация машины Корсакова

Сегодня предложены две равноправные трактовки корсаковского проекта, которые с учётом терминологических особенностей методологии ИИ правомерно назвать репрезентативной (символьной) и коннекционистской (нейральной) интерпретациями машины Корсакова. *Репрезентативная интерпретация* была предложена А.С. Михайловым в работе [Михайлов, 2009], где детально представлялось теоретико-множественное описание способа функционирования всех пяти машин. Ярко показано, что «интеллектуальные машины» Корсакова реализуют функции экспертных систем, обеспечивают эффективный многокритериальный поиск информации, фасетную классификацию, программное управление «базой знаний» и прочие функции современных систем репрезентативного ИИ.

Коннекционистская интерпретация как отмечалось выше, предложена в ряде работ автора и представлена выше. Перспективы применения машины Корсакова колоссальны. Дело в том, что изменение компьютерной метафоры, которая лежит в основании когнитивной науки, автоматически влечёт за собой коренную трансформацию методологии науки, её теоретических постулатов и практических перспектив. Поэтому автор предлагает дополнить традиционную (параллельно) линейно-символьную компьютерную метафору Бэббиджа – Тьюринга – Неймана более мощной, объёмно-нейросетевой компьютерной метафо-

рой Корсакова. Технологическую модернизацию предлагается осуществлять сквозным способом, от компьютеринговой системотехники элементарных логических вентилей до 3D-онтологии когнитивных феноменов.

В контексте э-культуры предпочтительно рассуждать о явной технологической детерминации социокультурного процесса. Современная базовая концептуальная детерминанта э-культуры – это метафора вычислимости когниций в соответствии с компьютерной парадигмой Тьюринга. Как мы показывали на протяжении данной работы, машина Тьюринга слишком груба для представления многообразия социокультурных феноменов. Многие феномены невозможно выразить словами, поэтому они не подпадают под символьную парадигму, т.е. не вычислимы в смысле компьютеринговой дефиниции Тьюринга. Однако они могут быть вычислимы в смысле Корсакова. Для доказательства данного утверждения требуется пересмотр фундаментальных оснований теории алгоритмов, к чему, собственно, и призывает автор данной статьи. Будет модификация теории алгоритмов, возникнут новые элементарные логико-цифровые схемы – электронная культура засверкает новыми, более жизнеутверждающими оттенками. По крайней мере, современная э-культура, из-за чрезмерного отвлечения человека на технологические формы обживания социальной реальности, отдаляет человека от других людей и от самого себя. Возможно, перспективный инструментарий э-культуры, в котором проявится роль машины Корсакова как фундаментальной когнитивной метафоры, заставляющая «пробивать» все новые и новые «окна в реальность», предоставит нам больше возможностей собственно человеческого развития, не связанного жесткими алгоритмами категоризации действительности.

Конечно, здесь мы фиксируем только призывы. Однако обрисованные нами контуры новой парадигмы вычислимости достаточно убедительны и диссертант не видит особых инженерных проблем в разработке конкретного проекта когнитивной машины, в которой алгоритм работы математика (машина Тьюринга) будет поддерживать квазиалгоритм работы художника (машина Корсакова), и, наоборот.

Художник и математик совместно участвуют в развитии электронной культуры.

Обозначим более четко те философские перспективы, которые сулит применение машины Корсакова-Тьюринга. Они связаны с рефлексией над возможностями компьютерного моделирования "смысла".

Роль машины Корсакова-Тьюринга в формировании объемной (3D) интенциональной семантики словаря искусственного интеллекта

Междисциплинарные исследования искусственного интеллекта принципиально невозможны, если не решён вопрос: каким образом термин, обозначающий когнитивный феномен (мышление, внимание, понимание, творчество, интуиция, любовь и пр.) способен совместно и синхронно (одномоментно) выражать, во-первых, личное *психическое переживание* (квалиа, осмысленность, осознанность, ясность и др.), во-вторых, естественнонаучное и социогуманитарное *описание и объяснение* этого феномена, и, в-третьих, *механизм* его возникновения и управляемого воспроизведения. Вопрос имеет не только узко научный когнитологический статус. Большее значение вопрос приобретает в социокультурном контексте. Современная культура – это электронная культура. В контексте становления и развития этой формы культуры традиционная психофизиологическая проблематика транслируется в *психотехнологическую*.

Возникает не только проблема объяснения взаимосвязи психики с нейрофизиологическими коррелятами, характерная для традиционной культуры: Что такое психика? Какую роль играет психика в жизни биологических существ, человека, общества, в мире в целом? Какова природа психики, состав, структура, функции психического? Можно ли психику описать, познать сущность, объяснить? Какие методы применимы для понимания психического? Как решить проблему психики других существ, в частности, подпроблемы intersубъективности, другого сознания?

В э-культуре встаёт интригующий вопрос о *реализации психических феноменов посредством электронных технологий*. Ранее на страницах журнала «Полигнозис» в статье «Проблема сознания в электронной культуре» [Алексеев, 2010-2] доказывалось следующее. Все технологии НБИКС-комплекса редуцируются к электронным технологиям. Электронная технология искусственного интеллекта (ИИ) – это метатехнология НБИКС-технологий, так как она спо-

способствует сложной интеллектуальной наукоёмкой деятельности¹. Поэтому психотехнологическая проблема – это задача преимущественно инженерного дела создания и развития интеллектуальной техники. Цель исследований ИИ – решить проблему, привлекая научно-теоретические и философские исследования, способствующие программной инженерии. Реализатором когнитивного феномена является не естественная биологическая система, как это предлагает Т. Полджер в одном из вариантов интегративного функционализма, призывая изучать психику с натуралистических позиций в рамках парадигмы функционализма [Polger, 2004]. «Реализатор» психического – это (нейро)компьютерная система реализации научных теорий как имитаций, моделей, репродукций и креаций² личностно-значимых психических феноменов. В свою очередь, так как принимается аэродинамическая методология¹ и построение инженерной конструкции преобладает среди других исследований, например, нейрофизиологических и психологических, то встает вопрос не только об искусственных реализаторах, адекватных психическим феноменам, но и *супервенторах* – тех психических феноменов, которые сопровождают конкретную структурно-функциональную организацию (нейро-)компьютерной системы. Поэтому в психотехнологической проблематике, как для компьютерных реализаторов психического, так и для психических супервенторов компьютерного, необходимо изучать «инженерный» интенционал когнитивного термина.

¹ Специально для решения вопроса о метатехнологическом статусе искусственного интеллекта относительно НБИКС-комплекса автор проводил ряд мероприятий. В рамках Московского форума культуры «Культура как стратегический ресурс Российской Федерации в XXI веке» (МГУКИ, 1 июля 2010 г.) состоялась работа секции «Искусственный интеллект и развитие электронной культуры России», на которой изучались вопросы: 1) электронная культура – стратегический ресурс развития технологической и духовной сфер общества; 2) междисциплинарные исследования искусственного интеллекта – ключевой фактор развития электронной культуры». Тематика секции была продолжена на XI Международной научной конференции «Модернизация России: ключевые проблемы и решения», 16-17 декабря 2010 г., ИНИОН РАН. Вопросы круглого стола центрировались на изучении междисциплинарных исследований ИИ как главного фактора интенсивного развития э-культуры.

² Если под *репродукцией* психического феномена понимается воспроизведение кодовых нейральных зависимостей на субстрате, отличном от мозга, то под *креацией* (творением «нового») понимается программно-управляемое возникновение таких когнитивных феноменов, для которых нет натуральных аналогов.

В связи с этим, в обозначенной выше работе, предлагалось исследовать феномены сознания средствами языка, обладающего трёхмерной интенциональной семантикой. Такой язык обеспечивает корреляцию смыслов и значений следующих суждений: 1) *феноменальных суждений*, характеризующих собственно факт сознания, переживаемый человеком (*первичный интенционал*); 2) *научных суждений*, рационально объясняющих когнитивный феномен методами и средствами естественных, гуманитарных, социальных наук (*вторичный интенционал*); 3) *инженерных суждений* о компьютерной реализации психического феномена, представленных в терминах проекта структурно-функционального, логико-математического, программного, информационного, лингвистического, методического устройства (нейро)компьютерной системы (*третичный интенционал*).

Сразу сделаем предупреждение. Может показаться, что наша идея копирует идею т.н. «дименсионализма». Данный термин стал достаточно часто употребляться в современной социально-философской литературе². По всей видимости, его впервые в русскоязычном сообществе употребил известный религиовед И.Н. Яблоков¹. Термин обозначает многомерность социокультурного явления, в нашем случае, когнитивного феномена. Для изучения сложного явления применяются различные инструментарии и, соответственно, различные языки. Например, обозначенное выше трёхмерное отношение можно было бы изучать инструментариями: 1) феноменологической интроспекции, формирующей автобиографический нарратив; 2) научной объяснительной теории; 3) компьютерного моделирования. В самом деле, получается многомерное исследова-

¹ «Аэродинамическая методология» означает экспериментальное создание инженерных конструкций, таких как самолет при условии отсутствия общей теории полёта. Так же поступают исследователи ИИ – они создают программные системы, которые проходят тест Тьюринга, не задумываясь при этом о некоторой общей, эссенциалистской теории «интеллекта». Такая теория, на мой взгляд, невозможна. Сравнение истории ИИ с развитием авиации имеет давнюю традицию. Недавно эту идею аккумулировал Б. Уитби в выражении «аэродинамика интеллекта» в контексте социокультурных исследований ИИ. На русском языке издан перевод работы (см.: [Уитби, 2004]). К сожалению, перевод выполнен неквалифицированно (например, метод продукций – это, по мнению переводчика, «метод постановки»). Несомненно, такая издательская работа вносит терминологическую путаницу и вредит аналогичным отечественным исследованиям.

² См., например, Аринин Е.И. Эзотеризм, дименсионализм, конфессиоцентризм, и религиозная идентичность//Третья международная научная конференция "Мистико-эзотерические движения в теории и практике: проблемы интерпретации эзотеризма и мистицизма", 3-5 декабря 2009 г., г.Владимир. Издательство Русской христианской гуманитарной академии, Санкт-Петербург. – 2010. – С. 15-23; Социальная эпистемология: идеи, методы, программы/Под редакцией И.Т. Касавина. — М.: «Канон+» РООИ «Реабилитация», 2010. — 712 с. — С. 299

ние феномена, характерное для любого междисциплинарного исследования. Однако данный подход никак не соответствует нашей идее. Дименсионализм, трактуемый таким образом, означает механическое объединение инструментальных средств и языков относительно конкретной предметной области. Смотрим с одной позиции – получаем такую картину. Используем другой словарь – картина меняется, хотя отражает ту же самую сущность, но по-иному. Наша задача иная. Мы стремимся создать *целостную концептуальную систему, которая выражает органичное, смысловое, интенциональное триединство*.

Ранее обоснование необходимости такого трёхмерного языка изучения когнитивно-компьютерной проблематики исходило от методологии изучения проблемы сознания [Алексеев, 2010-2, с. 138-140]. В настоящей работе мы подкрепим важность такого подхода рядом эмпирических соображений. Для более убедительной демонстрации выберем самую сложную на сегодняшний день сферу современных исследований искусственного интеллекта – проект *искусственного общества* [Алексеев, 2011-3].

Необходимость трехмерной семантики в проекте искусственного общества

Вначале подчеркнем универсальность компьютерного способа репрезентации социогумантарных и естественнонаучных теорий. В контексте методологии компьютерной науки все подобного рода теории принципиально исчислимы. Это подчёркивал Х. Патнэм, предлагая, во-первых, психологическую теорию представлять на ленте машины Тьюринга [Putnam, 1967], во-вторых, отождествлять собственные психические состояния с логическими состояниями машины Тьюринга [Putnam, 1960]. Исходя из этой двойкой позиции, выделим, соответственно, сильные и слабые компьютерные версии интерпретации когнитивных теорий.

В условиях *слабых интерпретаций* общие и частные теории, объясняющие когнитивные феномены, выступают в роли логико-алгебраических теорий компьютерной модели. Компьютерная (вычислительная) парадигма универсальна по причине синтаксической выразимости в числовом формате любой теории – физической, психологической, биологической, социологической и пр. Например, интересной интерпретацией физикалистской теории когнитивного феномена является пиксельная метафора [Stoljar, 2001]: пиксель на экране мо-

¹ Яблоков И. Н. Некоторые дискуссионные вопросы методологии религиоведения / Философско-методологические проблемы изучения религии. Материалы конференции (Москва, 28–29 октября 2003 г.). М.: РАГС, 2004. С. 65–66

нитора – это единичный физический процесс, а паттерны, образующиеся в результате сочетания пикселей в динамике формирования и трансформирования образов – это схемы витальных, психических, социальных феноменов. От пиксельной метафоры – один шаг к организации «искусственного общества» как однородной структуры и, далее, как многоагентной системы.

Так как современная наука и технология выражается неопозитивистским языком, по крайней мере, по стилю аргументации, логической связности, способу верификации результатов, поэтому все возможные научные высказывания принципиально редуцируемы (в грубом варианте) к целочисленной арифметике посредством, например, гёделевой нумерации (индексации) теоретических высказываний. Грубый вариант исключает автореферентные суждения (высказывания теории о себе самой) по известным причинам проблемы Гёделя. Так же нет возможности континуально охарактеризовать вход, выход, состояния функционалистской теории психического [Block, 1980]. Данная проблема послужила Н.Блоку основанием критики классического (патнэмовского) машинного функционализма. Он предложил психофункционализм¹. Термины психологической теории задают не конкретные элементы, а классы возможных входов, выходов и состояний системы. Имеется ряд других способов утончения целочисленной индексации терминов когнитологического словаря. Например, если поменять базовое множество обозначенной выше алгебраической системы, на котором интерпретируются термины социальной, психологической или другой теории, и индексировать суждения не целыми, а, скажем, комплексными числами, то получается возможность выражения достаточно сложных (в описательном плане) когнитивных феноменов. Т.е. компьютерно представимо то, что невыразимо словами. Далее, если поменять формально-логический язык описания этой алгебраической системы с языка предикатов второго порядка на квантово-логический язык², то получим интересные квантовые модели искусственного общества.

В рамках слабых интерпретаций когнитивных теорий целесообразно утверждать о том, что «искусственные общества» - это компьютерная имитационная модель некоторых социологических теорий. Такая теория обязательно включает представления об индивидах, способах и формах социализации и

¹ Психофункционализм стал сегодня классической функционалистской концепцией: см. [Levin, 2013].

² В классическом варианте основатель теории алгебраических систем АИ. Мальцев предлагал язык второго порядка – см. в [Мальцев, 1970]. Современное состояние «квантовой логики» представлено в [Васюков, 2005].

ориентируется на инструментарий имитации и/или моделирования социокультурных явлений. Например, В.А. Истратов, изучая параметры потребительской «надобности», использует инструментарий агентно-ориентированного моделирования и наделяет агентов (индивидов) широким спектром психологических и социологических параметров – «настроением», «эйфорией», «депрессией», «удовлетворённостью», «авторитетом», «счастьем» используя при этом разнообразные положения ряда научных теорий [Истратов, 2009]. К сожалению, на эти теории он смотрел с технической позиции, путем случайного подбора. Вопрос подбора социо-гуманитарных концепций должен решаться специалистом. (Позиция В.А.Истратова подробно рассмотрена в пп..1.2.2).

В школе В.Л. Макарова и А.Р. Бахтизина используется слабый подход к компьютерной интерпретации социологических моделей (см. [Макаров, 2006; 2010; Бахтизин, 2008]. Анализ работ, представленных в международном журнале «искусственных обществ» JASSS (Journal of Artificial Societies and Social Simulation), показывает, что такой подход доминирует в большинстве проектов. Биологические, психологические и социологические термины приписываются (атрибутируются) агентам и классам агентов. Речь идет лишь о компьютерной имитации, но не о компьютерной модели, тем более, о репродукции общества в искусственном формате, в котором Я выступаю в роли виртуального актора.

Сильные интерпретации когнитивных терминов не атрибутируют компьютерным моделям когнитивные феномены. Они непосредственно отождествляют «когнитивное» («психическое», «социальное» и пр.) со способом функционирования (нейро)компьютерной системы. Такой взгляд на проблему послужил Х. Патнэму основой построения концепции функционализма машины Тьюринга (1960 г.) [Putnam, 1960]. В последующих модификациях (психофункционализм, аналитический функционализм, контентный функционализм, интенциональный функционализм и пр.) сильный подход так же прослеживается. Компьютинг выступает в роли *собственного параметра* когнитивной теории, т.е. неотделим от неё¹. Согласно машинному функционализму, психические феномены вычислимы по своей природе и организованы по тому же принципу, что и компьютерная система, обеспечивающая обработку информации. Если, скажем, «настроение», «счастье» и пр. в упомянутой выше работе

¹ Доказательство редукции разновидностей современных «функционализмов» - количество их достаточно широко распространённых разновидностей превышает два десятка – к машинному функционализму выходит за рамки данной работы. Очевидно, что наиболее последовательным продолжателем идеи машинного функционализма является компьютеризм как наиболее развитая когнитивно-компьютерная психотехнологическая парадигма.

В.А. Истратова – это (нейро)компьютерно оформленные параметры модели соответствующей теории, то способы употребления когнитивных терминов применительно к компьютеру и к человеку не различаются. Таким образом, психика функционально инвариантна субстрату реализации – человеческому мозгу, компьютеру, тине марсианина. В работе «Философия и наша ментальная жизнь» Х. Патнэм наиболее чётко сформулировал идею функционализма [Putnam, 1975]: 1) человек — это машина Тьюринга; 2) психологические состояния человека — это состояния машины Тьюринга. При этом он ссылается на личные когнитивные состояния. Поэтому следует более чётко позиционировать сильный подход такими суждениями: «Я – машина Тьюринга», «мое психическое переживание боли – это логические состояния машины Тьюринга в соответствии с её специфицированными входами, выходами и состояниями». С учетом современной терминологии, следует выдвинуть тезис «Я – компьютер». И, соответственно, онтологический тезис будет таковым «*Computa ergo sum*» (компьютируем, следовательно, существую).

Обобщая эти идеи к проблематике искусственных обществ, сильный компьютеризм утверждает: 1) «социальное» инвариантно относительно субстрата реализации; 2) общество – это (нейро)компьютер; 3) способы, формы и содержание социального феномена – суть логические состояния (нейро)компьютера.

Однако! В такой трактовке компьютеризма отсутствуют теоретические положения психологии, нейрофизиологии, лингвистики, социологии и всех других наук. Компьютерная дефиниция терминов двумерная. Это – личные переживания, отождествляемые с функциональными состояниями компьютера.

В слабом подходе семантика терминов так же ограничивалась двумя измерениями, однако там отсутствовали суждения первого лица. Здесь отсутствуют суждения от третьего лица и контент компьютинга социальных явлений непосредственно (на функциональном уровне) отождествляется с психическим, духовным миром человека. Несомненно, это слишком сильный тезис. Сегодня такой подход, к сожалению, широко эксплуатируется. Его используют, например, исследователи «постчеловечества» (о чем будет кратко отмечено в заключении) и, скажем, отечественные идеологи оценки значимости учёного по арифметическим подсчетам числа публикаций и ссылок на эти публикации¹.

¹ Сегодня сырая, не прошедшая стандартной приёмки, компьютерная система РИНЦ, в приказном порядке, идущим от Минобрнауки, объявлена главным судьей значимости учёного. Грант выигрывает не авторитетный учёный, а выскочка, оперативно загрузивший малозначимые работы в базу данных www.elibrary.ru

Однако напомним этим исследователям то, что Х. Патнэм, пытаясь совместить социокультурную проблематику с машинным функционализмом, посчитал свою оригинальную концепцию глубокой ошибкой и скатился к солипсизму в решении вопроса множественной реализации тождественных психических состояний на различных физических субстратах: ментальные состояния – это мои личные состояния, укоренённые в социальных и бытовых отношениях между людьми, которые никакой (нейро)компьютер не способен воспроизвести. Более-менее правдоподобно объективные феномены компьютер способен симитировать или смоделировать с некоторой степенью адекватности. Но репродуцировать – никак не может.

Очевидны недостатки как слабого, так и сильного подхода к интерпретации когнитивных терминов [Horst, 2011]. Первый не отвечает требованиями экультуры – он имитирует социокультурные феномены и не вносит вклада в её развитие. Второй подход слишком груб и агрессивен.

Автор предлагает *умеренную версию* [Алексеев, 2013-2]: 1) двумерная семантика компьютерной дефиниции когнитивных терминов дополняется до трёх измерений, в результате чего образуются тернарные отношения: приватный когнитивный феномен – научная теория – компьютерный проект; 2) жесткий функционализм машины Тьюринга смягчается *функционализмом теста Тьюринга*. Отождествление на уровне функционального описания когнитивного феномена с компьютерной системой сменяется атрибутированием феноменов компьютеру. Но здесь нет возврата к слабой интерпретации. Атрибутирование осуществляется посредством комплексного теста Тьюринга, компьютерная система реализации которого *убеждает* наблюдателя, Меня, в том, что х-системе присущ феномен жизни, интеллекта, любви, понимания и пр. Таким образом, когнитивность х-системы, обладание жизнью, психикой, сознанием и пр., зависит от произвола и компетентности наблюдателя. Именно Я активно конституирую искусственное общество в соответствии с социокультурными смыслами, ценностями, нормами, регулятивами и пр. Такой подход изучения и конструирования социокультурной реальности сегодня принято обозначать термином «постнеклассическая рациональность». Но для реализации такого подхода инструментарий искусственного общества должен быть очень мощным.

На наш взгляд, мы достаточно убедительно продемонстрировали необходимость трехмерного языка построения искусственных обществ. В искусственном обществе Я и другие обладаем когнитивными феноменами посредством компьютеринга естественнонаучных и социогуманитарных теорий. Однако Я и

другие свободны в приписывании себе и другим свободы, жизни, творчества, интеллекта, понимания и пр. Мы не подчиняемся (нейро)компьютингу.

Вернёмся к философско-лингвистическим вопросам построения трёхмерного языка.

Философско-лингвистические аспекты трёхмерной семантики словаря искусственного интеллекта

В новейшей аналитической философии каноническим стилем употребления когнитивных терминов становится соотнесение их с двумерным (2D) лингво-логическим каркасом [Schroeter, 2012]. Двухмерность («two-dimensionality» по Д. Чалмерсу [Chalmers, 2006]) в общем случае характеризует многообразие отношений между варьируемыми смыслами и значениями терминов. 2D-семантика базируется на общих идеях Г. Фреге о различении смысла/значения слова [Фреге, 2000], интенционал-экстенциональной методологии Р. Карнапа [Карнап, 2007], крипкеанской многоаспектной соотносительной параметризации возможных миров [Kripke, 1972], патнэмовском интерпретативизме¹ и др. Применяется достаточно сложный инструментарий модальных и неклассических логик, аппарат интенционалов, экстенционалов, индексикалов, мягких и жестких десигнаторов, возможных миров и пр.

Выделим два способа использования двумерного каркаса в исследованиях ИИ: *методологический* и *эпистемологический*. Первый способ задает парадигму когнитивных исследований в плане совместного изучения лично переживаемых психических феноменов либо с intersubъективными способами их объяснения либо с механизмами их функционирования. В качестве примера мы приводили варианты использования двумерной семантики: 1) научная теория ↔ компьютерная реализация; 2) психический феномен ↔ компьютерная реализация.

Эпистемологическая двухмерность обслуживает цели корректного представления и кодифицирования способов слияния дифференцированных смыслов одного термина. Например, смысл слова «восприятие» связывается как с феноменом психики, так и с функциональной организацией мозговой активности.

В отечественной философской литературе недавно появились работы, посвященные двумерной семантике. Иногда употребляется трудно выговариваемый, как скороговорка, термин – «дидименсионализм». Одна из пионерских ра-

¹ См. гл.8 «Компьютерная психология и теория интерпретации» в [Putnam, 1983, pp.139-155].

бот принадлежит В.В. Горбатову [Горбатов, 2012]. По причине новизны тематики, в данной работе допущен ряд неточностей, которые мы отметим немного ниже.

Чтобы выделить особенности объёмной интенциональной семантики когнитивных терминов, последовательно рассмотрим возрастание семантической мерности словаря искусственного интеллекта.

0D-семантика. Историческим фактом является то, что ИИ вырос из метафоры отождествления психической деятельности с обработкой информации компьютером. Автор не является сторонником панриторической лакоффовской когнитологии, согласно которой мы живем метафорами, и полагает, что метафоры не обладают логико-лингвистическим статусом. Поэтому, например, стандартный среди ряда специалистов искусственного интеллекта экивок на то, что ИИ – это компьютерная метафора естественного интеллекта, для нас представляется запрещённым приёмом. Такой подход вносит невообразимую концептуальную путаницу среди исследователей когнитивных систем, их разработчиков, потребителей, сторонних обывателей. Риторико-ориентированные когнитивные термины обладают нулевой семантикой, они ничего не значат. Однако полезны, если служат целям убеждения (Д. Дэвидсон). Но если когнитивная наука претендует на статус строгой науки, то её словарь должен получить чёткое семантическое измерение.

1D-семантика. Когнитивистская терминология редуцируется к словарю конкретной дисциплины. Особенно показательны одномерные семантики (по инженерному «измерению») для проектов искусственного интеллекта, в которых «интеллект» - это программистские вариации на тему «мышления». Скупые методы моделирования интеллекта крайне затруднительно связать с научным «измерением», например продукционную модель знаний с мощными бихевиоральными теориями, фреймы – с гештальтпсихологией, а семантическую сеть – даже с простейшей бадеиной попперовской эпистемологией. Под флагом когнитивистики в одномерном формате работают не только программисты. За несколько последних лет как грибы после дождя выросли многочисленные центры когнитивных исследований психологической, социологической, нейрофизиологической, медицинской, экономической ориентаций. Например, в Российском экономическом университете им.Г.В. Плеханова до 2011 г. функционировала кафедра когнитивной экономики. Что означает «когнитивная экономика», автор затрудняется сказать. Однако когда эту кафедру присоединили к кафедре информатики, стало более понятно, что под «когнитивным» понимает-

ся процесс переработки формализованных «данных» и «знаний». В принципе, часто модный ярлык «когнитивных» присваивается традиционным информационным технологиям из-за наличия в их инструментарии баз «знаний» (Центр когнитивных исследований в МИСиС). Имеется и другая крайность. Она присуща нейрофизиологическим исследованиям – если такое-то вещество оказывает воздействие на такой-то участок головного мозга, то это частное исследование так же называют «когнитивным». Эмпирическое мелкотемье когнитивистики показали как крупные международные конференции по когнитивной науке (Казань-2004, Санкт-Петербург-2006, Москва-2008, Томск-2010, Калининград-2012), так и более мелкие конференции (например, «Когнитивная наука в Москве»-2013). На этих мероприятиях междисциплинарный статут исследований, как правило, декларируется, каждый исследователь или школа работает в обособленной парадигме, теоретическая составляющая нивелируется в хаосе бесчисленных прикладных исследований. Из-за объёмности и количества работ создается впечатление большой актуальности когнитивистики. Но это, скорее, форма проявления поп-науки: термин «когнитивный» - очень модный на сегодняшний день термин. Причина – неточно определённый словарь когнитивных терминов на фоне универсального статуса познавательной проблематики для любой дисциплины. С одномерной семантикой и модной «попсой» следует заканчивать.

2D-семантика. Двумерная семантика связывает интенционалы терминов возможных миров, каждый из которых формируется вдоль двух осей 2D-каркаса. Достаточно известный русскоязычному читателю пример использования в когнитологических исследованиях плоскостной семантико-онтологической схемы предложен Э. Вильянуэвой [Вильянуэва, 2006, с.153]. Двумерность понимается как методологическом, так и в эпистемологическом планах: 2D-каркас позиционирует по одной оси интенционалы высказываний о феноменальных свойствах сознания, референтами которых являются реальные личные переживания от первого лица (квалиа, феноменальный опыт); по второй оси – интенционалы научных объяснений феноменов сознания, референтами которых являются физические, психологические, нейрофизиологические, социологические реальные процессы и их модели. Продолжая данную схему, отметим, что интенционалы различных миров в общем случае вступают в реляционные отношения по схеме «многие ко многим». Например, для классического примера с термином «вода» (С. Крипке) в возможном мире субъективной реальности (первое измерение 2D-каркаса) интенционалом является «жидкое, бесцветное, утоляющее жажду, холодное, горячее и пр.». В возможном мире

химии (вторая ось 2D-каркаса) «вода=H₂O». Имеется возможный мир неземной химии, в котором «вода=XYZ» (как воображает этот случай Х. Патнэм¹). При этом этот объект обладает теми же феноменальными качествами, как и H₂O. Многообразие интенционалов слова «вода» значительно увеличивается, если 2D-каркас расширить возможным миром нейрофизиологической стимуляции психического феномена «водянистости». Но в этом случае, исходя из наших представлений о том, что смысл когнитивного термина включает характеристику технического модифицирования и/или репродуцирования психического феномена, следует ввести новую мерность – возможный мир нейрофизиологических компьютерных технологий. 2D-каркас становится объёмным.

Для более чёткого представления о двумерной ограниченности когнитивных исследований отдельно рассмотрим распределение интенционалов по осям трёхмерного семантического каркаса.

Двумерная семантика 1 и 3 измерений. Чётко проявилась в первой версии функционализма в отмеченном выше тезисе «Я – машина Тьюринга». К «Я-компьютеру» Х. Патнэм пытался присоединить научные теории – нейрофизиологическую и социологическую, т.е., по сути, создать 3D-семантику когнитивистской теории. Однако из-за концептуальной сложности задач он опустил руки и вообще отказался от функционализма машины Тьюринга. Тем не менее, его идею надо поддержать и попытаться преодолеть концептуальные сложности: когнитивная теория полна только тогда, когда *каждый когнитивный термин выражает смысловое триединство личного психического переживания, его научное объяснение и техническую реализацию*. Вряд ли кто-то способен опровергнуть данный тезис.

Двумерная семантика 1 и 2 измерений. Характерна для вариантов теоретического описания психических феноменов, в частности, для психофункционализма Н. Блока. Широко известна методика Рамсея-Льюиса. Она обеспечивает редукцию психических терминов к вариантам языка исчисления предикатов, кстати, очень грубым вариантам. В результате смыслового «слияния» с достаточно общим логическим языком (в стиле карнаповской программы построения универсального логического каркаса науки), психический феномен получает чёткий логический формат. Теперь с психологическими терминами можно производить строгие логические рассуждения и манипуляции. Однако,

¹ См. In Philosophical Papers, Vol. 2: Mind, Language and Reality. Cambridge University Press. – P.227

очевидна грубая неадекватность такого языка фактам субъективной и объективной реальности.

Двумерная семантика 2 и 3 измерений. Вернёмся к проекту «искусственного общества». Напомним, что это – компьютерная модель некоторой социологической теории, которая объясняет и предсказывает поведение индивидов в процессе их социализации. В двумерной интенциональной семантике термины естественнонаучных, социальных и гуманитарных теорий *не* интерпретируются на компьютерной модели, как это принято в одномерном, традиционном варианте. «Социальное» искусственного общества выражает «диагональный интенционал» (Д. Чалмерс) возможного мира социокультурной теории и возможного мира компьютерной системы, интенционалы «социальное» и «компьютерное» раздельно едины. Но в таком двумерном варианте искусственного общества нет Меня с присущими убеждениями, переживаниями, знаниями, целями, смыслами, ценностями и пр.

Подчеркнем, термины двумерного каркаса располагаются не в некотором метафорическом пространстве, но в чётко формализуемом и кодифицируемом пространстве двумерной логики. Сегодня насчитывается порядка десятка подходов к формированию единого интенционала термина средствами двумерной логики (ни в коем случае не путать с двузначной логикой!). Кратко перечислим несколько.

Эпистемологические способы формирования 2D-семантики.

Тематика семантической двухмерности возникла для обслуживания философско-лингвистических проблем.

1. *Двумерная семантика индексикалов.* Была предложена Д. Капланом (1979-1989 гг.). Индексикалы («Я», «здесь», «сейчас») и указательные слова - демонстративы – («это», «то») – это высказывания, которые дистрибутируют возможные миры и кодифицируют их индексами - конкретными субъектами, событиями, предметами, моментами и интервалами времени, выражаемыми в единичных суждениях. [1, 2]. Так «Я голоден» произносится, скажем, неким конкретным Андреем в момент времени t_1 и Катей в момент времени t_2 . Это высказывание выражает разные содержания и истинно, если и только если, соответственно, Андрей является голодным в своем возможном мире в момент времени t_1 , а Катя – в своем мире в момент t_2 . В значении высказываний различаются две составляющие: 1) «содержание» (*content*). Контент представлен модальным профайлом, репрезентирующим объекты, классы, отношения, операции. Он моделируется стандартной семантикой возможных миров; 2) «харак-

тер» высказывания (character) отражает семантические правила варьирования содержания в зависимости от контекста использования высказывания. Так, термин «такой-то Андрей, родившийся там-то, паспорт № такой-то и пр.» имеет постоянный профайл, который выделяет тот же самый объект во всех возможных контекстах. В то же время индексикал Я обладает различными профайлами, которые выделяют различные объекты в различных контекстах использования. Конвенциональные семантические правила, задающие индексикал Я, репрезентируются таблицей, в ячейках которой представлены обозначения интенционалов в различных контекстах употребления слова «Я». Наименования столбцов таблицы составляют имена (индексы) контекстов, а наименования строк – ссылки на конкретные профайлы. Табличное представление двумерной семантики сегодня стало каноническим [Schroeter, 2012].

2. *Двумерная семантика диагональных пропозиций* (Р. Сталнейкер, 2003-2004 гг.) ассоциируется с высказываниями как функция, которая отображает возможные миры в значения истинности этих высказываний. «Голодность» некоторого субъекта изучается инструментарием модальной логики. Т.е. если таблица Каплана характеризовала двухмерность путем совместного изучения ячеек таблицы, в которых фигурировали ссылки на профайлы конкретных «Я», то здесь в клетках имеется оператор модальности, определённый на двух элементах – «истина»/«ложь» (интенционал пропозиции – это вариации значений истинности). Для принятия интенционала требуется совместное рассмотрение как минимум двух клеток, образующих «диагональный интенционал».

3. *Двумерная семантика дескрипций*: Предлагается Г. Эвансом в 1977 г. для имен, референты которых фиксируются описательным путем. Двухмерность образуется на разных уровнях описания объектов – поверхностном и глубинном. Уровни связаны модальностью необходимости/возможности. Два имени, скажем «Андрей Алексеев» и «автор данной статьи» (аналог классического расселовского примера «'В. Скотт'='автор 'Веверлея''») имеют один (двумерный) интенционал, который соответствует различным уровням погружения в возможные миры, дескрипции которых с большей или меньшей степенью детализации характеризуют конкретного человека. Предикаторы типа «голодность» как имена феноменальных состояний и как имена гастрономических, нейрофизиологических, медицинских и др. отношений, так же предполагают дескриптивное описание с требуемой степенью детализации в соответствующих мирах. Сопоставлять такие дескрипции в табличном формате не удобно, поэтому предпочтителен аппарат семантических деревьев. Однако в этих

формах представления принципиальной разницы нет – любая сеть как разновидность графа выражима реляционными средствами.

4. *Двумерная семантика плавающих миров*. Предложена М. Дейвисом и И.Л. Хамберстоном (1981 г.). Двухмерность высказывания формируется путем применения оператора «актуальности» к возможному миру «Актуальность» – это способ модального представления реальности возможного мира. Каждое высказывание связывается с возможными мирами и степенью их актуальности. На возможные миры из «Андреев», «Екатерин», «голодностей» и пр. накладывается плавающая рамка, которая позволяет позиционировать актуальный мир «голода данного субъекта» и при этом учесть семантическое облако возможных миров, образовавшихся в процессе его актуализации. В ячейках двумерной таблицы теперь представляются параметры актуальных миров, а не профайлы или значения истинности, как это было характерно для предыдущих случаев.

5. *Обобщенная семантическая двухмерность*. Д. Чалмерс обобщает варианты 2D семантики, полагая наличие для каждого когнитивного термина или высказывания двух уровней анализа [Chalmers, 2006]. Одно семантическое значение ассоциируется с референтами и обычными условиями истинности суждения или высказывания. Второе значение – со способом зависимости референций и условий истинности от внешнего мира. Так, в общем случае, первичный интенционал конкретного высказывания – это типизируемая функция от сценария формирования экстенционала. В классической фрегевской терминологии, первичный интенционал – это «смысл» как функция от способа означивания «знаком» «значения» (экстенционала). Вторичный интенционал – это функция: «возможные миры» → «экстенционалы». Двумерный интенционал – это функция от упорядоченных пар: <сценарий, возможный мир> → экстенционал. Диагональные интенционалы чалмеровской двумерной семантики так же, как и выше, задаются путем совместного анализа как минимум двух ячеек таблиц. Для высказываний «Я голоден», «Андрей голоден», «Катя голодная» строятся возможные миры, в которых, во-первых, экстенционалы составных и атомарных терминов погружаются в некоторый возможный мир, и во-вторых, способы характеристики (параметризации, сингуляризации, фундирования, позиционирования) экстенционалов соотносятся с мирами в целом.

6. *Неинтенциональная 2D семантика когнитивных терминов*. Данный подход основательно раскрыт В.В. Горбатовым при анализе т.н. «метасемантического 2-дименсионализма» Р. Сталнейкера [Stalnaker, 2004]. Суть в следующем – диагональные интенционалы не играют конститутивной роли в формиро-

вании двумерного смысла. Их формирование предполагает интерналистский, зависимый от субъекта, опыт слияния «смыслов». Однако всё должно быть объективно-позитивно. Р. Сталнейкер лишает возможности субъекта «сливать» смыслы воедино. Т.е. позитивистская программа, приверженцем которой является Р. Сталнейкер, стремится исключить субъекта как носителя двумерного интенционала, например, индексикального термина. Диагональные интенционалы нужны лишь тогда, когда возникает случай недостаточной информативности горизонтальных интенционалов, как полагает Р. Сталнейкер. Горизонтальные, расположенные в строках таблицы – это базовые интенционалы. Диагональные интенционалы возникают лишь как «переинтерпретации» горизонтальных. Они не самостоятельны. Таким образом, для простейшего случая двумерной семантики индексикалов (см. выше) смысл высказывания «Я голоден» невозможен. Такой интенционал – это лишь косвенный способ переинтерпретации профайлов «Андрея» и «Кати», которые утверждают «Я голоден» в своих возможных мирах. Но для позитивиста – Р. Сталнейкера «Я голоден» не фиксируется даже в этих профайлах субъектов «голода». Он способен лишь объективно фиксировать профайлы «Андрей такой-то», «Катя такая-то» голодная. Таким образом, как двумерный, так и одномерный интенционал индексикала «Я» Р. Сталнейкер отбрасывает. Поэтому мы назвали такой подход «неинтенциональной» 2D-семантикой.

Удивительно то, что подобный случай борьбы со «смыслом» имел место век назад, когда позитивисты разрушили «смысловой мир» Г. Фреге. Вспомним, что в логико-смысловой конфигурации Г.Фреге «смысл» – это *способ задания значения, выражаемый знаком, обозначающим значение* (в реконструкции Б.В.Бирюкова) [Фреге, 2000]. Конфигурация обслуживала культуро-антропологические запросы Г.Фреге, связанные с идеей «Третьего Рейха ментального мира»¹. «Смыслы» собственных имён и предложений, знаков функций и имен понятий составляют объективное содержание мышления человека. «Третий ментальный Рейх» был разрушен позитивистами. Р. Карнап открыл

¹ Е.Д.Смирнова и П.В.Таванец считают, что Г.Фреге не создал законченной теории смысла (см. Смирнова Е.Д., Таванец П.В. Семантика в логике // Логическая семантика и модельная логика / Отв. ред. П.В.Таванец. М.: Наука - 1967. С.3-53). Мнение автора кардинально расходится с мнениями этих ученых и приближается к точке зрения Б.В.Бирюкова, который убедительно показал наличие теории смысла в трудах Г.Фреге в работе Б.В.Бирюков. В логическом мире // В кн. Фреге Готтлоб. Логика и логическая семантика: Сборник трудов/Пер.с нем. Б.В.Бирюкова под ред. З.А.Кузичевой: Учебное пособие для студентов вузов. – М.:Аспект Пресс, 2000 – 512 с. Более того, теория смысла имела для Г.Фреге мировоззренческий статус – благодаря ей он обосновал антропологию «Третьего Рейха» ментального мира. Анализ этой концепции выходит за рамки данной работы.

дурную бесконечность иерархии «смыслов». Однако придумал удачную замену – методологию интенционалов и экстенционалов. Б. Рассел был более жесток. Он обнаружил софистическую парадоксальность логико-математических построений Г. Фреге. В угоду логической атомистической механике, позволяющей «прямо и открыто глядеть в лицо миру» Б. Рассел устраняет «смысл» из фрегевской модели посредством мощной *программы элиминации смысла*: теория типов элиминировала «смысл» из косвенного предложения, теория пропозиций – из простого утвердительного предложения, а теория дескрипций – из собственного имени¹. Человек стал жить в мире фактов и значений, в котором «смысл» – это *неопределенность, неясность значения*. На наш взгляд, двумерный (диагональный) «смысл», который, согласно, Р. Сталнейкеру, всего лишь переинтерпретация «горизонтальных интенционалов» ничем не лучше расселовского «смысла» как «неясности значения».

Отметим, что элиминация смысла в наши дни имеет важное культурологическое значение – современная компьютерная технология, построенная, в основном, на базе позитивистской методологии, «бессмысленна» в силу собственно формальных соображений. В 80-е годы, как мы видим, смысловое многообразие пытаются ввести в логико-лингвистический каркас компьютерной технологии. Но вновь возникают новые позитивисты, которые не дают субъекту стать полноправным законодателем «смысла» в технологии. Диссертант – за постнеклассическую рациональность, где субъект – главное действующее лицо.

В.В. Горбатов идет далее 2-димиенсионализма Р. Сталнейкера. Он отбрасывает не только диагональные, но и горизонтальные интенционалы. Тем самым, полностью игнорируется интенционал-экстенциональная карнаповская методология. Лингво-логическая двумерность стала чисто экстенциональной, с двумя значениями «истина»/«ложь». Дело в том, что в архетипичном источнике 2-димиенсионализма – в концепции «смысла» Г. Фреге В.В. Горбатов заметил «реальность» лишь двух абстрактных объектов – «Истина» и «Ложь». Однако почему-то самое главное – фрегевского семантического треугольника он не заметил. Поэтому для меня было странно видеть значения истина/ложь (у В.В. Горбатова – «1» и «0») в качестве представителей горизонтальных интен-

¹ *Рассел Бертран*. Исследование значения и истины. Пер.с англ.Ледникова Е.Е., Никифорова А.Л. М. : «Идея-Пресс», 400 с.; *Б.Рассел*. Моё философское развитие// В кн. Аналитическая философия: Избранные тексты / Сост., вступ. ст. и коммент. А.Ф.Грязнова. – М.: Изд-во МГУ - 1993 – 181 с.; Рассел Бертран. Человеческое познание: Его сфера и границы: Пер. с англ. – К.: Ника-Центр, 1997 –560 с. ; *Lewis J. Bertrand Russel. Philosopher and Humanist*. London: Lawrgence and Wishart, 1968, 98 p., P.95; *Рассел Б.* Философия логического атомизма. – Томск: «Водолей», 1999. – 192 с.

сионалов¹. Там должны стоять обозначения интенционалов! Это по сути каноническое представление для двумерной семантики!²

Однако пора завершать с 2-димиенсионализмом. Применительно к чалмеровской концепции семантической двухмерности и исходя из обозначенной выше двумеристской идеи Э. Вильянуэвы, когнитивная тематика выделяет, как минимум, две вселенные. В одной вселенной – миры феноменальных событий и состояний конкретных Я с их априори истинными суждениями по поводу своей субъективной реальности. Во второй вселенной – возможные миры, состоящие из объектов, классов, операций, отношений, которые задают рациональные способы научных описаний, объяснений, предсказаний динамики миров первой вселенной. В этой схеме отсутствует инженерная идея образования миров – какими инструментами и методами конструируются миры и вселенные. Трёхмерная семантика разрешает эту проблему.

3D-семантика

Как утверждалось выше, для полновесных когнитологических исследований двухмерности Чалмерса-Вильянуэвы недостаточно. В рассмотренных вариантах формирования 2D-семантики отчётливо отсутствует референт технической реализации в форме имитации, репрезентации, репродукции, креации (3-е измерение) личностно переживаемого феномена (1-е измерение) и описываемого, объясняемого, предсказываемого естественными, гуманитарными, социальными науками (2-е измерение). Например, для классического крипкеанского примера с «водой» третье измерение, выделяет из возможного мира «чистых» компьютерных систем некоторое устройство: а) выработки наносубстанции, феноменально не отличимой от воды («вода=XRW», где X, R, W – некоторые искусственные наноэлементы) или б) реализации нейрофизиологической программы, вызывающей виртуальное ощущение воды при её физическом отсутствии («вода = ProgWater», где ProgWater – имя программы искусственной репродукции феномена «водянистости»). Теперь сценарий формирования экстенционала должен проходить через систему формирования трех интенционалов и через соответствующие фильтры 3D-каркаса, предназначенного для совместного изучения трёх интенционалов, выражаемых 1) *феноменологическим языком*, характеризующим приватный факт сознания, переживаемый человеком (*первичный интенционал*); 2) *научным языком*, рационально объясняющим данный факт сознания на теоретико-эмпирическом фундаменте современных естест-

¹ [Горбатов, 2012, с. 42].

² См. *Schroeter, L. Ibid.*

венных, гуманитарных, социальных наук (*вторичный интенционал*); 3) *инженерным языком* по поводу компьютерной или, в общем случае, машинной формы реализации феномена сознания (*третичный интенционал*). Инженерный *экстенционал* – это структурно-функциональное, логико-математическое, программное, информационное, лингвистическое, методическое и пр. устройство когнитивной системы. Инженерный *интенционал* – это концептуальный проект компьютерного устройства или механизма.

Предлагается очень простая идея обогащения двумерного каркаса: когнитивный термин характеризует семантическое триединство: 1) субъективный приватный психический феномен; 2) его рациональное интерсубъективное (объективное) описание и объяснение; 3) техническая реализацию посредством машин и механизмов (в частности, компьютеров). Графически семантическая трёхмерность изображается тремя осями «координат», по которым «образуются» возможные концептуальные миры. «Обитатели» этих миров – интенционалы соответствующих аспектов выражения когнитивных терминов.

Способы формирования 3D-семантики

Добавление новой, инженерной «вселенной» в виде третьей семантической оси приводит всего-навсего к механическому обогащению двумерной схемы. Такой подход обозначим термином: **«поверхностная трехмерная семантика»**. Подход значим, так как структурирует методологию полноценных когнитивных исследований. Однако в эпистемологическом отношении более важным представляется использование **глубинной трёхмерной семантики**. Суть её раскроем на примерах объемного расширения двумерных подходов.

1. Для индексикальной семантической двухмерности третье измерение характеризует способ формирования контекста. В 3D когнитологических исследованиях возможные миры не принимаются как нечто данное. Они творятся, поэтому третичный интенционал «Я голоден» включает в себя механизм формирования феномена «голодности», присущего некоторому конкретному субъекту, который, кстати, тоже конструируется.

2. Трёхмерная семантика диагональных пропозиций задаётся не плоской реляционной схемой (таблицей), но трёхмерной таблицей. В таблице третьего измерения указывается способ построения двумерной таблицы.

3. Дескриптивная двухмерность обогащается схемой формирования составных частей дескрипций и дескрипций в целом. Трёхмерность – это «описание описания» или, скажем «теория теории», «модель модели» некоторого объекта, класса объектов, отношений, сценариев. Данная конструкция имеет дина-

мический (не синергетический) статус и характеризует, на наш взгляд своеобразную энергийную характеристику как меру изменения частей некоторого выделенного возможного мира под воздействием операциональных изменений других частей этого мира как факторов динамики. Например, интенционал слова «голод голода» - это показатель степени интенсивности «моей голодности».

4. 3D-обогащение двумерной семантики плавающих миров усложняет схему формирования трехмерной пропозициональной семантики. Теперь элементами таблицы выступают не только пропозиции, но и характеристики миров в целом, которые актуализируются по мере акцентуации когнитологических целей и разработки когнитивистского инструментария.

5. Обобщение способов задания трёхмерной семантики мы связываем с обогащением чалмеровской схемы обобщённой двухмерности¹. *Трёхмерный интенционал* включает в себя, во-первых, все три семантические измерения, характерные для поверхностной 3D-семантики. Во-вторых, и это главное – 3D-каркас концептуализирует процессы «творения» возможных миров, эти миры не представляются априори данными. Их надо сконструировать.

Компьютерная «ось» объемного семантического каркаса

Наша идея конкретного компьютерного способа формирования 3D-каркаса состоит в следующем. Она связана с развитием идеи использования машины Корсакова как прототипа коннекционистско-репрезентативной системы². Для нашей работы интересно то, что машина функционирует на субуровне образования одинарных и композициональных терминов – с частными признаками предметов, классов и отношений, которые эти термины способны обозначать с определённой степенью модальности возможности/необходимости.³ Квазилингвистические схемы и сценарии, которые получаются на выходе данной машины – это термины индексикалов, дескрипторов, пропозиций, возможных миров и пр.

Например, программист на одной перфокарте машины Корсакова перфоратором прокалывает отверстия и формирует дескрипцию «голода» по частным признакам «пустота в желудке», «неприятное ощущение в животе», «подташнивание», «зацикленность мыслей на еде», «неспособность сосредоточиться», «плохое настроение» и пр. Далее подкладывает под неё другую перфокарту, на

¹ Chalmers, D. Ibid.

² См. ряд работ, посвящённых новым принципам компьютерации, основанной на исторической находке – машине Корсакова (1832 г.). См. в [Алексеев, 2011-6; 2012-2].

³ Модальности коннекций чётко прописаны в [Алексеев, 2013-4].

которой по частным признакам зафиксирован профайл «Андрей» и прокалывает в ней отверстия, соответствующие первой перфокарте, если эти признаки имеют место быть. Так же программист поступает и с перфокартой профайла «Катя». Далее табулятор проводится по стопке перфокарт. Если иголки попали во отверстия перфокарт-профайлов «Андрей» и «Катя», то это интерпретируется как совместное высказывание «Я голоден» с той или иной степенью модальности. Например, если Андрей менее голоден, то в профайл «Андрея» провалится меньшее количество игл табулятора, нежели чем в профайл «Катя». Теперь результат совместного суждения «Я голоден» будет несколько смягчен по отношению к абсолютной «голодности» Андрея и Кати.

Здесь конфигурация перфорированных отверстий и иглоков табуляторов задаёт третье измерение, инженерный интенционал, компьютерование. Конечно, можно привести текст более современной программы работы с трехмерными интенционалами. Тем не менее, мы выбрали машину Корсакова. Это – чистая коннекционистская машина, аналогичная чистой репрезентативной машине Бэббиджа-Тьюринга. Поэтому её применение характеризует наибольшую степень теоретической общности рассуждений. Диагональный интенционал – задаётся как функция от коннекции между перфокартами, т.е. когда иглока табулятора провалилась в отверстие перфокарты в процессе работы машины. И, наоборот, на этапе программирования машины, коннекция – это функция от диагонального интенционала, синхронного смыслового схватывания одного и другого признака. Интенциональная диагональ теперь проводится в трёхмерном пространстве, и одно из измерений характеризует способ (нейро)компьютинга.

Чёткое изложение трехмерной интенциональной семантики на базе машины Корсакова требует отдельного обстоятельного изложения.

Нам же следует завершить данную работу, в заключении высказав следующие, на наш взгляд, интересные идеи, вытекающие из концепции интенциональной семантической трёхмерности.

Человеку – интенциональное, машине – экстенциональное

В примере использования машины Корсакова мы видели, как она прекрасно работает с экстенционалами. В самом деле, налицо справедливость неинтенциональной семантики Р. Сталнейкера. Ведь машина эффективно работает с сопоставлением, с переинтерпретацией признаков, представленных в профайлах-перфокартах «Андрея», «Кати», «голода». Так же в некоторой степени верной оказалась концепция В.В. Горбатова – диагональ между экстенционалами как функциями оценки тех или иных возможных миров можно фиксировать

логико-семантическими значениями. Однако для машины Корсакова двужначные логические значения «истина», «ложь» – это крайне ущербный вариант семантики. Машина Корсакова работает как минимум с пятизначной логикой, булевый вариант не приемлем для коннекции. В самом деле, для простейшей коннекционистской схемы, (одной коннекции между двумя перфокартами) имеем следующее: 1) жесткое соединение перфокарт, когда иглока табулятора соединяет отверстия двух перфокарт; 2) соединение с первой перфокартой, отсутствие связи со второй (иглока проникла в отверстие первой перфокарты, но не вошла в отверстие второй); 3) соединение со второй перфокартой, отсутствие связи с первой (игла просквозила через отверстие первой перфокарты, но осталась в отверстии второй); 4) отсутствие связи между перфокартами, так как игла не протыкает отверстия; 5) отсутствие связи между перфокартами, так как отверстия отсутствуют.

Количество логических значений для одной коннекции между двумя перфокартами, конечно, больше, нежели чем пять выделенных. Отсутствие связи может быть обусловлено наличием отверстий в одной, отсутствием отверстия в другой; иглока одно отверстие прошла, а во второй перфокарте отверстие в «диагональном месте» отсутствует. Так же могут иметь место градуированные и градуальные коннекции, которые задают ряд дискретных значений для одной коннекции, и соответственно, задают непрерывные значения, которые определяются толщиной и заострённостью иглы, шириной отверстия в перфокарте и её толщиной. Как видим, «диагональный экстенционал» никоим образом нельзя описать двумя значениями – 1 и 0, как предлагал В.В. Горбатов.

Отметим важное положение – настройка машины и её функционирование определено смыслами, которыми владеет субъект. Для чисто физических условий работы машины эти «смыслы» не нужны, так как наличие/отсутствие признаков задается такими физическими параметрами, как светло/темно; наличие столкновения/свободное движение и пр. Поэтому возможны эффективные роботы, которые настраиваются без субъекта-учителя для манипуляции физическими предметами. Но это только для физической реальности. При компьютеризации когнитивных феноменов невозможно обойтись без субъекта, который отмечает на перфокарте, например, как в оригинальном корсаковском примере настройки машины, наличие *боли* в таком-то участке организма и чем эту боль снимать. Машина Корсакова превосходно работает с экстенциональной частью когнитивных суждений, но принципиально не работает со смыслом, с интенционалами суждений. Осмысленность – это чисто человеческая прерогатива. Способствовать формированию смысла, усиливать когнитивную деятельность

человека – компьютер способен. Но заменить человека в сфере творения смысла – никогда. Поэтому не надо устранять людей – экспертов из когнитивно-компьютерных технологий оценки, например, значимости учёного, знаний школьника по результатам ЕГЭ [Алексеев, 2009-1] или политического лидерства по результатам арифметического подсчёта голосов.

Парадокс семантического обогащения когнитивной терминологии

Обозначив третье, компьютерное семантическое измерение, мы получили возможность оперировать с более широким пространством смыслов. Теперь мы имеем лингво-логическое право инкорпорировать «смысл» технической реализации в «смысл» термина, выражающего психический феномен совместно с его научным объяснением.. Тем самым мы обогатили возможный мир нашей субъективной реальности. В этом мире, например, в рамках решения психофизической проблемы, достойное место занял интенционал технологической реализации. Тем самым проблематика приобрела психо-физиолого-технологический статус. Однако, с другой стороны, мы существенно ограничили возможные миры двумерных семантик. Дело в том, что трудности технической реализации накладывают ограничения на свободу употребления терминов в контексте двумерной когнитивной терминологии. Причем такого рода трудности имеют не только эмпирической, обусловленный, скажем, несовершенством современной техники, но и концептуальный статус, если в рассуждениях обсуждается принципиальный проект компьютера, такой, как машина Тьюринга или машина Корсакова. Например, в 3D-семантике высказывание Д. Чалмерса «Я – компьютерный зомби» не имеет смысла, так как, тот, кто имеет хоть малейшие представления о компьютерной инженерии, не способен помыслить бессознательную техническую систему, поведение и организация которой неотличимы от сознательной (Дж. Маккарти [McCarthy, 1995-3]). Напомним, что сложный аппарат семантической двухмерности Д. Чалмерс использует для обоснования натуралистического дуализма – парадоксальной и, в некотором смысле, постмодернистской конструкции, по причине ужасающей путаницы в представлениях Д. Чалмерса о психологических и феноменологических исследованиях¹. Трёхмерная семантика разрушает наиболее весомый аргумент этой теории, который заключается в концептуальной возможности мира, имеющего точнейший функциональный, физикальный, бихевиоральный дубли-

¹ Под психологией Д. Чалмерс понимает исключительно бихевиоризм и когнитивную психологию. Отсюда – источник некорректности последующих рассуждения о связи психического, феноменологического и физиологического. См. самую первую главу «Два понятия ментального» в [Чалмерс, 2013, с. 19-53].

кат реального Я–Чалмерса, реализованный посредством компьютеринга. В возможном мире двумерной семантики Чалмерс-зомби живет. В мире трёхмерной семантики он не мыслим. Правомочность неодуализма основательно подмочена. Для этого не потребовалась бесконечная метафизическая дискуссия.

Поэтому трехмерную семантику можно использовать как лингвологический инструмент аргументации в баталиях по проблематике философии сознания, которая сегодня инспирирована идеями компьютерной технологии, в частности, и электронной культуры, в общем.

Выводы: 3D-семантика как критерий демаркации исследований искусственного интеллекта

Исходя из достаточно простой идеи о том, что смысл когнитивного термина включает личное психическое переживание, научное объяснение и компьютерную реализацию, когнитивные исследования должны включать три необходимые компоненты: 1) феноменологическую дескрипцию мира субъективной реальности при наличии реального или виртуального факта присутствия в естественном или искусственном мире; 2) научной теории с подкрепляющими её эмпирическими фактами по поводу объяснения и предсказания феноменов субъективной реальности; 3) инженерного (компьютерного) проекта, обеспечивающего имитацию, моделирование, репродуцирование и/или креацирование психического феномена.

Если какая-то одна из составляющих отсутствует, то такие исследования не относятся к сфере ИИ. Можно сколь угодно долго говорить о нейрофизиологических особенностях организации сознательной деятельности, о кодах нейральной динамики. Но если к этому не прилагается инженерный проект компьютеринга данной кодовой зависимости, пускай, представленный в самой абстрактной форме, то такие исследования не обладают концептуальной полнотой. Они вообще выпадают из сферы ИИ. С другой стороны, если предлагается компьютерный проект системы, в котором декларируется возможность манипуляции со «знаниями», то необходимо дать сопроводительное пояснение, на основании какой психологической, лингвистической или нейрофизиологической теории «знаний» построен компьютерный механизм манипуляции «знаниями». Как правило, программисты интеллектуальных систем либо не способны дать вразумительных пояснений либо пользуются эклектичными терминами, которые наспех выбраны из прочитанных книжек.

Если обобщить этот критерий демаркации до уровня когнитивной науки, то так же следует утверждать о необходимости наличия всех трех составляю-

щих. Только в данном случае механизм реализации когнитивного феномена не обязательно представлять в формате компьютерной системы. Например, следует обозначать механизм работы мозга. В частности, вольные социально-эпистемологические исследования, в которых не учтен механизм познания, не следует относить к когнитивным исследованиям.

Следует ли считать исследования, предложенные в движении «2045», исследованиями в сфере ИИ? Вроде бы налицо все три составляющие – когнитивные феномены, которые реализуются компьютерными средствами искусственного бессмертного тела. Нет! Если для выше указанных некогнитивных исследования семантическая мерность не была учтена в достаточной мере, но здесь мы выходим в область семантической сверхмерности. Несомненно, исследования «2045» интересны с философской точки зрения, так как здесь изучаются трансцендентальные перспективы развития человека – электронного человека, е-Номо (А.С. Нариньяни), человечества – точнее, постчеловечества (В.И. Самохвалова), посредством НБИКС¹. Однако смыслы и значения терминов, которыми пользуются сторонники «2045» выходят за сферы очерченного нами трёхмерного семантического треугольника.

Одно дело – функциональная репродукция в виде протезов или функциональных заменителей участков головного мозга. Совсем иное – имплантация Я в искусственное тело. Интенционал Я априори очевиден. Но вот с экстенсионалом имеются проблемы. Его отчаянно искал Р. Декарт. Не нашел, но предположил, что Я является функцией механизма «сомнения» («Cogito ergo sum»). В принципе, мы тоже предположили, что Я является функцией когнитивно-компьютерного устройства («Computa ergo sum»). Но экстенционал – не функция чего-то. Это и есть само «что-то».

Идеологи «Россия-2045» (когнитологи бессмертия) обладают экстенсиональным расширением очевидного для всех интенционала слова «Я». Нам же не хватает для этого ни смелости, ни способности разуметь. По причине того, что экстенционал «Я» трансцендентен относительно трехмерного семантического пространства, выходит за его пределы, то проблематика «2045» занимается чем-то другим, «божественным», но не исследованиями ИИ.

Трёхмерная интенциональная семантика словаря искусственного интеллекта преследует более скромные задачи: отчётливо рассуждать по поводу

¹ Философским исследованиям НБИКС посвящен выпуск журнала «Философские науки», N 8, 2013 г.

междисциплинарных когнитивных исследований и служить лингво-логическим каркасом конвергентной метатехнологии НБИКС-комплекса.

Данная семантика является критерием демаркации исследований ИИ от исследований, которые таковыми не является. В теориях ИИ должно четко утверждаться триединство когнитивного феномена, научного объяснения этого феномена и способа компьютерной реализации (имитации, моделирования, ре-продуцирования, креацирования) этого феномена.

3.4. Конституирующая функция

В данном параграфе продолжается путь, который начался в Главе I, когда раскрывалась роль комплексного теста Тьюринга в позиционировании постнеклассического субъекта электронной культуры и в Главе II, когда мы начали "двигаться" от машины Тьюринга к машинному функционализму Х.Патнэма. Теперь пойдем в направлении к той форме функционализма, который предлагает комплексный тест Тьюринга: к тестовому функционализму, точнее, к тестовому компьютеризму.

3.4.1. Функционализм как доминирующая парадигма современной аналитической философии

Парадигма функционализма в современной философии сознания возникла в 1960-е гг. в ходе критического совмещения достижений бихевиоризма и физикализма. Функционализм подготовил плодородную почву для развития методологии описания, объяснения, предсказания, воспроизведения когнитивных феноменов. Источником функционализма явились идеи искусственного интеллекта. По сей день исследования в этой сфере являются самой продуктивной составляющей когнитивной технологии, так как снабжают когнитивную науку эмпирическим базисом — компьютерными реализациями когнитивных механизмов.

Другие крупные парадигмы философии сознания (физикализм, бихевиоризм, спиритуализм, эпифеноменализм, дуализм, дуаспектизм и пр.) вряд ли следует считать надежными эпистемологическими основами в решении проблемы сознания. Они способны нарисовать красивые блок-схемы психической деятельности (интеракционизм), предложить новые заманчивые теории информации (неодуализм), возбудить художественные интуиции (эмерджентизм в формате синергетического аутопоэзиса) и пр. Однако в условиях прохождения достаточно сурового фильтра компьютерной методологии и инженерии искусственного интеллекта нефункционалистские парадигмы раскрывают свою несостоятельность. Каузальные зависимости блок-схем лишаются номологического статуса. Натуралистический дуализм ведёт к эпифеноменалистическому отрицанию сознания даже у человека, не говоря уже о животных и искусственных системах. Синергетика, которая якобы сменила кибернетику в плане объяснения и конструирования программируемой социально-технической реальности, пользуется настолько невнятно артикулируемым словарём, что попытки его

лексического разбора дают повод усомниться в том, что синергетик сам понимает собственные утверждения.

Общая формула функционализма

В отношении главных оппонентов – физикализма и бихевиоризма – функционализм предлагает комплексный формат изучения психофизической реальности. Функционализм утверждает, что когнитивный феномен психического содержания (боль, мысль, желание, убеждение) зависит не только от внутреннего состава и строения системы (физикализм) и не только от внешней поведенческой диспозиции (бихевиоризм), но и от функции (роли), которую данный феномен выполняет в когнитивной системе в целом. Общее функционалистское суждение (*Фобщее*) относительно стандартного ментального состояния зарубежных функционалистов ("боли") сформулируем следующим образом:

Фобщее: Боль(*m*) – это функциональное состояние(*f*), возникающее из-за телесного повреждения(*p*); проявляется в столах(*b*) и вздрагиваниях(*b*); продуцирует знание (*m*) о том, что с телом что-то не в порядке (*p*), беспокойство(*m*) и желание(*m*) преодолеть (*p*) это состояние(*f*) при отсутствии конфликтующих интенций(*m*),

где термины суждения проиндексированы обозначениями: *m* – ментальное, *p* – физическое, *f* – функциональное, *b* – бихейвиоральное. В данной работе мы будем предлагать формулировки различных версий функционализма, соответствующие по стилю и содержанию формуле **Фобщее**. Формулы потребуются для выявления специфики тестового функционализма. Первоначальная задача на пути к этому – обосновать возможность концептуальной редукции «функционализмов» к исторически первой и оригинальной форме – к функционализму машины Тьюринга. Однако перед этим обозначим проблемы с изучением функционализма в отечественной науке.

Функционализму как философской парадигме сознания в нашей стране не повезло. Впрочем, не повезло и аналитической философии, неотъемлемой частью которой является функционализм. В пп. 1.1.1. мы отмечали, что впервые широкое, во всероссийском масштабе, обсуждение данной парадигмы как эпистемологического основания когнитивно-компьютерной науки было предложено на круглом столе «Функционализм и компьютерные науки» (ведущие – Ю.И. Александров, Б.В. Бирюков и Д.И. Дубровский) в рамках конференции «Философия искусственного интеллекта», г. Москва, 18-20 января 2005 г. [Алексеев, Смирнова, С. 9-10]. В междисциплинарном обсуждении участвовало свыше ста философов, психологов, биологов, филологов, математиков, физи-

ков, социологов, культурологов и др. Встреча носила во многом обзорный характер. Д.И.Дубровский представил разные интерпретации функционализма представителями аналитической философии, выделил нередукционистские варианты, которые могут служить основой продуктивной разработки проблемы сознания, теоретического объяснения природы субъективной реальности и создания перспективных направлений развития когнитивной науки. Функционалистский вариант двуаспектного решения психофизиологической проблемы был предложен Ю.И.Александровым – психические процессы (характеризующие организм и поведенческий акт как целое) и нейрофизиологические процессы (протекающие на уровне отдельных элементов), он предложил сопоставлять через функционально организованные информационно-системные процессы. Дискуссия продолжилась на круглом столе «Философско-методологические проблемы когнитивных и компьютерных наук» в рамках IV Российского философского конгресса «Философия и будущее цивилизации», г. Москва, 24-28 мая 2005 г. (руководители – В.А.Лекторский, В.А.Глазунов, А.Ю.Алексеев) [Философия и будущее цивилизации, 2005]. Обсуждения функционалистской проблематики чётко обозначили отечественные прорехи в знакомстве с зарубежными аналитико-философскими наработками функционализма.

За последние годы ситуация существенно не изменилась. Для отечественных исследователей функционализм до сих пор *terra incognita* и концентрированно не представлен в обзорных антологиях, учебных курсах, переводных изданиях, тематике постоянно действующих семинаров, конференций. Несомненно, при изучении психофизического функционализма опыт зарубежной аналитической философии очень важен. Однако и отечественная философская наука способна внести и вносит оригинальный вклад в развитие функционализма, отталкиваясь от собственных оригинальных идей.

Функционализм Х. Патнэма странным образом созвучен и в некоторых положениях тождественен функционалистским интерпретациям теории мозговой деятельности, предложенных в «Рефлексах головного мозга» И.М. Сеченовым [Сеченов, 1863], за сто лет до патнэмовской концепции. В данной работе мы не будем раскрывать, что аналогии просматриваются даже на уровне мысленных экспериментов, что отмечено в [Алексеев, 2014-11]. Вот только Х.Патнэм эти эксперименты провел век спустя. Сеченовская идея, эпохальная для психифизической проблематики, нашла развитие в психологической, нейрофизиологической, нейрокомпьютерной отечественных школах. Однако эта идея не нашла отражения в современной – аналитической, т.е. в большей мере, «островной» – философии сознания. Зарубежная философия об этой

работе И.М. Сеченова не знает – за 150 лет статья так и не была переведена с русского на английский. Возможно, поэтому советские летающие ракеты летали лучше американских: в нейрокомпьютерных механизмах полёта и наведения отечественных ракет (проекты В.Д. Цыганкова) был заложен принцип анохинской функциональной системы с активным «входом» (анохинская теория, в свою очередь, была инспирирована сеченовскими исследованиями). В американских ракетах использовался винеровский проект пассивной обратной связи, традиционный для бихевиористской парадигмы [Алексеев 2010-2,3]. В связи с этим американцами был упущен параметр «активности», крайне важный для самой сути самоорганизации – посредством генерации возможных траекторий формируется цель и план полета самонаводящейся ракеты, априори не предзаданные разработчиком. Как видим, просветы в методологии ИИ ведут к большим проблемам.

К протофункционалистским идеям, в которых доминирует идея компьютеринга психических феноменов, относится проект первой в мире коннекционистской машины – машины Корсакова, 1832 г. В предыдущем пп. и других работах мы показали то, что проект Корсакова является формальным определением понятия коннекционистского алгоритма, а машина Корсакова – протонейрокомпьютером [Алексеев, 2010-3; 2011-1; 2013-2].

К современным функционалистским идеям относится методологический проект разработки проблематики сознания, предложенный Д.И. Дубровским более 40 лет тому назад [Дубровский, 2007, с. 26-27]. Проект погружает исследователя в многомерное пространство функциональных отношений между когнитивными феноменами и их нейральными коррелятами, образуемое онтологическим, гносеологическим, аксиологическим и праксеологическим «измерениями». Информационный подход к изучению субъективной реальности явно позиционируется как *нередуктивная версия функционализма* [Дубровский, 2006-4].

Значительный вклад в развитие методологии функционализма способна внести даже отечественная парадигма постнеклассической рациональности, предложенная В.С. Стёпиным [Стёпин, 1992]. Несмотря на высокий уровень теоретико-методологической общности, данная парадигма категориально чётко предписывает роли и функции субъекта в современных полисистемных формированиях электронной культуры, в контексте социокультурных ценностей, норм, моральных регулятивов, эстетических канонов информационного общества.

Именно в рамках стёпинского подхода перспективно обозначить новую форму функционализма в философии сознания, которую автор назвал «тестовый функционализм» (или «функционализм теста Тьюринга») [Алексеев 2013]: постнеклассический субъект исполняет роли и выполняет функции в мире электронной культуры. В данном пп. тестовый функционализм рассматривается как начальная идея, которая требует проработки. Однако и эта нечетко сформулированная идея достаточна для демонстрации конститутивной функции, которую способен выполнить комплексный тест Тьюринга.

Поэтому диссертант и призывает обратить внимание на собственную историю развития функционалистской парадигмы, перестать «вслепую» копировать идеи зарубежной аналитической философии, и постараться органично включать эти идеи в контекст изучения «психики», «сознания», «мышления», «личности», «общества», сложившихся в социально-эпистемологическом контексте развития отечественной философии, науки, культуры.

Чтобы выявить функции тестового функционализма, упорядочим бесконечное разнообразие современных "функционализмов". Для этого выявим общий теоретический каркас «семейного сходства» его различных форм. Исторические факты позволяют судить о том, что этот каркас строится из тьюринговых идей в области компьютерной науки и философии искусственного интеллекта. Поэтому перед исследованием чистых истоков современного функционализма нарисуем пеструю диаграмму современных направлений данной парадигмы.

Современный функционализм как парадигма философии сознания многообразен и вариативен. Невозможно утверждать о некотором едином «функционализме» с чётко заданным словарем, методологией, программой. Следует утверждать лишь о «функционализмах». Причем количество разновидностей «функционализма» равно количеству функционалистов, умноженному на количество интервалов времени, в течение которых у функционалиста сохраняется некоторая устойчивая концепция функционализма.

Показательным примером «относительности теории функционализма от времени» являются различные редакции статьи «Functionalism» Джанет Лёвин (2004 и 2013 гг.), изданных в престижной Стэнфордской энциклопедии. В первой редакции (2004 г.) Дж. Левин выделила следующие разновидности: функционализм машинных состояний, психофункционализм, аналитический функционализм, функционализм тождества функциональных состояний. В редакции 2013 г., Дж. Лёвин немного изменила свои функционалистские взгляды.

Три первые разновидности функционализмов остались без особых изменений, но вместо четвёртого предложены новые формы: функционализм ролей (Role-functionalism) и функционализм реализаторов (Realizer-functionalism). В принципе, это объяснимо общемировыми натуралистическими тенденциями в изучении когнитивной проблематики. Так, сегодня достаточно комфортно среди квазифункционалистских теорий сознаний ужився натуралистический дуализм (Д. Чалмерс), который, на наш взгляд, является оттенком постмодернистской реформации научного знания. В нашумевшей недавно у нас в стране книге («Сознающий ум...» [Чалмерс, 2013]) явно обозначенные компьютеристские «идеи» неодуализма сложно связать с компьютерными технологиями. Например, главный аргумент неодуализма – аргумент мыслимости когнитивно-компьютерной системы «Я – зомби» не только расшатывается, но напроочь разрушается при попытке вообразить её инженерную реализацию (Дж. Маккарти, Д. Деннетт и др.). Ранее мы уже отмечали что Роберт Крик, основатель темы «философских зомби» (1978 г.), после знакомства с «Чалмерсом-зомби», в 2005 г. «убивает» своё концептуальное изобретение, так как из-за «зомби» возникла невероятная путаница в когнитивной науке (см. в [Алексеев, 2008-2]). Тем не менее, в работе Д. Чалмерса показательным является факт натурализации феноменологической составляющей когнитивной тематики, хотя непонятно, каким образом дуализм можно физикализировать. Апеллируя к мифической "новой теории информации" Д.Чалмерса?

Следуя натуралистической тенденции, Томас Полгер в объёмной работе «Естественное сознание» [Polger, 2004] попытался создать комплексный формат, единый для всех разновидностей функционализмов. Интегратором является *реализатор* когнитивного феномена. Реализатор – не что иное, как физикальная составляющая когнитивно-компьютерной конструкции. Однако в результате интеграции, на мой взгляд, получился *натуралистический функционализм*, созданный по аналогии с натуралистическим дуализмом. Для интеграции требуется концепция, выходящая за рамки системы понятий. Либо надо показать некий синергетический способ «самоорганизации» функционализмов посредством «странного аттрактора», в роли которого в теории Т. Полгера выступил «реализатор». Но реализатор, который описывается естественными науками, т.е. объективно (основной замысел Т. Полгера именно таков), не может иметь приватного феноменологического статуса, требуется субъект-наблюдатель, *Я*. Что же это за теория сознания, которая не способна описать своё собственное сознание? Возьмите любую нормальную парадигму философии сознания. В ней обязательно участвуют *Я, моё* сознание.

Вернёмся к работе Дж. Лёвин. Она выделяет следующие разновидности функционализма (в редакции 2004 г.).

Функционализм машинных состояний

Возникновение боли, пребывание в болезненном состоянии, стремление избежать боли – всё это описывается программой машины Тьюринга (см. формулу **Фмашинный**).

Психофункционализм

Эта разновидность рассматривается как методология когнитивной психологии, задающая исследовательскую программу описания и объяснения ментальных состояний и процессов в виде некоторых «сущностей», играющих роли в осуществлении поведения. Вспомним формулу Н.Блока (пп.2.1). Если T – психологическая теория с n ментальными состояниями, из которых, например, 17-ое является «болью», тогда психофункционалистская «теория боли» в классическом для этой парадигме формате задаётся следующим образом:

Фпсихофунк: Чувствовать боль = быть F_{17} -ым при условии: $\exists F_1 \dots \exists F_n (T(F_1 \dots F_n, i_1 \dots, o_1 \dots))$, где F_1, \dots, F_n – функциональные термы, обозначающие ментальные термины; i_1, o_1 и т.д. – это входы/выходы системы [Block, 1981].

Аналитический функционализм

Это - методология выработки, анализа, манипуляции и трансляции «тематически-нейтральных», функционально заданных терминов. Данные термины применяются для осуществления аналитических операций обобщения и ограничения когнитивных понятий, например, подведения отдельных психических феноменов под достаточно чётко фиксированный когнитивный тип, либо, напротив, выделение элементарных свойств и сингулярных событий. Данная разновидность функционализма позволяет, по замыслу их сторонников, создать некоторый промежуточный язык, который коррелирует когнитивные феномены с их физическими реализаторами, в частности, с компьютерными системами. Такой медиальный язык, «язык-посредник» по стилю похож на языки функционального программирования, например, аппликативных исчислений, для которых техническая реализация представляется делом второстепенной важности. Главное – предоставить полное и непротиворечивое описание функций, в которых степень общности задаётся рекурсивными определениями, подводимыми под математическое определение алгоритма А. Черча или лингвистические манипуляции в рамках определения А.И. Маркова. Онтологические проблемы для

аналитического функционализма «сняты», они переориентированы на решение лингвистической проблемы. Это позволяет чётко и ясно утверждать о психофизической тематике (вспомним, что аналитическая философия – это, в первую очередь, изучение значений и смыслов терминов используемого языка).

Функционализм тождества функциональных состояний.

Это – развитие программы аналитического функционализма на базе крупных методологических наработок физикалистской теории тождеств. Для анализа когнитивных феноменов применяется двумерный лингво-логический аппарат. Все суждения о психофизической проблематике «нанизываются» на этот логико-лингвистический каркас: по одному измерению позиционируются интенционалы ментальных терминов, по другому – интенционалы естественно-научных терминов. Экстенционалами терминов являются, в общем случае, возможные миры, характеризующие, соответственно, внутренний феноменальный мир человека и миры научных определений, регулярностей, законов и объяснений. Дж. Лёвин достаточно четко подчеркнула двумерность в формировании функционалистского языка – один язык участвует в образовании психологических дефиниций, другой – естественно-научных, в частности, нейрофизиологических. Возникает проблема смыслового тождества функционалистских дескрипций на уровне интенционалов терминов. Ещё одна проблема – выбор одного из двух альтернативных подходов к двумерности: некоторые исследователи считают, что психические феномены описываются языком второго порядка относительно нейрофизиологического языка, другие, напротив, порядок использования языков меняют.

Лингвистическая проблема двумерного описания когнитивных терминов чётко обозначилась в оригинальной для функционализма работе Х. Патнэма «Психика и машина»: 1. Боль Р тождественна стимуляции С-волокон [Putnam, 1960, р. 377]; 2) «Состояние S [боль – А.А.] тождественно включению триггера № 36» [Там же, р. 384]. Для решения данной проблемы был предложен диахронический язык, который чётко задал его двумерность. Орон Шагрир в работе «Взлет и падение компьютерного функционализма» [Shagrir, 2005, р. 229] отмечает два измерения, которые этот язык должен описывать:

1) отношение уникальной реализации F_{UR} : Любое ментальное событие М является так же некоторым функциональным событием F;

2) отношение супервенции (F_{SUP}): Любое функциональное событие F является так же ментальным событием М.

Как было отмечено выше, Т. Полгер [Polger, 2004] акцентирует на *реализаторе* – функциональной системе, которая фактически осуществляет (реализует) психическое, нейрофизиологическое или компьютерное состояние. В силу патнэмовского принципа изофункционализма субстрат реализации не существенен, тем не менее, для натуралистического объяснения сознания требуется конкретика. Поэтому следует различать химические, физические, нейрофизиологические и, добавим мы – компьютерные электронные реализаторы.

В связи с этим, «функционализмы» распределяются по типам отношения (реализаторы/супервенторы). Тогда возникают физические, химические, биологические, нейрофизиологические, психические, персоналистические, социальные и др. «функционализмы». Например, для психической системы (допустим, она описывается на языке второго порядка) реализатором является нейрофизиологическая функциональная система (язык первого порядка). Для социальной системы реализаторами являются функционально организованные личности. И пр.

Понятие *супервентора* вводится по аналогии с понятием «реализатора». Супервенция – это «сопровождение» ментальными феноменами некоторой физической организации. Имеются различные интерпретации «супервенции». Сопровождение понимается как следование перед низкоуровневым процессом, за ним, параллельно ему, или в непосредственной слиянности с ним. Логично предположить, что супервентные отношения осуществляют *супервенторы*. Например, психическая функциональная система является супервентором нейрофизиологической, персоналистическая система – супервентором социальной и т.п. Таким образом, порядок описания когнитивного феномена: супервентор → реализатор, либо реализатор → супервентор, по сути, определяется произволом исследователя, как это мы видели ранее на примере функционализма тождества функциональных состояний.

Теперь, если исходить из отношения «реализатор → супервентор», то супервентор – это функциональная система $(n+i)$ – го порядка, где n – порядок языка, описывающего мои психические феномены (например, «мне больно»), $i = 1, 2, \dots, N$ – порядок метаязыка функциональных спецификаций супервентора. Если исходить из обратного отношения, то реализаторы специфицируются функциональными системами $(n-j)$ – го порядка, где $j = 1, 2, \dots, N$ – порядок метаязыка функциональных спецификаций реализатора. При $i = j = 0$ метаязык вырождается в объектный язык. Например, для вырожденного случая, пусть a ,

x, y – это предметные переменные, тогда «ментальная причинность» в формате *народно-психологического функционализма*.

Народно-психологический функционализм

Раскрывается следующим образом в отношениях супервентора (s) и реализатора (r):

Народно-псих: Если агент a ($Я$), желает x (s) и убежден(s), что x можно достичь (s), выполнив y (r) и приписывает y (s) – лучший способ достижения x среди других возможных способов $\{y', y'', y''', \dots\}$ (r), тогда a устремляется к y (r) и к последующим y' (r) на основе данной интенции (s), при прочих равных условиях (r).

Здесь мы привели пример решения проблемы ментальной причинности, предложенной в [Robb, Neil, 2013]. Имеется некоторый n -ый уровень спецификации *моих* когнитивных феноменов. При $n = 0$ $Я$ имею феноменальное ощущение, «желание как таковое». При $n > 0$, как мы видим для приведённого примера, имеется языковое выражение для «моего желания». Например, « $Я$ желаю x » задаётся при $n = 1$; « $Я$ желаю x и убеждён в возможности удовлетворения x » задается при $n = 2$ и т.д. Порядок уровней описания феноменального качества (n), в принципе, произвольный, так как не совсем понятно, какое ощущение является первичным – «желания x » или «возможности x ». (Это положение, кстати, являлось для Н.Блока одним из аргументов в пользу вторичности интеллектуальных способностей относительно первичных психических переживаний при критике оригинального теста Тьюринга [Блок 2006]. В общем, Fsr задает стандартную функционалистскую схему удовлетворения потребности (по всей видимости, большинство современных теорий организации строятся по этой схеме).

Многообразие функционализмов значительно увеличивается при разрыве непосредственных связей супервентор-реализатор. Например, если супервенторы – это ментальные феномены, а реализатором является «тело» (т.е. здесь нейрофизиологическая функциональная система выпадает из рассмотрения), тогда возникает, скажем, психо/соматический функционализм. Однако вряд ли такое расширения многообразия функционализмов позволительно в когнитивной науке. Нарушение последовательности в отношениях супервенция-реализация – прямой путь к мистицизму, что ярко проявилось, например, при изучении темы «сознание-тело» в стиле а'la Блаватская у И.А. Бесковой [Бескова, 2005]. Отсюда - многообразные спекуляции на тему "когнитивной телесности", "сознательного тела" и пр.

Вернёмся к классификации функционализмов Дж. Лёвин, точнее, к тому, что она упустила. И в старой и в новой редакции статьи «Функционализм», она, по непонятным причинам, не упоминает компьютерный функционализм (computational functionalism).

3.4.2. Компьютеризм как высшая стадия функционализма

Компьютеризм (computationalism) – наиболее мощная парадигма современной когнитивно-компьютерной науки. Она задаёт парадигму изучения когнитивных феноменов самого широкого спектра – витальных, ментальных, персональных, социальных. Для компьютеризма встречаются иные транскрипции: «компьюшизм» [Юлина, 2004], «компьютационализм» [Алексеев, 2006-2], вычислительная теория сознания («The Computational Theory of Mind» [Hurst, 2011]). Все эти «кальки», на мой взгляд, неудобно произносить.

Компьютеризм – это функционализм всех перечисленных разновидностей, каждая из которых конкретизируется тогда, когда супервенторы и/или реализаторы представлены в компьютеризуемом формате. Например, боль как когнитивный супервентор – это не просто статическое психическое состояние или содержание. Это – компонента квазиалгоритмически функционирующей динамической системы, факторы которой задаются ментальными причинами и ролями «боли» в составе функций целостного организма. Реализатор боли – это компьютер, который трактуется двояко – в узком и широком смысле слова «компьютер». В узком смысле «компьютер» рассматривается в инженерном формате, т.е. со всеми прилагаемыми видами обеспечения вычислительного процесса: техническим, программным, информационным, лингвистическим и пр. Когда «компьютер» трактуется в широком смысле, то, к примеру, мозг, тело, общество рассматриваются как абстрактные машины исполнения алгоритма. Имеется ещё одна трактовка термина «компьютер», достаточно часто фигурирующая в концептуальных построениях современных функционалистов: «компьютер» – это некоторый проект, предполагающий реализацию в будущем, но отнюдь не проект компьютерной машины Бэббиджа-Тьюринга-Неймана: «искусственный интеллект на базе старой, дряхлой вычислительной машины – это фантастика дурного вкуса», - напомним известное высказывание Д. Деннета.

Компьютеризм строго продолжает идею машинного функционализма и является наиболее аутентичной разновидностью функционализма. В нем четко прописываются: экстерналистская/интерналистская программы изучения супервенции/реализации, проекты «воплощения» когний в компьютерной сис-

теме, эпистемологический базис в форме «знаний» когнитивно-компьютерной системы.

На наш взгляд, *компьютеризм – это высшая стадия развития функционализма*. Поэтому крайне странным представляется то, что Дж. Лёвин упустила его из виду в редакции статьи 2013 г. (хотя бы в плане ссылки на работы компьютеристов). В компьютеризме идея фикционализма машинных состояний обогащается идеей *содержания* этих состояний в виде теорий репрезентации и «внутреннего языка мысли» (это положение чётко прописал С. Хорст [Horst, 2011]). Дело в том, что машинный функционализм состояний Х. Патнэма был расширен его учеником Дж. Фодором сразу же после обнаружения проблем с тем, что машинные состояния не позволяют манипулировать с содержанием этих состояний. Скажем, для классического патнэмовского пример – «триггер № 36 включен» отнюдь не означает того, что Я обладаю квалиа, если «Я – это машина Тьюринга». Требуется репрезентация содержания «боли» и экспликация этого содержания, например, в психофункционалистском формате. Другой пример репрезентационного подхода: если изучаются когнитивные феномены социокультурного содержания, то в «культуре» модульно выделяются блоки «данных», «знаний» и пр. У функционалиста-культуролога такая поспешная технологическая структуризация "культуры" способна вызвать недоумение. Однако для когнитивно-компьютерного представления «культуры» концептуальных затруднений это как бы не вызывает.

Выше мы рассмотрели многообразие «функционализмов», обусловленное традицией зарубежной аналитической философии. На самом деле, это многообразие намного шире. Ключевой фактор, влияющий на неоднозначность функционалистских подходов, заключается в лингвистической полисемии термина «функция». Покажем влияние различных интерпретаций термина «функция» на невероятно широкое увеличение множества «функционализмов».

Под «функцией» часто понимается разновидность математического отображения, область отправления которой задаётся нейрофизиологическими субстратами и их диспозициями, а область назначения, которую составляют ментальные феномены. Для компьютеризма идеальным является вариант, когда это отображение вычислимо. Функциональная система опосредует данное отношение и может выражаться сложной системой кодов, характеризующих, например, информационную причинность между психическим/нейрофизиологическим. Расшифровать «код» информационной причинности – значит обрести возможность компьютеризовать феномены субъективной

реальности. Данную разновидность функционализма обозначим как *криптологический функционализм* (Д.И. Дубровский).

Если функция – это *роль*, которую выполняет то или иное когнитивное состояние в способе функционирования системы в целом, то получается *функционализм состояний*. Это – классический патнэмовский функционализм. Если принимаются во внимание содержательные характеристики «состояний», такие как «квалиа», «опыт», «переживание», то данная разновидность «ролевого функционализма» носит название *контентного функционализма* («функционализм содержаний»). Методологии их изучения принципиально различаются: если для функционализма состояний онтологический вопрос бессмысленен: «быть в определённом ментальном состоянии – значит находиться в определённом отношении входа к выходу и к другим состояниям системы в целом», то для контентного функционализма становятся необходимыми эссенциалистские соображения по поводу «сущности» конкретного или типового когнитивного состояния.

Далее, функционализм значительно усложняется при учёте того или иного подхода к трактовке термина «сознание». Унификация данного термина невозможна. Например, в аналитической традиции выделяется *«квалиативный» функционализм*, направленный на решение проблемы функционального описания и объяснения квалиа как психологического факта дифференцированной качественности того или иного состояния сознания (Д. Чалмерс). *«Акцессный функционализм»* описывает способы доступности феноменального содержания сознания (Н. Блок). *«Проективный функционализм»* стремится разрешить проблему «проективного сознания» с позиции приватного ощущения «какого быть» (например, каково быть летучей мышью? [Nagel, 1974]). *Нарративный функционализм*, очень схожий с когнитивно-компьютерной интерпретацией лингвистического бихевиоризма – парадигмальной основы оригинального теста Тьюринга, не испытывает проблем с «внутренним Я» и приватными психическими ощущениями. Психическое, самость, я, другие когниции – это иллюзии. Все люди – зомби. Мнимое «персональное» конституируется биографическими сказаниями и социокультурными мемами. Люди могут стать сознательными, если приобретут способность идентификации уникальной функции, которую они выполняют в системе глобального, гетерархически организованного искусственного интеллекта (см. подробный анализ работ Д. Деннета в [Юлина, 2004]). Однако любое «сказание» и мем, помимо формы и содержания «текста», должен иметь носителя (классический подход отечественной школы информатики к трактовке информации как триединству формы, содержания, носителя).

Поэтому даже «чистый» деннетовский функционализм состояний, обоснованный в формате глобального искусственного интеллекта (это – информационная система), немислим без реализаторов-носителей желания, боли, наслаждения и пр.

Понятие «функции» может иметь телеологическое значение, например, для чего и зачем, ради какой цели выполняется данная функция. *Телеофункционализм* знаменует возврат к классическому аристотелевскому определению функции как «быть полезным для чего-то». Стандартной интерпретацией телеофункционализма в социальной технологии является суждение «Если бы не родился Пушкин, Александр Сергеевич, то родился бы Фамилия, Имя Отчество, который выполнил бы функцию создателя русского литературного языка». Либо, скажем, для реализации некоторой неоевгенической социальной технологии индивиды должны выполнять определённые роли в соответствии с заданными когнитивными параметрами.

Термин «телеофункционализм» окружен метафизическими коннотатами. Поэтому удобнее было бы применять термин *контекстный функционализм*. Т.е. функция задается контекстом: для физико-химических процессов контекстом выступают способы функционирования биологических систем, для нейрофизиологических – сознательные системы, для индивидов – социальные системы. Контекстный функционализм характеризуется: 1) «эвристическим» подбором реализаторов когниций, адекватных программе системного развития; 2) холистической программой изучения и развития функциональных систем. Эти два положения значимы для организации конвергентного развития НБИКС-комплекса, в котором технологии выступают в качестве над- и под-систем, в зависимости от целей проработки некоторой комплексной проблемы. Например, биогерменевтика как технология «предпонимания» биологических систем, предполагает то, что биоинженерная конструкция является реализатором (квази)сознательной системы.

На разнообразие «функционализмов» оказывает влияние многоспектное представление того, что является «входом», «выходом», «супервентором», «реализатором». Например, Т. Полгер осуществил классификацию функционализмов по степени различия функциональной и физической организации сознательных систем, т.е. по множеству возможных характеристик «множественной реализации» (MR). Сознательными являются не только люди, но и следующие системы: а) у которых физическая организация незначительно отличается от человеческой, например обезьяны, кошки, собаки и пр. (*слабый функционализм*,

Weak MR); б) у которых физическая организация существенно отличается от человеческой, скажем инопланетяне (*инопланетный функционализм*, *SETI MR*, от «*Search for Extraterrestrial Intelligence*» – *поиск внеземного разума*); в) системы, с неопределённой физической организацией, например, черви, бактерии (*стандартный функционализм*, *Standard MR*); г) все в достаточной степени высоко функционально организованные системы вне зависимости от их физической организации, например, компьютеры (*радикальный функционализм*) [Polger, 2004, p. 6].

Другая классификация примерно по такому же основанию, но более грубая, предлагалась несколько десятилетий ранее Н. Блоком [Block, 1978; Блок, 2006]: а) *шовинистический функционализм*, который предполагает наличие сознания только у человека и, следовательно, все нечеловеческие реализаторы когний – это всего лишь метафоры; б) *либеральный функционализм*, который включает все классы функций, предложенные выше Т. Полгером.

Для построения классификации функционализмов Т. Полгер берёт в основу ранние методологические исследования Пола Чёрчленда (1988 г.), в которых различались четыре проекта философии сознания: онтологический, методологический, семантический и эпистемологический. К третьему классу Т. Полгер добавил интенциональный функционализм, а эпистемологический функционализм разбил на теоретический и объяснительный. Получилось следующее.

Метафизический функционализм

У П. Чёрчленда он обозначался термином «онтологический» функционализм. Предназначен для обоснования онтологических утверждений следующего вида:

Фметафиз: «быть в ментальном состоянии значит быть в функциональном состоянии относительно входа/выхода системы, других ментальных состояний и системы в целом».

Интенциональный функционализм

Стремится объяснить способность функционального состояния обладать интенциональным содержанием: как, что и почему некоторые состояния или функциональные элементы способны быть носителями значения или быть непосредственно некоторыми интенциональными объектами:

Финтенц: Для ментального состояния M системы S обладать интенциональным или репрезентативным содержанием C (или быть способным к

этому), значит выполнять для M функциональные роли $\{R1, \dots, Rn\}$ [Ibid, P. 87].

Семантический функционализм

Фсемант: Содержание или референция O ментального состояния M определяется функциональными ролями $\{R1, \dots, Rn\}$, выполняемыми M в системе S относительно (i) к внутренней когнитивной организации S либо (ii) к взаимодействию S с внешним окружением [Ibid, P. 90].

В отечественной науке данная разновидность функционализма использовалась в психосемантических исследованиях форм общественного сознания (см., например, [Петренко]) или, в более общем формате, в психосемиотике.

Теоретический функционализм

Это, как полагает Т. Полджер – наиболее проработанная версия семантического функционализма, согласно которой функционалистская теория суть психологическая теория, в которой все термы теории специфицированы их функциональными ролями в данной теории:

Фтеоретич: «Содержание, задаваемое ментальным термом или предикатом теории P определяется их функциональными ролями $\{R1, \dots, Rn\}$ которые P выполняет в T » [Polger 2004, P. 93]. Теоретический функционализм очень похож на рассмотренный ранее психофункционализм.

Объяснительный функционализм

Призван объяснить когнитивный феномен посредством референции когнитивного термина к функции (функциональным блокам), которую выполняет некоторый реализатор в системе в целом. Имеются классические объяснительные схемы: телеологическая, которая объясняет каузальные отношения стремлением к цели; адаптационная, которая объяснения строит исходя из понятия механизма приспособления и эволюции. Таким образом, объяснительный функционализм подразделяется на телеологический, этиологический, эволюционный, эволюционно-эпистемологический и др.

Фобъяснит: Объяснение наличия, сохранения или изменения ментального состояния M в системе S ссылается на (i) функциональные роли $\{R1, \dots, Rn\}$ которые M занимает в системе S , или (ii) на то, как некоторые свойства $\{P1, \dots, Pn\}$ способствуют конкретизации (демонстрации) посредством M функциональных состояний $\{F1, \dots, Fn\}$ в S .

Методологический функционализм

Изучает любой процесс или технологию, которые раскрывают понятие системы в терминах функций и функциональных ролей, системы, её подсистем, компонентов и элементов.

Методол: Метод понимания психического состояния M системы S заключается в определении функциональных ролей $\{R1, \dots, Rn\}$, которые M играет в системе S или в некоторой надсистеме S^* (если S играет в S^* функциональную роль); или в определении функциональных блоков $\{P1, \dots, Pn\}$, композиционно задающих M [Ibid, P. 107].

Выше обозначенные формы «функционализмов» базировались на методологических предпосылках исследования когнитивной тематики П. Чёрчлендом. Если принять во внимание подобные отечественные исследования, то получим иную классификацию функционализмов. Например, воспользуемся интересной концептуальной конструкцией – «философским квадратом», предложенной Д.И. Дубровским примерно в те же годы, что и методология П. Чёрчленда. Четырёхмерное пространство схемы когнитивных исследований задаётся пересечением онтологического, гносеологического, аксиологического, праксеологического. Так как эти категории, по мнению Д.И.Дубровского, инвариантны по отношению ко всем философским направлениям, то они соответственно конституируют и частную проблематику – проблему сознания (точнее, субъективной реальности) [Дубровский, 2009-2, С.8-44]: «Без основательной гносеологической рефлексии невозможна основательная онтология... Ценностные параметры и факторы активности непременно включены в любой акт познания и сказываются на его результате» [Дубровский, 2007, С.29]. Отношения, обозначенные на «квадрате», позволяют вести систематический и целостный учёт параметров сознания, не упускать из виду одно по причине изучения другого. При этом по аналогии с «функционализмами» Чёрчленда-Полгера, вполне основательно можно выделить *онтологический, гносеологический, аксиологический, праксеологический функционализмы*. Кстати, *этический функционализм* является традиционным направлением, идущим от «гномов» семи мудрецов. Так, «золотое правило поведения», или, скажем, категорический императив И. Канта четко предписывают *Я* функции и роли носителя, соответственно, справедливого и должного поведения.

Философский квадрат, однако, не способствует полноте охвата когнитивной проблематики. В нём не учитываются *культурологические* особенности трактовки «сознания». Так, например, Ван Гулик – представитель западной культуры – фиксирует не более десятка понятий «сознания», от которых берут начало упомянутые выше разновидности функционализмов: квалиативный, акцессный, проективный и пр. [Gulick, 2004]. Как ранее было отмечено в пп. 3.1.1, в контексте восточной культуры, например, в психологии раннего буддизма насчитывается, как минимум, 54 класса чувственных форм сознания, большинство которых «эстетически нагружено» (среди них – радость, грусть, боль). Осталь-

ные классы (общее число классов состояний сознания достигает 121) – это «этически» нагруженные состояния сознания. Очевидно, что такая детализация, инспирированная метафизикой буддизма, более «аналитична», нежели чем её западный аналог, рассмотренный немного выше. Несомненно и то, что все *социально-эпистемологические вариации трактовки «сознания» существенно влияют на расширение разнообразия «функционализмов».*

Складывается впечатление, что мы имеем некоторую разновидность *континуального функционализма*, который непрерывно вбирает в себя бесконечное множество всех других форм. Поэтому крайне актуальной и важной для развития философии сознания является разработка *интегрального функционалистского формата*. Так, Дж. Лёвин в упомянутой выше работе [Levin 2013] по сути предлагает пример комплексной формы функционализма, которая задается скрупулёзной систематизацией дефиниций термина «функционализм» средствами единого словаря. То есть априорные теоретические предположения исследователя по поводу многообразия "функционализмов" являются основой интеграции.

Т. Полгер идет противоположным путем, "снизу-вверх". Он выделяет важный, но, тем не менее, частный инвариант функционалистской дефиниции - проблему множественной реализации функциональных состояний. Все возможные «функционализмы» объединяются в «*Общем функционализме*» (Generic functionalism) посредством фиксации этого главного, по его мнению, функционалистского отношения. Натуралистическая тенденция проявилась в формировании естественнонаучных требований объективности и синхронности психических/(квази-)биологических феноменов.

Имеются другие способы интеграции функционализмов: не "сверху-вниз" (Дж.Лёвин), не "снизу-вверх" (Т.Полгер), но "сбоку". Для третьего способа правомочен стандартный методологического приёма редукции различных форм функционализма к некоторому прототипу. А так как все современные функционалистские концепции инспирированы идеей машинного функционализма, правомочно сформулировать *тезис редукции функционалистских парадигм к машинному функционализму*. Данную идею мы иллюстрировали ранее (п.п. 2.1.), выдвинув *Тезис эквивалентности теоретического функционализма и функционализма машины Тьюринга*. Кратко суть была в следующем. Обоснование тезиса осуществляется версиями теста Тьюринга, предложенными Н. Блоком: «Гомункулусная голова» (1978 г.) и «Китайская нация» (1980 г.). В первой работе [Block, 1978] представлена теория гомункулов, по сути, воспро-

изводящая современные когнитивные модели многоагентного программирования. Эти модели принципиально представимы машиной Тьюринга. Во второй работе [Block, 1980] устанавливается связь машинного функционализма с психофункционализмом посредством методики Рамсея-Льюиса, где референтами терминов психологической теории являются машинные состояния (см. подробно в [Алексеев, 2014-8, С. С.17-47]). Машинный функционализм - это чрезвычайно универсальная форма функционализма. В самом деле, имеется концептуальная возможность описать *мои* ментальные состояния как машинные состояния «мышления», где предметом "мышления" выступают во-первых, различные теории сознания, во-вторых, функционалистские теории, и, в-третьих, собственная теория машинного функционализма.

Выводы:

1) Принятие частного аспекта как основы исследования невероятно сложной психофизической проблематики приводит к априорной ущербности последующих концептуальных построений. Не надо формировать единую проблемную ось, вокруг которой будут раскручиваться различные бесчисленные «функционализмы», как это предлагает Т. Полгер относительно конкретного отношения множественной реализации.

2) Бесперспективен поиск единого универсального способа представления, своеобразного метаязыка, обеспечивающего экспликацию многообразия сложной проблемы сознания. Невозможна интегральная форма функционализма как механическое библиографическое объединение описаний функционализмов, по сути предложенная Дж.Лёвин.

3) Правомочна редукция "функционализмов" к базисной, архетипичной концепции, которой является идея машинного функционализма. Исторически сложилось именно так. Но при этом возникает проблема редукционистской полноты относительно фиксированного базиса редукции. Какая машина способна претендовать на работу мозга? Нам важен правдоподобный проект машины, который, "будучи приведен какими ни на есть причинами в движение, дает в окончательном результате тот ряд внешних явлений, которыми характеризуется психическая деятельность». Здесь мы воспользовались суждением И.М.Сеченова [Сеченов, 1863, с.9], для того, чтобы подчеркнуть давность идеи машинного функционализма. Она возникла не вдруг и не сразу в 60-е годы 20 века. Идея данной формы функционализма витала за сто лет до патнэмовского варианта.

4) Редукция осуществляется к системе атрибуций, которые *Я* приписываю некоторой *x*-системе. Эта *x*-система имитируется, моделируется, репродуцируется как переживающая, мыслящая, сознающая, самосознающая, любящая и пр. система. То есть интегральная форма функционализма – это то, что мы называли «тестовым функционализмом». В разнообразии «функционализмов» данная форма в наибольшей степени идентична патнэмовской идее. Однако эта идея кардинально модифицирована новой ролью компьютера в реализации психических феноменов. В тестовом функционализме восполняется то, что упустил Х. Патнэм при разработке своей версии. Машина Тьюринга (точнее, вероятностная УЦВМ) у него – на почетном месте, а вот идею теста Тьюринга он упустил.

3.4.3. От машинного функционализма к тестовому компьютеризму

Как отмечалось во Введении к данной главе, фактом своего рождения машинный функционализм обязан проблеме интерсубъективности, которая до 1990-х гг. носила название «проблема других сознаний». В самых первых абзацах своей первой функционалистской статьи «Сознание и машины» [Putnam, 1960], эпохальной для развития современного функционализма, Х. Патнэм обозначает логико-лингвистическую асимметрию концептуально аналогичных вопросов: «Откуда я знаю, что мне больно?»/ «Откуда я знаю, что Смиту больно?»; «Откуда некто знает, что ему больно?» / «Откуда некто знает, что кому-то другому больно?». Данные вопросы, полагает Х. Патнэм, подобны следующим: «Как машина Т устанавливает, что она находится в состоянии S?»/ «Как машина Т устанавливает, что другая машина Т' находится в состоянии S?». [Putnam, 1960, p. 362; Патнэм 1999, с. 23]. В свою очередь, появление искусственного интеллекта так же обязано проблеме интерсубъективности [Алексеев, 2014-8;9].

Социокультурную тематику мы также считаем коренным источником *тестового функционализма*: «Откуда я знаю, что общаюсь с человеком, а не с компьютером?»/ «Откуда некто знает, что общается с человеком, а не с компьютером?». И, более суровый вопрос, уже альтерсубъективного характера: «Откуда я знаю, что я не компьютер?»/«Откуда некто знает, что он не компьютер?». Если на первые два вопроса Х. Патнэм не обратил внимания – это проблематика теста Тьюринга, которую он почему-то игнорирует, то решение двух последних вопросов он связывает с эпистемологической проблемой удостоверения реальности собственного существования. Разрешение проблемы альтерсубъективности ищется в оппозиции экстернализма/интернализма с контексте «новой теории референции». Для этого разрабатывается знаменитый компьютерно-

ориентированный мысленный эксперимент «Мозги в сосуде». Он отражает возможность порождения реальности, осознания Я, внешнего и внутреннего мира человека посредством компьютерной стимуляции мозговой активности [Putnam, 1981, pp.2-21]. Следует отметить, что в русскоязычной литературе употребляется не совсем корректный перевод названия статьи: «Мозги в бочке». В бочках хранятся огурцы и селёдка. Сложно вообразить то, что в состав бочки входит компьютерный интерфейс стимуляции мозговой активности. Намного проще вообразить лабораторный сосуд – реторту с жидкостью, к горлышку которого подведёны интерфейсные провода. В этой реторте плавают мои и ваши мозги, стимулируемые компьютерной программой. Конечно, это – «дьявольский научный эксперимент» [Ibid, P. 5], однако он достаточно ярко обозначает компьютеристские проблемы альтерсубъективности.

Машинный функционализм, решая проблему интер-альтерсубъективности, ориентирован на компьютерную метафору психической деятельности: мозг - это машина Тьюринга. Теории, которые последовали вслед за патнэмовской концепцией сознания (психофункционализм, элиминативный материализм, интегральные версии функционализма и др.), потеряли идею мозга-компьютера, акцентировали внимание на более общей тематике "других сознаний" и компьютерную реализацию *другого* или его психических способностей явно не определяли.

Тестовый функционализм возвращает идею компьютеринга когнитивных феноменов, но значительно расширяет их сферу, не ограничиваясь явлениями ментального содержания. Тестовый функционализм, благодаря комплексированию частных тестов самого разного направления, ориентир на реалии информационного общества, компьютерно-коммуникационных технологий и, в общем, тенденций развития электронной культуры [Алексеев 2011-2; 2011; 2013-1] (см. пример "Девушка по переписке"). Человек снабжается концептуальным инструментарием комплексного тестирования компьютеризованной реальности. Инструментарий, в принципе, технически реализуем. Тогда он способен решать инженерные задачи проверки коммуниканта на предмет его «человечности». Однако не технологическая задача является главной в тестовом функционализме. Это - концептуальный инструментарий и он выполняет эпистемологическую функцию выработки рационального скептического мировоззрения к феноменам компьютеризованной реальности и к проектам электронной культуры. В связи с этим замысел комплексного теста связывается с общей целью применения интеллектуальных систем в обществе, метко подмеченный В.К. Финном [Финн, 2007, С.30]: интеллектуальные системы предназначены

для рационализации общественной жизни и усиления интеллектуальной деятельности человека, но отнюдь не для имитации и замены его во всех сферах деятельности, как это полагал А. Тьюринг.

Подчеркнём следующее: эпистемологический источник машинного и тестового функционализма тот же, а именно, проблема рационального изучения проблемы альтер- и интерсубъективности. То есть и тестовый и машинный функционализмы концептуально тождественны в плане решения одной и той же задачи. Другим источником отождествления является опора на фундаментальное определение понятия алгоритма, в соответствии с которым работает машина, продуцирующая психические феномены. Для машинного функционализма используется проект машины Тьюринга и концепция универсальной цифровой вычислительной машины. Сегодня эти машины морально устарели. Тестовый функционализм предлагает новый, более мощный концептуальный проект – машину Корсакова-Тьюринга. Как было показано в предыдущем пп., такая машина позволяет интегрировать коннекционистский и репрезентативный алгоритмы в одном формате, преодолеть детерминизм алгоритма.

Современный функционализм предлагает интегративный формат изучения и реализации когнитивных феноменов самого широкого содержания: физического, витального, персонального и социального. По крайней мере, в версиях Т.Полгера и Дж.Лёвин. Тестовый функционализм так же предполагает интегративный формат. Компьютируемая реальность организована в виде функциональных систем на различных уровнях её гетерархии. Однако он кардинально отличается от других интегративных разновидностей функционализма главной целью – он предназначен для решения собственно человеческих проблем конкретной личности, его свободы, смысла жизни, бессмертия, социального идеала. Я, выступая в роли тьюрингового судьи, конституирую окружающий меня компьютерный мир, в который я оказался "заброшен".

Тестовый функционализм

Обозначим еще одну формулу нашей работы. В Заключении работы она поможет сформулировать идею постнеклассического позиционирования субъекта компьютерной технологии и оценить праксеологические особенности преодоления человеком алгоритмических закономерностей мира электронной культуры:

Тестовый: Я – главное действующее лицо компьютерного мира, который конституирую в виде витальных, ментальных, персональных, социальных систем.

Вернёмся к тезису эквивалентности теоретического и машинного функционализмов, которые мы обозначили как "тезис Блока". Заодно покажем перспективы тестового функционализма. Тезис вполне правомочен, так как: 1) функциональными терминами представимы все иные разновидности терминов – номинативные, дескриптивные, диспозициональные, предикативные; 2) программа редукции ментального к функциональному позволяет реализовать (по крайней мере, имитировать) функции интеллектуальной системы, не вдаваясь в подробности её онтологии.

Однако! В процессе проведения операции редукции, как правило, отбрасывается некоторый сопутствующий «осадок». При формализации психических качеств, состояний, свойств, отношений, процессов и пр. в «осадок» выпадает качество субъективной реальности (Д.И.Дубровский): интенциональность, содержание, контент, квалиа. В результате редукции ментального к функциональному, аппроксимативно полного описания всех возможных входов / выходов / состояний системы из проблематики сознания изгоняется собственно «сознание», значение термина «сознание» переходит в область функциональной дескрипции. Вместе с этим перестают обладать качеством мои боли, надежды, мысли, моя свобода, вдохновение и, собственно, Я. «Сознание» утеряно. Его потерял теоретический \equiv машинный функционализм. Его потеряли все другие разновидности функционализмов, так как их оригинальная парадигма – это функционализм машины Тьюринга.

Тестовый функционализм возвращает «сознание» человеку, изучающему «сознание». Здесь отсутствует антиредукционистская программа организации отношений между ментальным и функциональным. Предлагается кардинальное решение: вообще покинуть стандартную схему рассуждений о психофизической проблеме и перейти в сферу постнеклассической методологии (В.С. Стёпин). Для этого и разрабатываются эвристически более богатая теоретическая конструкция, чем функционализм машины Тьюринга. Предлагается интегральная форма функционализма – *функционализм теста Тьюринга (тестовый функционализм)*. Насколько полно нам является «возвращённое сознание» индивида компьютеризованного общества, определяется феноменологическими особенностями восприятия результатов функционирования компьютерной системы или, в общем случае, х-системы, как «интеллектуальной», «болевой», «волевой», «сознательной», "любящей" и пр. Точнее, факт обладания «сознанием» задаётся убедительными, обоснованными атрибуциями данного факта. Система атрибуций строится наблюдателем – тьюринговым судьёй, в роли которого выступаю Я.

Позиция тьюрингового судьи задаёт четыре типовых способа интерпретации компьютерно-реализованного «сознания», которые подразумеваются в тестовом функционализме. «Компьютерное сознание» – это: 1) *имитация ментального феномена* – лично я убеждён в том, что современный компьютеринг может претендовать лишь на этот способ; 2) *модель ментального* – к этому склоняется большинство представителей когнитивных наук, для которых некоторый механизм сознания – это действительная модель когниции, с присущими методу моделирования онтологическими притязаниями на её адекватность; 3) *репродуцирование ментального*, т.е. фактическое воспроизведение компьютерными средствами феномена сознания: уверенность исследователя базируется, например, на принципе инвариантности информации относительно её носителя и на том, что онтология «сознания» имеет информационный статус (позиция Д.И. Дубровского); 4) *креация ментального* – это новые феноменальные качества, свойства, события, продуцируемые компьютером; они не имеют натуральных аналогов, это - "иное сознание" (сторонники сильного ИИ).

Мы очертили перспективы и кратко обозначили предмет тестового функционализма. А.М.Тьюринг претендовал исключительно на *имитацию* интеллектуального поведения. Наша задача намного амбициознее. С целью последующего решения этой задачи выдвинем еще один тезис тестового функционализма:

Тестовый функционализм обеспечивает рациональную атрибуцию компьютерным системам Я, сознания, интеллекта.

Выводы

В третьей главе, основной для диссертационной работы, решалась задача демонстрации метода использования комплексного теста Тьюринга как базового категориального аппарата философии искусственного интеллекта и как концептуального инструментария исследования когнитивных проблем электронной культуры.

Осуществлена экспликация *дефинитной функции* комплексного теста Тьюринга путем анализа способов приписывания компьютерным системам социальных, персональных и ментальных феноменов (п.3.1).

Экспликация *критической функции* комплексного теста Тьюринга осуществлена путем философской экспертизы технологий построения интеллектуальных компьютерных систем, проектов электронной культуры и программ развития искусственного интеллекта (п.3.2).

Экспликация *конструирующей функции* комплексного теста Тьюринга заключалась в разработке философско-методологических средств, пригодных для обоснования стратегии развития междисциплинарных исследований искусственного интеллекта, для модификации ориентиров научно-методологических исследований искусственного интеллекта и для реформирования фундаментальных принципов компьютерной технологии.

Экспликация *конституирующей функции* комплексного теста Тьюринга заключалась в постнеклассическом позиционирование субъекта электронной культуры, который пользуется концептуальным инструментарием теста в контексте оригинальной парадигмы «сознания» - в контексте тестового функционализма.

Заключение

Дальнейшие перспективы изучения роли комплексного теста Тьюринга в философии искусственного интеллекта имеют методологические и мировоззренческие ориентиры.

Методологические перспективы определяются тем, что комплексный тест становится основным концептуальным аппаратом наиболее значимого проекта искусственного интеллекта, который в современной науке и технологии обозначается термином "искусственные общества". Дело в том, что современные междисциплинарные исследования ИИ охватывают широкий спектр когнитивно-компьютерных феноменов. Философские, научные, инженерные исследования ИИ в последнее десятилетие достаточно четко дифференцированы проектами конвергентного развития комплекса НБИКС (нано-, био-, когни-, инфо-, социо-технологий). Развитию биотехнологий способствуют исследования «искусственной жизни» (artificial life), ориентированные на компьютеринг витальных феноменов. Когнитивная технология базируется на метафорах, подходах, концепциях, моделях «искусственного сознания» (machine consciousness, robot mind) и «искусственной личности» (artificial personality, computer person), соответствующих компьютерингу ментальных феноменов (квалиа, восприятие, представление, мышление, понимание, интуиция, доступность, проективность) и персональных феноменов (самость, личность, индивидуальность, моральная вменяемость, свобода, ответственность). Социокультурные явления в последнее десятилетие активно изучаются в «искусственном обществе» (artificial society) – чрезвычайно перспективной инструментарии информационных и социальных технологий, ориентированном на компьютеринг самых различных явлений общественной жизни – от материально-производственной сферы до религиозных верований. «Искусственный мир» (artificial world, synthetic world) в ряде футурологически окрашенных концепций представителей «цифровой философии» связывается с достижениями нанотехнологии, однако даже в презумпции редукционистского физикализма имеются методологически корректные идеи творить «искусственный мир» посредством интеграции достижений НБИКС. Следовало бы к этим технологиям добавить космическую, но тогда аббревиатура некрасивая получается, так как литера "С" - занята.

Подобного рода меганаучным проектам правомочно присвоить ярлык «исследования искусственного интеллекта». Во-первых, им имманентен слож-

ный интеллектуальный компьютеринг, в основе которого – схемы *творческой* деятельности человека. Во-вторых, и, пожалуй, наиболее значимо то, что фундаментальные способы исследования и реализации когнитивно-компьютерных феноменов НБИКС-технологий инспирированы методологическими и мировоззренческими идеями, концепциями, моделями, методами, средствами, наработанными за шестьдесят лет в философских исследованиях искусственного интеллекта. В связи с этим, скажем, «философия когнитивной науки», «философия высоких технологий» или так называемая «философия постчеловечества» – лишь некоторые вариации философии ИИ. Именно в философии ИИ впервые была обозначена идея компьютеринговой реализации самого существенного для *homo sapiens* когнитивного феномена – *интеллекта*. Все прочие когнитивные феномены – витальные, психические, социальные – производны от интеллекта как целостной интегративной способности человека упорядочивать чувственные данные, рационально рассуждать и нормативно регулировать активность.

Проект «искусственного общества»¹ представляется наиболее значимым инструментарием в ряду других возможных способов анализа состояния, планирования и прогнозирования путей общественного развития. В самой основной части этого инструментария – в методологической – особенно важными представляются следующие собственные и производные идеи, связанные с именем А.М. Тьюринга:

- 1) машина Тьюринга как компьютерная дефиниция «алгоритма» (1936 г.);
- 2) патнэмовский функционализм машины Тьюринга как компьютерная дефиниция «сознания», и вытекающая из данного определения, функционалистская психофизическая парадигма, изначально решающая проблему «дух/тело» путём отождествления психических феноменов с логическими состояниями машины Тьюринга и доминирующая в современной философии сознания в различных вариациях;
- 3) тест Тьюринга как компьютерная дефиниция «интеллекта» и формальное определение понятия «искусственный интеллект» (1950 г.);
- 4) разнообразные версии и интерпретации теста Тьюринга (с 1950-х годов по сегодняшний день) как компьютерные дефиниции многообразных физических, витальных, ментальных, персональных и социальных феноменов, таких

¹ Термин «искусственно общество» крайне полисемичен и его использование имеет давнюю историю. В рамках наличных у автора электронных ресурсов JSTOR и Wiley, первое употребление слова «artificial society» обнаруживается в обзорной работе 1863 г., описывающей способы конструирования социального поведения горилл и тропических негров [Owen, 1863, p.52].

как «жизнь», «творчество», «познание», «понимание», «любовь», «свобода», «общество» и пр., предложенные в работах Дж. Маккарти, Д. Деннета, Н. Блока, К. Колби, Дж. Серля, Р. Френча, Ю. Геновой, С. Ватта, А. Сломана, С. Брингсйорда, Дж. Лукаса, Р. Пенроуза, Дж. Мура, Дж. Баресса, С. Харнада, П. Швайзера, Дж. Поллока, Р. Кирка, Д. Чалмерса, В.К. Финна и др.

В данной работе предлагался комплексный тест Тьюринга, который в контексте этого перспективного проекта следует рассматривать как компьютерную дефиницию «общества» и формальное определение понятия «искусственное общество». По мнению автора, именно интегральное использование частных компьютерных дефиниций, которые предлагаются в различных версиях тестов Тьюринга, способствует динамичному охвату полидисциплинарной проблематики искусственных обществ – самой сложной проблематики на сегодняшний день в науке, культуре, технологии и наиболее перспективной для формирования общих и частных стратегий общественного развития.

В системе исследований искусственного интеллекта методология искусственных обществ (ИО) обладает фундаментальным и интегральным статусом. *Фундаментальная роль* приписывается в силу общих положений о роли и месте научной методологии в структуре знания современного человека. В исследованиях ИИ, в контексте которых рассматривается проблематика ИО, выделяются следующие уровни [Алексеев, 2005-8]: 1) *философский*, который формируется рефлексией над а) мировоззренческими ориентирами современного человека, связанные с компьютеринговой модификацией традиционных представлений о смысле жизни, свободе, социальном идеале; б) методологическими способами организации междисциплинарных исследований когнитивно-компьютерных феноменов; 2) *научный*, на котором а) выдвигаются теоретические конструкции (идеи, модели, системы) и б) исследуются прототипы когнитивно-компьютерных систем – эмпирические модели, которые верифицируют и эвристически обогащают теоретические положения; 3) *практический уровень*, на котором а) проектируются, внедряются, развиваются инженерные системы и технологии и б) используются компьютерные технологии в быту и повседневности. Система исследований ИИ целостна, так как на бытовом уровне действует конкретный индивид – носитель мировоззрения и поэтому уровень 3.б замыкается на уровень 1.а.

Методология искусственных обществ, формируя категориальные способы обживания современным человеком мира электронной культуры, цементирует все выделенные уровни. Фундируя общие параметры и специфические характеристики в философских, научных и практических методах, методология ИО

призвана обеспечить создание адекватного концептуального инструментария как «прибора добывания знания» (В.Л. Макаров), «культурной универсалии» (В.С. Стёпин), «аппарата отображения» (К. Лоренц), «воспринимающего аппарата» (К. Поппер), «априорной конструкции» (И. Кант). Используя такой когнитивно-компьютерный путеводитель по просторам информационного общества, человек рационально конституирует самого себя и конструирует собственное социокультурное окружение. Сегодня о полновесном инструментарии ИО можно судить лишь в сослагательном наклонении, однако автор не видит особых принципиальных трудностей в его построении.

Интегральная роль, приписываемая методологии ИО в исследованиях ИИ, объясняется предметно-содержательными тенденциями комплексного компьютеринга всех сфер жизни человека и общества, начиная от биологических проявлений и завершая трансцендентными феноменами религиозных верований. Функция интеграции различных направлений ИИ реализуется комплексным тестом Тьюринга, который обеспечивает конкретизацию, как правило, абстрактных когнитивно-компьютерных моделей под нужды субъекта – пользователя инструментарием ИО.

В методологических исследованиях ИО одна из главных задач состоит в выявлении фундаментальных компьютерно-ориентированных культурных универсалий, эпистемологических средств, моделей, метафор, концепций – всего того, на чём, собственно, базируется человеческое представление мира. Главная заслуга А.М. Тьюринга состоит в том, что он предлагает фундаментальные концепции компьютеринга в форме машины Тьюринга и теста Тьюринга. В данной работе эти идеи были продолжены и, по мнению автора, была охвачена более обширная сфера индивидуальной и общественной жизни, нежели чем «интеллект», для тестирования которого предназначался оригинальный тест.

Краткие методологические выводы, касающиеся перспектив применения комплексного теста для развития методологии искусственного интеллекта, таковы:

1. Комплексный тест Тьюринга – это когнитивно-компьютерная дефиниция общества электронной культуры.
2. «Искусственное общество» - лингво-нейрокомпьютерная реализация комплексного теста Тьюринга.
3. Комплексный тест Тьюринга не является метафизической концепцией «тестом тестов Тьюринга». Эпистемологический статус его прагматичен.

Наиболее значимые перспективы применения комплексного теста обусловлены мировоззренческим соображениями.

Мировоззренческие перспективы ранее рассматривались в ряде тем п.1.1. В заключение следует их более четко обозначить. Эти перспективы связываются с вписыванием концепта экзистенции *Я* в концепцию э-культурного бытия, в картину высокотехнологичного мира. Требуется в способы конвергентного развития и функционирования НБИКС-технологий, обеспечивающих и воспроизводящих *мою* э-культурную жизнь, включить *мое* сознание с его невероятно сложным комплексом мыслей, переживаний, болей, намерений и пр. Такое требование принципиально не согласуется с онтологическим тезисом репрезентативной проблематики исследований э-культуры (пп.1.2.3): самое поверхностное феноменологическое исследование альтер- и интерсубъективных отношений в контексте э-культуры раскрывает пустую формальность тезиса: «Существовать – значит быть представленным в интернете» и, более широко, «e–re–praesento ergo sum». Напомним, что при следовании репрезентативной парадигме, возникают такие курьёзы: если отсутствует сайт организации, то её как бы и нет; отсутствуют научные труды в РИНЦ – ученого вроде бы и не существует.

Для решения новой онтологической проблемы сознания с целью вписывания *Я, другого*, моего сознания, другого сознания в контекст электронной культуры наиболее адекватным представляется функционализм, соответствующий реалиям компьютерного мира – *компьютеризм*.

В персоналистических перспективах философии ИИ самым главным вопросом является праксеологический вопрос типа: Возможна ли свобода *Я* в контексте запрограммированной реальности? Для решения его предлагается постнеклассический формат функционализма – *тестовый функционализм (функционализм теста Тьюринга)*, исследующий альтер- и интер-субъективные функции *Я* и *другого* посредством методологии комплексного теста Тьюринга. Так как при атрибуции (приписывании) мышления, переживания, творчества, сознания, самосознания, любви и других феноменов сознания х-системам предполагается компьютерная реализация этих систем, то более корректно тестовый функционализм называется *тестовым компьютеризмом*. В разнообразии функционализмов тестовый функционализм в наибольшей степени идентичен машинному Ф. и восполняет то, что упустил Х.Патнэм – идею теста Тьюринга.

Я снабжается концептуальным инструментарием комплексного тестирования компьютерного мира. В принципе, возможна программно-техническая реализация этого инструментария, это – не существенно, здесь нет заявки на пре-

мию Лойбнера. Главное – исполнение эпистемологической функции выработки рационального скептического мировоззрения к феноменам компьютеризованной реальности. Для этого надо принять постнеклассическую позу: позиционировать себя как главного субъекта э-культуры, который принимает или отвергает её ценности и активно конституирует её феномены, тем самым расширяя альтер- и интерсубъективные смысловые пространства. По аналогии *Я* может принять позицию *другого*. *Я* освобождается от алгоритмической зависимости, осуществляя своеобразное компьютерное «эпохэ». Это означает следующее. При осуществлении компьютерной репрезентации когнитивного феномена, коррелирующего со мной или другим, т.е. при экспликации некоторого феномена в формате функционалистской дескрипции, *Я* редуцирует её до уровня квазиалгоритма и включает эту программу в «базу знаний» программ комплексного тестирования э-культуры. Эта программа принадлежит *Я*, она достаточно полно и точно, как утверждают функционализмы, характеризует *мое* сознание и сознание *другого*, но это – не *Я* и не *другой*. Программу реализации моего сознания следует по-гуссерлиански жестоко отбросить за «горизонт». Комплексно отбрасывая атрибуты, исключается очень многое, однако самое важное и приватное сохраняется «в остатке». Остается, по крайней мере, все то, с чем не справляется современный функционализм: самость, квалиа, проекция «каково быть собою», интенциональное содержание опыта и т.п.

В машинном и тестовом функционализмах роли *Я* и компьютера радикально различны. Если в первом утверждается: «*Я* – компьютер» (точнее, с учетом расширения машинного функционализма следовало бы утверждать: «*Я* – машина Корсакова-Тьюринга»), то во втором, напротив, «*Я* – не компьютер (не машина Корсакова-Тьюринга)». Поэтому тестовый компьютеризм, несмотря на значимость роли в развитии э-культуры в плане расширения её арсенала программами репрезентации когнитивных феноменов, в большей мере, предназначен для решения традиционных, отчасти компьютерно модифицированных, но собственно человеческих, проблем личности, свободы, смысла жизни, бессмертия, социального идеала.

Предложим онтологическую формулу *тестового компьютеризма*: *когнитивный феномен электронной культуры конституируется атрибуциями компьютерно независимого Я, исполняющего функции комплексного теста Тьюринга.*

В заключение выдвинем онтологический тезис существования *Я* в э-культуре с учетом обозначенных праксеологических коннотатов:

Non computato, ergo sum: не компьютеризуюсь, следовательно, существую.

Роль комплексного теста Тьюринга в изучении и развитии электронной культуры состоит в том, что тест является базисным концептуальным инструментом исследования ее феноменов. Такой тест рационально ориентирует человека на решение когнитивных проблем компьютерного мира.

Список сокращений и обозначений

ЕИ – естественный интеллект

ИВИ – игра в имитацию

ИИ – искусственный интеллект

ИО – искусственное общество

СР – субъективная реальность

ТТ – тест Тьюринга

ЭЭ – эволюционно-эпистемологический

Библиография

- [Алексеев, 1994-1] Алексеев, А.Ю. Персоналистическая концепция информатизации человеко-технических систем [Текст] // А.Ю. Алексеев, В.В. Деев // Информатика и науковедение. Труды 3 Междунар. науч. конференции 28-29 нояб. 1994 г. – Тамбов: МИНЦ - 1994 г., с.3-4
- [Алексеев, 1998-1] Алексеев, А.Ю. Гуманизм, персонализм и информатика (к общетеоретическим основам моделирования искусственной личности) [Текст] / А.Ю.Алексеев // Здравый смысл, 1998. № 3/7, С. 52-56
- [Алексеев, 1998-2] Алексеев, А.Ю. К осмыслению модели смысла [Текст] / А.Ю.Алексеев // Здравый смысл, № 7, 1998, С. 48-52
- [Алексеев, 1998-3] Алексеев, А.Ю. Искусственная личность. [Текст] / А.Ю. Алексеев // Труды конференции «Междисциплинарный подход к изучению человека», М.: Институт микроэкономики, 1998, С.28-31
- [Алексеев, 1998-4] Алексеев, А.Ю. Критерии целостности гуманистической информационной технологии. [Текст] / А.Ю. Алексеев // Труды Международной московской конференции гуманистов «Наука и здравый смысл в России. Кризис или новые возможности». – М.: 1998, С. 202-207
- [Алексеев, 1998-5] Алексеев, А.Ю. Программа информационно-технологического и организационно-структурного моделирования человека. [Текст] / А.Ю. Алексеев, В.В. Деев, В.В. Насыпный, В.В. Зивенко, А.Ю. Плюснин, А.С. Абаджан // Здравый смысл, № 7, 1998, С.45-48
- [Алексеев, 2001] Алексеев, А.Ю. Паранормализация искусственного интеллекта. [Текст] / А.Ю. Алексеев // Поругание разума: экспансия шарлатанства и паранормальных верований в российскую культуру XXI века / Тезисы к международному симпозиуму «Наука, антинаука и паранормальные верования. Москва, 3-7 октября 2001 г. М.: «Российское гуманистическое общество», 2001 г., С.8-11
- [Алексеев, 2002-1] Алексеев, А.Ю. Глобальные реальности XXI века. От переводчика. К статье П. Курца «Инфомедийная революция и перспективы глобального гуманизма» [Текст] / А.Ю. Алексеев // «Здравый смысл», № 3(24), лето 2002, М.: «Российское гуманистическое общество», С.3
- [Алексеев, 2002-2] Алексеев, А.Ю. Национальные особенности логико-смысловых конфигураций // [Текст] / А.Ю. Алексеев // Приладожье и русский Север: история, традиции, современность. Материалы научно-практической конференции в честь 300-летия г.Лодейное Поле, 8 – 12 июля 2002 г. / Научн. ред. Ю.Б. Сенчихина. Сост. А.Ю.Алексеев и С.Ю.Карпук. – М.: РГО, 2003 – 208 с. – С. 12-17
- [Алексеев, 2004] Алексеев, А.Ю. Компьютерное моделирование смысла (философско-антропологический анализ). Автореферат диссертации на соискание ученой степени кандидата философских наук, М.: МГУ им. М.В. Ломоносова, 2004 г.
- [Алексеев, 2005-1] Алексеев, А.Ю. К проблеме приписывания ментальных свойств компьютеру [Текст] / А.Ю. Алексеев, А.В. Бондарь, Н.С. Скришевская, С.А. Игнатов, Н.В. Карпук // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С. 88-92
- [Алексеев, 2005-2] Алексеев, А.Ю. Социокультурные ориентиры компьютерных моделей «смысла» [Текст] / А.Ю. Алексеев, А.Л. Дрозд, А. Кислюк, Н.В. Маклашевская, А.В. Скрябин // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С. 116-120
- [Алексеев, 2005-3] Алексеев, А.Ю. Проблема зомби и перспективы проектов искусственной личности и искусственного общества [Текст] / А.Ю. Алексеев, Т.А. Кураева, А.К. Тумасян // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С. 26-33
- [Алексеев 2005-4] Алексеев, А.Ю. Философия искусственного интеллекта [Текст] / А.Ю. Алексеев, Н.М. Смирнова // Журнал «Вопросы философии», № 9, 2005 - С. 68-80

- [Алексеев, 2005-5] Алексеев, А.Ю. К проблеме приписывания ментальных свойств системам [Текст] / А.Ю. Алексеев, А.В. Гидулянова, Н.В. Карпук, А.К. Тумасян, М. Шпёррле // Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. Под ред. В.А. Лекторского и Д.И. Дубровского – М.: ИФ РАН, 2005. – 400 с. – С. 21 – 24
- [Алексеев, 2005-6] Алексеев А.Ю., Дрозд А.Л., Кислюк А., Маклашевская Н.В., Скрыбин А.В. 2005. Социокультурные ориентиры компьютерных моделей «смысла» //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.116-120
- [Алексеев, 2005-7] Алексеев, А.Ю. Уровни изучения искусственного интеллекта [Текст] / А.Ю. Алексеев // Методологические и теоретические аспекты искусственного интеллекта. – М.: МИЭМ, 2005. – 192 с. – С. 23-35
- [Алексеев, 2005-8] Алексеев, А.Ю. Мировоззренческо-методологическая функция философии искусственного интеллекта [Текст] / А.Ю. Алексеев // Преподавание философии в вузе: проблемы, цели, тенденции: Сб. статей Всероссийской научно-методической конференции / Под ред. А.М. Арзамасцева. – Магнитогорск: МГТУ, 2005. – 320 с. – С. 187 – 193
- [Алексеев, 2005-9] Алексеев А.Ю., Кураева Т.А. 2005. Проблема зомби и перспективы проекта искусственной личности //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.5-10
- [Алексеев, 2005-10] Алексеев, А.Ю. О конференции «Философия искусственного интеллекта» [Текст] / А.Ю. Алексеев // Вестник Российского философского общества, 1(33), 2005 - С. 105- 109
- [Алексеев, 2005-12] Алексеев А.Ю., Кураева Т.А., Тумасян А.К. Проблема зомби и перспективы проектов искусственной личности и искусственного общества //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.26-33
- [Алексеев, 2006-1] Алексеев, А.Ю. Возможности искусственного интеллекта: можно ли пройти тесты Тьюринга? [Текст] / А.Ю. Алексеев // Искусственный интеллект: Междисциплинарный подход / под ред. Д.И. Дубровского и В.А. Лекторского. – М.: ИИнтелЛ, 2006. – 448 с. – С. 223-243
- [Алексеев, 2006-2] Алексеев, А.Ю. Методологические и теоретические аспекты искусственного интеллекта [Текст]/ А.Ю. Алексеев – [ред.] – М.: МИЭМ, 2006. – 192 с.
- [Алексеев, 2006-3] Алексеев, А.Ю. Проблемы искусственной личности: развитие междисциплинарного подхода к искусственному интеллекту [Текст] / А.Ю. Алексеев // Труды XXV межрегиональной научно-технической конференции «Проблемы эффективности и безопасности функционирования сложных технических и информационных систем. Ч 2. – Серпухов: Серпуховский ВИ РВ, 2006 г. – 303 с. – С. 255-260
- [Алексеев, 2006-4] Алексеев, А.Ю. Тест Тьюринга. Роботы. Зомби. Пер. с англ. [Текст] / А.Ю. Алексеев - [ред.] - М.: МИЭМ, 2006. – 112 с.
- [Алексеев 2006-5] Алексеев, А.Ю. Роль Теста Тьюринга в методологии искусственного интеллекта //Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 26 апр 2006 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2006 г.. URL : <http://www.scmi.ru/seminar/1-11>
- [Алексеев, 2007] Алексеев, А.Ю. Когнитивно-компьютерные модели сознания и тест Тьюринга [Текст] / А.Ю. Алексеев // Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара. Под ред. Е.В. Серёдкиной. – Пермь: Изд-во Перм.гос.техн.ун-та, 2007 г. – 210 с., С. 40-53
- [Алексеев, 2008-1] Алексеев, А.Ю. Методологические проблемы проекта искусственной личности // Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара. Под ред. Е.В. Серёдкиной. – Пермь: Изд-во Перм.гос.техн.ун-та, 2008 г. – 228 с., С. 50-74.
- [Алексеев, 2008-2] Алексеев, А.Ю. Определение философских зомби [Текст] / А.Ю. Алексеев // Философские науки, № 2, 2008, С. 126-149

- [Алексеев, 2008-3] Алексеев, А.Ю. Проблема другого в компьютерных коммуникациях [Текст] / А.Ю. Алексеев // Философские науки, № 6, 2008, С. 114-136
- [Алексеев, 2008-4] Алексеев, А.Ю. Трудности и проблемы проекта искусственной личности [Текст] / А.Ю. Алексеев // Журнал «Полигнозис», № 1, 2008 г., С. 20-44
- [Алексеев, 2009-1] Алексеев, А.Ю. Компьютер оценивает человека? Комплексный тест Тьюринга против комплексного компьютерного тестирования в вузовском образовании [Текст] / А.Ю. Алексеев // Полигнозис, 4(37), 2009, С.129-141
- [Алексеев, 2009-2] Алексеев, А.Ю. Электронную культуру — в жизнь! [Текст] / А.Ю. Алексеев // Полигнозис, 4(37), 2009, С.150-151
- [Алексеев, 2009-3] Алексеев, А.Ю. Понятие зомби и проблемы сознания [Текст] / А.Ю. Алексеев // Проблемы сознания в философии и науке / под ред. проф. Д.И. Дубровского. — М.: «Канон+» РООИ «Реабилитация», 2009. — 472 с., С.195-214
- [Алексеев, 2009-4] Алексеев, А.Ю. Электронная культура: трансляция в социокультурной и образовательной среде [Текст] / А.Ю. Алексеева, С.Ю. Карпук — [ред.] — М.: Изд-во МГУКИ, 2009. — 260с.
- [Алексеев, 2009-5] Алексеев, А.Ю. Исследования искусственного интеллекта — главный фактор развития электронной культуры [Текст] / А.Ю. Алексеев // Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук — М.: Изд-во МГУКИ, 2009. — 260 с. - С.30-36
- [Алексеев, 2009-6] Алексеев, А.Ю. Российский студент в «Китайской комнате». К вопросу о тестировании в системе образования [Текст] / А.Ю. Алексеев // Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук — М.: Изд-во МГУКИ, 2009. — 260 с. - С.152-157
- [Алексеев, 2009-7] Алексеев, А.Ю. Методологические контуры науки об электронной культуре [Текст] / А.Ю. Алексеев // Вузы культуры и искусств в мировом образовательном пространстве: новые пути наук о культуре: Материалы международной научно-практической конференции. Ч.2 (Москва, 28-29 мая 2009 года): сборник статей/ Научно-исследовательский институт МГУКИ. — М.: МГУКИ, 2009. — 246 с., С. 6-16
- [Алексеев, Бахтизин, 2009-8] Алексеев А.Ю., Бахтизин А.Р. Проблемы проекта искусственной личности //Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 03 июня 2009 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2009. URL : <http://www.scmair.ru/seminar/1-38>
- [Алексеев, 2010-1] Алексеев, А.Ю. Четыре лика электронной культуры [Текст] / А.Ю. Алексеев // Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук — М.: Изд-во МГУКИ, 2010. — 180 с., С. 50-68
- [Алексеев, 2010-2] Алексеев, А.Ю. Проблема сознания в электронной культуре [Текст] / А.Ю. Алексеев // Полигнозис, 3(39), 2010, С.129-141
- [Алексеев 2010-3] Алексеев, А.Ю. Роль нейрокомпьютера в электронной культуре [Текст] / А.Ю. Алексеев // Журнал: Нейрокомпьютеры: разработка и применение, № 8, 2010, С.14-26
- [Алексеев, 2010-4] Алексеев, А.Ю. (ред.) Электронная культура: феномен неопросветительства. [Текст] / А.Ю. Алексеева, С.Ю. Карпук — [ред.] — М.: Изд-во МГУКИ, 2010. — 180 с.
- [Алексеев, 2010-5] Алексеев, А.Ю. Когнитивные исследования в электронной культуре [Текст] / А.Ю. Алексеев // Материалы Второй международной научно-практической конференции "Актуальные проблемы современной когнитивной науки" — Иваново: Изд-во Ивановского госуниверситета, 2010. — 240 с. — С.102-108
- [Алексеев, 2011-1] Алексеев, А.Ю. Корсаков С.Н. — основатель когнитивной науки [Текст] / А.Ю. Алексеев, Фатуева Л.А. // Актуальные проблемы современной когнитивной науки. Материалы четвертой всероссийской научно-практической конференции с международным участием (20-21 октября 2011 года). Иваново: ОАО «Изд-во «Иваново»», 2011. — 350 с. — С. 124-131
- [Алексеев, 2011-2] Алексеев, А.Ю. Развитие проблематики сознания в электронной культуре [Текст] / А.Ю. Алексеев // Науки о культуре: инновации и приоритетные разработки: [сборник статей преподавателей МГУКИ] : в 2 ч./научн.ред. Л.С. Жаркова, В.М. Беспалов, В.Г. Иванова. — М.: МГУКИ, 2011. — Ч.1. — 176 с. - С.134-146

- [Алексеев, 2011-3] Алексеев, А.Ю. Роль комплексного теста Тьюринга в методологии искусственных обществ [Текст] / А.Ю. Алексеев // Ежеквартальный Интернет – журнал «Искусственные общества», Том 6, номер 1-4, I-IV квартал 2011 г. – С. 18-64
- [Алексеев, 2011-4] Алексеев, А.Ю. Проблемы электронной культуры: когнитивные исследования в образовательной среде [Текст] / А.Ю. Алексеев // Актуальные проблемы современной когнитивной науки. Материалы четвертой всероссийской научно-практической конференции с международным участием (20-21 октября 2011 года). Иваново: ОАО «Изд-во «Иваново»», 2011. – 350 с. – С. 160-170
- [Алексеев, 2011-5] Алексеев, А.Ю. Идея Комплексного Теста Тьюринга [Текст] / А.Ю. Алексеев // «Тест Тьюринга: философские интерпретации и практические реализации»: Материалы научно-практической конференции, посвящённой 60-летию публикации статьи Алана Тьюринга «Computing Machinery and Intelligence»; Под общ.ред. А.А. Костиковой, редакция и составление Н.Ю. Ключевой – М.: Алькор Паблишерс, 2011. – 136 с. – (Серия кафедры философии языка и коммуникации). – С. 8 - 19
- [Алексеев, 2011-6] Алексеев, А.Ю. Социокультурные параметры нейрокомпьютинга [Текст] / А.Ю. Алексеев // Полигнозис, 2(42), 2011, С.144-156
- [Алексеев, 2011-7] Алексеев, А.Ю. Нейрокомпьютер и творчество [Текст] / А.Ю. Алексеев, С.А. Игнатов, Т.А. Конькова // Журнал: Нейрокомпьютеры: разработка и применение, № 12, 2011, С. 3-18
- [Алексеев, 2011-8] Алексеев, А.Ю. Интеллектуализация электронной культуры – ключевое направление модернизации России [Текст] / А.Ю. Алексеев // Регионы России: Стратегии и механизмы модернизации, инновационного и технологического развития. Тр. Седьмой междунар. научн.-практ. конф. / РАН. ИНИОН. Отд. науч. сотрудничества и междунар. связей; Отв. ред. Ю.С. Пивоваров. – М., 2011. – Ч. 2. – 588 с. – С. 477 – 483.
- [Алексеев, 2012-1] Алексеев, А.Ю. Нейрокомпьютинг в комплексном тесте Тьюринга [Текст] / А.Ю. Алексеев, С.А. Игнатов, Т.А. Конькова // Журнал: Нейрокомпьютеры: разработка и применение, № 5, 2012, С.41-47 (1 п.л./0.5 п.л.)
- [Алексеев, 2012-2] Алексеев, А.Ю. Очерк о новом способе когнитивных исследований посредством корсаковских машин сравнения идей [Текст] / А.Ю. Алексеев // Воин, ученый, гражданин. К 225-летию со дня рождения С.Н. Корсакова. – М.: Техполиграфцентр, 2012. – 132 с. – С. 69-103 (2 п.л.)
- [Алексеев, 2012-3] Алексеев, А.Ю. Нейрокомпьютер и электронная культура [Текст] / А.Ю. Алексеев // Нейрокомпьютерная парадигма и общество / под ред. Ю.Ю. Петрунина. — М.: Издательство Московского университета, 2012. — 304 с. — (Научные исследования). – С. 107-133 (коллективная монография) (1.8 п.л.)
- [Алексеев, 2013-1] Алексеев, А.Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. [Текст] / А.Ю.Алексеев. – М.: ИИнтелл, 2013 г. – 304 с. – ISBN: 978-5-88387-090-2 (20 п.л.)
- [Алексеев, 2013-2] Алексеев, А.Ю. Проблема творчества в исследованиях искусственного интеллекта// Эпистемология креативности / Отв.ред. Е.Н. Князева. – М.: «Канон+» РООИ «Реабилитация», 2013. – 512 с. – С. 374-414 (колл. монография) (2 п.л.)
- [Алексеев, 2013-3] Алексеев, А.Ю. Идеи эволюционной эпистемологии в построении искусственного интеллекта// Эволюционная эпистемология: современные дискуссии и тенденции. Е.Н. Князева. – Рос.акад.наук, Ин-т философии; Отв.ред. Е.Н. Князева. – М.: ИФ РАН, 2012. – 236 с. – С. 209-231 (коллективная монография) (1 п.л.)
- [Алексеев, 2013-4] Алексеев, А.Ю. Протонейрокомпьютер Корсакова [Текст]// А.Ю.Алексеев // Нейрокомпьютеры: разработка и применение, № 7, 2013. – С. 6-17. (1.2 п.л.)
- [Алексеев, 2013-5] Алексеев, А.Ю. Комплексная, коннекционистско-репрезентативная интерпретация машины Корсакова//Материалы Шестой международной научно-практической конференции «Актуальные проблемы современной когнитивной науки» – Иваново: ОАО «Издательство Иваново», 2013. – 330 с. – С.53-64 (1 п.л.)
- [Алексеев, 2014-1] Alekseev, A.Yu. Philosophy of Artificial Intelligence: neurocomputing realizers of cognitions / A.Yu. Alekseev // Neurocomputers, #4, 2014, p.9 (0.2 п.л.)

- [Алексеев, 2014-2] Alekseev, A.Yu. Computing dance vs Neurocomputing dance / A.Yu. Alekseev, Badmaeva T.B. // Neurocomputers, #4, 2014, p.71 (0.2 п.л.)
- [Алексеев, 2014-3] Alekseev, A.Yu. Connections as an engineering methodology of biomedical radioelectronics / A.Yu. Alekseev // Biomedicine Radioengineering, #4, 2014, p.10 (0.2 п.л.)
- [Алексеев, 2014-4] Алексеев, А.Ю. Новейший философский поворот в нейронауках [Текст] // А.Ю.Алексеев, А.В.Савельев, Е.А.Янковская // Журнал «Философия науки», 1(60), 2014. – С.114-135. (1.5 п.л./1 п.л.)
- [Алексеев, 2014-5] Алексеев, А.Ю. Нейрофилософия [Текст] // А.Ю.Алексеев, Ю.Ю.Петрунин, А.В.Савельев, А.В.Чечкин, Е.А.Янковская // Международный научно-технический журнал «Нейрокомпьютеры: разработка и применение», № 8, 2014, С.6-15 (0.5 п.л./ 1.5 п.л.)
- [Алексеев, 2014-6] Алексеев А.Ю. Нейрофилософия: впервые в России [Текст] // А.В.Чечкин, А.Ю.Алексеев, Ю.Ю.Петрунин, А.В.Савельев, А.В.Чечкин, Е.А.Янковская / Нейрокомпьютеры: разработка и применение. – Москва. – Радиотехника. – 2014. – №10. – С. 58-69.
- [Алексеев, 2014-7] Алексеев, А.Ю. Направления исследований электронной культуры [Текст] // А.Ю.Алексеев, Е.А.Янковская / Полигнозис, 3-4(47), 2014, С.93-106. (0.9/1.4 п.л.).
- [Алексеев, 2014-8] Алексеев, А.Ю. Концепт интерсубъективности в когнитивных моделях искусственного интеллекта //Интерсубъективность в науке и технике / Под ред.Н.М.Смирновой. – М.: «Канон+» РООИ «Реабилитация», 2014. – 416 с. – С.142-182. (2.5 п.л.)
- [Алексеев, 2014-9] Алексеев, А.Ю. Когнитологические проекты искусственной личности [Текст] //Человек: Образ и сущность. Гуманитарные аспекты. Ежегодник. – Л.В.Скворцов [редактор-составитель] – М.: Человек. Гуманитарное знание, 2014. – 400 с. – С. 156-174. (1 п.л.)
- [Алексеев, 2014-10] Алексеев, А.Ю. 150 лет «Рефлексам головного мозга» [Текст] / А.Ю.Алексеев, Ю.Ю. Петрунин, А.В. Савельев, Е.А. Янковская - отв.ред. – Сборник научных трудов, посвященных изданию статьи И.М.Сеченова (23 ноября 1863 г.) / – М.: ИИнтелл, 2014. – 432 с. – ISBN 978-5-98956-006-6 – С. 6 – 26 (25 п.л./ 15 п.л.)
- [Алексеев, 2014-11] Алексеев, А.Ю. Функционализм машины Сеченова vs функционализм машины Тьюринга [Текст]// А.Ю. Алексеев / 150 лет «Рефлексам головного мозга». Сборник научных трудов, посвященных изданию статьи И.М.Сеченова (23 ноября 1863 г.) / Отв.ред.А.Ю. Алексеев, Ю.Ю.Петрунин, А.В.Савельев, Е.А.Янковская. – М.: ИИнтелл, 2014. – 432 с. – С. 138 – 185. (2.4 п.л.)
- [Алексеев, 2014-12] Алексеев, А.Ю. Тестовый компьютеризм как высшая форма функционализма [Текст] //А.Ю.Алексеев /Актуальные проблемы современной когнитивной науки. Материалы седьмой всероссийской научно-практической конференции с международным участием (16-18 октября 2014 года). Иваново: ОАО «Изд-во «Иваново»», 2014. – 130 с. – С. 9-11. (0.7 п.л.)
- [Алексеев, 2014-13] Алексеев, А.Ю. Когнитивно-антропологические проблемы исследования электронной культуры [Текст] // А.Ю.Алексеев /Вестник гуманитарного факультета Ивановского государственного химико-технологического университета, выпуск 7, 2014. - С.35-40 (0.7 п.л.)
- [Алексеев, 2014-15] Алексеев, А.Ю. Электронная культура в контексте постнеклассической методологии [Эл.ресурс]//А.Ю.Алексеев / Электронный научный журнал «Культура: теория и практика», N 2, 2014; [URL] - <http://theoryofculture.ru/issues/29/623/>(1п.л.)
- [Алексеев, 2015-1] Алексеев, А.Ю. Актуальные вопросы нейрофилософии [Текст] / А.Ю.Алексеев, В.Г.Кузнецов, Ю.Ю.Петрунин, А.В.Савельев, Е.А.Янковская // Нейрокомпьютеры: разработка, применение. – Москва. – Радиотехника. – 2015 № 4, с. 9-11 (0.1 п.л./0.6 п.л.).
- [Алексеев, 2015-2] Алексеев А.Ю. Нейрофилософия как концептуальная основа нейрокомпьютинга / А.Ю.Алексеев, Ю.Ю.Петрунин, А.В.Савельев, А.В.Чечкин, Е.А.Янковская // Нейрокомпьютеры: разработка, применение. – Москва. – Радиотехника. – 2015. – №5. – С. 69-77 (0.4 п.л./0.8 п.л.).

- [Алексеев, 2015-3] Алексеев, А.Ю. Всероссийский междисциплинарный семинар «Нейро-философия» [Текст] // А.Ю.Алексеев / Нейрокомпьютеры: разработка, применение. – Москва. – Радиотехника. – 2015. – №5. – С. 100 (0.1 п.л.).
- [Алексеев, 2015-4] Алексеев, А.Ю. Нейрокомпьютерный тест Черчлендов [Текст] // А.Ю.Алексеев / Нейрокомпьютеры: разработка, применение. – Москва. – Радиотехника. – 2015. – №5. – С. 12-13 (0.2 п.л.).
- [Алексеев, 2015-5] Алексеев, А.Ю. 3D-семантика словаря когнитивной биомедицины и машина Корсакова–Тьюринга [Текст] // А.Ю.Алексеев / Журнал «Биомедицинская радио-электроника» – 2015.– № 4. – С. 9-11 (0.2 п.л.).
- [Алексеев, 2015-6] Алексеев, А.Ю. Перспективы развития электронной культуры России [Текст] // А.Ю.Алексеев / Россия: тенденции и перспективы развития. Ежегодник. Вып. 10. Часть II. Редкол.: Пивоваров Ю.С. (отв. ред.) и др. – М.: ИНИОН РАН, 2015. – 644 с. – С.398-410.
- [Алексеева, И.Ю. 1992] Алексеева И.Ю. Человеческое знание и его компьютерный образ. – М.: Наука, 1992. – 209 с.
- [Антопольский, 2010] Антопольский А.Б. Принципы создания единого информационного пространства вуза//Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С. 82-86
- [Баксанский, 2011] Баксанский О.Е. Когнитивная наука и проблемное поле философии образования // Актуальные проблемы современной когнитивной науки. Материалы четвёртой всероссийской научно-практической конференции (с международным участием (20-21 октября 2011 года). Иваново: ОАО "Изд-во "Иваново", 2011. – 350 с. – С. 12-25
- [Баксанский, Кучер, 2005] Баксанский О.Е., Кучер Е.Н. Моделирование социального интеллекта в когнитивных науках //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.120-124
- [Барендрегт, 1985] Барендрегт, Х. Ламбда-исчисление. Его синтаксис и семантика: Пер. с англ. — М.: Мир, 1985. — 606 с.
- [Бахтизин, 2005] Бахтизин А.Р. Моделирование социально-экономической системы с помощью технологии искусственного интеллекта //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.232-237
- [Бахтизин, 2006] Бахтизин А.Р. Искусственный интеллект и его приложения в экономико-математических моделях //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.347-378
- [Бахтизин, 2008] Бахтизин А.Р. (2008) Агент-ориентированные модели экономики. М.: Экономика, 2008. – 304 с.
- [Бахтизин, 2009] Бахтизин А.Р. Подход к разработке агент–ориентированных моделей// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.
- [Бахтизин, 2011] Бахтизин А.Р. Агентное моделирование на базе технологий искусственного интеллекта// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 281-310
- [Бескова, 2005] Бескова И.А. Вариант прочтения интенциональной семантики //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.190-192
- [Бирюков, 1989] Бирюков Б.В. Кибернетика, информатика, вычислительная техника, автоматика: проблемы становления и развития. Вклад течественной науки// Кибернетика: прошлое для будущего. Этюды по истории отечественной кибернетики. Теория управления. Автоматика. Биокибернетика. М., 1989.
- [Бирюков, 2004] Бирюков Б.В. Жар холодных числ и пафос бесстрастной логики. Формализация мышления от античных времен до эпохи кибернетики. Изд. 3-е, переработанное и дополненное. М., 2004.
- [Бодякин, 2005] Бодякин В.И. Концепция построения искусственного разума на базе нейро-семантического подхода //Новое в искусственном интеллекте. Методологические и теоре-

- тические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.192-195
- [Брингсйорд и др., 2006] С.Брингсйорд, П.Беллоу, Д.Феруччи. Творчество, тест Тьюринга и улучшенный тест Лайвлейс. Пер.с англ. А. Ласточкина в кн. //Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 129 с. – С. 62-83 – 120 с.
- [Брукшир, 2004] Брукшир, Дж. Информатика и вычислительная техника. 7-е изд. - СПб.: Питер, 2004. - 620 с: ил.
- [Будущее ИИ, 1991] Будущее искусственного интеллекта. - М.: Наука, 1991. – 302 с.
- [Булычев, 2000-1] Булычев И.И. Категория "субъект" в свете методологии универсального логико-философского алгоритма. Вестник Тамбовского университета. Серия: Гуманитарные науки. 2000. № 2. С. 3-18.
- [Булычёв, Кирюшин, 2007] Булычёв И.И., Кирюшин А.Н. О базисных логико-философских константах теории коммуникации. Философия и общество. 2011. № 4. С. 97-116.
- [Булычев, Шутов, 2002] Булычев И.И., Шутов Р.В. О сущности естественного и искусственного интеллектов. Вестник Тамбовского университета. Серия: Гуманитарные науки. 2002. № 1. С. 21-26.
- [Васюков, 2005]. Васюков В.Л. Квантовая логика. - М.: ПЕР СЭ, 2005. - 192 с.
- [Васюков, 2006] Васюков В.Л. Формальная онтология и искусственный интеллект //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.305-324
- [Вельтман, 2002] Вельтман Ким. Электронная культура: достижения и перспективы. Пер. с англ. Н.В. Браkker //Информационное общество, 2002, вып. 1, С. 24-30.
- [Вильянуэва, 2006] Вильянуэва Э. Что такое психологические свойства? Метафизика психологии. М.: Идея-Пресс, 2006. – 256 с.
- [Вильянуэва, 2006] Вильянуэва Э. Что такое психологические свойства? Метафизика психологии. М.: Идея-Пресс, 2006. – 256 с.
- [Винер, 1983] Винер, Н. Кибернетика, или управление и связь в животном и машине. 2-е издание. М.: Наука; Главная редакция изданий для зарубежных стран, 1983. – 344 с.
- [Глазунов, 2010] Глазунов В.А. Методологические проблемы робототехники. Междисциплинарность, бифуркации, многокритериальность. Антропоморфные и параллельные роботы//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 20 янв. 2010 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2010 г.. URL : <http://www.scmairu.ru/seminar/1-41>
- [Глазунов, Алексеев 2007] Глазунов В.А., Алексеев А.Ю. Методологические вопросы робототехники//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 25 апр. 2007 г., М:ЦЭМИ РАН / НСМИИ РАН, 2007 г.. URL : <http://www.scmairu.ru/seminar/1-19>
- [Глазунов, Глазунова, 2005] Глазунов В.А., Глазунова О.В. Моделирование творческого акта средствами робототехники //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.91-92.
- [Горбатов, 2012] Горбатов В. В. Логико-онтологические предпосылки двумерной семантики // Известия Уральского государственного университета. Серия 3: Общественные науки. 2012. Т. 100. № 1. С. 37-44.
- [Грэхэм, 1991] Грэхэм Л.Р. Естествознание, философия и науки о человеческом поведении в Советском Союзе: Пер. с англ. – М. Политиздат, 1991. – 480 с.
- [Гук, 2010] Гук А.А. Утилитарно-практическое и художественно-эстетическое в экранном творчестве//Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелЛЛ, 2010. – 228 с. – С. 133-139
- [Декарт, 1989] Декарт Р. Сочинения в 2 т.: Пер. с лат. и франц. Т. I/Сост., ред., вступ. ст. В. В. Соколова. - М.: Мысль, 1989. – 654 с. – С. 315 (см. «Первоначала философии» - С.297-422, пер. с лат. С. Я. Шейнман-Топштейн, с франц. Н. Н. Сретенского)
- [Делицын, 2010] Делицын Л.Л. Элементы префигуративного типа культуры в феномене неопросветительства: способ преодоления внутреннего цифрового разрыва//Электронная

- культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С. 158-163
- [Деннетт, 2004] Деннетт Дэниел С. Виды психики: на пути к пониманию сознания. Перевод с англ. А. Веретенникова. Под общ.ред. Л.Б. Макеевой. – М.: Идея-Пресс, 2004. – 184 с. – С.
- [Дмитриевская, 1995] Дмитриевская И.В. Принцип интенциональности в логике Г.Фреге // Философия языка и семиотика. Иваново, 1995
- [Дубровский, 1971] Дубровский Д.И. Психические явления и мозг. Философский анализ проблемы в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики. М., «Наука», 1971
- [Дубровский, 1980] Дубровский Д.И. Информация, сознание, мозг. М., «Высшая школа», 1980
- [Дубровский, 1994-1] Дубровский Д.И. Психика и мозг. Результаты и перспективы исследований. // Мозг и разум. М.: Наука, 1994 – 176 с.
- [Дубровский, 1994-2] Дубровский Д. И. Обман. Философско-психологический анализ. М., 1994
- [Дубровский, 2002-1] Дубровский Д.И. Проблема идеального. Субъективная реальность. — М., 2002.— 368 с.
- [Дубровский, 2002-2] Дубровский Д.И. Новая реальность: человек и компьютер / Полигнозис № 3 (23), С. 20-32
- [Дубровский, 2003-1] Дубровский Д.И. Проблема духа и тела: возможности решения (в связи со статьей Т.Нагеля «Мыслимость невозможного и проблема духа и тела»). Вопросы философии, 2003 г.
- [Дубровский, 2003-2] Дубровский Д.И. Новое открытие сознания? (По поводу книги Дж. Серла «Открывая сознание заново») // Вопросы философии, 2003, №7
- [Дубровский, 2004-1] Дубровский Д.И. Тест Тьюринга, парадигма функционализма и проблема сознания// Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции “Философия искусственного интеллекта”, МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексеева – М.: МИЭМ-ИИнтелл, 2006. – 224 с. – С. 126-128 – С. 109
- [Дубровский, 2004-2] Дубровский Д.И. Гносеология субъективной реальности. К постановке проблемы // Эпистемология и философия науки, 2004, № 2.
- [Дубровский, 2005-1] Дубровский Д.И. 2005. Искусственный интеллект и проблема сознания //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.26-32
- [Дубровский, 2005-2] Дубровский Д.И. Сознание, мозг, искусственный интеллект//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 18 мар 2005 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2005 г.. URL : <http://www.scmairu/seminar/1-1>
- [Дубровский, 2005-3] Дубровский Д.И. 2005. Искусственный интеллект и проблема сознания //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелл, 2005. – 280 с. – С.42-48
- [Дубровский, 2006-1] Дубровский Д.И. Бессознательное (в его отношении к сознательному) и квантовая механика. Философские науки, № 8, 2006. – С. 52 - 73
- [Дубровский, 2006-2] Дубровский Д.И. 2006. Приветственное слово участникам конференции //Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции «Философия искусственного интеллекта», МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 192 с. – С.8-11
- [Дубровский, 2006-3] Дубровский Давид 2006. Тест Тьюринга, парадигма функционализма и проблема сознания. Подведение итогов работы секции № 1 //Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции «Философия искусственного интеллекта», МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 192 с. – С.93-97

- [Дубровский, 2006-4] Дубровский Д.И. Сознание, мозг, искусственный интеллект //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.75-89
- [Дубровский, 2007] Дубровский Д. И. Сознание, мозг, искусственный интеллект: сб. статей. – М.: 2007. – 272 с.
- [Дубровский, 2009-1] Дубровский Д.И. (ред.) Проблемы сознания в философии и науке. Под ред. проф. Д.И. Дубровского. – М.: «Канон+» РООИ «Реабилитация», 2009. – 472 с., С.195-214
- [Дубровский, 2009-2] Дубровский Д.И. Проблема сознания: опыт обзора основных вопросов и теоретических трудностей //«Проблемы сознания в философии и науке» / под ред. проф. Д.И. Дубровского. – М.: «Канон+»РООИ «Реабилитация», 2009. – 472 с., С.8-44
- [Дубровский, 2009-3] Дубровский Д.И. Мозг: фундаментальные и прикладные проблемы. Философские подходы к проблеме «мозг и психика»: к вопросу о расшифровке нейродинамических кодов явлений субъективной реальности // Доклад на научной сессии Общего собрания Российской Академии Наук 15 декабря 2009 г., Президиум РАН, г.Москва
- [Дубровский, 2010] Дубровский Д.И. Субъективная реальность, мозг и развитие НБИК-конвергенции: эпистемологические проблемы// Эпистемология вчера и сегодня/ Рос.акад.наук, Ин-т философии; Отв.ред. В.А. Лекторский. - М.: ИФРАН, 2010. – 188 с. – С. 78
- [Дубровский, 2011-1] Дубровский Д.И. Актуальные проблемы интерсубъективности// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 129-148
- [Дубровский, 2011-2] Дубровский Д.И. Трудная проблема сознания (в связи с книгой В.В.Васильева)// Вопросы философии, № 9, 2011, С. 136-148. URL : <http://elibrary.ru/item.asp?id=16860687>
- [Дубровский, Лекторский, 2006] Дубровский Д.И., Лекторский В.А. 2006. От редакторов //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.3
- [Дэвидсон, 1990] Дэвидсон Д. Что означают метафоры // Теория метафоры. М., 1990. С. 172-193. Пер. М.А. Дмитриевской. URL : <http://kant.narod.ru/davidson.htm>
- [Егоров, 2005] Егоров А.Г. Информационно-поисковая система экспертного типа по гуманитарным наукам как аналог машины Раймонда Луллия //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.339-343
- [Жижек, 2002] Жижек С., 2002. Обойдемся без секса, ведь мы же пост-люди! Пер. с англ. Артема Смирнова. URL : http://www.ruthenia.ru/logos/kofr/2002/2002_12.htm#_edn1#_edn1
- [Зотов, 2005] Зотов А.Ф. 2005. Что значит «мыслить»? //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.163-167
- [Зотов, 2006] Зотов А.Ф. Роль феноменологии в организации междисциплинарных исследований в области искусственного интеллекта //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.22-32
- [Иваницкий, 2006-1] Иваницкий А.М. Думающий мозг и искусственный интеллект – встречное движение // Труды XXV межрегиональной научно-технической конференции «Проблемы эффективности и безопасности функционирования сложных технических и информационных систем. Ч. 4. – Серпухов: Серпуховский ВИ РВ, 2006 г. – 303 с.
- [Иваницкий, 2006-2] Иваницкий А.М. Физиологические основы сознания и проблема искусственного интеллекта //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.90-99
- [Иванов, 1999] Иванов Е.М. К проблеме «вычислимости» функции сознания, 1999; <http://filosof.historic.ru/books/item/f00/s00/z0000899/index.shtml>
- [ИИ, 1990] Искусственный интеллект – В 3-х кн. Кн.2. Модели и методы: Справочник/Под ред. Д.А. Поспелова – М.: Радио и связь, 1990. – 304 с.

- [ИИ, межд.подход, 2006] Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с.
- [ИИ: философия, методология, инновации, 2006] Искусственный интеллект: философия, методология, инновации. Материалы Первой Всероссийской конференции студентов, аспирантов и молодых учёных, г. Москва, МИРЭА, 6-8 апреля 2006 г. Под ред. Д.И. Дубровского и Е.А. Никитиной - М.: ИИнтелЛЛ, 2006. – 440 с.
- [ИИС, 1991] Интеллектуальные информационные системы. М.: Военное издательство, 1991
- [Истратов, 2009] Истратов В.А. «Агентно-ориентированная модель поведения человека в социально-экономической среде». Автореферат диссертации, представленной на соискание ученой степени кандидата экономических наук по специальности 08.00.13. – Математические и инструментальные методы экономики. М: ЦЭМИ РАН, 2009
- [Калинин, 2011] Калинин П.Е. Становление статико-динамичного и континуально-дискретного единства в деятельности сознания (онтологический и гносеологический аспекты). Автореферат на соискание ученой степени кандидата философских наук по специальности 09.00.01 – онтология и теория познания. – Иваново: ИГУ, 2011.
- [Карнап, 2007] Карнап Р. Значение и необходимость. Исследование по семантике и модальное логике. Пер. с англ./ Общая редю Д.А.Бочвара. Предисл. С.А. Яновской. М.: Издательство ЛКИ, 2007. – 384 с.
- [Колин, 2005] Колин К.К. Философские основы информатики //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.248-252.
- [Колин, 2006-1] Колин К.К. Россия и мир на пути к информационному обществу//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 01 фев 2006 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2006 г.. URL : <http://www.scmai.ru/seminar/1-8>
- [Колин, 2006-2] Колин К.К. Становление информационного общества в России и национальная безопасность //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.379-392
- [Колин, 2010-1] Колин К.К. Системная модернизация России и проблемы развития информационного общества//Государственная служба, № 2, 2010, С.32-37
- [Колин, 2010-2] Колин К.К. Философские проблемы информатики. – М.: БИНОМ. Лаборатория знаний. 2010. – 264 с.
- [Колин, 2011] Колин К.К. Электронная культура в информационном обществе// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 224-252
- [Колмогоров, 1964] Колмогоров А. Автоматы и жизнь // Возможное и невозможное в кибернетике. М., 1964
- [Конькова, 2009] Конькова Т.А. Электронная культура? Критика тотальной вычислимости// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.
- [Корсаков (Михайлов), 2009] Корсаков С.Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / Пер. с франц. под ред. А.С. Михайлова. – М.: МИФИ, 2009. – 44 с.
- [Корсаков (Михайлов), 2009] Корсаков С.Н. Начертание нового способа исследования при помощи машин, сравнивающих идеи / Пер. с франц. под ред. А.С. Михайлова. – М.: МИФИ, 2009, 44 с.
- [Корсаков (Сыромятин), 2009] Корсаков С. Н. (1832 г.). Очерк о новом способе исследования посредством машин для сравнения идей. (Перевод с франц. А.В. Сыромятина) // Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.
- [Корсаков (Сыромятин), 2009] Корсаков С.Н. (1832 г.). Очерк о новом способе исследования посредством машин для сравнения идей. (Перевод с франц. А.В. Сыромятина) //Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: МГУКИ, 2009. – 260 с. - С.15-26.

- [Кочергин, 2005] Кочергин А.Н. Искусственный интеллект и мышление //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.37-39
- [Кочергин, 2006] Кочергин А.Н. 2006. Искусственный интеллект, психика, творчество //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.209-222
- [Кубрякова, 1996] Краткий словарь когнитивных терминов. Под общ.редакцией Е.С. Кубряковой. М.: МГУ, 1996 – 245 с.
- [Кувакин, 1998] Твой рай и ад. Человечность и бесчеловечность человека: Философия, психология и стиль мышления гуманизма. —СПб.: «Алетейя», —М.: «Логос», 1998. — 360 с.
- [Кудряшова, 2005] Кудряшова Т.Б. Онтология языков познания. Ч.I. - Иваново: ОАО "Изд-во "Иваново", 2005. - 316 с.
- [Кудряшова, 2010] Кудряшова Т.Б. Пути возрождения инженерной культуры через развитие творческого когнитивного потенциала субъектов высшего образования //Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С. 86-91
- [Кудряшова, 2011] Кудряшова Т.Б. Роль пространственного воображения в формировании и развитии способности к категоризации// Актуальные проблемы современной когнитивной науки. Материалы четвертой всероссийской научно-практической конференции (с международным участием (20-21 октября 2011 года). Иваново: ОАО "Изд-во "Иваново", 2011. - 350 с. - С. 32-36
- [Кузичев, 2006] Кузичев А.С. Программа Колмогорова, интеллектуальные системы и теоремы Гёделя о неполноте //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.330-346
- [Кузнецов, 1989] Кузнецов В.Е. Представление в ЭВМ неформальных процедур: производственные системы. М.: Наука. – 1989 – 160 с.
- [Кузнецов, 1991] Кузнецов В.Г. Герменевтика и гуманитарное познание. М.: Изд-во Моск.ун-та, 1991-192 с.
- [Кузнецов, Сергиенко, Холодная, 2006] Кузнецов О.П. Может ли искусственный интеллект развиваться независимо от исследований естественного интеллекта?//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 01 ноя 2006 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2006 г.. URL : <http://www.scmair.ru/seminar/1-14>
- [Лахути, 2009] Лахути Д.Г. О пяти кругах искусственного интеллекта и о дискуссии Поппера с Бернайсом // Вопросы философии, № 10, 2009, С.116-120. URL : <http://elibrary.ru/item.asp?id=13214944>
- [Лейбниц, 1982] Лейбниц. В.Г. Сочинения в четырех томах: Том I. / Ред. и сост., авт. вступит. статьи и примеч. В.В.Соколов; перевод Я.М.Боровского и др. – М.: «Мысль», 1982. – 636 с.
- [Лекторский, 2005] Лекторский В.А. 2005. Философия, когнитивные науки и искусственный интеллект //Философия и будущее цивилизации: Тезисы докладов и выступлений IV Российского философского конгресса (Москва, 24-28 мая 2005 г.): В 5 т. Т.1. - М.: Современные тетради, 2005. - 768 с. – С.728
- [Лекторский, 2006] Лекторский В.А. Философия, искусственный интеллект и когнитивная наука //Искусственный интеллект: междисциплинарный подход. Под ред.Д.И.Дубровского и В.А.Лекторского – М.: ИИнтелЛ, 2006.– 448 с.– С.12-22
- [Лекторский, 2010] Лекторский В.А. Эпистемологическое исследование когнитивных процессов. Эпистемология вчера и сегодня/ Рос.акад.наук, Ин-т философии; Отв.ред. В.А. Лекторский. - М.: ИФРАН, 2010. – 188 с.
- [Лекторский, 2010] Лекторский В.А. Эпистемологическое исследование когнитивных процессов. Эпистемология вчера и сегодня/ Рос.акад.наук, Ин-т философии; Отв.ред. В.А. Лекторский. - М.: ИФРАН, 2010. – 188 с.
- [Лекторский, 2011] Лекторский В.А. Исследование интеллектуальных процессов в современной когнитивной науке: философские проблемы// Естественный и искусственный ин-

- теллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 3-16
- [Лещёв, 2005] Лещёв С.В. Искусственное сознание и проблема интерпретации //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.143-145
- [Лещёв, 2010] Лещёв С.В. Археология электронной культуры: о харизме новых форм коммуникации//Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С.54-58
- [Линдсей, Норман, 1974] П. Линдсей, Д. Норман. Переработка информации у человека. Введение в психологию. Пер. с англ. Под ред. А.Р. Лурия. М.: Мир, 1974 – 550 с.
- [Лобовиков, 2005] Лобовиков В.О. Естественное право как математика свободы и морально-правовое программирование автономных адаптивных роботов //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.105-107
- [Макаров, 2006] Макаров В.Л. Получение нового знания методом компьютерного моделирования //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И.Дубровского и В.А.Лекторского – М.:ИИнтелЛЛ, 2006.– 448 с. – С.5-12.
- [Макаров, 2006] Макаров В.Л., Бахтизин А.Р., Бахтизина Н.В. (2006) CGE модель социально-экономической системы России со встроенными нейронными сетями. М: ЦЭМИ, 2006. – 240 с.
- [Макаров, 2010] Макаров В.Л. (2010) «Artificial Societies: A new Tool to understand how a Society works». Искусственные общества, Том 5, номер 1-4, I-IV квартал 2010.
- [Макаров, 2010] Макаров В.Л. Социальный кластеризм. Российский вызов / М.: Бизнес Атлас, 2010. – 272 с.
- [Макаров, 2011] Макаров В.Л. Становится ли человеческое общество стабильнее?// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 211-223
- [Макаров, Бахтизин, 2005] Макаров В.Л., Бахтизин А.Р. Компьютерное моделирование искусственных миров//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 20 апр 2005 г., г.Москва: ЦЭМИ РАН / НСММ РАН, 2005 г.. URL : <http://www.scmi.ru/seminar/1-2>
- [Макеева, 1996] Макеева Л.Б. Философия Х.Патнэма. – М., 1996. – 190 с.
- [Мальцев, 1970] Мальцев А.И. Алгебраические системы. М., Наука, Физматлит., 1970. - 392 с.
- [Марков, 1984] Марков А.А., Нагорный Н.М. Теория алгорифмов - М.: Наука, 1984
- [Мельчук, 1995] Мельчук И.А. Русский язык в модели «Смысл-Текст» - Москва-Вена: Школа «Языки русской культуры», 1995. – 682 с.
- [Меркулов, 2005] Меркулов И.П. Современная эпистемология: синтез информационных и эволюционных представлений //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.211-215
- [Меркулов, 2006] Меркулов И.П. Эпистемологический анализ когнитивной природы математических и логических формализмов //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.289-304
- [Микешина, Опенков, 1997] Микешина Л.А., Опенков М.Ю. Новые образы познания и реальности. – М.: Российская политическая энциклопедия (РОССПЭН), 1997. – 238 с.
- [Миронов, 2006] Миронов В.В. Информационное пространство: диалог культур //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.392-398
- [Миронов, 2011] Миронов В.В. Размышления о реформе российского образования: Доклад на международной научной конференции «Философия и образование в процессе трансформации культур», посвящённой 70-летию воссоздания философского факультета в структуре МГУ имени М.В. Ломоносова. – М.: Издатель Воробьев А.В., 2011. – 64 с.

- [Михеенкова, Финн, 2005] Михеенкова М.А., Финн В.К. Проблемы создания интеллектуальных систем для анализа социального поведения //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.366-368
- [Михеенкова, Финн, 2010] Михеенкова В.А., Финн В.К. Интеллектуальный анализ данных и проблемы когнитивной социологии//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 24 фев 2010 г., г.Москва: ЦЭМИ РАН / НСММИ РАН, 2010 г.. URL : <http://www.scmair.ru/seminar/1-42>
- [Мищенко, 2012] Мищенко В.С., Бондаренко Г.Н., Фатуева Л.А. Семён Николаевич Корсаков – страницы жизни и судьбы//Воин, ученый, гражданин. К 225-летию со дня рождения С.Н. Корсакова. – М.: Техполиграфцентр, 2012. – 132 с. – С. 18-33
- [Моисеев, 2005] Моисеев В.И. Интервал Тьюринга и имитация интеллекта //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2005. – 280 с. – С.215-217
- [Нагуманова, 2007] Нагуманова С.Ф. Аргумент мыслимости против материализма// Философские науки, № 3, 2007 г.
- [Нариньяни, 2005-1] Нариньяни А.С. Путь России в информационной технологии XXI века. // Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2005. – 280 с. – С. 11-25
- [Нариньяни, 2005-2] Нариньяни А.С. еНОМО – два в одном (Homo Sapience в ближайшей перспективе) //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2005. – 280 с. – С.259-278
- [Нариньяни, 2005-3] Нариньяни А.С. Перспективы развития интеллектуальных информационных технологий нового поколения//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 26 май 2005 г., г.Москва: ЦЭМИ РАН / НСММИ РАН, 2005 г.. URL : <http://www.scmair.ru/seminar/1-3>
- [Нариньяни, 2006] Нариньяни А.С. Между Знанием и Незнанием – Наивная Топография 2 //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2006. – 448 с. – С.49-74
- [Нитусов, 2012] Нитусов А.Ю. Семён Николаевич Корсаков//Воин, ученый, гражданин. К 225-летию со дня рождения С.Н. Корсакова. – М.: Техполиграфцентр, 2012. – 132 с. – С. 104-124.
- [Новое в ИИ, 2005] Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2005. – 280 с.
- [Огурцов, 2005] Огурцов А.П. 2005. Достижения и трудности в моделировании интеллектуальных актов //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.56-60
- [Огурцов, 2006] Огурцов А.П. Возможности и трудности в моделировании интеллекта //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИНТЕЛЛ, 2006. – 448 с. – С.32-48
- [Осипов, 2010] Осипов Г.С. Основные направления и тенденции развития работ в области искусственного интеллекта в Европе//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 29 окт 2010 г., г.Москва: ЦЭМИ РАН / НСММИ РАН, 2010 г.. URL : <http://www.scmair.ru/seminar/1-44>
- [Пассмор, 2002] Пассмор, Дж. Современные философы. Перевод с англ. Л.Б.Макеевой. - М.: Идея-Пресс, 2002. - 192 с.
- [Патнэм, 1999] Патнэм Хилари. Философия сознания. Перевод с англ. Макеевой Л. Б., Назаровой О. А., Никифорова А. Л.; предисл. Макеевой Л. Б. — М.: Дом интеллектуальной книги, 1999. — 240 с.

- [Пенроуз, 2005] Пенроуз Р. Новый ум короля: о компьютерах, мышлении и законах физики: Пер. с англ./ Общ.ред. В.О. Малышенко – М.: Едиториал УРСС, 2005 – 400 с.
- [Петренко, Осипов, 2005] Петренко В.Ф., Осипов Г.С. Междисциплинарный подход к исследованию сознания и личности: экспериментальная психосемантика//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 05 окт 2005 г., г.Москва: ЦЭМИ РАН / НСММИ РАН, 2005 г.. URL : <http://www.scmairu/seminar/1-5>
- [Петрунин, 1996] Петрунин Ю.Ю. От тайного знания к нейрокомпьютеру: Очерки по истории искусственного интеллекта. М.: Университетский гуманитарный лицей. 1996
- [Петрунин, 2001] Петрунин Ю.Ю. Искусственный интеллект//Новая философская энциклопедия в 4 томах. М. «Мысль». 2001. Т. 2.
- [Петрунин, 2002] Петрунин Ю.Ю. Искусственный интеллект: история, методология, философия. М.: Звездапад. 2002
- [Петрунин, 2010] Петрунин Ю.Ю. Модели и методы искусственного интеллекта в управлении знаниями// Государственное управление. Электронный вестник. Выпуск № 24. Сентябрь 2010 (в соавторстве). URL : http://e-journal.spa.msu.ru/24_2010Petrinin_Petrulina.html
- [Петрунин, 2010] Петрунин Ю.Ю., Рязанов М.А., Савельев А.В. Философия искусственного интеллекта в концепциях нейронаук. М.: МАКС Пресс, 2010.–83с.
- [Петрунин, Алексеев, 2011] Петрунин Ю.Ю., Алексеев А.Ю. Искусственный интеллект как феномен современной культуры//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 08 дек 2011 г., г.Москва: ИФ РАН / НСММИ РАН, 2011 г.. URL : <http://www.scmairu/seminar/1-50>
- [Поваров, 2005-1] Поваров Г.Н. С.Н. Корсаков – русский пионер искусственного разума //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.198-199
- [Поваров, 2005-2] Поваров Г.Н. Диалектический материализм и проблема искусственного интеллекта //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.196-198
- [Полосухин, 1993] Полосухин Б.М. Феномен вечного бытия. Некоторые итоги размышлений по поводу алгоритмической модели сознания. – М.:Наука, 1993 – 176 с.
- [Портнов, 1994] Портнов, А.Н. Язык и сознание: основные парадигмы исследования проблемы в философии XIX-XX вв. Издательство: Иваново: Ивановский Государственный Университет, 1994. - 336 с.
- [Поспелов, 1982] Поспелов Д.А. Фантазия или наука: на пути к искусственному интеллекту – М.: Наука, 1982. – 224 с., С. 5- 42
- [Рассел, 1999] Рассел Бертран. Исследование истины и значения. Пер. Ледникова Е.Е., Икифорова А.Л. – М.: Идея-Пресс, Дом интеллектуальной книги, 1999.–440с.
- [Рассел, Норвиг 2006] Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд. – М.: Вильямс, 2006. - с. 1408.
- [Реале, Антисери, 1997] Реале Дж., Антисери Д. Западная философия от истоков до наших дней. Том 2. Средневековье. – ТОО ТК «Петрополис», 1997. – 368 с. – С.206 – 207.
- [Редько, 2006] Редько В.Г. Задача моделирования когнитивной эволюции //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛ, 2006. – 448 с. – С.243-258
- [Редько, 2009] Редько В.Г. Об исследовании когнитивной эволюции// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. – С. 99-104
- [Резник, 1999-1] Резник Ю. М. Введение в социальную теорию. Социальная онтология. Пособие. — М.: Институт востоковедения РАН, 1999. — 514с.
- [Резник, 1999-2] Резник Ю. М. Введение в социальную теорию. Социальная эпистемология. Пособие. — М.: Институт востоковедения РАН, 1999. — 327 с.
- [Резник, 2003] Резник Ю. М. Введение в социальную теорию: Социальная системология / Ю. М. Резник; Ин-т человека. — М.: Наука, 2003. — 525с.

- [Розин, 2001] Розин Вадим. Семиотические исследования. – М.: ПЕР СЭ; СПб: Университетская кни-га, 2001 - 256 с.
- [Розин, 2005] Розин В.М. 2005. Мышление и искусственный интеллект //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.157-164
- [Розин, 2011] Розин В.М. Концепция искусственного интеллекта и общество как эвристический источник идей для нейрокомпьютера нового поколения// Нейрокомпьютер, № 1, 2011 г., С.4-10
- [Рыжов, 2005] Рыжов В.П. Художественное творчество как информационный процесс //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.254-267
- [Савельев, 2004] Савельев А.В. Нейрокомпьютеры в изобретениях // Нейрокомпьютеры: разработка и применение, 2004, № 2-3, С. 33-49
- [Савельев, 2005] Савельев А.В. Искусственный интеллект или искусственный социум? //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.210-212
- [Савельев, 2006] Савельев А. В. Самоорганизационное нейроуправление на реверберационных нейропроцессах // в сб. тр. МИФИ: "Нейроинформатика-2006", М.: МИФИ, 2006, ч. I, с. 172-180
- [Савельев, 2008] Савельев А.В. Нейрологические аспекты клеточной нейроматематики // журнал "Искусственный интеллект", НАН Украины, Донецк, 2008, № 4, с. 612-623
- [Савельев, 2009] Савельев А.В. Нейроэкзистенциальное моделирование распространения спайков – путь к новой концепции субстрата долговременной нейронной памяти// журнал «Искусственный интеллект» НАН Украины, г. Донецк, 2009 г. С.329-335
- [Савельев, 2010] Савельев А.В. Каузальность информационной культуры и «нейрокомпьютеризация» сознания//Электронная культура: феномен неопросветительства. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: ИИнтелл, 2010. – 228 с. – С.46-53
- [Самохвалова, 2005] Самохвалова В.И. Предполагает ли машинное творчество возможность машинного вдохновения? //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.62-66.
- [Самохвалова, 2006] Самохвалова В.И. «Человеческое, слишком человеческое», или процессор в экстазе //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.423-442.
- [Самохвалова, 2007] Самохвалова В.И. Творчество: божественный дар; космический принцип; родовая идентичность человека: Научное издание. – М.: РУДН, 2007. – 538 с.
- [Сергиенко, Холодная, 2006] Сергиенко Е.А., Холодная М.А. Истоки познания: революция в когнитивной психологии развития//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 29 мар 2006 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2006 г.. URL : <http://www.scmairu/seminar/1-10>
- [Серёдкина, 2007] Серёдкина Е.В. Общая теория рациональности и артефакты в проекте OSCAR Дж. Поллока. // Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара. Под ред. Е.В. Серёдкиной. – Пермь: Изд-во Перм.гос.техн. ун-та, 2007. – 210 с. – С. 108-122.
- [Серль, 2002] Серль Дж. Открывая сознание заново. Пер. с англ. А.Ф. Грязнова. М.: 2002. – 256 с.
- [Серль, 2006] Серль Джон Р. Разумы, мозги, программы (Перевод с англ. Д. Родионова) // Тест Тьюринга. Роботы. Зомби. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с. – С.6–20.
- [Сеченов, 1863] Сеченов И.М. Рефлексы головного мозга/Медицинский вестник, 1863, № 47, стр. 461—484; № 48, стр. 493 — 512. (Текст приводится по кн.: Сеченов И.М. Избран-

- ные произведения. Том первый. Физиология и психология. Редакции и послесловие Х.С.Коштыяца. М.: Издательство Академии Наук СССР, 1952 – 753 с. – С. 7-128).
- [Сеченов, 2001] Сеченов И.М. Элементы мысли// Сеченов И. М. Элементы мысли. — СПб.: Питер, 2001. — 416 с. — (Серия « Психология-классика») – С. 209-355.
- [Скворцов, 2005] Скворцов А.А. Нравственные ценности виртуального мира //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.164-166.
- [Скворцов, 2006] Скворцов А.А. 2006. Мировоззренческие и теоретические основы этических исследований виртуальности //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2006. – 448 с. – С.411-433.
- [Сломэн, 2006] Сломэн Арон. 2006. Что значит быть камнем? (Перевод В. Крючкова) //Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с. – С.86-102.
- [Смирнов Д.Г., 2008] Смирнов Д.Г. Семиософия ноосферного универсума [Текст] : ноосфера и семиосфера в глобальном дискурсе / Д. Г. Смирнов ; [науч. ред. А. Н. Портнов]; М-во образования и науки Рос. Федерации, Федер. агентство по образованию, ГОУ ВПО "Иван. гос. ун-т". - Иваново : Иван. гос. ун-т, 2008. - 370 с.
- [Смирнова, 2005-1] Смирнова Н.М. Трансцендентально-феноменологическая теория конституирования смыслов и ее эвристические возможности для философии искусственного интеллекта //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.217-220.
- [Смирнова, 2005-2] Смирнова Н.М. Методология феноменологического конституирования и её значение для философии искусственного интеллекта //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛ, 2005. – 280 с. – С.166-169
- [Смирнова, 2006] Смирнова Н.М. Трансцендентальная intersубъективность, проблема «чужих сознаний», искусственный интеллект// В кн. Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского. М.: ИИнтелЛ, 2006. – 448 с. – С. 163 – 179.
- [Смирнова, 2011] Смирнова Н.М. Intersубъективность и проблема «Других сознаний»// Естественный и искусственный интеллект: методологические и социальные проблемы/ Под ред. Д.И. Дубровского и В.А. Лекторского. – М., 2011. – 352 с. – С. 106-128.
- [Стёпин, 1991] Стёпин В.С. Научная рациональность в гуманистическом измерении. // В кн. О человеческом в человеке/ Под общ.ред. И.Т.Фролова – М., 1991. – 384 с.
- [Стёпин, 1992] Стёпин В.С. Философская антропология и философия науки. - М.: Высшая школа, 1992. - 191 с.
- [Стёпин, 2007] Стёпин В.С. Постнеклассическая рациональность и информационное общество//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 31 окт 2007 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2007 г.. URL : <http://www.scmair.ru/seminar/1-22>
- [Тест Тьюринга, 2006] Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с.
- [Трушкина, 2005] Трушкина Н.Ю., Михейкин И.В., Кундин Л.В. Решение проблемы сознания в рамках концепции сильного искусственного интеллекта //Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с. – С.79-82
- [Трушкина, 2006] Трушкина Н.Ю. «Отношение теоретических концепций и компьютерных моделей в исследованиях искусственного интеллекта». Диссертация на соискание ученой степени кандидата философских наук по специальности 09.00.08. – философия науки и техники. М.: МГУ им. М.В. Ломоносова
- [Тьюринг, 1960] Алан М. Тьюринг. Может ли машина мыслить? (Под ред. Б.В. Бирюкова). М., 1960. С. 19-158
- [Уёмов, 1963] Уёмов А.И. Вещи, свойства и отношения. М., 1963

- [Уитби, 2004] Уитби Б. Искусственный интеллект: реальна ли Матрица. – М.: ФАИР-ПРЕСС, 2004. – с. 224.
- [Управление информация интеллект, 1976] Управление, информация, интеллект. Под ред. А.И.Берга и др. М.: Мысль, 1976.–383 с.– С. 282.
- [Философия и будущее цивилизации, 2005] Алексеев, А.Ю., Глазунов, В.А., Лекторский В.А. (ред.). Круглый стол «Философско-методологические проблемы когнитивных и компьютерных наук»// Философия и будущее цивилизации: Тезисы докладов и выступлений IV Российского философского конгресса (Москва, 24-28 мая 2005 г.): В 5 т. Т.1. – М.: Современные тетради, 2005. – 768 с.
- [Философия ИИ, 2005] Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с.
- [Философия ИИ, 2004] Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции «Философия искусственного интеллекта», МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 192 с.
- [Философско-методологические проблемы ИИ, 2007] Философско-методологические проблемы искусственного интеллекта: материалы постоянно действующего теоретического междисциплинарного семинара/ Под ред. Е.В.Серёдкиной. - Пермь: Изд-во Перм.гос.техн.ун-та, 2007. - 210 с.
- [Финн, 2007] Финн В.К. Интеллектуальные системы и общество: Сборник статей/ Предисл. Д.А. Поспелова, Д.Г. Лахути, В.Б. Тарасова. – М.: КомКнига, 2007. – 352 с.
- [Финн, 2009] Финн В.К. К структурной когнитологии: феноменология сознания с точки зрения искусственного интеллекта //Вопросы философии №1, 2009
- [Фланаган, Полджер, 2006] Фланаган Оуэн, Полджер Томас. 2006. Зомби и роль сознания (Перевод Т. Кураевой) //Тест Тьюринга. Роботы. Зомби. Перевод с англ. Под ред. А.Ю. Алексеева – М.: МИЭМ, 2006. – 120 с. – С.103-116
- [Фреге, 2000] Фреге Готтлоб. Логика и логическая семантика: Сборник трудов/ Пер. с нем. Б.В.Бирюкова под ред. З.А.Кузичевой: Учебное пособие для студентов вузов. - М.: Аспект Пресс, 2000. - 512 с., С.493.
- [Холодная, Сергиенко, 2005] Холодная М.А., Сергиенко Е.А. Структура и функции естественного интеллекта в контексте проблемы искусственного интеллекта//Философско-методологические проблемы искусственного интеллекта: материалы междисциплинарного научно-теоретического семинара, 26 ноя 2005 г., г.Москва: ЦЭМИ РАН / НСМИИ РАН, 2005 г.. URL : <http://www.scmair.ru/seminar/1-6>
- [Хофштадтер, 2001] Хофштадтер Д.Р. Гёдель, Эшер, Бах: эта бесконечная гирлянда. - Самара: Издательский Дом «Бахрах-М», 2001 – 752 с.
- [Чалмерс, 2013] Чалмерс Давид. Сознательный ум: В поисках фундаментальной теории.
- [Чарнецки, Айзенкер, 2005] Чарнецки К., Айзенкер У. Порождающее программирование: методы, инструменты, применения. Для профессионалов. - СПб.: Питер, 2005. - 731 с.
- [Черниговская, 2006] Черниговская Т.В. 2006. Зеркальный мозг, концепты и язык: цена антропогенеза //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИИнтелЛ, 2006. – 448 с. – С.127-148
- [Шалютин, 1985] Шалютин, С.М. Искусственный интеллект [Текст] : гносеологический аспект / С. М. Шалютин. - Москва : Мысль, 1985. - 198
- [Шалютин, 2006] Шалютин С.М. Машины, Люди. Ценности.- Курган: Изд-во Курганского гос.университета, 2006 -393 с.
- [Шлыкова, 2009] Шлыкова О.В. Электронная культура: дефиниции и тенденции развития//Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. - С.30-36
- [Шульга, 2002] Шульга Е.Н. Когнитивная герменевтика, М., ИФ РАН, 2002
- [Шульга, 2008] Шульга Е.Н. Понимание и интерпретация. М.: Наука, 2008
- [Шэнк, 1980] Шэнк Р. Обработка концептуальной информации. М.: Энергия, 1980
- [Электронная культура, 2009] Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с.

- [Юлина, 2004] Юлина Н.С. Головоломки проблемы сознания: концепция Дэниела Деннета. – М.: Канон+, 2004. – 544 с. – С. 168
- [Юлина, 2005] Юлина Н.С. 2005. Обретение разумности: «сократический диалог» и интеракция с компьютером //Новое в искусственном интеллекте. Методологические и теоретические вопросы. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2005. – 280 с. – С.82-85
- [Юлина, 2006] Юлина Н.С. 2006. Улучшение качества мышления: «сократический диалог» и интеракция с компьютером //Искусственный интеллект: междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского – М.: ИИнтелЛЛ, 2006. – 448 с. – С.180-193
- [Яблоков, 2003] Яблоков И. Н. Некоторые дискуссионные вопросы методологии религиоведения / Философско-методологические проблемы изучения религии. Материалы конференции (Москва, 28–29 октября 2003 г.). М.: РАГС, 2004. С. 65–66.
- [Янковская, 2008] Янковская Е.А. Гетерархический принцип устройства познавательного опыта. Автореферат на соискание ученой степени кандидата философских наук. – Иваново, ИГУ, 2008
- [Янковская, 2011] Янковская Е.А. Гетерархическая модель интересубъективного взаимодействия // Актуальные проблемы современной когнитивной науки. Материалы четвертой всероссийской научно-практической конференции с международным участием (20-21 октября 2011 года). Иваново: ОАО «Изд-во «Иваново»», 2011. – 350 с.
- [Ярославцева, 2009] Ярославцева, Е.И. Искусственный интеллект в цифровом мире// Электронная культура: трансляция в социокультурной и образовательной среде. Под ред. А.Ю. Алексеева, С.Ю. Карпук – М.: Изд-во МГУКИ, 2009. – 260 с. – С. 48-54
- [Ясницкий, 2005] Ясницкий Л.Н. Введение в искусственный интеллект. – М.: Издательский центр «Академия», 2005. – 176 с.
- [Abelson, 1968] Abelson, RP. 1968, Simulation of Social Behavior, in G. Lindzey and E. Aronson, eds. Handbook of Social Psychology Reading, MA.: Addison Wesley, pp. 274–356.
- [Balog. 1999] Balog, K. 1999. Conceivability, possibility, and the mind-body problem. Philosophical Review 108:497-528.. URL : <http://humanities.ucsc.edu/NEH/balog.htm>
- [Barresi, 1987] Barresi, J. 1987. Prospects for the Cyberiad: Certain Limits on Human Self-Knowledge in the Cybernetic Age, Journal for the Theory of Social Behavior 17, pp. 19–46.
- [Barresi, 1990] Barresi, J. 1999. On becoming a person. Philosophical Psychology 12:79-98.
- [Barresi, 1995] Barresi, John (1995) Building Persons: Some Rules for the Game, Psycology: 6,#12 Robot Consciousness (7). URL : <http://psycprints.ecs.soton.ac.uk/archive/00000459/>
- [Blackmore, 2003] Blackmore, Susan J. 2003. Consciousness : An Introduction. US Edition Published by Oxford University Press, New York 2004. URL : <http://www.susanblackmore.co.uk/Books/Consciousness/Consciousness%20Chap%202.pdf>
- [Block, 1978] Block, N. 1978. Troubles with functionalism. Minnesota Studies in the Philosophy of Science 9:261-325. Reprinted in Readings in the Philosophy of Psychology (MIT Press, 1980).. URL : <http://perso.univ-rennes1.fr/pascal.ludwig/troubles.pdf>
- [Block, 1980] Block, N. 1980. Functionalism. In (N. Block, ed) Readings in the Philosophy of Psychology, Vol. 1. MIT Press. (a revised version of the entry on functionalism in The Encyclopedia of Philosophy Supplement, Macmillan, 1996). URL : <http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.pdf>
- [Block, 1980] Block, N. 1980. Functionalism. In (N. Block, ed) Readings in the Philosophy of Psychology, Vol. 1. MIT Press: <http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.pdf>.
- [Block, 1981-1] Block, N. 1980. Troubles with functionalism. In (N. Block, ed) Readings in the Philosophy of Psychology, Vol 1. Harvard University Press.. URL : <http://perso.univ-rennes1.fr/pascal.ludwig/troubles.pdf>
- [Block, 1981-2] Block, N. 1981, Psychologism and Behaviorism, Philosophical Review 90, pp. 5–43.
- [Bringsjord, 1994] Bringsjord, Selmer (1994) What Robots Can and Can't be , Psycology: 5,#59 Robot Consciousness (1). URL : <http://psycprints.ecs.soton.ac.uk/archive/00000418/#html>

- [Bringsjord, 1996] Bringsjord, S. 1996, The Inverted Turing Test is Provably Redundant. *Psycoloquy* 7(29).. URL : <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.29>.
- [Bringsjord, 1999] Bringsjord, S. 1999. The zombie attack on the computational conception of mind. *Philosophy and Phenomenological Research* 59:41-69.
- [Bringsjord, 2001-1] Bringsjord, S. 2001 The Zombie Attack on the Computational Conception of Mind. URL : <http://www.rpi.edu/~brings/SELPAP/zombies.ppr.pdf>
- [Bringsjord, 2001-2] Bringsjord, S. 2001. Creativity, the Turing test, and the (better) Lovelace test. *Minds & Machines* 11:3-27.. URL : <http://www.rpi.edu/~faheyj2/SB/SELPAP/DARTMOUTH/lt3.pdf>.
- [Butler, 1865] Butler, Samuel. Erehwon, London, 1865. (Chapter 23-25: The Book of the Machines). URL : <http://www.hoboes.com/FireBlade/Fiction/Butler/Erehwon/erewhon23/>
- [Chaitin, 2003] Chaitin, G.J. *Algorithmic Information Theory*, Cambridge University Press, 1987, Third Printing, 2003, 234 p.
- [Chalmers, 1993] Chalmers, D. J. 1993. The Conscious Mind: In Search of a Fundamental Theory. URL : <http://jamaica.u.arizona.edu/~chalmers/book/tcm.html>
- [Chalmers, 2006] Chalmers, D. Two-Dimensional Semantics (2006), “Two-Dimensional Semantics”, in *Oxford Handbook of Philosophy of Language*, E. Lepore and B. Smith (eds.), Oxford: Oxford University Press, pp. 575–606, URL = <http://consc.net/papers/twodim.pdf>
- [Chalmers, 2011] Chalmers, D. J. Zombies on the web. URL : <http://jamaica.u.arizona.edu/~chalmers/zombies.html>
- [Chalmers, Bourget, 2010] Chalmers, D. Bourget. A bibliography of work in the philosophy of mind, the philosophy of cognitive science. URL : <http://consc.net/mindpapers>
- [Church, 1936] Church, Alonzo. An Unsolvable Problem of Elementary Number Theory *American Journal of Mathematics*, Vol. 58, No. 2. (Apr., 1936), pp. 345-363.. URL : <https://www.fdi.ucm.es/profesor/fraguas/CC/church-An%20Unsolvable%20Problem%20of%20Elementary%20Number%20Theory.pdf>
- [Churchland, 1996] Churchland, P. S. 1996. The hornswoggle problem. *Journal of Consciousness Studies* 3:402-8. Reprinted in Shear 1997,
- [Churchland, P.M., Churchland, P.S., 1990] Churchland, Paul M., Churchland, Patricia Smith. Could a Machine Think? // *Nature Publishing Group*. – Vol.262. – №. – 1. – P. 32-37.
- [Clifton, 2003] Clifton A., 2003. The Introspection game. Or, Does The Tin Man Have A Heart. URL : [http://cogprints.ecs.soton.ac.uk/archive/00003483/01/Introspection_game\(@](http://cogprints.ecs.soton.ac.uk/archive/00003483/01/Introspection_game(@)
- [Colby, 1981] Colby, K.M. 1981, Modeling a Paranoid Mind, *Behavioral and Brain Sciences* 4(4), pp. 515–560.
- [Colby, Hilf, Weber, 1971] Colby, K.M. Hilf, F.D. and Weber, S. 1971, Artificial Paranoia, *Artificial Intelligence* 2, pp. 1–25.
- [Colby, Hilf, Weber, Kraemer, 1972] Colby, K.M. Hilf, F.D., Weber, S. and Kraemer, 1972, Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes, *Artificial Intelligence* 3, pp. 199–222.
- [Cole, 2009] Cole, David. Chinese Room Argument. *Stanford Encyclopedia of Philosophy*, 2009. URL : <http://plato.stanford.edu/entries/chinese-room/>
- [Crick, 1994] Crick, F. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. Scribners.
- [Cummins, 2000] Cummins, Robert. *Minds, brains, and computers : the foundations of cognitive science ; an anthology*. Malden, Mass. [u.a.] : Blackwell, 2000.. URL : <http://www.worldcat.org/title/minds-brains-and-computers-the-foundations-of-cognitive-science-an-anthology/oclc/246213097>
- [Dennett 1995-1] Dennett Daniel C. The Unimagined Preposterousness of Zombies. Commentary on T. Moody, O. Flanagan and T. Polger, *Journal of Consciousness Studies*, vol. 2, no. 4, 1995, pp. 322-326.. URL : <http://ase.tufts.edu/cogstud/papers/unzombie.htm>
- [Dennett 1995-3] Dennett, D. C. 1995. The Practical Requirements for Making a Conscious Robot. URL : <http://www.ecs.soton.ac.uk/~harnad/Papers/Py104/dennett.rob.N@>
- [Dennett 1995-4] Dennett, D. Review on Searle J.R. «The Rediscovery of Consciousness»// *Cambridge MA //Journal of Philosophy*. V. XC. P. 198-199

- [Dennett, 1984] Dennett, D. C. 1984. The Age of Intelligent Machines: Can Machines Think? In (M. Shafto, ed) *How We Know*. Harper & Row.. URL : <http://www.kurzweilai.net/articles/art0099.html?printable=1>
- [Dennett, 1985] Dennett, D. C. 1985. Can machines think? In *How We Know* (Shafto).
- [Dennett, 1994-1] Dennett, D. C. 1994. The practical requirements for making a conscious robot. *Philosophical Transactions of the Royal Society A* 349:133-46.
- [Dennett, 1995] Dennett, D. C. 1995. Cog: Steps toward consciousness in robots. In (T. Metzinger, ed) *Conscious Experience*. Ferdinand Schoningh.
- [Dennett, 1995-5] Daniel C. Dennett. Cog as a Thought Experiment. FINAL DRAFT for Monte Verità issue, 1995. URL : <http://ase.tufts.edu/cogstud/papers/cogthogt.htm>
- [Dennett, 1999] Dennett Daniel C. (1999), "The zombic hunch: Extinction of an intuition?". URL : <http://www.nyu.edu/gsas/dept/philo/courses/consciousness/papers/DD-zombie.html>
- [Dietrich, 1988] Dietrich, E. 1988. Computers, intentionality, and the new dualism. Manuscript.
- [Dowe, Oppy, 2003] Dowe D., Oppy G. 2003. The Turing Test. *Stanford Encyclopedia of Philosophy*. URL : <http://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=turing-test>
- [Flanagan, 1991] Flanagan, O. (1991), *The Science of the Mind* (Cambridge, MA: MIT Press), p. 309
- [Flanagan, Polger, 1995] Flanagan Owen & Polger Tom (1995), "Zombies and the function of consciousness", *Journal of Consciousness Studies* 2:313-21.. URL : <http://homepages.uc.edu/~polgertw/Polger-ZombiesJCS.pdf>
- [Flanagan, Polger, 1998] Flanagan, O. & Polger, G. 1998. Consciousness, adaptation, and epiphenomenalism. In (G. Mulhauser, ed) *Evolving Consciousness*.
- [French, 1990] French, R. 1990, Subcognition and the Limits of the Turing Test, *Mind* 99(393), pp. 53–65.
- [French, 1995] French R. (1995), The Inverted Turing Test: A Simple (Mindless) Program that Could Pass It, *Psychology* 7(39). URL : <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.
- [French, 1995-1] French, R. M. 1995. Refocusing the Debate on the Turing Test: A Response. *Behavior and Philosophy* 23, pp. 59–60.
- [French, 1995-2] French, R. (1995), The Inverted Turing Test: A Simple (Mindless) Program that Could Pass It, *Psychology* 7(39). <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.. URL : <http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?7.39>.
- [French, 1996] French, R. M, 1996, THE INVERTED TURING TEST: HOW A MINDLESS PROGRAM COULD PASS IT (In *Psychology* 7(39) turing-test.6.french.). URL : +
- [French, 2000] French, R. M. 2000. The Turing Test: The First Fifty Years In *Trends in Cognitive Sciences*, 4(3), (2000), pp. 115-121.. URL : http://www.ulg.ac.be/cogsci/rfrench/TICS_turing.pdf
- [From ICT to E-culture, 2003] From ICT to E-culture. Advisory report on the digitalisation of culture and the implications for cultural policy. Netherlands Council for Culture, 2003. URL : http://www.cultuur.nl/files/pdf/advies/E-cultuur_engels.pdf
- [Garson, 2002] Garson James, 2002, Connectionism. URL : <http://plato.stanford.edu/entries/connection#@>
- [Garson, 2010] Garson, James (2010). Connectionism. *Stanford Encyclopedia of Philosophy*. First published Sun May 18, 1997; substantive revision Tue Jul 27, 2010 [URL]: <http://plato.stanford.edu/entries/connectionism/>
- [Genova, 1994-1] Genova, J. Response to Anderson and Keith, *Social Epistemology*, 1994, 8(4), pp. 341–343
- [Genova, 1994-2] Genova, J. Turing's Sexual Guessing Game, *Social Epistemology*, 1994, 8(4), pp. 313–326.
- [Godel, 1931] Godel, Kurt. On Formally Undecidable Propositions of Principia Mathematica and Related Systems. Dover, 1962. URL : <http://www.research.ibm.com/people/h/hirzel/papers/canon00-goedel.pdf>
- [Gulick, 2004] Gulick, Robert Van. Consciousness. *Stanford Encyclopedia of Philosophy*, 2004. URL : <http://plato.stanford.edu/entries/consciousness/>
- [Gunderson, 1964] Gunderson, K. (1964), The Imitation Game, *Mind* 73 pp. 234–245.
- [Harnad, 2001] Harnad, Stevan (2001) *Minds, Machines and Turing: The Indistinguishability of Indistinguishables*. *Journal of Logic, Language, and Information* 9(4):pp. 425-445.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00002615>

- [Harnish, 2001] Harnish, Robert M. *Minds, brains, computers: an historical introduction to the foundations of cognitive science*. Oxford : Blackwell Publishers, 2001. URL : <http://www.worldcat.org/title/minds-brains-computers-an-historical-introduction-to-the-foundations-of-cognitive-science/oclc/223151027>
- [Hauser, 1995] Hauser Larry (1995), "Revenge of the zombies". URL : <http://members.aol.com/lshauser/zombies.html>
- [Hayles, 1999] Hayles, N. Katherine. *How We Became Posthuman*, Chicago: The University of Chicago Press, 1999
- [Hodges, 1997] Hodges, A. (1997). *Turing. A Natural Philosopher*. Phoenix. URL : http://www.amazon.co.uk/exec/obidos/ASIN/0753801922/alanturingwebsit#reader_0753801922
- [Hodges, 2010] Hodges, Andrew (2010). *The Alan Turing Home Page*. URL : <http://www.turing.org.uk/turing/index.html>
- [Hofstadter, 1981] Hofstadter, D. R. 1981. A coffee-house conversation on the Turing test. *Scientific American*.
- [Hopkins, 1998] Hopkins, Patrick D. *Sex/machine: readings in culture, gender, and technology*. Indiana University Press, 1998
- [Horst, 2003] Horst Steven, 2003. *The Computational Theory of Mind*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/computational-mind>
- [Horst, 2011] Horst, Steven, "The Computational Theory of Mind", *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/computational-mind/>.
- [Hyslop, 2005] Hyslop, A. 2005. *Other Minds*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/other-minds/>
- [Jefferson, 1949] Jefferson, Geoffrey. *The Mind of Mechanical Man*, *British Medical Journal*, 1(4616):1105–1110, 25 June 1949. URL : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2050428/>
- [Karelis, 1986] Karelis, C. 1986. *Reflections on the Turing Test*, *Journal for the Theory of Social Behavior* 16, pp. 161–172.
- [Karsakof, 1832] Karsakof S. *Aperçu d'un procédé nouveau d'investigation au moyen de machines à comparer les idées*. – St. Petersburg, 1832. 22 p., 2 pl
- [Kirk, 1974-1] Kirk, R. 1974. *Sentience and behaviour*. *Mind* 81:43-60.
- [Kirk, 1974-2] Kirk, R. 1974. *Zombies vs materialists*. *Aristotelian Society Supplement* 48:135-52.
- [Kirk, 2005] Kirk, R. (2005) *Zombies and Consciousness*. University of Nottingham, November, 2005. URL : <http://www.oup.co.uk/isbn/0-19-928548-9>
- [Kirk, 2006] Kirk, R. 2006. *Zombies*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/zombies/>
- [Kripke, 1972] Kripke, S.A. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press, 1972.
- [Kugel, 2002] Kugel Peter. 2002. *Computing Machines Can't Be Intelligent (...And Turing Said So)*. URL : <http://www.cs.bc.edu/~kugel/Publications/Hyper.pdf>
- [Kurzweil, 1999] Kurzweil, Ray. *The Age of Spiritual Machines*, London: Phoenix 1999
- [Lassegue, 1988] Lassegue, J. 1988, *What Kind of Turing Test did Turing Have in Mind?*, *Tekhnema* 3, pp. 37–58.. URL : <http://tekhnema.free.fr/3Lasseguearticle.htm>
- [Lassègue, 1996] Lassègue Jean, 1996. *What Kind of Turing Test Did Turing Have in Mind?* *Tekhnema* 3 / "A Touch of Memory" / Spring 1996. URL : <http://tekhnema.free.fr/3Lasseguearticle.htm>
- [Leiber, 1989] Leiber, J. 1989. *Shanon on the Turing Test*, *Journal for the Theory of Social Behavior* 19(2), pp. 257–259.
- [Levin, 2004] Levin, J. 2004. *Functionalism*. Stanford Encyclopedia of Philosophy. URL : <http://plato.stanford.edu/entries/functionalism>
- [Levin, 2013] Levin, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2013/entries/functionalism/>.
- [Lucas, 1961] Lucas, J.R. (1961) *Minds, Machines and Goedel*. *Philosophy* 36:pp. 112-127.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000356>

- [Lycan, 1987] Lycan, W. G. 1987. Homunctionalism and qualia. In *Consciousness*. MIT Press.. URL : <http://philpapers.org/rec/LYCHAQ>
- [Marton, 1998] Marton, P. 1998. Zombies vs. materialists: The battle over conceivability. *South-west Philosophy Review* 14:131-38.. URL : <http://www.brown.edu/Departments/Philosophy/zoqwvCTmbie.html>
- [Masrou, 1998] Masrou Farid S. Are Zombies Positively Conceivable? Institute for Studies in Theoretical Physics and Mathematics (IPM). URL : http://www.geocities.com/f_masrur/zombies.html
- [McCarthy, 1969] McCarthy John, 1969, SOME PHILOSOPHICAL PROBLEMS FROM THE STANDPOINT OF ARTIFICIAL INTELLIGENCE. URL : <http://www-formal.stanford.edu/jmc/>
- [McCarthy, 1979] McCarthy, John (1979) Ascribing Mental Qualities to Machines.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000416>
- [McCarthy, 1982] McCarthy John 1982, The Common Business Communication Language. . URL : <http://www-formal.stanford.edu/jmc>
- [McCarthy, 1984] McCarthy, J (1984) Some Expert Systems Need Common Sense.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000423>
- [McCarthy, 1987] McCarthy, John (1987) Generality in Artificial Intelligence; URL : <http://cogprints.ecs.soton.ac.uk/archive/00000425>
- [McCarthy, 1990] McCarthy John, 1990, ARTIFICIAL INTELLIGENCE, LOGIC AND FORMALIZING COMMON SENSE. URL : <http://www-formal.stanford.edu/jmc/>
- [McCarthy, 1995-1] McCarthy, John (1995) Artificial Intelligence and Philosophy.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000420>
- [McCarthy, 1995-2] McCarthy, John (1995) Making Robots Conscious of their Mental States.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000422>
- [McCarthy, 1995-3] McCarthy, John (1995) Todd Moody's Zombies.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000426>
- [McCarthy, 1997] McCarthy, John (1997) The Well-Designed Child.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000428>
- [McCarthy, 1998] McCarthy, John (1998) What is Artificial Intelligence?.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000412>
- [McCarthy, 2001] McCarthy John, 1999, THE ROBOT AND THE BABY. URL : <http://www-formal.stanford.edu/jmc/>
- [Michie, 1974] Michie, D. On Machine Intelligence, New York: Halsted Press, 1974
- [Miłkowski, 2007] Miłkowski, Marcin. Is computationalism trivial?// In: Gordana Dodig Crnkovic and Susan Stuart (eds.), *Computation, Information, Cognition – The Nexus and the Liminal*, Cambridge Scholars Publishing, 2007, pp. 236-246.
- [Millar, 1973] Millar, P.H. (1973), On the Point of the Imitation Game, *Mind* 82, pp. 595–597.
- [Moody, 1995] Moody, Todd (1995), "Conversations with zombies". URL : http://www.imprint.co.uk/Moody_zombies.html
- [Moor, 1976] Moor, J.H. (1976), An Analysis of the Turing Test, *Philosophical Studies* 30, pp. 249–257.. URL : <http://www.springerlink.com/content/x336q61282151813/fulltext.pdf>
- [Moor, 1978] Moor, J. H. 1978. Explaining computer behavior. *Philosophical Studies* 34, pp. 325–327
- [Nagel, 1974] Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 4:435-50. Reprinted in *Mortal Questions* (Cambridge University Press, 1979).. URL : <http://hps.elte.hu/~gk/Books/cog-reader/nagel.htm>
- [New Yorker, 1993] On the Internet, Nobody Knows You're a Dog. Page 61 of July 5, 1993 issue of *The New Yorker*, (Vol.69 (LXIX) no. 20). URL : <http://www.unc.edu/depts/jomc/academics/dri/idog.html>
- [O'Hara, Scutt, 1996] O'Hara, K. & Scutt, T. 1996. There is no hard problem of consciousness. *Journal of Consciousness Studies* 3:290-302. Reprinted in Shear 1997.
- [Oppy, Dowe, 2008] Oppy, Graham., Dowe, David (2008). The Turing Test. *Stanford Encyclopedia of Philosophy*. URL : <http://plato.stanford.edu/entries/turing-test/>

- [Owen, 1863] Owen on the Limbs of the Gorilla. Comparison of the Bones of the Limbs of the Troglodytes Gorilla. Troglodytes niger, and of Different Varieties of the Human Race; and on the General Character of the Skeleton of the Gorilla by Professor Owen/ Anthropological Review, Vol. 1, No. 1 (May, 1863), pp. 149-152
- [Pei, 1995] Wang Pei, 1995. On the Working Definition of Intelligence. URL : <http://www.cogsci.indiana.edu/pub/wang.intelligence.ps>
- [Penrose, 1989] Penrose, R. 1989. The Emperor's New Mind. Oxford University Press.
- [Piccinini, 2004] Piccinini, Gualtiero. Functionalism, Computationalism and Mental content // Canadian Journal of Philosophy, Volume 34, Number 3, September 2004, pp.375-410
- [Platt, 1995] Platt, C. 1995, What's It Mean To Be Human, Anyway?, Wired.. URL : <http://www.wired.com/wired/archive/3.04/turing.html>
- [Polger, 2004] Polger, Thomas W. Natural minds. Massachusetts Institute of Technology: "A Bradford book", 2004. - 294 p.
- [Polger, 2004] Polger, Thomas W. Natural minds. Massachusetts Institute of Technology: "A Bradford book", 2004. - 294 p.
- [Polger, 2010] Polger, Tom (web). Computational functionalism. In J. Symons & P. Calvo (eds.), Routledge Companion to the Philosophy of Psychology. Routledge. URL : <http://homepages.uc.edu/~polgertw/Polger-CompFunctionalism.pdf>
- [Polger, 2011] Polger Tom, Zombies. URL : <http://host.uniroma3.it/progetti/kant/field/zombies.htm#Top>
- [Povarov, 2001] G. N. Povarov. Semen Nikolayevich Korsakov: Machines for the Comparison of Ideas // G. Trogemann, A.Y. Nitussov , W. Ernst , Eds. Computing in Russia. Wiesbaden: Vieweg & Son, 2001
- [Povarov, 2001] G. N. Povarov. Semen Nikolayevich Korsakov: Machines for the Comparison of Ideas // G. Trogemann, A.Y. Nitussov , W. Ernst , Eds. Computing in Russia. Wiesbaden: Vieweg & Son, 2001
- [Purthill, 1971] Purthill, R.L. (1971), Beating the Imitation Game, Mind 80, 290–294
- [Putnam, 1960] Putnam, H. (1960). Minds and machines. In (S. Hook, ed) Dimensions of Mind. New York University Press. Reprinted in Mind, Language, and Reality (Cambridge University Press, 1975).
- [Putnam, 1960] Putnam, H. 1960. Minds and machines. In (S. Hook, ed) Dimensions of Mind. New York University Press. Reprinted in Mind, Language, and Reality (Cambridge University Press, 1975).
- [Putnam, 1967] Putnam, H. (1967). The mental life of some machines. In (H. Castaneda, ed) Intentionality, Minds and Perception. Wayne State University Press. Reprinted in Mind, Language, and Reality (Cambridge University Press, 1975).
- [Putnam, 1975] Putnam, H. (1975). Philosophy and our mental life. In Mind, Language, and Reality. Cambridge University Press.
- [Putnam, 1975] Putnam, H. 1975. Philosophy and our mental life. In Mind, Language, and Reality. Cambridge University Press.
- [Putnam, 1981] Putnam, H., 1981, Reason, Truth, and History, Cambridge: Cambridge University Press. – 222 p.
- [Putnam, 1983] Putnam, H. (1983) Realism and Reason. Philosophical Papers, Volume 3. - Cambridge: University Press. 332 p.
- [Putnam, 1985] Putnam, Hilary. Realism and reason. Cambridge University Press, 1985, 332 p.. URL : <http://books.google.com/books?id=HAjfSA3ir3kC&hl=ru&source=gbssimilarbooks>
- [Putnam, 1997]. "Functionalism: Cognitive Science or Science Fiction?" In D. M. Johnson and C. E. Erneling (eds.), The Future of the Cognitive Revolution. Oxford: Oxford University Press, pp. 32–44.
- [Raiser, Colby, Faught, Parkinson, 1981] Jon F. Raiser, Kenneth Mark Colby, William S. Faught, and Roger C. Parkinson, "Can Psychiatrists Distinguish a Computer Simulation of Paranoia from the Real Thing? The Limitations of Turing-like Tests as Measures of the Adequacy of Simulations," in Journal of Psychiatric Research 15, no. 3 (1980):149-162. Colby discusses PARRY and its implications in "Modeling a Paranoid Mind," in Behavioral and Brain Sciences 4, no. 4 (1981): 515-560.

- [Rapoport, 2000] Rapoport William.J. (2000) How to Pass a Turing Test/ Journal of Logic, Language, and Information 9: 467-490, 2000. 467. URL : <http://www.cse.buffalo.edu/~rapoport/Papers/TURING.pdf>
- [Robb, Heil, 2013] Robb, David and Heil, John, "Mental Causation", The Stanford Encyclopedia of Philosophy (Spring 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2013/entries/mental-causation/>.
- [Ronchi, 2008] Ronchi, Alfredo M. eCulture. Cultural Content in the Digital Age, 2008, Approx. 425 p.. URL : <http://www.springer.com/humanities/cultural+studies/book/978-3-540-75273-8>
- [Sampson, 1973] Sampson, G. (1973), In Defence of Turing, Mind 82, pp. 592–594.
- [Sayers 1941] Sayers, Dorothy L. (1941). The Mind of the Maker (1st Ed. ed.). London: Methuen. URL : <http://www.worldinvisible.com/library/dlsayers/mindofmaker/mind.c.htm>
- [Saygin, Cicekli, Akman, 2000] Saygin, A. P. , Cicekli, I. & Akman V. 2000. Turing test: 50 years later. Minds and Machines 10:463-518.. URL : <http://www.cs.bilkent.edu.tr/~akman/jour-papers/mam/tt50.ps>
- [Scheffler, 1950] Scheffler, I. 1950. The new dualism: Psychological and physical terms. Journal of Philosophy.
- [Schroeter, 2012] Schroeter, L. "Two-Dimensional Semantics", The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2012/entries/two-dimensional-semantics/>.
- [Schwarz, 2006] Schwarz, Michiel. e-Culture: Crossovers and Challenges. Publiziert auf eCulture Factory, 2006. URL : <http://eculturefactory.de/download/schwarz.pdf>
- [Schweizer, 1998] Schweizer, P. 1998. The truly total Turing Test. Minds and Machines 8, pp. 263–272.
- [Searle, 1980] Searle, J. R. 1980. Minds, brains and programs. Behavioral and Brain Sciences 3, pp. 417–424.. URL : <http://www.cs.ucsb.edu/~cs165a/articles/Searle.html>
- [Shagrir, 2005] Shagrir, O., 2005. "The Rise and Fall of Computational Functionalism", in Y. Ben-Menahem (ed.), Hilary Putnam, Cambridge: Cambridge University Press, 220–250.
- [Shanon, 1989] Shanon, B. 1989. A Simple Comment Regarding the Turing Test, Journal for the Theory of Social Behavior 19(2), pp. 249–256.
- [Shoemaker, 1975] Shoemaker, S. 1975. Functionalism and qualia. Philosophical Studies 27:291–315. Reprinted in Identity, Cause, and Mind (Cambridge University Press, 1984).. URL : <http://www.springerlink.com/content/w76571qu83636261/>
- [Sloman, 1978] Sloman, A. 1978. The Computer Revolution in Philosophy. Harvester.. URL : <http://www.cs.bham.ac.uk/research/cogaff/crp/>
- [Sloman, 1996] Sloman, Aaron (1996) What sort of control system is able to have a personality?, in Trapp, Robert and Petta, Paolo, Eds. Creating Personalities for Synthetic Actors: Towards Autonomous Personality Agents, chapter n/a, pages pp. 16-208. Springer.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000702/>>
- [Sloman, 1998] Sloman, Aaron (1998) What Is It Like To Be a Rock?. URL : http://www.cs.bham.ac.uk/~axs/misc/like_to_be_a_rock/rock.html
- [Sloman, 1998] Sloman, Aaron (1998) What Sort of Architecture is Required for a Human-Like Agent?, in Foundations of Rational Agency. Kluwer Academic Publishers.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00000411/>>
- [Sloman, Croucher, 1981] Sloman, A. and Croucher, M. (1981) 'Why robots will have emotions.' Proceedings IJCAI 1981, Vancouver.. URL : <http://cogprints.org/n@>
- [Smolensky, 1988] Smolensky, P. 1988. On the proper treatment of connectionism. Behavioral and Brain Sciences 11:1-23.. URL : <http://cogprints.ecs.soton.ac.uk/archive/00001675/>
- [Smolensky, 1991] Smolensky, P. 1991. Connectionism, constituency and the language of thought. In (B. Loewer & G. Rey, eds) Meaning in Mind: Fodor and his Critics. Blackwell.
- [Smolensky, 1995] Smolensky, P. 1995. On the projectable predicates of connectionist psychology: A case for belief. In (C. Macdonald, ed) Connectionism: Debates on Psychological Explanation. Blackwell.
- [Stalker, 1978] Stalker, D. F. 1978. Why machines can't think: A reply to James Moor. Philosophical Studies 34, pp. 317–320.

- [Stalnaker, 2004] Stalnaker, R. 2004. "Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics", *Philosophical Studies*, 118: 299–322.
- [Stevenson, 1976] Stevenson, J.G. (1976), *On the Imitation Game*, *Philosophia* 6, pp. 131–133.
- [Stoljar, 2001] Stoljar, D. (2001). *Physicalism*. *Stanford Encyclopedia of Philosophy*; <http://plato.stanford.edu/entries/physical>.
- [Traiger, 2000] TRAIGER, SAUL. Making the Right Identification in the Turing Test. *Minds and Machines* 10: 561-572, 2000.. URL : <http://www.traiger.net/publications/making-the-right-identification.pdf>
- [Turing, 1936] TURING, A. M. On Computable Numbers, with an Application to the Entscheidungsproblem. [Received 28 May, 1936.—Read 12 November, 1936.]. URL : http://www.thocp.net/biographies/papers/turing_oncomputablenumbers_1936.pdf
- [Turing, 1936] Turing, A.M. (1936). On Computable Numbers, with an Application to the Entscheidungsproblem [Received 28 May, 1936.— Read 12 November, 1936.]; http://www.thocp.net/biographies/papers/turing_oncomputablenumbers_1936.pdf
- [Turing, 1950] Turing, A. (1950), 'Computing Machinery and Intelligence', *Mind* 59(236), pp. 433–460.
- [Turing, 1969] Turing, A. (1969), 'Intelligent Machinery', in ed. *Machine Intelligence* 5, Edinburgh University Press, pp. 3–23. Originally, a National Physics Laboratory Report, 1948.
- [van Gulick, 1999] van Gulick, R. 1999. Conceiving beyond our means: The limits of thought experiments. In (S. Hameroff, A. Kaszniak, & D. Chalmers, eds) *Toward a Science of Consciousness III*. MIT Press.. URL : http://cognet.mit.edu/posters/TUCSON3/Van_Gulick.html
- [Watt, 1996] Watt, S. 1996. Naive Psychology and the Inverted Turing Test, *Psychology* 7(14). URL : <http://psycprints.ecs.soton.ac.uk/archive/00000506>
- [Watt, 1998-1] Watt, S. 1998. Seeing this as people: anthropomorphism and common-sense psychology, KMI. URL : <http://citeseer.nj.nec.com/cache/papers/cs/21274/http:zSzzSzkmi.open.ac.ukzSztrSzpaperszSzkmi-tr-61.pdf/watt98seeing.pdf>
- [Watt, 1998-2] Watt, S. 1998. Artificial societies and psychological agents, KMI. URL: <http://citeseer.nj.nec.com/cache/papers/cs/3443/http:zSzzSzkmi.open.ac.ukzSztechreportszSzpaperszSzkmi-tr-33.pdf/watt97artificial.pdf>
- [Weizenbaum, 1974] Weizenbaum, Joseph. *CACM*17, no. 9 (September 1974): 543.
- [Weizenbaum, 1976] Weizenbaum, J. (1976), *Computer Power and Human Reason: From Judgment to Calculation*, San Francisco, CA: W.H. Freeman.
- [Weizenbaum, 1996] Weizenbaum, J. (1996), 'ELIZA—A Computer Program for the Study of Natural Language Communication Between Men and Machines', *Communications of the ACM* 9, pp. 36–45.