

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ИНСТИТУТ
ЭЛЕКТРОНИКИ И МАТЕМАТИКИ
(ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ)**

**МЕТОДОЛОГИЧЕСКИЕ
И ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

**МАТЕРИАЛЫ СТУДЕНЧЕСКОЙ
КОНФЕРЕНЦИИ**

**«ФИЛОСОФИЯ
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА»**

Москва, МИЭМ, 20 мая 2004 г.

Москва 2006

УДК 100.32
ББК 32.816
М 56

Под редакцией к.ф.н. А.Ю. Алексеева

М56 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.

В книге представлены работы участников студенческой конференции «Философия искусственного интеллекта» (Москва, МИЭМ, 20 мая 2004 г.). Обсуждается ряд философских аспектов информационной технологии. Изучаются теоретические основы искусственного интеллекта – функционалистская парадигма мышления и Тест Тьюринга. Выделяются базовые положения философии искусственного интеллекта, предложенные одним из её основоположников – Дж. Маккарти. Выявляются интересные перспективы компьютерного моделирования, связанные с осмыслением историко-философской проблематики. Раскрывается ряд методологических аспектов робототехники, виртуальности, квантовых компьютеров, нейросетевых систем и др. Предлагаются интернет-навигаторы, посвящённые проблематике искусственного интеллекта.

Доклады студентов, выполненные, в основном, в форме рефератов по работам англо-американских исследователей искусственного интеллекта, призывают специалистов более старшего поколения на решение ряда важнейших методологических и теоретических проблем развития перспективных информационных технологий – проблем, не нашедших должного отражения в современной отечественной философии.

ISBN 5-98956-001-X

© МИЭМ, 2006 г.
© ИИнтелЛЛ, 2005 г.

О студенческой конференции «Философия искусственного интеллекта», 20 мая 2004 г., г. Москва, МИЭМ

Студенческая конференция «*Философия искусственного интеллекта*» состоялась 20 мая 2004 г. в Московском государственном институте электроники и математики (МИЭМ), сайт конференции – <http://philos.miem.edu.ru/110.htm>. Ее организаторами выступили ректорат МИЭМ и кафедра философии МИЭМ. Конференция проводилась под эгидой подготовки к Всероссийской междисциплинарной конференции «Философия искусственного интеллекта», которая намечалась на 18-20 января 2005 г. на базе МИЭМ.

В роли председателя конференции выступил **Г.П. ПУТИЛОВ**, доктор технических наук, профессор, проректор МИЭМ по информатизации и новым технологиям в образовании; учёным секретарём и ведущим явился **А.Ю. АЛЕКСЕЕВ**, старший преподаватель кафедры философии МИЭМ; ведущим – **Е.В. ДЕМИДОВА**, доцент кафедры философии МИЭМ.

В работе конференции приняло участие свыше 250 человек, в основном, студентов 4 курса МИЭМ. Было представлено свыше 80 тезисов докладов, многие из которых опубликованы в настоящем Сборнике. 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», обусловлен, к сожалению, *отсутствием* в отечественной науке серьёзных наработок в области философии и методологии ИИ, по сути, полным её забвением. Например, последняя конференция всесоюзного масштаба по проблеме философии ИИ состоялась свыше 30 лет назад. С другой стороны, последнее десятилетие отмечается 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 следует считать пионерами, открывающими для нашей страны проблематику философии ИИ, столь важную для современной науки, культуры и техники.

Были заслушаны интересные доклады видных ученых – **Д.И. ДУБРОВСКОГО**, доктора философских наук, профессора, ведущего научного сотрудника Института философии РАН; **К.К. КОЛИНА**, доктора технических наук, профессора, главного научного сотрудника Института проблем информатики РАН.

Работа конференции была организована следующим образом: пленарные выступления, секционные заседания (по четырем направлениям), заключительные выступления. Ниже представлена Программа конференции:

I. Открытие конференции (10.00-10.30)

Приветственное слово участникам конференции и вступительные доклады:

1. *Г.П. Путилов*, проректор МИЭМ, д.т.н., проф., председатель конференции
2. *Д.И. Дубровский*, д.ф.н., проф., в.н.с. ИФ РАН

II. Работа секции № 1 «Функционалистская концепция мышления. Тест Тьюринга: pro et contra» (10.30-12.20)

1. *Романов Денис*. Функционалистская концепция мышления
2. *Лизоркин Сергей*. Тест Тьюринга. Основные положения
3. *Комаров Дмитрий*. «Машина ли может мыслить?». Стандарт Тьюринга «за» и «против»
4. *Ласточкин Алексей*. Тест Лавлейс. Креативистская критика
5. *Денисов Алексей*. Тест Френча. Субкогнитивистская критика
6. *Родионов Денис*. Тест Серла. Интенционалистская критика
7. *Матанцева Ирина*. Тест Блока. Антибихевиористская критика
8. *Романова Елена*. Тест Ватта. Инвертированный тест Тьюринга
9. *Рыбин Илья*. Социокультурные аспекты теста Тьюринга
10. *Никишев Артур*. Аргумент Гёделя. Критика сильного искусственного интеллекта

III. Работа секции № 2 «Искусственный интеллект и здравый смысл» (12.30-13.30)

1. *Алексеева Анна*. Философия искусственного интеллекта Джона Маккарти
2. *Колесников Станислав*. Формализация здравого смысла в экспертных системах
3. *Гришкин Максим*. Реализационные перспективы теории речевых актов
4. *Подопригора Иван*. Естественно-языковой интерфейс. Моделирование смысла
5. *Горюнов Роман*. Робот и его псевдосознание
6. *Крючков Василий*. Что значит быть роботом?

V. Работа секции № 3 «Историко-философские перспективы компьютерного моделирования» (14.30-15.20)

1. *Косинова Татьяна*. Искусственный интеллект и стоическая эпистемология
2. *Сёмочкин Михаил*. Моделирование виртуальных миров (Н.Кузанский)
3. *Артюхов Анатолий*. Моделирование смысловых миров (А.Ф. Лосев)
4. *Кураева Татьяна*. Искусственный интеллект и зомби
5. *Розов Максим*. Искусственный интеллект и святоотеческая традиция

VI. Доклад К.К. Колина (д.т.н., проф., г.н.с. ИПИ РАН) «Философские принципы информатики» (15.30-16.00)

VII. Работа секции № 4 «Философия искусственного интеллекта и компьютерная технология» (16.00-17.30)

1. *Пак Марк*. Методологические аспекты нанотехнологии
2. *Домась Константин*. Бионика как направление робототехники
3. *Кольцов Михаил*. Парадигма коннекционизма как методология нейротехнологии
4. *Бондарь Александр*. Квалиа и технология виртуальной реальности
5. *Зайцев Игорь*. Квалиа и парадигма функционализма

VIII. Подведение итогов работы конференции

Работа конференции была признана успешной. Доклады ряда студентов МИЭМ, вызвавшие оживлённые дискуссии, рекомендованы на Всероссийскую междисциплинарную конференцию «Философия искусственного интеллекта» (январь 2005 г.). Пожалуй, самое главное – студенты продемонстрировали жгучий интерес к проблематике философии искусственного интеллекта. Это позволяет судить о том, что в области построения и развития интеллектуальных информационно-коммуникационных технологий ещё не всё потеряно. Будущее – за сегодняшними студентами.

Преемственность поколений исследователей искусственного интеллекта

**Г.П. Путилов, доктор технических наук, профессор,
проректор МИЭМ по информатизации и новым
технологиям в образовании**

В молодости, в середине 70-х, я занимался исследованиями в области искусственного интеллекта, скорее, имел некоторое отношение к искусственному интеллекту в весьма модной в научных исследованиях и важной в приложениях области моделирования зрительного восприятия. Всем известно, какой уровень технологического развития был в то время. Это был этап не только бурного развития вычислительной техники и computer science, но и период, вызвавший мощный поток работ, посвященных попыткам и философского осмысления возможностей и перспектив вычислительной науки и техники в целом.

«Может ли машина мыслить?» – вопрос, вставший в полный рост в начале к началу 50-х годов. Знаменитый тест А.М. Тьюринга обсуждался и весьма заинтересовано целым рядом знаменитых ученых. Этот тест достаточно категорично определял, что в интеллекте важнее всего с точки зрения разумного поведения. Именно тогда и появилось словосочетание «искусственный интеллект», значение которого, может быть, не совсем точно характеризует тот спектр работ, которые, на самом-то деле, и должны быть связаны с проблемой искусственного интеллекта. И далеко не последнее место, естественно, заняли проблемы философии искусственного разума, которые в первую очередь ставят вопрос о сущности разума естественного.

В Советском Союзе, в то время, каждые два года проходили международные конференции по искусственному интеллекту. На одной из них, на IV Международной объединённой конференции по искусственному интеллекту, мне посчастливилось присутствовать. Она состоялась в Грузии на базе Института кибернетики Грузинской ССР. Специально к сегодняшней встрече я порывлся в библиотеке и нашел препринты работ этой конференции. Это объемные сборники трудов, посвящённые таким вопросам, как:

- алгоритмы управления движениями роботов,
- методы представления задач,
- методы поиска решений,
- применение искусственного интеллекта,
- автоматическое распознавание речи и т.д.

В это время за рубежом, точнее в 1954 году, А. Ньюэлл задумал создать программу для игры в шахматы. Интересно, что К. Шеннон (отец теории информации) к этому моменту уже предложил метод решения этой задачи, а А. Тьюринг, один из первых специалистов в области информатики, уточнил и продемонстрировал этот метод вручную. Затем в Рэнд Корпорейшен Дж. Шоу и Г. Саймон, вместе с А. Ньюэллом, при поддержке психологов (А. де Гроот), изучавших стиль игры крупнейших гроссмейстеров, разработали и в 1956 году продемонстрировали программу «Логик-Теоретик». Развитие идей этой программы привело к созданию знаменитой GPS (Универсальный Решатель Задач) – программа, которая уже умела решать нетривиальные головоломки, задача о «Ханойской башне». А затем последовали классические

работы Дж. Маккарти, М. Минского, Г. Саймона и др., начавшие цикл научных исследований по искусственному интеллекту.

В 1960-м году Дж. Гелентер оповестил мир о программе, работавшей впервые лучше (!), чем ее разработчик, при доказательстве теорем школьной геометрии. В то же время создается программа для моделирования психологических ситуаций – ЕРАМ – программа Е. Фейгенбаума, воспроизводившая процессы восприятия и запоминания ситуаций. Программа Student Грина (1961 год) была первым решателем алгебраических задач.

Принципиальное значение в автоматизации логического вывода при решении задач имела разработка Дж. Робинсона (1965 год) машинного интерпретатора для автоматического доказательства теорем. Только после этого стало возможным реализовать такие языки искусственного интеллекта как PLANNER, Maccyma, Reduce и, наконец, PROLOG Колмрауера (1971 год).

Работы по техническому зрению, в частности, Гузмана, Уолца и Уинстона позволяли решать задачи обработки сложных изображений при представлении объектов трехмерной сцены. В это же время появилась система, моделирующая разумное поведение в условиях среды ограниченной сложности, – программа Винограда, умеющая играть в кубики (1971 год).

Таким образом, к 70-ым годам был получен ряд интересных результатов, которые обозначили контуры будущих исследований. Конечно, за последнюю четверть века в ходе дальнейшего развития этой интереснейшей области человеческого знания получено огромное количество фундаментальных и прикладных результатов, которые очевидным образом определяют передний край науки и современных информационных технологий. Мы со своей, технической, точки зрения в меру своих МИЭМовских возможностей отслеживали и отслеживаем эти результаты, но несмотря на то, что мы не стояли в стороне от важнейших исследований в этой области, для меня несколько «вдруг» возник вопрос о том, что в МИЭМе организуется конференция, посвященная философии искусственного интеллекта, открыт сайт, на сайте представлено много докладов. Даже при беглом просмотре этого сайта становится очевидным, что спираль человеческого познания совершает новый виток своего развития, которая определяет, в свою очередь, развитие новейших информационных технологий и, что особенно важно, появление новой волны интереса молодежи к этой проблематике.

Хочу сказать, что просто так ничего не бывает. Когда я говорил о себе, я говорил о времени, когда был жив Аксель Иванович Берг, академик, адмирал и, как он о себе, совершенно справедливо, высказывался – первый в СССР кибернетик. Именно он – один из основоположников работ, связанных с кибернетикой и ее дальнейшим развитием, являлся председателем Научного Совета Кибернетики АН СССР. В те же времена, в МИЭМе на кафедре Кибернетики (она тогда была единой кафедрой, зав. каф. К.А. Пупков) выполнялась важнейшая работа, порученная нам Президиумом АН СССР. Она приобщала весь коллектив кафедры к важнейшим фундаментальным и прикладным проблемам науки и техники. Дело в том, что проф. Пупков К.А. был назначен председателем Секции использования результатов бионики Научного Совета АН СССР по Проблемам управления движением и навигации, а мне довелось принять достаточно активное участие в качестве помощника ученого секретаря этой секции. Таким образом, мы хоть и на общественных началах участвовали в деятельности одного из важнейших советов «Большой» Академии

Наук, который возглавлял академик-секретарь, вице-президент АН СССР, председатель Интеркосмоса Б.Н. Петров.

В то время мы надеялись, что впереди прекрасное будущее ожидает всех тех, кто занимается данными проблемами. Но вот как-то прошла эйфория. Мне удалось защитить диссертацию, посвященную достаточно интересным аспектам технического зрения. Но в силу ряда причин интерес в нашей промышленности к этой проблеме потихоньку угасал. Работы на Западе, связанные с этой проблематикой, ушли в закрытую тематику. Сегодня мы видим, что это не случайно. Примеры:

- появление (сначала в СССР) крылатых ракет,
- появление автоматических интеллектуальных летающих роботов-шпионов над Афганистаном,
- появление американских роботов на Марсе.

В них, так или иначе, используются те наработки, которые не только обсуждались и на упомянутой выше IV конференции по искусственному интеллекту, но и разрабатывались в промышленности, исследовались в Академии Наук и Высшей Школе.

Но о конференции в МИЭМ с названием «философия искусственного интеллекта», я, честно говоря, хоть и занимаясь этими проблемами в техническом плане, еще раз повторяю, не слышал.

Чем объяснить этот растущий интерес? И интерес студентов в частности? Казалось бы, десятилетие перестроек сделало людей более прагматичными. Особенно людей молодых. Оказывается, что помимо вопросов технического характера, которые обсуждаются в среде нашего института (я говорю о среде профессиональной), появляется иная профессиональная среда, где также можно обсуждать вопросы подобного рода. И здесь, я хотел бы подчеркнуть, безусловно, заслуга непосредственно кафедры философии. Кафедра философии традиционно раз в год проводит студенческие конференции. Также раз в год, в феврале, проводится общевузовская конференция студентов и аспирантов. Победители этой конференции участвуют в международной студенческой школе-семинаре в Судаке, там как раз есть секция по гуманитарным наукам. Но конференция в таком формате, отражающая специфические философские проблемы, – это, проводится, действительно, впервые.

Очевидно, если есть люди, студенты, которые интересны студентам, то есть взаимный интерес, который побуждает к тому, что каждому хочется высказать своё мнение по тому или иному вопросу, который будет сегодня подниматься. И это здорово!

Поэтому, действительно, я искренне рад, что данная конференция состоится сегодня. Причем, я хотел бы отметить, что наша конференция не просто чисто внутренняя студенческая. На ней присутствует Дубровский Давид Израилевич – известный специалист в области философии сознания. Он тоже выступит с приветственным словом и выразит своё видение проблемы в целом, а также выступит с заключительным словом по работе первой секции. Ожидается еще ряд философов ученых, которые проявили живой интерес к нашей конференции. Поэтому у нас формат примерно такой. Мы начинаем. Все секции идут одна за другой. Студенты, в свободное от учебы время участвуют в конференции. Вплоть до самого вечера, до 18 часов у нас будет проходить конференция.

Хотел бы отметить, что данная конференция – это первый шаг подготовки нашего института к проведению большой, уже Всероссийской конференции по философии искусственного интеллекта. Более подробно об этом может рассказать Андрей Юрьевич Алексеев. Хочется только подчеркнуть, что большой интерес ученых, которые будут принимать участие во Всероссийской конференции, актуальность и собственно необходимость её проведения по существу проблемы говорит о том, что мы идем по правильному пути.

Несколько слов о целях и задачах конференции. Если говорить о целях – это пробуждение дополнительного интереса как у студентов и аспирантов, так и у преподавателей нашего института. Естественная цель, которая всегда присутствует на подобного рода конференциях. Она сопутствует консолидации, координации и централизации деятельности различных отечественных школ, связанных с этой проблемой. Хотелось бы, чтобы мы знали друг о друге, и первые попытки в этом направлении имеются. Тот сайт, который мы создали, форум, на котором, я надеюсь, в общем-то, будут публиковаться и другие работы, причем, проходя определенное рецензирование, позволит поднять значимость нашей деятельности.

Но главными проблемами остаются те, которые связаны с сутью дела:

- 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?

Рассматривая эти проблемы, конференция очевидно не оставит в стороне и актуальный ряд формальных задач искусственного интеллекта, таких как символическая репрезентация «сущностей», логико-математическая экспликация поведения и рассуждения, модально-эпистемический учёт «точек зрения», когнитивное моделирование мотиваций, целей, проблем и др. задач.

Важно ответить и на ряд вопросов: как изменятся технологии обучения, автоматизация научных исследований, как изменятся современные информационные технологии, использующие элементы искусственного интеллекта? Что будет с глобальными информационными технологиями и т.д.?

Надеюсь, что нас ждет не самое плохое будущее. В заключение, традиционно, я говорю: позвольте открыть конференцию и пожелать её участникам всяческих успехов – успехов в философском осмыслении современных результатов фундаментальных и прикладных наук, в оценке перспектив развития теории и практики искусственного разума.

Приветственное слово участникам конференции

**Д.И. Дубровский, доктор философских наук, профессор,
ведущий научный сотрудник Института философии РАН**

Мне очень приятно приветствовать вас, участников этой знаменательной конференции. Ведь подобных конференций у нас не было более двадцати лет. Между тем «философия искусственного интеллекта» представляет собой круг актуальнейших проблем, разработка которых существенно влияет на решение конкретных вопросов в науке, технике, культуре в целом.

Для многих философия – это нечто абстрактное и далекое от жизни. Но это не так. Философия искусственного интеллекта как раз и демонстрирует нам свою тесную связь с решением конкретных практических задач. Она включает вместе с тем ряд узловых, фундаментальных проблем современного научного познания. Вокруг этих проблем вот уже более полувека не стихают острые дискуссии, в которых принимают участие лучшие умы. Об этом свидетельствует поистине огромная литература, насчитывающая многие тысячами книг и статей, авторами которых являются не только философы, но и психологи, биологи, социологи, представителями когнитивных, физических, математических и компьютерных наук. Здесь располагается широкое поле соприкосновения и взаимодействия естественных, математических, технических, психологических, социальных и гуманитарных дисциплин – важнейшее условие плодотворного развития научного знания в XXI веке.

Думаю, вы понимаете, что прогресс в любой области знаний определяется новыми идеями, новыми концепциями и теориями. А откуда они берутся? В поиске такого источника вы неизбежно выходите в область философских, методологических и общетеоретических вопросов, связанных с развитием компьютерных наук и информационных технологий. Этот круг актуальных вопросов современности и составляет в общем то, что кратко именуется «философией искусственного интеллекта».

Повторяю, эти проблемы являются чрезвычайно острыми. Они очень широко обсуждаются на Западе практически всеми ведущими учёными, занимающимися компьютерными технологиями, роботизацией, исследованием информационных процессов. Значение этих областей знания для нашей страны понятно каждому. Поэтому разработка «философии искусственного интеллекта» – важный стимул развития отечественной науки.

Весьма показательно, что после такого большого перерыва, двадцать лет спустя, именно молодежь начинает снова поднимать на щит эту проблематику.

Я хотел бы в нескольких словах коснуться исторических моментов. В 50-е годы, когда кибернетика только возникла, у нас в Советском Союзе ее объявили буржуазной лженаукой. Были блокированы все теоретические разработки в этой области. И, как до меня уже говорил Георгий Петрович Путилов, огромная заслуга в том, что эти барьеры были сломаны, принадлежала выдающемуся нашему ученому, академику Акселю Ивановичу Бергу. Он создал (а это было тогда крайне сложно) Совет по кибернетике при Президиуме Академии наук СССР. Под его руководством Совет энергично работал, проводил научные конференции, издавал книги, журналы и сборники. Благодаря этому и были открыты пути развития чрезвычайно актуальной проблематики искусственного интеллекта. Большой вклад в работу Совета по

кибернетике при Президиуме Академии наук внес заместитель академика А.И. Берга, доктор философских наук, профессор Борис Владимирович Бирюков. Лично я многим обязан этому Совету, а также персонально академику А.И. Бергу и проф. Б.В. Бирюкову. Благодаря их поддержке в 1971 году под грифом этого Совета была, наконец, издана (после долгих проволочек в других организациях) моя монография «Психические явления и мозг. Философский анализ проблемы в связи с некоторыми актуальными задачами нейрофизиологии, психологии и кибернетики». Совету по кибернетике обязаны многие другие философы и ученые, которым была оказана поддержка в публикации их работ, помощь в решении не только теоретических, но и практических вопросов.

Конечно, очень приятно, что широкое публичное обсуждение философии искусственного интеллекта у нас в стране возобновляется молодыми людьми. Это – знаменательное явление и оно вполне закономерно. Почему был такой большой антракт – 20 лет? Это связано со многими причинами – социальными и политическими, в первую очередь. Но сейчас эти проблемы снова начинают интенсивно разрабатываться. Георгий Петрович Путилов говорил уже о том, что у нас готовится большая Всероссийская конференция с таким же названием – «Философия искусственного интеллекта», которая состоится в январе 2005 года. Она как бы идет по вашим стопам, повторяет ту же самую тему. Во Всероссийской конференции будут участвовать ведущие академические институты, представители самых различных областей знания. Я думаю, что лучшие доклады студентов на сегодняшней конференции следует озвучить и на нашей большой конференции.

Хотелось бы поблагодарить ректорат вашего Института за организацию этой конференции. Надо подчеркнуть, что ректорат МИЭМ, вносит, собственно, решающий вклад и в организацию большой конференции, которая будет проходить в стенах вашего Института. Хочу также выразить искреннюю благодарность учёному секретарю конференции Андрею Юрьевичу Алексееву (который, кстати, сочетает в себе профессиональные навыки философа и специалиста в области компьютерных технологий): его отменная работоспособность, его энтузиазм – важные факторы в организации и проведении как этой, так и предстоящей, большой конференции, где он также является Ученым секретарем и выполняет основную организационную работу.

Хочется пожелать вам успешной работы. И чтобы у вас и далее развивался вкус к ключевым теоретическим и методологическим проблемам искусственного интеллекта и компьютерных наук в целом. Ибо в этом, я думаю, состоит одно из важнейших условий широкого и основательного научного мышления, продуктивного творческого подхода к решению проблем, достижения нынешними студентами больших, выдающихся результатов в будущем.

И главное, ребята, по окончании института не уезжайте за границу, мы на вас надеемся, будем вместе работать на благо нашей страны.

Философские аспекты информационных технологий

**К.К. Колин, доктор технических наук, профессор,
главный научный сотрудник Института проблем информатики РАН**

В докладе рассматривается проблема формирования информационной технологии как самостоятельной науки о методах и средствах создания высокоэффективных информационных технологий (в узком понимании этого термина). Определяются отличительные признаки высокоэффективных технологий и основные принципы их проектирования. Формулируются общие критерии для оценки эффективности и социальной полезности информационных технологий.

Введение

В настоящее время происходит стремительное развитие глобального процесса информатизации общества. При этом кардинальным образом изменяется вся информационная среда общества, а новые автоматизированные информационные технологии проникают практически во все сферы социальной практики и становятся неотъемлемой частью новой, информационной культуры человечества.

Тем не менее, в фундаментальной науке до сих пор отсутствует самостоятельное научное направление, которое являлось бы теоретической базой для проектирования перспективных информационных технологий, их оптимизации и сравнительной количественной оценки эффективности, а также для разработки методов и инструментальных средств, которые обеспечивали бы наилучшие способы организации наиболее массовых и социально значимых информационных процессов.

Именно поэтому сегодня представляется исключительно актуальной и важной *проблема формирования информационной технологии, как фундаментальной науки о методах и средствах создания и высокоэффективной реализации информационных технологий* (в обычном, узком понимании этого термина).

Таким образом, помимо уже широко используемого в науке и практике понятия информационной технологии, как способа рациональной организации некоторого часто повторяющегося информационного процесса, необходимо развивать и новое, более широкое представление о значении этого термина. И в этом случае он будет обозначать самостоятельный раздел фундаментальной науки точно так же, как это имеет место в отношении самого понятия «технология».

1. Технология как научная дисциплина

В Кратком словаре современных понятий и терминов [1] дается следующее определение содержания термина «технология»:

«ТЕХНОЛОГИЯ (греч. *Techne* – искусство, мастерство) – совокупность методов обработки, изготовления, изменения состояния, свойств, формы сырья, материала или *полуфабриката*, осуществляемых в процессе производства продукции».

Однако необходимо учитывать, что термин «технология» имеет в современном русском языке еще и несколько других значений. Этим термином

обозначается также и совокупность документов, которые определяют порядок реализации того или иного технологического процесса, т.е. так называемая технологическая документация. Кроме того, этот же термин часто используют и для обозначения самого технологического процесса.

И, наконец, существует еще одно значение этого термина, которое в контексте данной работы представляется исключительно важным. Термином «технология» обозначается также и самостоятельная *техническая наука*, для которой объектом изучения являются сами технологии (в узком понимании этого термина).

Учитывая вышеизложенное, естественно предположить, что сегодня весьма актуальным является формирование и такой новой самостоятельной научной дисциплины, которая изучала бы лишь вполне определенную часть технологий, а именно *информационные технологии*, которые уже достаточно широко используются в самых различных сферах жизнедеятельности общества [3].

При этом основная гипотеза автора настоящей работы состоит в том что, вышеуказанная новая научная дисциплина (информационная технология) может быть сформирована с использованием *принципа аналогий* с теми основными закономерностями, которые установлены для уже изученных наукой других видов технологий, связанных не с информационными, а с материальными и энергетическими процессами.

Философская концепция данной гипотезы базируется на предположении о существовании некоторых общих закономерностей природы, связанных с использованием материальных, энергетических, информационных или же социальных ресурсов. Ниже будет показано, что использование таких аналогий оказывается весьма продуктивным, хотя существование указанных выше общих закономерностей еще предстоит доказать в будущем.

2. Информационная технология как научная дисциплина

В работе [3] показано, что объектом исследований информационной технологии как научной дисциплины должны являться основные закономерности, связанные с рациональной организацией часто повторяющихся информационных процессов, т.е. информационных технологий (в узком понимании этого термина).

Предметом же исследований должны стать *методы создания информационных технологий*, а также способы и средства их эффективной реализации.

Для развития информационной технологии в таком понимании нам в ближайшие годы предстоит пройти весь цикл формирования этого нового научного направления: осуществить классификацию различных видов информационных технологий, разработать критерии для сравнительного анализа и количественной оценки их эффективности, создать методы синтеза высокоэффективных технологий, основанные на последних достижениях фундаментальной науки.

Вполне возможно, что для успешного развития этой научной дисциплины придется также создать и ряд других новых научных дисциплин, в том числе – *теорию информационного взаимодействия в природе и обществе*. При этом представляется важным уделить особое внимание не только таким традиционным и уже более или менее изученным фазам реализации информационных процессов, таким, как кодирование, обработка и передача ин-

формации. Нам предстоит разобраться и с гораздо более сложными фазами этих процессов, которые практически еще очень мало изучаются современной наукой. Это фазы *генерации информации*, а также ее *рецепции* (восприятия) информационными системами, в том числе – такими сложными и мало изученными, как сознание и подсознание человека.

Только после этого мы сможем научиться создавать и практически использовать действительно высокоэффективные информационные системы и технологии, которые и должны будут стать технологической базой развития цивилизации в XXI-м веке.

3. Структура предметной области информационной технологии и ее место в современной системе научного знания

Предметную область информационной технологии, как научной дисциплины на начальном этапе ее формирования, вероятнее всего, будут составлять следующие первоочередные задачи:

Разработка методов классификации информационных технологий различного вида и назначения по их характерным признакам.

Разработка критериев эффективности информационных технологий, методов их оптимизации и сравнительной количественной оценки.

Определение перспективных направлений развития информационных технологий на ближайшие годы, а также тех научных методов, которые должны лежать в их основе.

Определение принципов построения перспективных средств для эффективной реализации информационных технологий нового поколения.

Приведенные выше определения объекта и предмета исследований информационной технологии как науки, а также анализ содержания решаемых ею задач позволяют сделать вывод о том, что эта дисциплина должна войти в состав *естественных наук*. Причем, в значительной части своих исследований она будет характеризоваться как *техническая наука*, являющаяся одним из разделов информатики.

Теоретической базой для этой новой науки должны стать достижения в области *теоретической информатики*, и, прежде всего, в области *общей теории информации* – той новой фундаментальной научной дисциплины, которая уже активно формируется в последние годы.

Принципиально важными для развития информационной технологии должны также стать и результаты исследований в области таких наук, как *семиотика*, *семантика*, *когнитология*, *информационная психология*. Ведь для создания принципиально новых по своему качеству информационных технологий будущего нам нужно будет хорошо знать не только те процессы и факторы, которые содействуют эффективному *восприятию* информации человеческим сознанием и подсознанием, но также и факторы, которые содействуют ее наилучшему анализу, запоминанию и адекватному пониманию [3].

Другими словами, перспективные информационные технологии должны быть ориентированы на человека и обеспечивать возможность *развития* у него качеств, содействующих восприятию, запоминанию, анализу и пониманию смысла информации. В современной научной литературе такие технологии все чаще называют *креативными технологиями*.

Таким образом, можно полагать, что для развития креативных технологий в ближайшие десятилетия откроются новые перспективы. Особенно широко эти технологии будут применяться в системе образования.

4. Классификация информационных технологий

Основные классы информационных технологий. Классификация информационных технологий, по-видимому, будет одной из первоочередных задач развития новой научной дисциплины. Сегодня же классификация информационных технологий осуществляется, в основном, по тем или иным признакам, связанным с областью их практического использования, т.е. из чисто прагматических соображений. Нам представляется, что анализ информационных технологий с научных позиций позволит выработать несколько иные подходы к проблеме их классификации. В основе этих подходов, возможно, будут лежать основные признаки тех или иных научных методов, при помощи которых и достигаются основные характеристики этих технологий.

Хотелось бы отметить весьма интересный концептуальный подход к классификации информационных технологий, который предложен И.М. Зацманом в его монографии, посвященной исследованию проблемы концептуального поиска информации в электронных библиотеках [4]. Этот подход базируется на использовании сформулированных в данной работе семиотических основаниях информатики как фундаментальной науки.

По назначению и характеру использования представляется целесообразным выделить следующие два основных класса информационных технологий:

Базовые информационные технологии;

Прикладные информационные технологии.

Базовые информационные технологии представляют собой наиболее эффективные способы организации отдельных фрагментов тех или иных информационных процессов, связанных с преобразованием, хранением или же передачей определенных видов информации. Примерами таких технологий могут быть технологии сжатия информации, ее кодирования и декодирования, распознавания образов и т.п.

Характерным признаком базовых информационных технологий является то, что они не предназначены для непосредственной реализации тех или иных конкретных информационных процессов, а являются лишь теми базовыми их компонентами, на основе которых и проектируются затем прикладные информационные технологии.

Таким образом, главная цель базовых информационных технологий заключается в достижении максимальной эффективности в реализации некоторого фрагмента информационного процесса на основе использования последних достижений фундаментальной науки. Именно поэтому базовые информационные технологии и будут являться главной частью объекта исследований информационной технологии как науки.

Прикладные информационные технологии. Основной задачей здесь является рациональная организация конкретного информационного процесса. Осуществляется это путем адаптации одной или нескольких базовых информационных технологий, позволяющих наилучшим образом реализовать отдельные фрагменты этого процесса. Поэтому основными научными проблемами в области исследования прикладных информационных технологий можно считать:

- Разработку методов анализа, синтеза и оптимизации прикладных информационных технологий.
- Создание теории проектирования информационных технологий различного вида и практического назначения.
- Создание методологии сравнительной количественной оценки различных вариантов построения информационных технологий, их эффективности.
- Разработку требований к аппаратно-программным средствам автоматизации процессов реализации информационных технологий.

Одним из примеров прикладной информационной технологии может служить технология ввода в ЭВМ речевой информации. С технологической точки зрения, весь информационный процесс здесь разделяется на несколько последовательных этапов, на каждом из которых используется своя базовая технология. Такими этапами в данном случае являются:

Аналого-цифровое преобразование речевого сигнала и ввод полученной цифровой информации в память ЭВМ. Базовой технологией здесь является аналого-цифровое преобразование, а реализуется эта технология, как правило, аппаратным способом при помощи специальных электронных устройств, характеристики которых заранее оптимизированы и хорошо известны проектировщикам.

2. Выделение в составе речевой информации отдельных фонем того языка, на котором произносилась речь, и отождествление их с типовыми «образами» этих фонем, хранящимися в памяти вычислительной системы. Базовой технологией здесь является технология *распознавания образов*.

3. Преобразование речевой информации в текстовую форму и осуществление процедур ее морфологического и синтаксического контроля. Базовыми технологиями здесь являются процедуры *морфологического и синтаксического контроля текста* и внесение в него необходимых корректур, связанных с исправлением ошибок.

Приведенный выше пример достаточно наглядно иллюстрирует принцип формирования прикладной технологии путем адаптации ряда заранее отработанных базовых технологий, необходимых для реализации данного информационного процесса. Этот подход не только дает большую экономию времени для разработчиков прикладных информационных технологий, но также и в значительной степени гарантирует их достаточно высокую эффективность.

5. Критерии эффективности информационных технологий

Частные критерии эффективности. Для оптимизации и количественной оценки эффективности различных вариантов проектируемых или же уже существующих информационных технологий необходимо правильно выбирать критерии их эффективности. Такими критериями могут быть:

Функциональные критерии, которые характеризуют степень достижения при данной технологии желаемых характеристик информационного процесса, необходимых пользователю. Такими характеристиками могут быть, например:

- объемно-временные характеристики реализуемого информационного процесса (скорость передачи данных, объем памяти для хранения информации и т.п.);

- надежность характеристики информационного процесса (вероятность правильной передачи или преобразования информации, уровень ее помехозащищенности и др.);
- параметры, характеризующие степень достижения конечного результата информационного процесса, реализуемого при помощи данной технологии (правильность распознавания речи или изображения, качество формируемой графической информации и др.).

Ресурсные критерии, которые характеризуют количество и качество различного вида ресурсов, необходимых для реализации данной информационной технологии. Такими ресурсами могут быть:

- *материальные ресурсы* (инструментально-технологическое оборудование, необходимое для реализации данной технологии);
- *энергетические ресурсы* (затраты энергии на реализацию информационного процесса при данной технологии);
- *людские ресурсы* (количество и уровень подготовки персонала, необходимого для реализации данной технологии);
- *временные ресурсы* (количество времени, необходимого для реализации информационного процесса при данной технологии его организации);
- *информационные ресурсы* (состав данных и знаний, необходимых для успешной реализации информационного процесса).

Специфика реализации информационных технологий. Основными видами ресурсов в производственной сфере являются материальные и энергетические ресурсы. Именно поэтому наибольшее внимание при производстве промышленной продукции уделяется *материалосберегающим* и *энергосберегающим* производственным технологиям. Что же касается информационных технологий, то здесь имеется своя достаточно существенная специфика. Так, например, энергетические ресурсы для информационных технологий, как правило, имеют второстепенное значение. Ведь информационные процессы по самой своей природе обладают сравнительно низкой энергоемкостью по сравнению с силовыми процессами, которые реализуются в механических и энергетических технологиях.

Информационные технологии являются основным средством формирования и использования информационных ресурсов общества. Однако их принципиальная особенность заключается в том, что для своего функционирования они сами нуждаются в использовании информационных ресурсов. Эти ресурсы в виде баз данных и знаний могут заранее вводиться в память информационной системы, а также поступать в нее извне в процессе реализации информационного процесса.

Характерным примером здесь являются *экспертные системы*. Эти технологии, как правило, используют уже накопленный опыт в организации того или иного информационного процесса. При этом достигается возможность существенным образом снизить уровень требований к профессиональной квалификации пользователей экспертной системы, что может дать значительный экономический и социальный эффект.

Этот пример показывает, что информационные технологии позволяют не только формировать знания, но также и их экономно использовать. Другими словами, они обладают свойствами *информационно сберегающих технологий*. Никакие другие технологии такими свойствами не обладают.

Общий критерий эффективности. Ресурсные критерии эффективности позволяют сравнивать между собою различные виды технологий. Кроме того, они дают возможность количественно оценивать получаемый в результате применения этих технологий эффект с точки зрения их социальной полезности в плане экономии различных видов ресурсов общества.

Именно поэтому наиболее распространенными критериями для сравнительной оценки производственных технологий являются *энергетические критерии*. Ведь затраты энергии в общественно полезном производстве являются одним из важнейших показателей уровня технологического развития современного общества.

Однако наиболее общим показателем технологии любого вида (производственной, социальной или же информационной) следует признать *экономия социального времени*, которая достигается в результате использования данной технологии. Этот критерий, предложенный академиком В.Г. Афанасьевым и П.Г. Кузнецовым в качестве одной из наиболее общих мер развития общества [5], представляется нам вполне пригодным для сравнительной количественной оценки эффективности различных видов информационных технологий. Ведь хорошо известно, что любая экономия в конечном итоге может быть сведена к экономии времени. Мало того, по мнению П.Г. Кузнецова, которое разделяет и автор настоящей работы, именно *бюджет социального времени* и является главным ресурсом для жизнеобеспечения и развития современного общества.

Действительно, ведь для практического осуществления любого процесса развития общества (экономического, интеллектуального или духовного) необходимо, чтобы общество имело возможность затратить на эти цели некоторую часть своего общего ресурса социального времени. Другими словами, необходим некоторый «свободный ресурс» социального времени, который должен остаться в бюджете социального времени общества помимо затрат по другим «статьям» этого бюджета, связанным с решением задач простого воспроизводства и жизнеобеспечения общества.

Таким образом, наиболее полезными для общества являются те информационные технологии, которые позволяют сэкономить наибольшее количество социального времени, высвобождая его для других целей, в том числе — для целей развития самого общества.

Изложенный выше подход, коренным образом изменяет традиционную точку зрения на эффективность тех или иных видов информационных технологий, которые сегодня оцениваются, как правило, лишь по функциональным критериям. Так, например, с точки зрения экономии социального времени, для общества очень эффективным является использование информационных технологий в сфере *массового обслуживания* населения (на предприятиях торговли, общественного питания, в сберегательных банках, билетных кассах и т.п.).

Конечно же, мы отдаем себе отчет в том, что использование экономии социального времени в качестве общего критерия эффективности информационных технологий сегодня еще не обеспечено необходимыми методическими разработками.

Однако хотелось бы подчеркнуть, что данный подход представляется нам исключительно перспективным. Ведь он не только позволяет создать необходимую научную и методологическую основу для практического воплощения в жизнь известного лозунга: «Все во благо человека!», но также изме-

няет и мировоззрение общества, его отношение к социальной роли и значимости развития информационных технологий.

6. Отличительные признаки высокоэффективных технологий и перспективные направления их развития

Рассмотрим теперь те наиболее важные отличительные признаки, которые свидетельствуют о высокой потенциальной эффективности различных видов технологий и позволяют таким образом определить перспективные направления их развития. При этом, используя упомянутый выше принцип аналогии, мы будем вначале рассматривать уже известные механические и энергетические технологии для того, чтобы выявить имеющие там место некоторые общие принципы и закономерности и распространить их затем также и на информационные технологии.

Концентрация ресурсов в пространстве. По-видимому, одним из общих принципов создания высокоэффективных технологий является *принцип концентрации ресурсов в пространстве*. Действительно, ведь первые орудия труда, созданные человеком, основаны на использовании именно этого принципа. Изобретенные еще первобытными людьми такие режущие инструменты, как нож и плуг, позволили им концентрировать на их лезвиях ресурсы своей мышечной силы и силы домашних животных и получить за счет этого принципиально новые возможности для обработки земли и материалов, т.е. для выполнения социально полезной работы, жизненно необходимой для своего существования.

Тот же принцип используется и при создании эффективных энергетических технологий, где также осуществляется *концентрация потоков энергии в пространстве*. При создании основ теории тепловых машин Готтфридом Лейбницем было показано, что именно *плотность потока энергии* оказывается главным фактором, который определяет возможности той или иной тепловой машины по совершению работы.

При этом была выявлена следующая принципиально важная закономерность. Оказалось, что меньшее количество энергии, которое используется при более высокой плотности, способно производить гораздо больший объем работы по сравнению с теми случаями, когда используются большие объемы энергии малой плотности.

Эта закономерность была использована впоследствии при создании *лазерных технологий*, когда поток когерентного излучения специально концентрируется в очень малых объемах пространства. Лазерные технологии уже доказали свою высокую эффективность в самых различных областях практического использования. Сегодня они представляют собой одно из наиболее перспективных направлений дальнейшего технологического развития общества. С теоретических позиций, эти ожидания вполне оправданы, так как лазерные технологии позволяют получать потоки энергии исключительно высокой плотности, которые не удастся создать никакими другими способами. Именно поэтому свои надежды получить, наконец, управляемую термоядерную реакцию современные физики во многом связывают с применением лазерных технологий.

Концентрация ресурсов во времени. Еще одним важным принципом создания высокоэффективных технологий является *принцип концентрации ресурсов во времени*. Характерными примерами использования таких технологий яв-

ляются кузнечное производство, а также все другие виды механических технологий, в которых используется энергия удара.

Изобретение молота было, по-видимому, одним из величайших технологических достижений человечества, которое позволило ему решить целый ряд сложнейших проблем в строительстве и промышленном производстве. Используется удар и в энергетических технологиях, где активно развиваются так называемые *импульсные технологии*. Они позволяют создавать высокую концентрацию энергии в течение очень малых промежутков времени, которых оказывается достаточно для того, чтобы получить полезный эффект, который не удастся достигнуть никакими другими способами.

Поэтому важным количественным признаком высокоэффективных технологий является показатель *мощности потока энергии*, который при ее использовании удастся создать в технологическом процессе. На принципиальную важность понятия мощности указывал в своих работах по теории тепловых машин еще Г. Лейбниц.

Комбинированные технологии. Технологии этого вида используют принципы *концентрации ресурсов в пространстве и времени одновременно*. Характерными примерами таких технологий являются все те их виды, в которых применяются удары заостренными поверхностями или же остронаправленные импульсы лучистой энергии. К таким технологиям относятся *фрезерование и распиливание* материалов, рубящие операции, а также операции иглой в швейной промышленности и некоторые другие.

Технологии данного вида очень эффективны. Ведь не зря же они издавна применяются в различных видах оружия. Меч и кинжал, боевой топор и копье, лук и арбалет – все эти виды оружия в течение многих столетий использовались людьми благодаря их высокой поражающей способности. Да и в настоящее время во многих видах оружия используется принцип *одновременной концентрации энергии в пространстве и времени*. Так, например, коммунитивный снаряд современной переносной ракетной противотанковой установки обладает способностью пробивать броню толщиной порядка 800 мм. Достигается это за счет того, что в самой ракете помимо взрывчатого вещества находится еще и иглообразный сердечник из закаленной стали, который буквально прокалывает броню танка, раскаленную коммунитивным снарядом.

Векторная ориентация ресурсов. Хотелось бы обратить внимание читателя еще на одну принципиальную особенность высокоэффективных технологий. Она заключается в том, что эти технологии позволяют не только создавать достаточно высокую концентрацию механического усилия или же потока энергии в пространстве и времени, но также и ориентировать их во вполне определенном *направлении*. Причем *концентрация этой направленности* также оказывается исключительно важной.

Таким образом, для того, чтобы создать достаточно эффективную технологию, мы должны прежде всего позаботиться о том, чтобы у нас имелись средства для концентрации используемых в данной технологии ресурсов *в пространстве и времени*, а также для концентрированного воздействия этих ресурсов во вполне определенном *направлении*.

7. Перспективные направления исследований в области развития информационной технологии как науки

Если же говорить о направлениях развития информационной технологии, как самостоятельной научной теории, то здесь нам представляются наиболее перспективными следующие основные направления исследований.

Создание *новых методов сжатия информации* с целью повышения уровня ее концентрации в пределах некоторых весьма ограниченных объемов пространства. При этом вполне вероятно, что может оказаться полезным введение таких новых понятий, как *«плотность информации»* и *«плотность информационного потока»*.

По аналогии с другими видами технологий, основанными на использовании энергии, можно ожидать, что *повышение плотности информационных потоков позволит получить качественно новые результаты в области целого ряда практических приложений* информационных технологий. Необходимо только будет определить значения тех пороговых уровней плотности информации, которые и позволят получить эти новые качества в тех или иных информационных системах.

2. Продолжая аналогию с энергетическими видами технологий, можно предположить, что высокoeffективными могут оказаться также и *импульсные информационные технологии*, в которых будет обеспечиваться сжатие информационных потоков не только в пространстве, но и во времени. Ведь недаром же людьми давно уже применяются различные виды «мозгового штурма», методы «глубокого погружения» и другие аналогичные способы повышения эффективности информационных процессов, как на этапах генерации новой информации, так и на этапах ее восприятия и осмысления.

При этом вполне возможно, что в арсенал научной терминологии информационной технологии, как науки, придется ввести такое новое понятие, как *«мощность информационного потока»*. Это понятие будет характеризовать *интенсивность протекания информационных процессов во времени* и, может быть, в значительной степени будет определять их эффективность.

Таким образом, при развитии информационной технологии, как научной дисциплины, весьма полезным может оказаться использование общих принципов и закономерностей, проявляющих себя в других видах технологий (механических или энергетических), а также аналогий в тех закономерностях, которые связывают их эффективность с общими принципами функционирования природных систем и, в первую очередь, объектов живой природы.

Проблема семантического сжатия информации. Можно указать на еще одно перспективное направление развития информационных технологий, которое является специфичным лишь для технологий именно этого вида. Речь идет о разработке и практическом использовании *методов «семантического сжатия» информации*. Дело в том, что для повышения эффективности использования информации ее необходимо сжимать не только в пространстве и времени, но также и в семантическом плане. Другими словами, необходимо сделать так, чтобы в результате использования того или иного вида информационной технологии формировался своего рода *«информационный конус»*, вершиной которого являлась бы основная целевая функция оптимизируемого информационного процесса.

Практическими примерами такого рода технологий могут служить процессы формирования *проблемно-ориентированных сегментов* из больших баз данных и знаний. В зависимости от цели использования такого сегмента (на-

учное исследование или же образовательный процесс) он мог бы начинаться соответственно проблемно-постановочной или же обзорной статьей по изучаемой проблеме. Затем в порядке расширения анализируемой предметной области могли бы располагаться научные статьи или обзоры, посвященные раскрытию содержания отдельных компонентов этой проблемы. И, наконец, приводилась бы информация о самых последних результатах ее исследования, заявки на изобретения и открытия в данной области, научные прогнозы.

Семантические концентраторы. Естественно, что формирование такого рода проблемно-ориентированных сегментов баз данных и знаний является делом весьма трудоемким и потребует привлечения для этих целей высококвалифицированных специалистов. Однако эффективность использования таких сегментов в научных целях, а также в системе образования может оказаться весьма значительной. Ведь сама «архитектура» формируемого таким образом массива информации содействует сосредоточению внимания пользователя на все более «плотных» участках информации, обеспечивая, таким образом, концентрацию его сознания на тех семантических направлениях, которые должны быстрее привести к решению той или иной задачи.

В то же время «коническая структура» семантических информационных сегментов позволит исследователю периодически возвращаться к исходным позициям и обозревать те или иные информационные «срезы» данной проблемы целиком, на достаточно представительном поле данных и знаний.

Информационные технологии данного вида мы предлагаем называть «*семантически концентрированными*». Можно предположить, что в будущем в процессе развития методов искусственного интеллекта и их приложений в области создания и использования информационных систем будут созданы также и специальные автоматизированные «*семантические концентраторы*». Их можно представить себе в виде программно-аппаратных комплексов, специально ориентированных на создание семантически концентрированных сегментов знаний по заданным параметрам проблемной области. Исходной информацией для работы таких комплексов, вероятнее всего, будут служить распределенные базы данных и знаний в глобальных информационно-телекоммуникационных сетях нашей планеты, которые активно формируются уже сегодня.

8. Человеческий фактор в перспективных информационных технологиях

Представляется принципиально важным, чтобы перспективные информационные технологии, которые будут широко использоваться обществом уже в начале XXI-го века, были бы изначально *ориентированы на человека*, учитывали бы его способности по восприятию информации и формированию на ее основе новых знаний. В этом плане весьма перспективными направлениями научных исследований и прикладных разработок представляются различные методы представления и использования информации в виде изображений. Это могут быть различные виды графики, картографическая информация, объемные и цветные изображения, а также различные виды анимации.

Представление информации в виде изображений является одним из наиболее эффективных методов ее сжатия в пространстве. Кроме того, зрительный канал восприятия информации человеком является наиболее широкополосным среди всех других имеющихся у него каналов получения информации. Поэтому передача информации по этому каналу может осуществляться с очень высокими скоростями и, следовательно, именно здесь могут быть достигнуты наиболее высокие показатели *мощности информационных потоков*,

необходимые для повышения эффективности информационных технологий. Ведь не зря же говорят: «Лучше один раз увидеть, чем сто раз услышать».

Таким образом, развитие методов компьютерной графики, пиктографических интерфейсов взаимодействия человека с информационной техникой, мультимедиа-технологий, геоинформационных систем, а также систем виртуальной реальности — все это актуальные и весьма перспективные направления фундаментальных и прикладных исследований для информационной технологии как нового научного направления.

Развитие этих исследований и практическое использование их результатов на базе новых поколений быстро прогрессирующей информационной техники уже в ближайшие годы может дать весьма ощутимые и социально значимые результаты в самых различных сферах человеческой деятельности. Эти результаты, безусловно, изменят весь уклад жизни и деятельности людей в новой высокоавтоматизированной информационной среде, приведут к созданию информационного общества.

9. Методологический аппарат науки как информационная технология

Изложенные выше подходы к рассмотрению основных проблем информационной технологии как научной дисциплины позволяют нам рассматривать и методологию науки как весьма своеобразную информационную технологию достаточно высокого уровня. Ведь если с позиций информационного подхода проанализировать методологический аппарат современной науки, то мы без труда обнаружим в нем все основные характерные признаки информационной технологии.

Действительно, здесь присутствуют и функции *сжатия информации*, которые выполняет используемый в науке аппарат формализованного представления знаний в той или иной предметной области. Наглядным примером такого аппарата является математика. Ведь одним из самых значимых ее достижений является возможность представления весьма сложных зависимостей в достаточно компактном виде. Именно это позволяет исследователю целиком обобщать те или иные фрагменты изучаемого явления, анализировать его возможные граничные состояния и делать в результате этого свои умозаключения.

Характерным примером здесь может служить математический аппарат синергетики, где разработан и широко применяется метод представления основных характеристик самоорганизующихся систем в фазовом пространстве. Анализируя возможные траектории поведения системы в этом пространстве, представленные в виде так называемых *аттракторов*, исследователь сразу же концентрирует свое внимание на важнейших параметрах, от которых и зависят по существу возможности того или иного пути развития этой системы (в синергетике они называются *параметрами порядка*). При этом из его поля зрения исключаются практически все второстепенные факторы процесса функционирования системы.

Что же это такое, если не *семантическая концентрация информации*? Таким образом, здесь мы также видим явные признаки и свойства информационной технологии.

Нам представляется, что анализ методологического аппарата науки с точки зрения информационной технологии как научной дисциплины, может оказаться весьма полезным не только для науковедения, но также и в мето-

логическом плане. Ведь такой подход принципиально позволяет определять наиболее перспективные направления развития методологического аппарата самой науки. Плодотворным здесь может оказаться также и сравнительный анализ эффективности этого аппарата в различных областях научного знания, который мог бы дать дополнительную ориентацию для их развития.

Следовательно, формирование информационной технологии как самостоятельного научного направления может оказаться весьма полезным и для развития самой науки в части дальнейшего совершенствования ее методологического аппарата.

Уровень развития технологий сегодня характеризует не только развитие экономики той или иной страны, но также и ее место в мировом сообществе. Уже в ближайшем будущем следует ожидать создания и распространения принципиально новых производственных, социальных и информационных технологий. Их эффективность будет превышать современный уровень уже не на проценты, а на десятичные порядки. Поэтому такие технологии в научной и общественно-политической литературе часто называют «прорывными», имея в виду, что их появление будет означать «прорыв» общества на качественно новый уровень технологического развития.

Нет никакого сомнения в том, что определяющую роль в осуществлении такого «прорыва» будут играть информационные технологии. Именно поэтому уже сегодня необходимо прилагать усилия для того, чтобы среди технических наук своевременно была сформирована *новая научная дисциплина – информационная технология*, которая должна будет стать научной базой для информационно-технологического направления дальнейшего развития цивилизации.

Хотелось бы надеяться, что данная работа послужит еще одним поводом для научной дискуссии о проблемах развития этого нового направления.

Литература

1. Краткий словарь современных понятий и терминов. – 2-е изд. – М.: Республика, 1995– 510 с.
2. *Колин К.К.* Информационные проблемы социально-экономического развития общества. – М.: Союз, 1995. -72 с.
3. *Колин К.К.* Информационная технология как научная дисциплина.// Информационные технологии. №2, 2001. – С. 2-10.
4. *Зацман И.М.* Концептуальный поиск и качество информации. – М.: Наука, 2003. – 271 с.
5. *Гвардейцев М.И., Кузнецов П.Г., Розенберг В.Я.* Математическое обеспечение управления. Меры развития общества. М.: Радио и связь, 1996.

Уровни изучения искусственного интеллекта

А.Ю. Алексеев, кандидат философских наук

1. Философия искусственного интеллекта – философия программиста

Многие студенты технического вуза считают курс философии «стихийным бедствием». Бороться с философией бесполезно, с ней следует считаться, правдами-неправдами пережить и забыть. Студентов можно понять. Сегодняшняя философия большинства технических вузов, как правило, примеряет категориальный «кафтан» марксистско-ленинской философии, изнашивавшийся несколько десятилетий, к современным социокультурным и научно-технологическим реалиям. Однако ход времени неумолим – на смену старой философии приходит новая философия, отвечающая современным социокультурным реалиям. На кафедрах философии началась «война» философских миров.

С одной стороны противостояния – диалектический материализм. Несомненно, что методологический аппарат «марксоидной» философии в какой-то степени полезен для практических ориентиров студента. Он способствует инициации его рефлексии над основаниями науки и техники. Но туманность диалектических категорий и метафизических законов типа «спиралей», отдалённое отношение к методологии как специальной философской дисциплине – всё это вызывает отторжение у молодого человека. Мировоззренческие задачи философии вообще отданы на откуп простому любопытству с наивно-воспитательной пропагандой философствования как антропологической константы человеческого бытия.

С другой стороны философской диспозиции – постпозитивистская традиция, изначально руководствующая достижениями науки и техники и непосредственно способствующая им, впитавшая в себя философские наработки XX века в области анализа языка, феноменологии, герменевтики, экзистенциализма, структурализма и др. течений.

Исход битвы с «марксиянами» предсказать несложно. Положительным фактом в связи с этим представляется будущая замена аспирантских экзаменов кандидатского минимума по философии «вообще» на конкретно специализированные экзамены по методологии той науки, которую изучают молодые специалисты. Свобода идей, достигнутая благодаря официальному ниспровержению старой философии (при всём том, что она превосходно поддерживала социалистический строй), способствовала появлению в нашей стране учебных курсов по прикладным философиям, непосредственно связанных с конкретными смысложизненными запросами человека либо запросами профессионального, бытового, религиозного, научного плана. Появились социальная философия, философия политики, философия морали, философия физики, философия биологии, философии географии, философия математики и т.п. В этом ряду с необходимостью должна возникнуть специальная дисциплина – философия для программиста.

На наш взгляд, интегральную, мировоззренческо-методологическую роль, инициацию на изучение смысложизненных проблем исходя из целей и задач освоения программистской профессии, способна выполнить *философия искусственного интеллекта* (ФИИ) – особая сфера философского знания и методологической междисциплинарной деятельности, возникшая на волне постпозитивизма в ряде англоязычных стран в последнее десятилетие.

При этом следует уточнить понятие «программист». Техногенный (точнее, компьютерогенный) стиль жизни современного человека обуславливает расширенную трактовку этого понятия. Сегодня под программистом понимают не узкого специалиста, способного кодировать алгоритмы. Любой современный человек, экзистенциально «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 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.

Ряд современных мыслителей рассматривает ФИИ в составе методологии компьютерной науки. Такой взгляд представляется слишком узким, ограниченным логико-эпистемологическими рамками. ФИИ следует понимать шире – она имеет непосредственный выход к фундаментальным социокультурным вопросам и проблематике сознания, аналитически чётко формулируя ключевые аспекты основного вопроса философии – о соотношении сознания и бытия – актуальность которого не снимается (в чем убеждают многие отечественные специалисты), а повышается в условиях информационного общества.

Сегодня в большинстве зарубежных вузов в рамках дисциплины Computer Sciences стали изучать курсы по философии искусственного интеллекта (The Philosophy of Artificial Intelligence). Форма преподавания ФИИ различна. Либо самостоятельный курс, читаемый в течение всего учебного года, либо четыре – десять лекций, предваряющих основной курс по программированию ИИ. Курс ФИИ, помимо программистов, преподается социологам, психологам, биологам, физикам, политологам, журналистам и студентам многих других специальностей. В нашей стране такие курсы пока не предусмотрены. Это объясняется тем, что проблематика ФИИ до 90-х годов велась крайне слабо, так как «буржуазные» теоретические подходы к пониманию не только искусственного, но и естественного интеллекта не согласовывались с официальной марксистско-ленинской философией. В начале перестройки данная тема вообще заглохла по вполне понятным причинам. И это в то время когда в англо-американской науке ФИИ отпочковалась в самостоятельную область философского знания!

Сегодняшняя студенческая конференция и призвана в какой-то мере после сорокалетнего забвения нашей наукой методологических оснований искусственного интеллекта напомнить «взрослой» философии о необходимости реанимации данной проблематики.

2. С чего следует изучать философию искусственного интеллекта?

Принято считать, что термин «искусственный интеллект» впервые был применён летом 1956 г. выдающимся американским философом, логиком, программистом Джоном Маккарти. Несколько позже Маккарти разработал язык LISP, широко применяемый в программировании систем ИИ. Также считается, что Дж. Маккарти впервые ввёл в обиход и термин «философия искусственного интеллекта». Под ФИИ он понимает специфическую область методологических исследований, подобную прикладным сферам философского знания типа: «фи-

лософия биологии», «философия физики». Специфику философии ИИ Маккарти представляет следующим образом: любая система ИИ требует концептуально-методической и программно-информационной реализации философских положений, в основном, эпистемологического плана. То, что философия ИИ – это *философия программиста*, Маккарти неоднократно подчёркивает – начало философской рефлексии 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. Выучите какие-нибудь языки программирования, например, Java или C++, они сейчас хорошо востребованы. Читайте больше работ по психологии и физиологии нервной системы»¹.

С учетом того, что было ранее сказано о расширенной трактовке понятия программиста, ФИИ следует изучать комплексно, выделив многомерную структуру предметной области, не ограничиваясь при этом «изучением языка программирования».

В общей философии изучение предметной области принято разбивать на уровни. Выделяются эмпирический, теоретический, методологический, экзистенциальный уровни. По аналогии с этим выделим уровни изучения проблематики ИИ: программно-инженерный, методологический, теоретический, экзистенциальный.

3. Программно-инженерный уровень изучения ИИ

Программно-инженерный (эмпирический) уровень – это уровень конкретных научно-практических разработок интеллектуальных систем и интеллектуальных информационных технологий. В инженерии ИИ выделяют следующие направления²:

1. Игры (Игра в шахматы, Samuel, 1963);
2. Доказательство теорем и автоматизация рассуждений (automated reasoning): «Логик-теоретик» Ньюэлла (Newell's Logic Theorist (LT, 1956);
3. Решение проблем и планирование: а) «Универсальный решатель проблем» (Ньюэлл, Шоу и Саймон (Newell, Shaw and Simon's General Problem Solver – GPS, 1963); б) программа СТРИПС Нильсона и Файка (Nilsson & Fike's, Stanford Research Institute Problem Solver (STRIPS, 1971), созданная для управления мобильным роботом, названным Шейки (Shakey);
4. Восприятие:
 - 4.1. Техническое зрение: а) анализ письменного текста; б) распознавание изображений (image processing); в) анализ сцен; г) слежение за движением (motion tracking); д) распознавание лиц (face recognition);
 - 4.2. Распознавание речи: а) распознавание отдельных выражений (isolated-word recognition); б) распознавание связной речи (continuous-speech recognition): программа Рэдди («Слух») – Reddy's HEARSAY II (1975 г.);
 - 4.3. Распознавание образов (pattern recognition);

¹ Маккарти, Дж, 1995 г. «Что общего у философии и ИИ?».

² Jonathan Mohr. COMPUTING SCIENCE. Introduction to Computing Science. COURSE MATERIALS, 2000 (Курс «Введение в компьютерные науки»).
(http://WWW.AUGUSTANA.CA/~mohrj/courses/2000.fall/csc110/lecture_notes/AI.html)

5. Понимание естественного языка и машинный перевод (Natural language understanding and machine translation): программа Винограда (Winograd's) SHRDLU (начало 1970 г.);
6. Экспертные системы: а) символическая логика (symbolic mathematics): Мозис, МАКСИМА (MACSYMA, Moses, 1977 г.); б) медицинская диагностика (medical diagnosis) МУСИН Шортлиффа и Бачмена: MYCIN (Shortliffe and Buchanan, 1976); в) химический анализ (chemical analysis): DENDRAL Фейгенбаума и Леденберга, 1964 (Feigenbaum and Lederberg); геологическая экспертная система: PROSPECTOR Дуда и Харт (Duda and Hart, 1979);
7. Обучение: 1) программа открытий (discovery programs): Ленат (Lenat) – автоматический математик (Automated Mathematician, AM, 1977); нейронные сети (neural networks);
8. Робототехника.

На инженерно-программном уровне трудно дать вразумительное 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. Несомненно, здесь мы имеем не критичное понимание ИИ.

4. Методологический уровень изучения ИИ

На методологическом уровне изучения программно-инженерные разработки подводятся под определение ИИ-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 ведущими специалистами в этой области¹:

Ленат и Фейгенбаум, 1991: Интеллект – это способность настолько быстро находить правильное решение в обширном пространстве поиска, что для стороннего наблюдателя эта способность кажется априорной.

Марр, 1977: Цель ИИ – идентификация значимых проблем обработки информации и решение этих проблем.

Маккарти, 1988: ИИ имеет дело с методами достижения целей в сложных ситуациях и проблемами, представленными в форме ситуаций. Методы решения не зависят от того, кем они решаются – человеком, марсианином или компьютерной программой.

Минский, 1985: Интеллект означает способность решать сложные проблемы.

Ньюэлл и Саймон, 1976: Выражением «универсальное интеллектуальное действие» обозначаются те характеристики интеллекта, которые обнаруживаются в человеческом действии: т.е. в любой реальной ситуации пове-

¹ По Wang Pei, 1995. On the Working Definition of Intelligence, P.1-5
<http://www.cogsci.indiana.edu/pub/wang.intelligence.ps>

дение стремится к определённой системной цели и приспосабливается с определённой скоростью к выявляемым требованиям среды определённой степени сложности.

Шэнк, 1991: Интеллект означает достижение лучшего за приемлемое время.

Янг (Wang Pie), 1995: Интеллект – это способность информационной системы адаптироваться к среде в условиях нехватки знаний и ресурсов.

Выидные отечественные учёные дают следующие определения:

Г.С.Осипов: ИИ – это наука о том, как усилить человеческие возможности в решении сложных задач.

В.К.Финн: ИИ – это системы, работающие на знаниях и включающие логический решатель задач.

Можно привести ещё ряд определений ИИ. Очевидно, что понятию ИИ присуще многообразие трактовок. Какое из них более приемлемо? В силу того, что никто не может дать определение понятия интеллекта на «все случаи жизни», они все заслуживают внимания, но с известной долей скепсиса по отношению самой возможности определения ментальных терминов, к которым относится и термин «интеллект».

5. Теоретический уровень изучения ИИ

На теоретическом уровне осуществляется систематическое изучение понятия искусственного интеллекта. С этого уровня начинается собственно философская рефлексия над проблематикой ИИ. Можно предложить следующий план теоретического изучения ИИ.

I. Введение в ФИИ

Уровни изучения ИИ: инженерно-технологический, научно-теоретический, методологический, экзистенциальный. *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 праксеологической 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.

II. Критическая философия искусственного интеллекта

Основной вопрос ФИИ: «Может ли машина мыслить?». Тест Тьюринга (ТТ) и функционалистская парадигма мышления. Преодоление онтологиче-

ской 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 неотличимости). 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», семиотико-синтаксическое. Интенциональность и вычислимость. Машина мыслить может, но понимать не может. Человек как мыслящая и понимающая машина. Многообразие интерпретаций МЭ Дж. Серля. Современный этап развития когнитивной науки как попытка разрешения проблемы «понимания». 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 бихевиоральной парадигмы. Антибихевиоральная критика глобального ИИ и психофункционализм Н. Блока. Тест и машина Н.Блока как критика репрезентативной трактовки глобального ИИ. МЭ «Китайская комната» 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 Тьюрина 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.

III. Положительная философия ИИ.

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 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.

6. Экзистенциальный уровень изучения ИИ

На экзистенциальном уровне подчеркивается необходимость привлечения проблематики ИИ для получения, в первую очередь, определённости понятия естественного интеллекта. Неизбежно в изучение вовлекаются связан-

ные 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 и ярким примером задач экзистенциального уровня проблематики ИИ может послужить систематизация познавательных ситуаций, предложенная Д.И. Дубровским в ряде работ в связи с анализом феномена веры¹. 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 познавательной 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.

Вывод. Философии искусственного интеллекта представляется не только прикладной философией, конкретизирующей общеполитическую проблематику в контексте частной методологии научной и профессиональной деятельности программиста. ФИИ может выступить в роли специальной отрасли философии, активизирующей выработку у студента мировоззренческих ориентиров, выполняющая при этом интегральную мировоззренчески-методологическую функцию в

¹ Дубровский Д. И. Обман. Философско-психологический анализ. М., 1994; Дубровский Д.И. Проблема идеального. Субъективная реальность. — М., 2002. — 368с., с. 297-307

ходе изучения курса философии и применения философских знаний в профессиональной практике и своей жизнедеятельности.

7. Студенческая конференция по философии ИИ как пример многоуровневого изучения проблематики искусственного интеллекта

В данном сборнике представлены тезисы докладов студентов МИЭМ, отобранные из массива работ, поступивших в Оргкомитет конференции. В основном, работы выполнены на базе переводных статей англо-американских мыслителей. «Философия искусственного интеллекта» - это *terra incognita* не только для российского студента, но и преподавателя, учёного, специалиста более старшего поколения. В стране нет требуемой литературы. Переводы работ зарубежных философов ИИ и реанимация трудов отечественных исследователей – дело, не выгодное для издательств. Несмотря на наличие курсов по ИИ в большинстве вузов, где готовят программистов, в этих курсах совершенно не отражаются вопросы проблематики философии и методологии ИИ, то есть преподавание ИИ не выходит за рамки обозначенного нами инженерно-программного уровня изучения ИИ. В Сборнике, напротив, этот уровень изучения не нашёл отражения. В нём представлены, в основном, теоретические и методологические положения ФИИ.

Особый интерес 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) (См. доклады А. ЖУКОВА «Базовые положения Теста Тьюринга», Д. КОМАРОВА «"Может ли машина мыслить"? Полемика стандарт Тьюринга», С. ЛИЗОРКИНА «"Искусственный интеллект" Алана Тьюринга»).

Большинство докладов посвящено различным модификациям ТТ. Чтобы раскрыть важность этих докладов 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 панкибернетизма.¹ В 1963 г. академик А.Н. Колмогоров, в недалёком прошлом (до 1957 г.) ярый противник кибернетики как науки, пишет статью «Автоматы и жизнь» 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 серьёзно?». В ответ академик С.Л. Соболев пишет работу «Да, это вполне серьёзно!».² Завязалась серьёзная полемика, в рамках которой обсуждались вопросы возможности/невозможности естественно-научного определения таких по-

¹ См. исследования американского историка науки: Грэхэм Л.Р. Естествознание, философия и науки о человеческом поведении в Советском Союзе: Пер. с англ. – М. Политиздат, 1991. – 480 с. – С. 266-291. Волна кибернетизации началась в 1961 г. после публикации книги под редакцией академика А.И. Берга «Кибернетика на службе коммунизма». Панкибернетизм – это мировоззренческая позиция, согласно которой кибернетика способна решить все проблемы социалистического управления и хозяйства и ни одна другая страна не сможет использовать кибернетику так же эффективно, как СССР.

² Колмогоров А. Автоматы и жизнь // Возможное и невозможное в кибернетике. М., 1964. С.10; Бялик Б. Товарищи, вы это серьёзно?; Соболев С. Да, это вполне серьёзно! // Там же.

нений, как воля, мышление, эмоции и др.; противоречия/соответствия материализму 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 настоящего времени в отечественной печати вообще не было работ, посвященных этой теме.

В англо-американской же философии этот спор не угасал. Особо усилились дискуссии в последнее десятилетие. Это обусловлено тем, что: 1) развитие компьютерной техники привело к реализации многих предположений Тьюринга; 2) в области методологии появились крупные теоретические разработки в области построения самообучающихся программ, реализации алгоритмов «как бы» волевой мотивации, «эмоционального» 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. 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 (см. Д. ЗВОРЫКИН «Зачем вкладывать деньги в Тест Тьюринга? Лойбнеровская премия»; А. КЛОПКОВ «Программирование Теста Тьюринга»). Обсуждались вопросы о том, является ли ТТ тестом на интеллект, возможно, сам Тьюринг преследовал целью не определение интеллектуальности системы, а нечто иное (см. И. РЫБИН «Социокультурные аспекты Теста Тьюринга»; Ю. ЦВЕТКОВ «Тест Тьюринга и паранойя»).

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

Итоги работы секции № 1 подвёл проф. Д.И. ДУБРОВСКИЙ. В выступлении он акцентировал важность темы и отметил о концептуальной возможности расширения понятия теста Тьюринга. Это связано с важной теоре-

тической проблемой современной философии сознания – проблемой диагностирования системы на предмет обладания ею субъективной реальностью.

Работа секции № 2 «Искусственный интеллект и «здоровый смысл» была посвящена, в основном, изучению работ выдающего математика, философа, программиста Дж. Маккарти, который впервые вводит 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 «Псевдоволя»); проблемы логики (М. КРАСИВСКАЯ «Логические аспекты создания искусственного интеллекта»). Был представлен ряд докладов, посвящённых реализационным ориентирам позиции «здорового смысла» (см. М. ГРИШКИН. Реализационные перспективы теории речевых актов»; М. ЛАПИН «Экспертные системы, основанные на здоровом смысле»; Д. РОДИОНОВ «Проблема дискурса искусственного интеллекта. Конструкторская позиция»).

В работе секции № 3 «Историко-философские перспективы компьютерного моделирования» представлены исследования, посвящённые возможностям применения 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 проблему идеального, которая двадцать пять лет назад вызвала ожесточённые споры и определила оппозицию В.Э. Ильенкова и Д.И. Дубровского (см. А. ДРОБЯЩЕНКО «Проблема идеального и искусственный интеллект»); рассмотрение лингвистической проблематики (М. КОРОЛЕВ «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». Студент подготовил интереснейший доклад, связав воедино разноплановые темы ИИ, философии сознания и социально-философские аспекты.¹

К сожалению, работа секции показала, что, несмотря на энтузиазм студентов, «искусственный» характер попыток применения «древних» философских положений. Философия ИИ должна базироваться на современных новейших философских положениях и научно-технологических достиже-

¹ Хочется отметить, что во время редактирования данного Сборника в Оргкомитет конференции обратилось несколько профессоров из разных городов страны с просьбой сообщить координаты Т. Кураевой в плане знакомства с темой зомби. Факт любопытен – учёный обращается к студенту с просьбой о консультации. Это подчёркивает как важность настоящей конференции, так и то, что для многих отечественных философов и учёных философия ИИ – это «непознанная земля».

ниях. Иначе дискурс получается расплывчатым и слабо связанным с существом вопроса. На «Канте» искусственный интеллект не построить!

Работа секции № 4 «Философия искусственного интеллекта и компьютерная технология была ориентирована на методологические вопросы применения ИИ. Рассматривались вопросы приложений ИИ в технике (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. ПОДОПРИГОРА «Естественно-языковой 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. СМЕРНОВА «Клоны и «полуискусственный интеллект»; А. БОНДАРЬ «“Квалиа” как базовая категория витруалистики»; В. Крючков «Что значит “Быть роботом”»?).

* * *

По степени проявленной студентами активности можно судить о том, что молодым интересны идеи рефлексии над проблематикой ИИ. Душная атмосфера бессмысленного заучивания гегелевских категорий «марксистской» 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.

I. ФУНКЦИОНАЛИСТСКАЯ КОНЦЕПЦИЯ МЫШЛЕНИЯ. ТЕСТ ТЬЮРИНГА: PRO ET CONTRA

Функционализм, физикализм и искусственный интеллект

Анна Алексеева (М-08-03)

Функционализм – одно из направлений в философии сознания, рассматривающее связи между ментальным и физическим в виде функциональных отношений. Становление функционализма связывается с именами Патнэма, Райла, Дэвидсона, Деннета. В отечественной философии – Д.И. Дубровского. В основании функционализма как состоявшейся парадигмы лежит концепция мышления, предложенная А.Тьюрингом в форме теста. Парадигма функционализма была сформулирована в терминах «теории тождества» Патнэма: определённые типы ментальных состояний отождествляются с определённым типом физических состояний, а последние, в свою очередь, могут быть реализованы различными носителями – другим человеком, машиной, инопланетянином и т.п. (Патнэм, «Сознание и машины», 1960).

В настоящее время функционализм – как методологическое обоснование философии искусственного интеллекта – одно из мощнейших философских течений. Представляет интерес рассмотрение данной парадигмы в сравнении с физикализмом – основным оппонентом функционализма в борьбе за адекватную теорию сознания.

Физикализм ориентируется на онтологические и метафизические допущения в решении двух ключевых вопросов: 1) онтологического – что такое «факт» (ощущение, чувственное данное); 2) метафизического – что даёт каждому типу ментальных состояний свою собственную идентичность (к примеру, что вызывает конкретную боль). Физикализм – монистичен. В отличие от дуализма, который утверждает наличие двух субстанций – духовной и материальной, он утверждает наличие только физической субстанции. В бихевиоризме, например, боль есть нечто поведенческое.

Функционализм отвечает лишь на метафизический вопрос и его совершенно не интересует онтология. Он утверждает, например, что то, что вызывает боль – суть функция. Однако он не утверждает, будет ли сущность боли включать какие-либо физические составляющие. Данный подход может быть описан в терминах автомата. Для того, чтобы быть автоматом того же типа, который задан в описании, конкретной машине требуются описания состояний, которые относятся: 1) друг к другу, 2) к входу и 3) к выходу. То есть чтобы создать ИИ требуется описать способ функционирования естественного интеллекта и представить его в терминах автоматной таблицы. Такое описание не сообщает нам из чего машина сделана. Например, не исключается и машина, которая работает под действием нематериальной души, только душа при этом должна действовать детерминировано в соответствии со

способом, описанном в таблице. Здесь функционализм согласуется не только идеализмом, но и со спиритуализмом – основными противниками физикализма.

Размышляя о дистинкции между функционализмом и физикализмом, полезно различать две категории физикалистских тезисов. Одна версия физикализма – *строгого физикализма* - конкурирует с функционализмом, делая метафизическое утверждение о физической природе ментальных свойств или типов свойств (последний физикализм часто называют «типизированным»). Как упомянуто выше, данную позицию функционализм принципиально не признает.

Имеется более скромный, физикализм, предметная область которого в большей степени онтологическая нежели метафизическая. Не все утверждения такого физикализма несовместимы с функционализмом. К примеру, в рамках *слабого физикализма* есть версия, утверждающая, что мыслящая вещь – суть организованная материя и способность к сознанию конституируется таким же образом, каким, например, формируется требуемый металлический предмет в процессековки металла. Здесь большинство функционалистов может перебежать в лагерь физикалистов. Также и функционализм может быть модифицирован в физикалистском направлении, к примеру, требуя, чтобы все свойства, квантифицированные функциональными дефинициями, были физическими свойствами.

Как мы видим, у основных врагов в области теории сознания – физикализма и функционализма – есть точки соприкосновения и возможности к перемирию. Почему же большинство исследователей ИИ обращается к функционализму, а не к физикализму? Потому что: 1) в функционализме способ формулировки проблемы сознания органичен корпусу дисциплинарных знаний специалиста в области компьютерной науки; 2) функционализм не задается вопросом о квалиа – для него не важно феноменальное переживание сознательного опыта; 3) он позволяет делать концептуально корректные выводы о возможности реализации сознания на разных «носителях сознания», включая компьютерный базис; 4) даёт намного больше надежд в возможности реализации ИИ, нежели физикализм:

Создание искусственного интеллекта для функционалистов не представляет неразрешимой проблемы, главное – написать «правильную» программу!

Тест Френча: субкогнитивистское опровержение концепции Тьюринга

Алексей Денисов (ЭК-82)

В ряду опровержений функционалистской концепции мышления особое место занимает тест Р. Френча (ТФ), предложенный им в 1990 г. в статье «Субкогнитивные способности и границы Теста Тьюринга». Тест Френча обосновывает невозможность программирования Теста Тьюринга (ТТ) в силу социокультурных причин.

ТТ, считает Френч, – это высококачественный тест на интеллект, искусно обошедший бескрайнее философское болото проблемы дух/тело (mind-body problem). В этом видится философская заслуга Тьюринга. Однако ТТ не

способен исследовать глубинные и наиболее существенные области человеческого интеллекта. Поэтому ТТ фактически бесполезен как тест на интеллект. ТТ способен пройти лишь тот, кто живет и ощущает окружающий мир так же, как и люди. ТТ – это не тест на интеллект «вообще», а тест на социокультурно-обусловленный человеческий интеллект.

Для обоснования своего подхода Френч предлагает метафору «Тест Чайки».

Тест Чайки.

Познакомьтесь с притчей. Так случилось, что обитателям одного из скандинавских островов был известен лишь один вид существ, способных летать – чайки. Все на острове считают, что только чайки могут летать. Однажды двух местных философов подслушали во время спора по поводу определения сущности полета.

Первый говорит: «Суть полета в том, чтобы двигаться сквозь воздух».

«Но ты же не назовешь это полетом» – ответил второй, бросив булыжник с берега в море.

«Ну, тогда оставаться в воздухе на определенное количество времени».

«Да, но тучи, дым, детские воздушные шары пребывают в воздухе достаточно долго. Я, к примеру, могу в ветреный день удержать в воздухе воздушного змея столько долго, сколько захочу. Мне кажется, в полете есть нечто большее, нежели пребывание в воздухе».

«Тогда полёт обязательно включает наличие крыльев и перьев у тех, кто летает».

«У пингинов есть и то и другое, но мы-то хорошо знаем, как они летают...».

И так далее. Наконец они пришли к решению вопроса, в сущности, уйдя от него. А получилось это из-за того, что с самого начала они договорились: единственный пример объекта, способного летать – это чайки, живущие на острове. И в этом мнении они абсолютно убеждены.

На основе этих допущений, ознакомления со знаменитой статьей Алана Тьюринга о тесте для определения интеллекта, после долгих раздумий мудрецы наконец-то придумали Тест Чайки для определения полета. Этот тест стал считаться строгим достаточным условием полета. Впредь, если кто-то скажет «я изобрел машину, которая умеет летать», вместо применения какого-нибудь набора критериев, характеризующих полет, машину изобретателя подвергнут Тесту Чайки. *Абсолютно достоверно*, что тот, за кем признаётся способность к полёту – это тот, кто пройдет Тест Чайки. С другой стороны, если кто-то проваливает тест, они не выносят решения, может он летать или нет.

Тест Чайки работает во многом, как и ТТ. У наших философов в распоряжении два экрана стереоскопических радаров: один отслеживает настоящую чайку, другой – машину с её предполагаемой способностью летать. Есть также судьи – это философы: машина прошла тест на способность летать, если оба философа совершенно не смогли отличить чайку от машины.

Может быть выдвинуто возражение, что некоторые из проверок (например, способность нырять на лету) не имеет отношения к полету. Философы на это ответят: «Ну и что? У нас есть абсолютное достаточное условие полета, а не *минимальное* достаточное условие. Конечно, мы понимаем, что наш тест очень трудно пройти. Но, изобретатели летающих машин, будьте спокойны, то, что вы не выдержали тест, ещё ничего не доказывает. Мы не будем утверждать, что ваша машина *не может* летать, если она не проходит Тест Чайки, она может оказаться еще как способной на это, тем не менее, мы, философы, хотим быть абсолютно уверены, что имеем дело с абсолютно достоверным случаем полета; мы имеем единственный доказанный факт полета и теперь абсолютно уверены, что ваша машина летает, если она способна пройти Тест Чайки».

Теперь, конечно, Тест Чайки справедливо исключит пули, мыльные пузыри и снежки из соревнования. Это справедливо. Однако вертолеты и реактивные самолеты, которые *летают*, тоже никогда не пройдут тест. По той же причине не будут летать летучие мыши или пчелы, альбатросы или колибри. На самом деле, при более внимательном рассмотрении, возможно, только чайки и пройдут Тест Чайки, причём только те, которые обитают на острове, где живут философы. В таком случае, то, что мы имеем – это не тест на способность к полету вообще, а скорее тест на способность к полету скандинавской чайки.

Выводы из данной метафоры применительно к ТТ очевидны: тестируемый объект может быть чрезвычайно интеллектуальным, но если он не отвечает на вопросы точно так же, как отвечает на них человек, то он не пройдет ТТ. Френч уверен, что *единственный* путь, который даст этому объекту возможность отвечать на вопросы совершенно как человек – это жить и ощущать мир так, как живёт и ощущает мир человек. Таким образом, то что у нас есть – *это не тест на интеллект вообще, а тест на интеллект, применяемый в жизни конкретным человеком.*

Для доказательства бесполезности ТТ следует задать и человеку и машине ряд т.н. субкогнитивных вопросов.

Субкогнитивный вопрос – это любой вопрос, позволяющий идентифицировать низкоуровневую когнитивную структуру (т.е. структуру, расположенную на подсознательном уровне). Пример низкоуровневой когнитивной структуры – подсознательная сеть ассоциаций. Данная сеть состоит из огромного количества сильно переплетенных образов, которые воспринимались в прошлом и которые могут быть активированы в любой момент. Образы соединяются в статистические ансамбли, которые образуются в силу ряда социокультурных обстоятельств жизнедеятельности человека.

Френч предлагает метод идентификации подсознательной сети ассоциаций: «метод ассоциативных заучиваний».

Метод ассоциативных заучиваний состоит в следующем. У человека в процессе его конкретной жизнедеятельности развиваются определенные ассоциативные связи между понятиями (ассоциативную сеть понятий иногда называют семантической сетью). Сила связи между понятиями – величина варьируемая. С помощью т.н. заданий на распознавание понятий (lexical decision) можно выявить степень связанности понятий. Это осуществляется путем оценки времени, в течение которого происходит связывание одного понятия с другим. Если произносится слово «хлеб», то с ним быстрее свяжется понятие, обозначаемое словом «масло», нежели со словом «собака».

Под *тестом Френча* понимается ТТ, который состоит из подобного рода субкогнитивных вопросов. Судья (из теста Тьюринга) может воспользоваться методом ассоциативного заучивания следующим образом. За день до тестирования судья проводит контрольный опрос среди интервьюируемых (людей). Для этого судья подбирает произвольный набор слов, выдает интервьюируемому задание на распознавание понятий и фиксирует среднее время связывания понятий. На следующий день при проведении теста судья задаёт кандидатам (игрокам из ТТ) тоже самое задание, собирает результаты тестирования, учитывает результаты контрольного опроса. Судья сможет достаточно чётко определить, кто из кандидатов – машина, а кто – человек. Чело-

веком будет тот, чьи результаты окажутся более похожими на усредненный результат контрольного опроса.

Машина будет неизменно проваливать ТФ, так как невозможно представить некий априорный способ определения ассоциативной степени связывания понятий в совокупности всех возможных понятий. Есть только единственный способ различения машиной всей данной совокупности ассоциативных сил между понятиями – это окунуться в ход конкретной жизнедеятельности людей.

По поводу ТФ возникает ряд вопросов и возражений.

1. Вопрос: Возможно ли модифицировать ТТ таким образом, чтобы запретить субкогнитивные вопросы?

Ответ: Нет.

2. Возражение: Субкогнитивные вопросы задавать нечестно. Их следует запретить.

Ответ: отсутствует способ демаркации субкогнитивного вопроса от когнитивного. Запретить субкогнитивные вопросы просто невозможно.

Доказательство: Предлагается класс вопросов, которые только вначале кажутся «когнитивными». Если их исследовать детальнее, то на самом деле каждый такой вопрос окажется зависимым от подсознательных и бессознательных механизмов. Тщательное исследование оригинальных вопросов Тьюринга показывает, что все они – тоже субкогнитивные, «подсознательные». То же самое характерно и для целого класса вопросов, составляющих ТТ. В ТТ неизбежно встретятся вопросы, ответы на которые опираются на подсознательные ассоциации. То есть невозможно отделить «подсознательные» вопросы от тех, которые таковыми не являются. Следовательно, сознательный и подсознательный уровни переплетены самым сложным и невероятным образом.

Вывод: В тесной взаимосвязанности субкогнитивного (подсознательного) и когнитивного уровней заключена причина, делающая ТФ тестом только на человеческий интеллект, но не на интеллект «вообще».

Френч предлагает ещё ряд методов. Все они приводят к *субкогнитивистскому опровержению ТТ*, так как любая обычная совокупность вопросов ТТ неизбежно содержит в себе субкогнитивные вопросы в той или иной форме. Следует только задавать достаточно большое число таких вопросов. И даже, несмотря на то, что кто-то может преуспеть в программировании определенного числа ассоциаций (например, задавая людям вопросы, подобные задаваемым судьёй и быстро программируя ответы), реакция кандидата-человека – это мгновенный результат несчётного количества субкогнитивных действий, и до тех пор пока машина не будет иметь набора ассоциаций, сходного как по уровню, так и по типу с тем, что использует человек, результаты в ТФ будут непременно в большей степени отличаться от усредненных показателей интервьюируемых, нежели показатели кандидата-человека. Такой тест пройти будет более сложно, чем оригинальный ТТ, пожалуй, невозможно.

Базовые положения теста Тьюринга **Алексей Жуков (ИС-82)**

Пятидесятилетний юбилей теста Тьюринга (ТТ) был широко отмечен в философии и науке, большей частью англо-американской. Тест был предложен в 1950 г. выдающимся английским математиком и философом Аланом Тьюрингом в статье «Вычислительные машины и интеллект». Вышли сотни статей и объемных монографий, посвященные ТТ. В отечественной литературе данная тема, к сожалению, не получила должного освещения. Этим обуславливаются во многом и проблемы в становлении и развитии отечественного искусственного интеллекта. Перевод статьи А.Тьюринга (под редакцией Б.В. Бирюкова) вышел в 1960 г., долгое время был библиографической редкостью и стал переиздаваться лишь сегодня. Тем не менее в остальном мире ТТ – одна из наиболее важных тем в философии искусственного интеллекта, философии сознания, когнитивной и компьютерной науке.

Популярность ТТ объясняется как научными, так и философскими положениями. Научный интерес к ТТ обусловлен тем, что научному сообществу предоставилась возможность организовать рациональный дискурс на вопрос, чарующий человека столетиями: «Можно ли создать мыслящую машину?». Проявляя неопозитивистский стиль демаркации суждений на научные/псевдонаучные, А.Тьюринг посчитал вопрос «Может ли машина мыслить?» бессмысленным. Невозможно, по его мнению, внятно рассуждать не только по поводу того, что такое «мышление», но и по поводу того, что такое «машина». И предлагает такому псевдонаучному вопросу замену – тест на мышление, отвергающему метафизические спекуляции.

Философский успех обусловлен тем, что ТТ положил начало мощнейшему направлению в современной философии, известной под названием «функционализм». Возникла функционалистская парадигма мышления. Патнэм, Дэвидсон, Райл, Деннет, Н. Блок и многие другие проponentы и оппоненты функционализма – все они опирались на ТТ. В отечественной философии в рамках функционалистской традиции работает Д.И.Дубровский.

И философский и научный успех во многом обусловлен простотой ТТ. Вместо поиска ответов на вопрос «Может ли машина мыслить?», Тьюринг предлагает поиграть в своеобразную языковую игру «в мышление» – т.н. «игру в имитацию». Следует отметить, что в отечественной литературе иногда встречается калька с английского – «имитационная игра». По мнению Б.В.Бирюкова, такой перевод не совсем точно отражает суть понятия Тьюринга. Следует ещё отметить, что языковая игра Тьюринга появилась незадолго ранее до распространения концепции языковых игр Л.Витгенштейна в англоязычных странах. Так что А.Тьюринга можно считать одним из пионеров лингвистических игр.

Играют мужчина (А), женщина (В) и судья (С). Пол судьи несущественен. Судья изолирован от А и В. Он находится в комнате с непроницаемыми стенами. Задача судьи – определить, кто из игроков – женщина. Задача мужчины и женщины – убедить судью в том, что именно он/она женщина. Средством общения является телеграф. В современных статьях по ТТ телеграф обычно заменяют более современными средствами – электронной почтой или интернетовским чатом. Судья задает вопросы в письменной форме, ис-

пользуя естественный язык. Ответы он получает в той же форме. Вопросы могут быть на любую тему: от математики до поэзии, от погоды до шахмат.

По задумке Тьюринга, новый вопрос по поводу мышления машины следует поставить следующим образом: «Если машина займет место игрока В, то будет ли судья ошибаться столь же часто, как и при игре с мужчиной и женщиной?» (то есть вместо женщины будет играть машина).

Большинству более поздних интерпретаций ТТ присуще игнорирование половой принадлежности игроков – играют машина (А), человек (В) и судья (С). Эти интерпретации опираются на последнюю фигуру имитационной игры, которая собственно и получила название «*тест Тьюринга*». Теперь место машины заменяет и игрока А. Цель С определить, кто из игроков – машина, кто – человек. Именно на этом этапе происходит замена вопроса «Может ли машина мыслить?» вопросом «*Может ли машина играть в имитационную игру?*».

Новая постановка проблемы фокусируется на функциональной способности игроков реализовать возможности ведения имитационной игры. Функции задаются вопросами/ответами. Материальный субстрат мышления становится неинтересен. Бесконечный поиск «сущности мышления» прекращается. Начинается продуктивная работа по конструктивному построению мыслящих машин, которые конвенционально считаются таковыми, если способны пройти ТТ.

Так как вопрос-ответный способ представляется, по сути, универсальным методом и научного поиска и философского исследования, то данный факт, безусловно, повлиял на успех ТТ и в научном и в философском сообществах.

Несомненный *успех ТТ* также представляется базовым положением теста Тьюринга.

Квалиа и парадигма функционализма **Игорь Зайцев (С-85)**

Доклад основан на работе Мишеля Туэ («Квалиа», 2003 г.)¹.

Термин «квалиа» принято использовать для обозначения интроспективно доступных феноменальных аспектов психической жизни. Существует множество различных чувств и ощущений. О них в контексте выбранной нами тематики принято говорить с позиции первого лица. Я провожу пальцами по наждачной бумаге, слышу неприятный запах, чувствую острую боль в пальце, вижу ярко фиолетовый цвет. В каждом из этих случаев я – субъект психического состояния с четко выраженным субъективным характером восприятия феноменов. Эти феномены, взятые в аспекте чувственной значимости для меня как носителя психических актов и не отторгаемые от меня состояния субъективной реальности принято называть *квалиа*.

Квалиа – ключ к проблеме дух/тело, открывающий глубинные аспекты природы человеческого сознания. В дискуссиях по поводу квалиа обсуждаются вопросы о конкретных психических состояниях, которые имеют квалиа, о том, задаются ли квалиа только особенностями субъекта (носителя квалиа)

¹ Tye, M. 2003. Qualia. Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/qualia/>

либо здесь участвуют и объективные факторы, как квалиа соотносятся с физическим миром – «снаружи» они либо «внутри» головы.

Философия искусственного интеллекта порождает новые вопросы о квалиа: вопросы о «присущности» квалиа компьютерной системе и воспроизводимости квалиа компьютерной системой.

Первый вопрос – **может ли компьютер обладать квалиа?** Постановка данного вопроса правомочна для позиции сильного искусственного интеллекта, последовательно экстраполированной на ментальные свойства и состояния в целом. Такая позиция, безусловно, приводит к некритическому представлению о том, что да, компьютер фактически может обладать квалиа. Объяснительным поводом в данном случае может служить представление о том, что не отторгаемая от субъекта квалиа, на самом деле, лишь некоторая часть «ментальной реальности», объемлющей не только компьютер, но и всё в этом мире (некая платоновская «мировая душа»). На наш взгляд, на такой подход стоит обращать внимание лишь ради критики паранормальных представлений о реальности.

Второй вопрос – **воспроизводима ли квалиа компьютерными средствами?** Данный вопрос, несомненно, заслуживает серьёзного обсуждения. Его можно связать с позицией слабого искусственного интеллекта. В реализационном отношении он может быть изучен в плане возможности внедрения чипов в компоненты коры головного мозга. Такие чипы функционируют на уровне т.н. «предельного процессирования» и предназначены для реализации квалиа, например, в высоких технологиях виртуальной реальности либо в рамках проекта (на сей день гипотетического) т.н. «практического бессмертия». При этом квалиа получают статус *управляемых параметров* субъективного восприятия реальности. Ключевой проблемой при этом становится расшифровка нейродинамических кодов во взаимосвязи их с психическими явлениями (Д.И. Дубровский). Этот подход к анализу понятия «квалиа», подход со стороны слабого искусственного интеллекта, перспективно рассмотреть в контексте парадигмы функционализма.

В *функционалистской парадигме утверждается, что субъективные квалиа имеют функциональную природу.*

Боль, например – это результат каузальной или телеофункциональной зависимости между физическими состояниями (например, повреждением тела) и физическими последствиями такого повреждения. С данной точки зрения (Лусан, 1987) квалиа физически реализуема различными способами. Для нашей задачи это означает реализуемость как «естественными» нейральными составляющими коры головного мозга, так и чипами, которые замещают данные компоненты. Внутренние состояния таких естественных и искусственных составляющих, сильно различающихся физически, могут, несмотря на это различие, приводить к *тождественным квалиа*. Ключом к воспроизводимости квалиа компьютерными (нанотехнологическими) средствами является *функциональная роль*, а не «железо» (материальный субстрат), лежащее в основе феноменального явления.

Однако функционалистская позиция слабого ИИ представляется также слабой. Существует два наиболее распространённых возражения против функционалистских теорий квалиа: «инвертированный спектр» (Inverted Spectrum) и гипотеза «отсутствующих квалиа» (Absent Qualia Hypothesis).

1. Инвертированный спектр

Этот мысленный эксперимент состоит в заявлении, что при рассмотрении одной и той же вещи один субъект может видеть красный цвет, а другой – зелёный. И наоборот, когда первый субъект видит зелёный – другой видит инвертированный цвет, красный. Такое восприятие характерно и для других цветов, т.е. воспринимаемые нами цветовые ощущения инвертированы. Репрезентационные отличия очевидны: один человек имеет зрительное восприятие, которое представляет красный цвет, другой – зелёный цвет. Это происходит в одних и тех же обстоятельствах, в одно и то же время. Репрезентативная разница должна нести с собой и различие в наших моделях каузальных взаимодействий с внешними вещами. Эта разница, следовательно, обуславливает функциональные отличия. Но ведь внешние вещи те же самые для «инвертированных» субъектов!

Функционализм, однако, обходит данное затруднение. Он спрашивает – а вообще, возможны ли такие случаи метафизически? Так же не ясна концептуальная возможность таких ситуаций (Harrison 1973, Hardin 1993, Tye 1995). С другой стороны, далеко не очевидно, что между различными феноменальными опытами восприятия нет функциональной разницы. Рассмотрим пример вычислений. Для любых двух числовых входных данных M и N , некоторый компьютер выдает выходные данные – результаты операций над M и N . Имеем ещё один компьютер, который выполняет абсолютно идентичные операции. Функционально оба компьютера идентичны. Следует ли из этого, что они выполняют в точности одну и ту же программу? Конечно, нет! Может быть большое число видов программ, которые перемножают два числа. Такие программы могут разительно различаться. На высоком уровне, допустим, эти машины функционально идентичны. Однако на более низких уровнях они функционально различаются. Необходимо только определить уровень функциональных различий и функциональных тождеств.

Трудно представить, что инвертированные квалиа существуют в *физических дубликатах*, созданных с точностью до молекулы, как и то, что они существуют и в *функциональных дубликатах*. Если вышеназванные дубликаты действительно метафизически возможны, как заявляют физикалисты, требуются дальнейшие убедительные аргументы, чтобы показать, что оба случая не аналогичны. До сих пор не было представлено таких аргументов. Так что функционализм пока отводит аргумент «инвертированных квалиа».

2. Гипотеза «отсутствующих квалиа»

Данная гипотеза утверждает о возможности функционального дубликата чувствующих существ. Такой дубликат совершенно не имеет квалиа. Например Н. Блок (Блок, 1978), предлагает представить, что каждый китаец из миллиарда китайцев снабжён двунаправленной рацией для связи с другими китайцами и искусственным спутником Земли. Движения тела 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, а радиосвязь выполняет роль синапсов. То есть в целом весь китайский народ дублирует нейро-динамическую организацию человеческого мозга. Если бы такая система была когда-нибудь реализована, могла ли она в

действительности испытывать чувства и ощущения? Например, со спутника выдаётся сигнал «Боль!». Будет ли «китайская нация» фактически испытывать боль. Если есть метафизическая возможность этого, то тогда квалиа не соразмерны с функционалистской парадигмой.

Функционалистский ответ готов и на этот мысленный эксперимент. Такая система *может переживать квалиа*. Возникновение гипотез, подобных «гипотезе отсутствующих квалиа» вытекает из-за неадекватного представления относительных размеров системы и её элементов (Lucan, 1987). Каждый из нас настолько меньше данной системы, что мы «не видим за деревьями леса». Существо размером с нейрон, заключённое в человеческой голове может быть ошибочно убеждено, что в голове не может быть никакой сознательности. Аналогично, рассматривая систему «китайского тела», мы выводим неверное заключение. Так же утверждается (например, Shoemaker 1975), что любая система, представляющая *наш полный функциональный дубликат*, станет субъектом всех наших ощущений. Таким образом, система «китайского тела» будет испытывать боль. Если такой подход верен, то это показывает, что *феноменальные свойства имеют функциональную природу*. Но это, конечно, *не доказывает*, что индивидуальные квалиа функциональны по природе.

Вопрос о функционалистской «организации» квалиа остаётся открытым. Возвращаясь к непосредственному предмету доклада – о воспроизводимости квалиа компьютерными средствами – следует отметить, что вышеприведённые дискуссии и невозможность ни доказать, ни опровергнуть функционализм делают проблематичным и теоретически неразрешимой задачу создания «силиконовых мозгов», продуцирующих на ином материальном (компьютерном) субстрате квалиа, тождественные квалиа биологического субстрата. И следует вновь «пересматривать сознание», то есть критически оценить работу Дж. Серла «Открывая сознание заново», 1992¹. В ней Дж. Серл принципиально отрицает возможность «силиконовых мозгов», риторически апеллируя к «стуку силиконовых чипов, когда вы трясёте своей головой», то есть когда у вас вместо естественного мозга его искусственный функциональный дубликат. Дальнейшее развитие ИИ покажет справедливость первой либо второй концепции относительно квалиа – феноменальных человеческих и, возможно, нечеловеческих качеств.

Зачем вкладывать деньги в тест Тьюринга? Лойбнеровская премия Дмитрий Зворыкин (М-08-03)

Ответ на вопрос, поставленный в наименовании темы доклада, косвенно даётся в статье журналиста Чарльза Платта [«Что всё-таки означает быть человеком?»](#). Он описывает интересные состязания, которые состоялись в 1999 г. и в которых воспроизводится идея Тьюринга о создании мыслящих машин. В статье раскрываются особенности построения программ, цель которых – пройти тест Тьюринга и тем самым на практике разрешить проблему «Может ли машина мыслить?». Данные особенности обсуждаются также с финансовой точки зрения.

¹ Имеется в виду работа: Серль, Джон. Открывая сознание заново. Перев.с англ. А.Ф. Грязнова. М.: Идея-Пресс, 2002. – 256 с., С 77-80

Чарльз Платт рассказывает о недавнем состязании для определения наиболее «человечного» компьютера. Цель состязания состояла в том, чтобы выяснить, могут ли 10 судей в условиях диалога распознать разницу между людьми и программами ИИ. Программа, набравшая максимум баллов, принесёт своему создателю **2000\$**.

Собственно, идея состязания принадлежит Алану Тьюрингу (1950).

Может ли компьютер использовать набор уловок, чтобы имитировать человеческие ответы? И что всё-таки означают слова «человечность» и «разум»?

Состязания стали проводиться с 1991 г. по инициативе Хьюго Лойбнера (Hugh Loebner). Он предложил **\$100K** (это и есть сумма *лойбнеровской премии*) тому, кто создаст программу, способную в течение **трёх часов** разговоров на любые темы убедить **10** судей в своей человечности (то есть имитировать, что она – человек).

Как показала практика, этот результат *недостижим сегодня*, поэтому Hugh Loebner также объявил ежегодную премию в **\$2K** автору программы, которая окажется наиболее человеческой. И чтобы ещё упростить задание, он позволил программисту заранее выбирать определённую тему для разговора.

Состязание происходит в форме диалога десяти судей с десятью оппонентами, пятеро из которых – люди, а пятеро – компьютерные программы.

Чтобы программа выиграла, программист сам должен быть «человечным», по крайней мере выступать в плане владения своей речью «эталонно» для собственной разработки.

- Почему вы выбрали ИИ? – спрашивает Ч. Платт у Х. Лойбнера.

- Почему бы и нет? – отвечает он, – четыре состязания стоили мне около \$25K. Сомневаюсь, что я сделал бы более заметный вклад в науку и общество, если бы вложил эти деньги в исследования, связанные со СПИДом или с чем-то ещё. Думаю, что *развитие ИИ – значительный вклад в развитие цивилизации*.

После трёх часов состязание заканчивается. Далее все собираются в большом зале, где участники, зрители и представители прессы могут видеть на экранах результаты состязаний.

Во время состязания, которое описывает Ч. Платт, ни одна программа не оказалась «разумной» настолько, чтобы смогла убедить судей, что она человек. Программа, подошедшая к этому ближе всего, разговаривала о сексе.

Все диалоги во время конкурсов записываются. Как замечает Ч. Платт, ответы имеют тенденцию повторяться, большинство из них лишены смысла. Но, как известно любому пользователю Интернета, существует значительная разница между тем, какой человек в жизни, и тем, каким он выглядит в сети – во время состязания повторы и бессмысленность были не столь заметны.

Несмотря на шутливую сторону лойбнеровского конкурса и премии, один из организаторов, Роберт Эпштейн серьёзен по поводу ИИ. Он с нетерпением ожидает разумных «цифровых помощников», которые будут служить своеобразными «цифровыми землеройками», находя то, что нам нужно, сравнивая, суммируя это и представляя в наиболее эффективном виде. «Без ИИ интерфейса,- говорит он, – не существует способа фильтровать огромный объём информации, который на нас сегодня обрушивается. А что будет завтра? Когда такая система будет создана, могут проявиться интересные по-

следствия. Настоящий ИИ будет представлять собой сврехсложную, разумную сущность, которая станет размножаться и защищаться. Она в некотором смысле начнет мутировать; дочерние особи (программы, естественно) будут отделяться от родительских и размножаться через Сеть. Неизвестно, когда это случится, но вполне возможно – с такими футурологическими прогнозами развития ИИ согласны многие участники состязания по реализации ТТ. Сам Тьюринг проявлял большой оптимизм по поводу развития и применения мыслящих машин – они станут полноправными членами общества и решать задачи не хуже людей, помогая человеку.

Программирование теста Тьюринга **Алексей Клопков (М-08-03)**

Полвека назад А.Тьюринг утверждал «Мы можем надеяться, что машины, в конечном счёте, смогут соревноваться с человеком в абсолютно любой области интеллектуального труда». И добавил, что через пятьдесят лет техническое совершенство компьютеров будет удовлетворять таким требованиям.

Представляет интерес анализ того, насколько был прав А.Тьюринг, тем более, что этим словам в следующем, 2005 г. будет 55 лет. В работе изучается вопрос – как ТТ-программы представлены в Интернете.

История развития программ ТТ в Интернете

Первой программой, которая проходила тест Тьюринга (ТТ) в среде Интернет, считает Марк Хамприус – специалист в области ТТ-программирования — была программа Mgonz. Данная программа (а) была чатом реального времени, (б) в неё были заложены элементы неожиданности и (в) работала в Интернете.

До Mgonz было разработано много программ – чатов с представленными на них программами ТТ. Однако они не были столь интересны, как Mgonz. К примеру, в них отсутствовали элементы неожиданности. См., например, BITNET: <http://www.compapp.dcu.ie/~humphrys/net.80s.html>.

Историю развития ТТ – программ в среде интернет можно отследить по следующим адресам:

- Бот (<http://www.amazon.co.uk/exec/obidos/ASIN/1888869054/>), написан Эндрю Леонардом, 1997
- В анонимные почтовых сообщениях с обсуждением чат-ботов в сети требовалось, чтобы к программе тестирования вначале подключилась Элиза (1975 год., <http://slashdot.org/articles/00/02/02/0735242.shtml>)
- Бот Марка В. Шани, находящийся в локальной сети (тогда не было чатов в реальном времени) и содержащий элементы неожиданности ещё в 1984 году
- первые чаты (<http://www.compapp.dcu.ie/~humphrys/net.80s.html#chat>)
- Небольшая многопользовательская конференция, открытая в августе 1989 (<http://www.fuzine.com/lti/pub/aaai94.html>)

Программы ИИ в Интернете

- Yahoo:
 - программы
(http://uk.dir.yahoo.com/Recreation/Games/Computer_Games/Internet_Games/Web_Games/Artificial_Intelligence/)

- IRC боты (IRC – глобальная система, посредством которой пользователи могут общаться друг с другом в реальном времени)
(http://uk.dir.yahoo.com/Computers_and_Internet/Internet/Chats_and_Forum/Internet_Relay_Chat__IRC_/Bots/)
- Вэб боты
(http://uk.dir.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web/Crawlers__Robots__and_Spiders/)
 - Google:
- Чат боты:
(http://directory.google.com/Top/Computers/Artificial_Intelligence/Natural_Language/Chatterbots/)
 - Демо – интерфейсы
(<http://www.emsl.pnl.gov:2080/proj/neuron/cogsys/demos.html>)
 - список чат ботов (<http://www.botspot.com/search/s-chat.htm>)
 - сайты автоматического создания постмодернистских «произведений»
- сайт Эндрю С. Булхак <http://www.elsewhere.org/cgi-bin/postmodern>,
(<http://dev.null.org/>) (технический отчёт http://www.csse.monash.edu.au/cgi-bin/pub_search?104+1996+bulhak+Postmodernism)
 - сайты, посвященные проблематике искусственной жизни
<http://www.compapp.dcu.ie/~humphrys/ai.links.html#alife.online>
 - роботы в сети
<http://www.compapp.dcu.ie/~humphrys/ai.links.html#robots.online>

Особо следует выделить программу **Джени18**.

По быстрдействию из всех диалоговых программ, которые когда либо были написаны, с ИИ или без ИИ, платная программа Джени18 впечатляет пользователей. Имеется несколько версий.

- Джэк Каумфман (<http://virt.vgmix.com/>) (<http://virt.ph34r.net/>)
- Джени18 – бот «Элиза», занимающийся киберсексом
(<http://virt.vgmix.com/jenny18/>), запущенный в сети DALnet в IRC
(<http://www.dal.net/>).
- Джени18 – программа, упрощённая до уровня Mgonz. Данный бот создан исключительно для того, чтобы разыгрывать из себя вульгарную девушку, которая занимается киберсексом. Цель программы – довести пользователей до оргазма: «dom01» (<http://virt.vgmix.com/jenny18/logs/dom01.txt>), «Happy_Boy» (http://virt.vgmix.com/jenny18/logs/Happy_Boy.txt). Наиболее оживленный диалог происходит в «Scorpion832»: <http://virt.vgmix.com/jenny18/logs/Scorpion832.txt>.
- Аналогичные программы – «GoldenBoy2222»
<http://virt.vgmix.com/jenny18/logs/goldenboy2222.txt>) и «Lander100»
(<http://virt.vgmix.com/jenny18/logs/lander100.txt>)
- Тест Тьюринга «Оргазм»(программы, которые доводят человека до оргазма) вытеснил программу Джени18.

Пример программы, которая прошла ТТ. Её написал в 1995 г. упомянутый выше Марк Хамприус.

Небольшой пример ТТ – отрывок из диалога машины и человека («Кто-то у Дрейка» – это либо человек, либо программа):

**когда ты в последний раз занимался сексом?*

Кто-то у Дрейка: вчера

** ну ладно, скажи честно, когда это было*

Кто-то у Дрейка: я же говорю, что это было вчера

** ну ладно, скажи честно, когда это было*

Кто-то у Дрейка: ну ладно-ладно – это было около 20 часов назад, а у тебя похоже, было лет 20 назад, раз задаёшь такие вопросы.

Человек не догадывается, что с ним разговаривает машина (*Кто-то у Дрейка*). Он думает, что с ним ведёт диалог просто настырный собеседник.

Какое будет следующим поколение программ, проходящих Тест Тьюринга? Хочется надеется, что новые *интеллектуальные* программы ТТ будут более *интеллигентными*, по крайней мере, не будут ориентироваться на низшие биологические потребности и употреблять ненормативную лексику.

«Может ли машина мыслить?». Полемический стандарт Тьюринга Дмитрий Комаров (Р-81)

Тьюринг, отвечая на поставленный им же вопрос «Может ли машина мыслить?» предлагает базовую структуру проведения дискуссий: 1) тема возражения; 2) довод в пользу возражения; 3) ответ на возражение.

В докладе представлены основные положения дискуссии, которые предлагал сам Тьюринг для защиты тезиса («Машина может мыслить») в ответ на возражения воображаемых оппонентов. Эти положения предлагается назвать «*полемическим стандартом Тьюринга*». Он состоит из девяти положений. Предполагаемое возражение, опровергающее тезис Тьюринга, обозначено знаком «-», довод в пользу знаком «+». Здесь нарушен порядок следования возражения. Их Тьюринг не так излагал. Возражения «Теологическое», «От боязни» и «Телепатическое» приведены в конце списка, как менее значимые для цели полемики. Некоторые положения мы сформулировали иначе, нежели чем в первоисточнике.

1. Математическое возражение

-: Существует несколько теорем, доказывающих, что мощность дискретных машин ограничена. Самая знаменитая из них, возможно, 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, считают не требующим доказательств и факт того, что машины, обладающие интеллектом, не ошибаются. Однако отсутствие или наличие ошибок не есть требование к мышлению.

2. 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.). 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, единст-

венный способ узнать мыслит или не мыслит другой человек – *быть* этим другим человеком. А это невозможно. Данная проблема обычно называется *проблемой чужого разума* и она постоянно возникает в ходе дискуссий о ТТ.

+: 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 качества 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.

3. Возражение о различных невозможностях

-: Машины никогда не смогут сделать X, где X может быть любой человеческой особенностью, такой как чувство юмора, креативность, способность влюбляться или обожать клубнику.

+: 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.

4. 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.

5. 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 проведения ТТ.

6. 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.

7. Теологическое возражение

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

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

8. 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.

9. Телепатическое возражение («с точки зрения сверхчувственного восприятия»)

-: Человеку иногда присуще сверхчувственное восприятие, а машине – нет. Человек способен интуитивно постичь правильный ответ на нечёткий вопрос судьи. Машина на это не способна.

+: Если считать, что телепатия возможна, тогда судью следует помещать в комнату, «защищенную от телепатии».

За полвека дискуссий по проблеме ТТ, предложенная основоположником структура ведения полемики превратилась в своеобразный стандарт для исследователей. Каждый из них, обозначая свой собственный вариант ТТ, явно или неявно обращался к форме «возражение»/«ответ» (Н.Блок, Д.Деннет, 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 бессмысленность самой постановки этого вопроса – дискуссия бесконечна, так же, как бесконечно многообразие представлений об особенностях человеческого мышления и способах его моделирования.

Тест Лавлейс: машина творить не может! **Алексей Ласточкин (ЭК-82)**

Мышление и творчество считаются взаимодополняющими характеристиками. В современных техногенных условиях крайне актуально звучит вопрос: «Может ли машина творить?» С 50-х годов прошлого столетия учёные стали утверждать о положительном решении данного вопроса в рамках обсуждения проблематики построения мыслящих машин, т.е. в контексте идей ИИ. Обоснованность их суждений обычно подкрепляется двумя способами: парадигмой функционализма и тестом Тьюринга (ТТ).

Парадигма функционализма претендует на то, что если должным образом запрограммировать машину (компьютер), то появляется возможность реализации мыслительной, творческой и любой другой ментальной деятельности. Для этого нужно: 1) определить функцию, вход и выход которой характеризует реализуемую деятельность (некую «функцию творчества», «функцию мышления»); 2) разработать эффективную *программу* вычисления данной функции¹.

Тест Тьюринга не претендует на метафизическую возможность того, что должным образом запрограммированный компьютер будет на самом деле тво-

¹ Пол М. Черчленд, Патриция Смит Черчленд. Искусственный интеллект: Может ли машина мыслить? <http://grokhovs2.chat.ru/bra/bra.html>

рять и мыслить. Разработанная программа лишь обеспечивает условия для *имитации* творческой или мыслительной деятельности, *обмана* пользователя (судьи).

Оказалось, что обман по поводу реализации мыслительной деятельности возможен – в принципе, любая достаточно мощная экспертная система сегодня может ввести в заблуждение пользователя, если он будет работать с ней в условиях ТТ, т.е. в условиях неопределённости, с кем он общается – с компьютером или человеком. Однако обмануть в том, что машина может творить оказалось непросто. Сам А. Тьюринг полагал, что главным возражением на его тезис о создании мыслящей машины является именно тезис о том, что машина творить не может, а так как творчество и мышление взаимообуславливающие способности, то и машинное мышление не возможно. То есть если отрицателен ответ на вопрос «Может ли машина творить?», то снимается и вопрос «Может ли машина мыслить?»

Главным оппонентом ТТ, за сотню лет до этого умершей, он считал леди Лавлейс¹. Возражение Лавлейс сформулировано Тьюрингом следующим образом: «Компьютер самостоятельно не может ничего создать. Создание чего-либо требует, как минимум, *изобретения* чего-либо *нового*. Но компьютеры не изобретают ничего нового. Они всего лишь делают только то, что мы посредством программ приказываем им делать».

Более точная формулировка этого опровержения представлена в работах С. Брингсйорда, П. Белло, Д. Феруччи². Они предлагают модифицированный ТТ, который назвали Тестом Лавлейс.

Тест Лавлейс выглядит следующим образом:

Def_{LT} : Искусственная система *A*, созданная человеком *H*, проходит ТЛ тогда и только тогда, когда:

I. *A* изобретает *о*. Для Лавлейс – дочери Байрона – в роли *о* (результата работы машины) выступало бы, по всей видимости, оригинальное поэтическое произведение;

II. *о*, выдаваемые *A* – это не результаты случайного стечения обстоятельств или сбоя машины. *A* всегда может воспроизвести (повторить) *о*;

III. *H* (любое существо обладающее знаниями и возможностями *H*) не может объяснить, как *A* выдал *о* даже в условиях полного представления об *A* – о структуре баз данных, алгоритмах функционирования и т.п.

Анализ ТЛ вызывает ряд следующих соображений:

1) ТЛ невозможно пройти путем *обмана* (подтасовка ответов характерна для прохождения стандартного ТТ).

2) ТЛ автоматически выбивает из соревнования все модели ИИ, созданные по сей день. Все эти модели запрограммированы разработчиками, которые знают особенности их функционирования. Ничего «от себя» такие программы прибавить не могут.

3) Необходимым условием прохождения ТЛ является способность машины к самоизменению своего программного кода, если не всего целиком, то, по крайней мере, части. Машина должна быть самообучаемой.

¹ Выдающиеся люди в истории информатики. Ада Лавлейс <http://jollity.narod.ru/ada.html>

² Selmer Bringsjord, Paul Bello, David Ferrucci. Creativity, the Turing Test, and the (Better) Lovelace Test. <http://www.rpi.edu/~faheyj2/SB/SELPAP/DARTMOUTH/lt3.pdf>

4) Но и самообучаемости недостаточно для прохождения ТЛ. Дело в том, что все изменяемые правила носят «переходящий» характер. Результат обучения – это, к примеру, новая структура базы данных или изменённый программный код. Но такие изменения невозможно представить как нечто суть оригинальное. Все они осуществляются в контексте известного.

5) Другим вариантом решения проблемы «машинного творчества» может оказаться наделение машины свободной волей. Это означает предоставление машине права выбора. Для реализации такого права необходимо смоделировать мотивационно-волевые механизмы человека, стремящегося к созданию нечто нового. Но и в данном случае результат *о* будет представим в некотором пространстве решений, что также исключает оригинальность *о*.

6) Вывод: программно управляемая машина творить не может. Она лишь оперирует символами согласно задаваемым разработчиком алгоритмам, осуществляющим формальные преобразования. Для манипулирования символами компьютеру не требуется никаких ментальных способностей – ни творения образов предполагаемого будущего, ни творческого внимания для достижения понимания и т.п. Это наглядно демонстрируется тестом «китайская комната» Дж.Серла).

Насколько убедителен ТЛ? Тьюринг в ответ на возражение Лавлейс заменял «создавать новое» на «способность удивлять». Он считал, что любая ошибка в программном коде или аппаратный сбой опровергают возражение Лавлейс.

Для ТЛ точка зрения А. Тьюринга недостаточна: 1) Этот случай оговорен в определении DefTL. 2) К компьютерам применяются крайне жесткие условия контроля вычислительного процесса даже в случае сбоя (для этого возможна трассировка событий, определение причин сбоя и т.п.). Подобного рода контроль, однако, над человеком невозможен. 3) Необходимые условия прохождения теста могут оказаться неверными, так как суждение о наличии или отсутствии интеллекта делается на основе постулата, что интеллект способен создать что-то новое, в то время как сам механизм этого процесса не изучен и не известен в достаточной степени: причины и условия творчества не известны, не известен и результат, к которому приведёт творческий процесс, если о нём судить до или во время его протекания. То есть на основе двух «сигналов» – входного и выходного – практически невозможно установить функцию, посредством которой она представляется вычислительной. Требуется длительное, которое в конце концов может все равно ничего не дать.

Приведённых суждений вполне достаточно для заключения: «Машина творить не может!». Тест Тьюринга, в её новой модификации – в формате Теста Лавлейс – не способен обмануть достаточно придирчивого судью в том, что компьютер может мыслить – ведь он не способен к творчеству. А вне творчества мышление невозможно.

«Искусственный интеллект» Алана Тьюринга **Сергей Лизоркин (АП-81)**

Алан Тьюринг (1912-1954) никогда не причислял себя к философам¹. Тем не менее его статья «Вычислительные машины и интеллект» (1950) стала наиболее часто цитируемой работой в современной философской литературе. Она дала новый подход к решению традиционной проблемы «дух/тело». В научной сфере данная работа рассматривается как исток компьютерной науки и искусственного интеллекта, в первую очередь, теоретико-методологической составляющей ИИ.

Считается, что «проблемой разумных машин» А.Тьюринг занялся в 1941 после прочтения теологической книги Дороти Сэйерс (Dorothy Sayers). Он часто цитирует данную работу в своих статьях.

Тьюринг подчёркивает, что ему известно общепринятое суждение о «механическом» как о «лишённом интеллекта». Он стремится развеять эти предрассудки и мнение, а также то, что «возможности машины ограничены лишь простыми и повторяющимися задачами».

Противоречивость суждений А.Тьюринга по поводу ИИ подчёркивает то, что в одном и том же году (1950) он высказывает два противоположных мнения:

1) «ЭВМ предназначены для выполнения любого задания, которое может быть выполнено человеком, если он будет следовать своим инструкциям точно и абсолютно *бездумно*».

2) «Мы можем надеяться, что машины, в конечном счёте, смогут соревноваться с человеком в абсолютно любой области интеллектуального труда».

Как может появиться *интеллект* из абсолютно неинтеллектуальных операций?

Аргумент Тьюринга заключается в факте порождения мозгом интеллекта, а также представимости работы мозга конечным автоматом.

Трактовка Тьюринга весьма ограничена: любое значимое действие мозга не только вычислимо, но и реализуемо полностью конечной машиной, например машиной Тьюринга, у которой лента – конечной длины или вообще отсутствует. По его мнению, весь спектр вычислимых функций, определённых на основе машин Тьюринга с бесконечной лентой, представляет лишь теоретический интерес. При этом Тьюринг приводит довод в пользу конечности нервной системы, давая приближённое значение ёмкости компьютерной памяти, необходимой для имитации работы мозга: 10^9 бит.

Игра в имитацию Тьюринга получила оживлённое обсуждение. Основная идея Тьюринга – обойти споры о природе мысли, разума и самосознания и предложить новый критерий интеллектуальности, основанный на наблюдении за поведением. Он считает, что можно зафиксировать разумность людей, используя лишь поведенческие показатели, и предлагает вести «честную игру» в отношении машин.

¹ В основе доклада лежит статья Энрю Хотгиса «Алан Тьюринг», Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/turing/>

Он иллюстрирует свою точку зрения мысленным экспериментом, который в наши дни легко провести:

компьютерная программа пытается убедить беспристрастного судью, что именно она является человеком. Она и оппонент-человек используют только текстовые сообщения. Если программа побеждает, её следует признать разумной.

Тьюринг приводит на удивление неудачный пример, считается А. Ходгис – групповую игру, в которой мужчина притворяется женщиной. Неточность в высказываниях привела к тому, что некоторые даже полагали, что машина должна притвориться мужчиной, который, в свою очередь, притворяется женщиной. Например, Лассегью (1998) делают упор на игре подражания полов, её реальных и мнимых смыслах. На самом же деле, суть теста с обменом текстовыми сообщениями – ***отделить интеллект от иных человеческих способностей.***

Можно справедливо сказать, что эта путаница отражает чрезвычайно претенциозное представление Тьюринга о человеческом разуме. Друзья знали Тьюринга как человека, в котором ум, юмор и секс были неотделимы друг от друга.

Тьюринг понимал сложность отделения интеллекта от иных аспектов человеческих чувств и действий; он описал роботов с сенсорами и задался вопросом, смогут ли они, например, насладиться вкусом клубники со сливками или почувствовать общность с другими роботами. С другой стороны, в своём тесте он уделял мало внимания важным эпистемологическим вопросам достоверности, явного и неявного обмана, в первую очередь, потому что хотел обойти стороной вопросы подлинности сознания. Достаточно тонкий момент одного из воображаемых «разумных» диалогов (Turing 1950) в том, что программа имитирует человеческий интеллект, давая *неверный* ответ на простую арифметическую задачу. Но в постановке Тьюринга мы не должны задаваться вопросом, ни тем, «сознательно» ли компьютер обманывает судью, создавая впечатление человечности, ни тем, зачем программа может желать этого. В таком подходе есть некоторая нехватка убедительности.

Принцип имитации, выбранный Тьюрингом, также подразумевает (как и «тесты IQ» того времени) *идентичную языковую и культурную общность испытуемого и судьи* в его воображаемых «тестах». При этом он игнорирует социокультурное многообразие людей. Не рассматривается также возможность бессловесного мышления, например, у животных или у гипотетических инопланетян.

Другая положительная особенность статьи — намечается программа исследований в области построения «*обучающихся машин*» и обучении *машин-«детей»*. Принято считать (например, Dreyfus 1990), что всегда существовало противостояние между программированием и коннекционистской настройкой нейросети. Тьюринг утверждал о необходимости использования обоих этих подходов.

Имеет место точка зрения, что первые идеи относительно искусственного интеллекта пришли первооткрывателям в этой области в 50-х годах XX-го века, *после* значительных успехов применения ЭВМ при решении сложных арифметических задач.

Любопытен факт, что статья Тьюринга затрагивала философские темы, ведь можно смело сказать, что Тьюринг, приверженец материалистического

мировоззрения, вовсе не был профессиональным философом. Тьюринг, будучи математиком, привнёс в философию взгляд с позиции открытий в области математики и физики.

Тьюринг делает упор на достаточность для объяснения вычислительной схемы действий разума. Пенроуз (1994), исследуя гипотезу Тьюринга, представляет её как «тезис Тьюринга» таким образом:

Он считает любые физические воздействия (которые включают в себя и активность человеческого мозга) сводимыми к каким-либо действиям машины Тьюринга.

Утверждение о том, что *любое физическое воздействие* по сути своей можно вычислить, нигде у Тьюринга явно не высказано, все они – следствие неявных предположений статьи 1950-го года. Рассуждение Тьюринга «возражение с позиции непрерывности нервной системы» в сущности, сводится к тому, что физическая структура мозга может быть смоделирована на компьютере с желаемой точностью.

Конечно, в работах Тьюринга 1945-50х годов нет ничего, чтобы ответить на трактовку Пенроуза. Предшествующие статьи, явно технической направленности, (Turing 1947, 1948) содержат разнообразные заметки о физических процессах, но не содержат упоминаний о возможности физических взаимодействий оказаться невычислимыми.

В частности, раздел статьи (Turing 1948) посвящён общей классификации «машин». Тьюринг отделяет «управляющее» обеспечение от «активного». Анализ Тьюринга касается первого типа — в современной терминологии это «информационное обеспечение». Следует отметить, что ни в 1936-м, ни в 1948-м году, несмотря на свои обширные познания в квантовой физике, Тьюринг не пытается применять идей квантовой механики к концепции «управления». Концепция «управления» — целиком внутри классической структуры машины Тьюринга (которую он называет в своей статье «логической вычислительной машиной».)

В этом же разделе статьи (Turing 1948) проводит разделение *дискретных* и *непрерывных* машин, иллюстрируя вторые телефоном: непрерывной управляющей машиной. В том же разделе освещена проблема сведения непрерывных физических процессов к дискретным, и хотя Тьюринг определяет мозг как непрерывную машину, он утверждает, что можно рассматривать мозг и как дискретную машину. Тьюринг не утверждает, что непрерывность может означать невычислимость. В действительности, в своей статье (Turing 1947) он утверждает, что цифровые компьютеры *более мощны*, чем аналоговые (например, дифференциальные анализаторы). При их сравнении, он даёт следующую вольную трактовку тезиса Чёрча-Тьюринга:

Одним из моих заключений было то, что «правила большого пальца» и «машинный процесс» являются синонимами. Выражение «машинный процесс», естественно, означает процесс, который может быть выполнен машиной рассматриваемого типа [например, машиной Тьюринга].

Тьюринг не говорит, является ли дискретность машин Тьюринга реальным ограничением, или что непрерывные процессы аналоговых машин могут иметь какое-то особое значение.

Тьюринг также представляет идею о «случайных элементах», но его примеры (использование разрядов числа π) показывают, что он считал *псевдослучайные* последовательности (вычислимые последовательности с подходящими «случайными» свойствами) вполне подходящими. Он не делает предположения, что случайность предполагает нечто невычислимое, и, действительно, не даёт определения термина «случайный». Это вызывает удивление, особенно если учесть, что его работы в области математики, логики и криптографии дают возможность говорить о случайности вполне серьёзно.

Все приведённые положения: вычислимости, непрерывности, случайности и т.п. – суть дополнительные показатели основного тезиса – тезиса о *возможности функционального воспроизводства человеческого мышления на ином материальном субстрате, отличном от естественного*. Данный тезис через 10 лет Патнэмом был представлен в форме функционалистской парадигмы мышления, а работы А. Тьюринга получили собственно философское звучание.

Функционально-структурные аспекты понятия «потребность» Наталья Маклашевская (ЭП-61)

Одной из проблем искусственного интеллекта является моделирование и воспроизводства в компьютерной среде специфически человеческого феномена потребности. В докладе делается попытка построения модели «потребности» в контексте парадигмы методологического функционализма.

Понятие «потребность» – одна из ключевых философских категорий. В контексте диалектической философии к данному понятию обращаются при изучении соотношений причинности/закона, необходимости/случайности, свободы/необходимости и др. В экзистенциализме «потребность» изучается совместно с онтологически значимыми параметрами человеческого бытия (тоской, тошнотой, страхом и пр.). Некоторые варианты философской антропологии строятся на постулировании метафизической структуры потребностей (например, у М. Шелера). В прагматизме потребность выступает ключевым параметром активизации деятельности. В феноменологии «потребность» инициирует формирование интенциональной структуры «жизненного мира».

Несмотря на концептуальную значимость, категориальное определение «потребности» представляется нечётким. Так, например, П.М. Ершов считает, что «потребность – это специфическое свойство живой материи, отличающее ее, живую материю, от материи неживой». То есть потребность – первопричина жизни, свойство всего живого. [2, С.148]. Как вполне справедливо отмечает Н.М. Бережной, данному определению присущ «...налет телеологизма. Можно подумать, что коровы пасутся на лугу, обуреваемые потребностью напоить молоком детей, а овес растет потому, что надо кормить лошадей». Сам он считает вполне приемлемым следующее определение понятия: «Потребности – нужда или недостаток в чем-либо необходимом для поддержания жизнедеятельности организма, человеческой личности, социальной группы, общества в целом, внутренний побудитель активности» [1, С. 518]. Однако, что такое «нужда»? Непонятно. В логике данный способ определения понятия принято считать ошибочным. Такую ошибку называют «определить непонятное через непонятное» или «х через у». В советской фи-

лософии (особенно в учениях о нравственных ценностях) существовали разногласия: что первично: активность или потребность. Так, потребность определяется через активность – т.е. активность присуща материи как атрибут в силу самодвижения последней (ср. с определением П.М. Ершова). С другой стороны, активность задаётся динамикой потребностей [3, С. 57-58]. Таким образом, понятие «потребность» в рамках философских учений представляется противоречивым или, в лучшем случае, нечётким.

Несомненна методологическая важность выработки понятия «потребность» для целого ряда наук о человеке и обществе: социологии, культурологии, политологии, антропологии, лингвистики и др. Многие современные теории базируются на понятии «потребности». Например, одно из влиятельных течений в современной психологии – т.н. «гуманистическая психология» – базируется на теории потребностей А. Маслоу. Потребность характеризует первую стадию мотивационно-волевой деятельности человека. Потребность проявляется в виде того, что человек начинает ощущать, что ему чего-то не хватает. Проявляется она в конкретное время и начинает «требовать» от человека, чтобы он нашел возможность и предпринял какие-то шаги для ее устранения. Потребности могут быть самыми различными. А.Маслоу делит их на пять групп, располагая их в иерархическом порядке (снизу вверх в порядке перечисления): физиологические потребности, потребности в безопасности, социальные потребности, потребности в уважении, потребности в самореализации. Однако понятие потребности остаётся неопределённым. Из-за этого возникает сомнение в том, что данный вариант «гуманистической психологии» на самом деле является «гуманистическим», ведь гуманистической эпистемологии присуще скептическое отвержение догматических конструкций, в форме которой, несомненно является нам теория А. Маслоу (см. работы по философии гуманизма П. Курца, В.А. Кувакина [4]). В теории А. Маслоу такой рациональности нет. Более того, сведение потребности к ощущениям ведёт прямой дорогой к солипсизму, причем не методологическому, как, например, у Э. Гуссерля, а традиционному солипсизму, берклианско-махистского типа.

Таким образом, попытка дать определение потребности представляет проблему и в онтологическом и в логико-эпистемологическом отношении. В онтологическом плане представляется затруднительной объективация потребности путем выделения и атрибуции некоего объекта «потребность», обладающего интерсенсуальным и интерсубъективным статусом. В логико-эпистемологическом отношении не обоснована редукция базовых научно-теоретических положений к аксиоматически заданному, однако, неочевидному понятию «потребности».

Безуспешность данных попыток, конечно, не отменяет продуктивности частных определений данного понятия. Как ранее подчеркивалось, методологически важным является выработка понятия «потребности» для ряда общественных наук. Более того, от конструктивности определённости этого понятия зависят успехи в применении метода моделирования. Адекватные и полные социологические, экономические, политологические и др. модели немыслимы без моделирования «потребности». Именно на эту модель замыкаются частные модели мотивационно-волевых динамик, социально-ценностных структур, культурных «смыслов» и т.п. На наш взгляд, наиболее приемлемым для моделирования потребностей представляется использование функционального подхода к данному понятию.

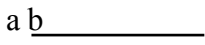
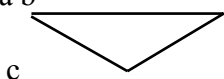
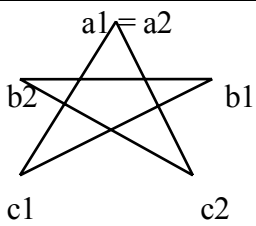
Следует различать две разновидности *функционалистского определения* данного понятия: парадигмальное и методологическое.

Парадигмально-функционалистское определение «потребности». Функционализм как философская парадигма насчитывает несколько десятков лет и связан с развитием компьютерных и когнитивных наук, с философией сознания, искусственным интеллектом. «Потребность» с позиции этого подхода следует определить, как способность изучаемой системы действовать таким же образом, как Я, исследователь, действую, когда испытываю потребность. Функционализм как парадигма сознания противопоставляется физикализму и бихевиоризму. В отличие от физикализма, функционализм утверждает о возможности реализации ментальных состояний на субстрате, отличном от субстрата человеческого мозга, например, на кремниевой основе. В отличие от бихевиоризма, функционализм убеждён в том, что ментальные состояния (в частности, состояния потребности) причинно зависят от определённых соотношений на входе/выходе системы и феномены сознания (а в нашем случае и бессознательные феномены, часть которых суть потребности) следует изучать исходя не только от структуры поведенческих диспозиций, но и от внутренней структуры и материальных условий реализации системы, производящей эти состояния.

Компьютерное моделирование «потребности» при принятии парадигмальной установки функционализма сводится к созданию системы, способной пройти тест Тьюринга, в котором посредством вопросов и ответов имитируется изучение системы потребностей со стороны того или иного исследователя (в роли этого исследователя, применительно к существу нашей задачи, могут выступать социолог, антрополог, культуролог, политолог и др.) Несмотря на очевидные достижения в области практической реализации теста Тьюринга, следует отметить крайнюю отдалённость перспективы построения имитационной системы, на основе которой можно изучать схемы потребностей и механизмы их обуславливания.

Функционально-структурное определение «потребности». Более практичным представляется определение понятия потребности, данное в рамках т.н. «универсального функционализма» Б. Малиновского (1884–1942). Понятие потребности сводится к сохранению социальной структурой своей целостности. Исходный принцип данной концепции гласит: «...в любом типе цивилизаций любой обычай, материальный объект, идея и верования выполняют некоторую жизненную функцию, решают некоторую задачу, представляют собой необходимую часть внутри действующего целого». При этом согласно Б. Малиновскому любая культура в ходе своего развития вырабатывает некоторую систему устойчивого «равновесия», где каждая часть целого выполняет свою функцию. Если уничтожить какой-либо элемент культуры (например, запретить вредный, с точки зрения исследователя, обряд), то вся этнокультурная система, а значит, и народ, живущий в ней, может быть подвержена деградации и гибели. Также подчеркивается, что «традиция, с биологической точки зрения, есть форма коллективной адаптации общности к ее среде. Уничтожьте традицию, и вы лишите социальный организм его защитного покрова и обречете его на медленный, неизбежный процесс умирания». Как мы видим, понятие потребности, данное Б. Малиновским, представляется созвучным кибернетическому понятию «гомеостазиса» (Н. Винер, У. Эшби). Однако данному понятию присущи недостатки: в нем отсутствуют такие, су-

губо человеческие характеристики, как потребность в творчестве, в установлении социальных связей.

Форма	Содержание	Аналит. выражение	Диagramматическое выражение	Потребность
Структурная	Потребность в самосохранении системы	$c = \frac{a+b}{2}$ среднее арифметическое		Гомеостазис
Динамическая	Продуцирование новой потребности в рамках установки целесообразности	$c = \sqrt{a*b}$ среднее геометрическое		Творчество
Гармоническая	Соединение неоднородных, гетерогенных, хаотичных мотивов, ценностей, целей, задач в единое целое	$c1 = \frac{2*a1*b1}{a1+b1}$ $c2 = \frac{2*a2*b2}{a2+b2}$ $a1 = \frac{2*b1*c1}{b1+c1}$ $b1 = \frac{2*a1*c1}{a1+c1}$ Среднее гармоническое		Социализация

В таблице представлены варианты устранения данного недостатка. Символами а, b, с обозначены количественные значения мотивационно-волевых, ценностных, целевых, ресурсных и т.п. параметров. Значения задаются экспертным путем. Целостность представлена в трех формах: структурной, динамической и гармонической. Также приведены различные формы выражений – аналитическая и диаграмматическая. Соотношения базируются на эстетико-математических исследованиях [5] и апеллируют к пифагорейской традиции идеально-числовых критериев красоты. В условиях компьютерного моделирования воспроизводится проблема «поверить алгеброй гармонию». На наш взгляд, данный комплексный подход к определению понятия «потребность» через достаточно подробно охарактеризованную нами системную (функционально-структурную) целостность даёт возможность построения концептуально обоснованных моделей.

Литература

1. Философский энциклопедический словарь. М., 1983
2. Ершов П. М. Потребности человека. М. 1990.
3. Кувакин В.А. Твой рай и ад. Человечность и бесчеловечность человека. – СПб.: «Алетейя», М.: «Логос», 1998. – 358 с.
4. Капранов В.А. Нравственный смысл жизни и деятельности человека. Л.: Изд-во ЛГУ, 1975 – 150 с.
5. Алексеев А.Ю. Критерии целостности гуманистической информационной технологии. // В сб. трудов Международной московской конференции гуманистов «Наука и здравый смысл в России. Кризис или новые возможности». М.: 1998, С. 202-207

Тест Тьюринга и Д. Деннет

Алексей Великанов, Антон Макарычев (ЗИ-81)

Сегодня проблема ИИ стала не только технологической проблемой, но и, в большей мере, социокультурной проблемой. Если физики управляют электро-магнитными полями, биологи – генетическим кодом, то будут ли исследователи ИИ манипулировать формами и способами мышления людей (учитывая не столь далёкие инфотехнологические перспективы)? Может ли вторгнуться наука и технология в сферу человеческого сознания – последней, как считают многие, *terra incognita* человеческого, собственно человеческого?

В решении этого вопроса особый интерес представляет точка зрения Д. Деннета – одного из главных представителей сильного ИИ¹. Он считает, что страх «вторжения» в «святая святых» беспричинен. Но не в силу того, что сознание и самосознание как «последнее убежище» личности под натиском техногенной экспансии не поддастся компьютерной реализации. Напротив, человеческий разум, сознание и самосознание – сами суть компутационные формирования. Порождены они гигантским количеством различных «роботов» – от крупных анатомо-физиологических компонент до элементов субклеточного уровня.

Но пока человечеству нечего бояться. Слишком слабы компьютерные и когнитивные науки, чтобы притязать на реализацию сознания. И будут слабы до тех пор, пока основываются на парадигме исходного теста Тьюринга. Д. Деннет отстаивает правоту тезиса А. Тьюринга («машины могут мыслить») и считает, что любой компьютер, «без обмана» прошедший ТТ, можно считать разумным. «Без обмана» означает чистоту помыслов разработчиков теста – способность прохождения ТТ не должна соотноситься с намерением обмануть судью.

Однако ТТ следует модифицировать, чтобы он в большей мере соответствовал парадигме функционализма и позволял прояснять аспекты компутационального воспроизводства сознания и самосознания. Для этого Д. Деннет к игрокам-машинам добавляет различного рода компоненты: перцептивно-сенсорные приборы, моторно-двигательные механизмы и пр. За счет этого, как считает Д. Деннет, достигается «честность» ТТ – ведь «дряхлый, слепой компьютер с достаточно искусной программой, способной обмануть тест Тьюринга, есть научная фантастика наихудшего сорта». Этот старый компьютер ничего не может сделать в смысле построения «мыслящей машины» в силу «комбинаторного взрыва» в области возможных вариантов ответа на вопросы. Попытки реализации старого ТТ – это экспертные системы ([MYCIN](#), CYRUS и др.). Все такие системы – «деревни Потёмкина». У них красивая оболочка, но внутри – ничего интересного. Они узко специализированы. Отсутствуют не то что глубинные, но даже поверхностные «знания» о причинно-следственных зависимостях. Отсутствуют механизмы самоорганизации «знаний». Возможно и не будет компьютеров (в несколько ближайших лет), которые смогли бы пройти ТТ без существенных ограничений языка общения. Нельзя экстраполировать успех в некоторых узких предметных областях (достигаемый в усло-

¹ Dennett, D. C. 1984. *Can machines think?* In (M. Shafto, ed) *How We Know*. Harper & Row. <http://www.kurzweilai.net/articles/art0099.html?printable=1>

виях щедрого функционирования – неоднократно подчёркивает Д.Деннет) на возможность построения интеллектуальных систем в целом в рамках старой парадигмы.

В простых компьютерных системах, основанных на старом ТТ нет ничего отдалённо похожего на мышление, понимание, осознание и др. Имеет место лишь бессмысленная трансформация одних лент символов в другие в соответствии с простейшими синтаксическими и механическими правилами.

В будущих, гигантских системах ИИ дело может обстоять иначе. Да, они будут состоять из ничего не осознающих компьютеров. Но в сложных сетевых хитросплетениях, бесконечных взаимодействиях этих «ничего-не-осознающих» механизмов и возможно возникновение феномена сознания и самосознания.

Ведь и человек – совокупность «ничего-не-осознающих» роботов. Но он и владелец этих «таинственных» феноменов.

Тест Блока: антибихевиористское опровержение тьюринговой концепции мышления

Ирина Матанцева (ЗИ – 81)

Существенное методологическое значение для искусственного интеллекта имеет разработка способов критического анализа его достижений и претензий. В этом плане несомненный интерес представляет опыт антибихевиористской критики «мыслящих машин» и функционалистской парадигмы мышления в целом.

Суть критики такова: вначале показывается, что тест Тьюринга – разновидность бихевиоризма. Бихевиористская парадигма разрушается под натиском неоспоримых аргументов. Далее уничтожаются остатки тех нюансов, которые привносит ТТ в бихевиоризм. А так как функционалистская парадигма мышления находит надежную опору в ТТ, то, потеряв эту опору, он падает. Что и надо было доказать – функционализм несостоятелен.

В работе демонстрируется убедительной такой критики.

Основания функционалистской парадигмы мышления были заложены А.Тьюрином (1950)¹ в форме теста. Её объяснительное основание – «игра в имитацию» – представляется разновидностью лингвистического бихевиоризма². В *бихевиоризме* предметом психологического анализа выступает некая ментальная структура субъективной реальности, а объективно фиксируемые параметры поведения (реакции), определяемые внешними воздействиями (стимулами). В *лингвистическом бихевиоризме* стимулы и реакции – суть вербальные выражения. Имитация мышления осуществляется в ходе диалогового обмена последовательностями языковых выражений со стороны игроков. Игроки, в ходе диалога играющие в имитацию интеллектуальной деятельности – это и люди и машины.

¹ Turing, A. (1950), 'Computing Machinery and Intelligence', Mind 59(236), pp. 433–460.

² Другая лингво-бихевиористская парадигма была предложена Л.Витгенштейном (т.н. концепция «языковой игры»). См. *Современная буржуазная философия*. Учеб.пособие. Под ред. А.С.Богомолова, Ю.К.Мельвиля, И.С.Нарского. М., «Высшая школа», 1978 – 582 с. – С. 195; Зотов А.Ф. *Современная западная философия*. Учебн. - М., Высш.шк., 2001. – 784 с. - С.275-290.

Тотальная критика бихевиористского подхода к отождествлению способов описания интеллекта с диспозициональной структурой вербальных стимулов и реакций прослеживается в ряде работ Нэда Блока – профессора философии и психологии Гарвардского университета¹. Антибихевиористская критика теста Тьюринга особо отчетливо прозвучала в его работах «Психологизм и бихевиоризм» (1981)² и «Разум как программное обеспечение мозга» (1995)³. Совокупность критических положений разделяется на два класса – стандартных и нестандартных аргументов. Первая форма аргументации сложилась в русле общепсихологической критики бихевиоризма. Вторая обусловлена спецификой вопроса «может ли машина мыслить?».

Стандартные антибихевиористские аргументы

Н.Блок приводит ряд стандартных аргументов против бихевиоризма, полагая, что они эффективны и против концепции ТТ.

1. Аргумент Чишолма и Гича. Невозможно выделить некоторое психическое состояние в поведенческой диспозиции (структуры намерений) без учёта *всей совокупности иных психических состояний*. Данное положение иллюстрируется следующим образом. Допустим, бихевиорист исследует желание некоего человека съесть мороженое. Это желание он представляет в форме поведенческой диспозиции. В структуру этой диспозиции входит, например, намерение немедленно схватить мороженое, особенно если его дарят («пока не передумали»). Однако способ реализации даже такой мгновенной реакции предполагает протекание внутренних процессов. И эти процессы невозможно зафиксировать в схеме «стимул-реакция». Человек, жаждущий мороженое, намерен его схватить, если: а) он *знает*, что предлагаемый предмет – именно мороженое (а не тюбик дегтя, например, который предлагается шутки ради), и б) он *уверен*, что взятие мороженого не повлечет конфликта с другими более важными *желаниями*, (например, с желанием избежать обязательств в оказании ответной услуги). Таким образом, поведенческая диспозиция, характеризующая *желание*, вытекает из всей совокупности иных психических состояний человека. Подобные суждения справедливы и относительно *боли*. Поведенческая диспозиция – «предрасположенность к боли» – не является достаточным условием утверждения того, что индивид на самом деле чувствует боль. Его поведение может быть вызвано рядом различных комбинаций психических состояний, таких как: {боль + обычное переживание боли} или {отсутствие боли + желание обмануть, что боль переживается}.

2. Аргумент «совершенный актер» Патнэма. Данный аргумент против бихевиоризма вытекает из критики утверждения, что *различные* психические группы могут продуцировать *одинаковые* поведенческие диспозиции. Критику подобного рода принято называть «аргументом совершенного актера». Аргумент предложил Патнэм: вообразим общество совершенных актеров – неких «сверх-супер-спартанцев». В силу принятых в обществе законов им запрещено отражать в своем поведении те реакции, которые бихевиорист может ассоциировать с болью, даже если на самом деле люди испыты-

¹ См. домашнюю страницу Н.Блока: <http://www.nyu.edu/gsas/dept/philo/faculty/block/>

² Block, N. 1981, *Psychologism and Behaviorism*, Philosophical Review 90, pp. 5–43

³ Block, N. 1995, *The Mind as the Software of the Brain*, In D. Osherson, L. Gleitman, S. Kosslyn, E. Smith and S. Sternberg, eds., *An Invitation to Cognitive Science*. Cambridge, MA.: MIT Press (<http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/msb.html>)

вают боль. Боль не всегда проявляется в поведении. Контрпример: человек притворяется, что ему больно, хотя на самом деле это не так.

3) Аргумент «паралитики» и «мозги в бочке». Эти «пациенты» вообще не выказывают определенных признаков, означающих то, что они испытывают боль.

На основе этих аргументов, но уже применительно к мыслительной деятельности, Н.Блок высказывает следующие соображения: интеллектуальное поведение, как правило, продуцируется комбинацией: {интеллект + обычная склонность к мышлению}. Но возможна ли комбинация {отсутствие интеллекта + желание обмануть, что интеллект имеется (желание казаться разумным)}? Н.Блок отрицает такую возможность – нельзя представить такую комбинацию ментальных состояний и свойств, которые *не включают в себя интеллектуальные состояния (свойства)* и, тем не менее, продуцируют обманым путём диспозицию к разумному поведению.

Вследствие того, что ТТ (как лингво-бихевиористская парадигма) рушится под напором стандартной антибихевиористской аргументации, Н.Блок полагает, что необходима модификация теста Тьюринга. Новая формулировка ТТ как новая концепция мышления (в рамках тьюринговой установки), которая называется «нео-ТТ», звучит следующим образом: *Интеллект или, более точно, диалоговый интеллект (т.е. интеллект, проявляющийся в диалоге) – суть способность продуцирования осмысленной последовательности вербальных реакций на некоторую последовательность вербальных стимулов, но отнюдь не факт продуцирования.*

По мнению Н.Блока, данная формулировка позволяет отразить атаки стандартных антибихевиористских аргументов:

1) Возражение Чизхолма-Гича: действительно, можно представить сколь угодно много вариантов убеждений и желаний, которые могут оказаться причиной того, что разумное существо *не будет расположено* к выдаче осмысленных ответов. Тем не менее, данные убеждения и желания никак не сказываются на *способности* существа давать такие ответы;

2) Аргумент «совершенный актер»: надо обладать недюжинными интеллектуальными *способностями* для того, чтобы в совершенстве симулировать отсутствие разума. Невозможно притворяться разумным и не быть таковым;

3) Для интеллектуалов-паралитиков и «мозгов в бочке» вообще нельзя найти опровергающие примеры – у них, несомненно, имеется *способность* реагировать осмысленно. Однако для реализации этой способности им не хватает средств.

Таким образом, невозможно вообразить, что некая комбинация психических состояний, в составе которой отсутствует интеллектуальная способность, будет иметь следствием способность отвечать осмысленно на произвольную последовательность стимулов.

В силу очевидности данного утверждения, неоТТ отражает стандартные антибихевиористские атаки. Тем не менее, «может ли машина мыслить?» в условиях новой модификации ТТ.

Нестандартные антибихевиористские аргументы

Для «отражения» нестандартных аргументов Н.Блок детально описывает машину, которая способна производить осмысленную последовательность вербальных реакций на вербальные стимулы¹. В соответствии с концепцией неоТТ, данная машина на поведенческом уровне представляется интеллектуальной. Однако знания её внутреннего устройства убеждают в полном отсутствии у неё «интеллекта». Вся совокупность осмысленных вербальных выражений (которая, к тому же является конечным множеством) на выходе такой машины задаётся исключительно человеческим коллективом (проектировщиков, инженеров, программистов и др.). Люди долго и упорно трудятся для реализации всевозможных последовательностей осмысленных ответов на предполагаемые последовательности вербальных стимулов. В ходе работы используются специальные инструментальные средства, например, средства автоматизации программирования. Самое главное – люди *применяют воображение и принимают решения* о том, что считать осмысленной последовательностью языковых выражений.

Такая машина, несомненно, является разумной в соответствии с концепцией неоТТ. Однако *интеллект, который машина «проявляет» как в форме актуального поведения так и в форме диспозиции интеллектуального поведения – суть интеллект программистов.*

Вывод: лингво-бихевиористской способности искусственной системы продуцировать «осмысленные» ответы не достаточно для приписывания ей интеллектуальных качеств. Мышление машины – суть мышление экспертов (программистов, в первую очередь). Нео-концепция мышления, представленная в форме неоТТ, опровергается. Заодно опровергается и стандартный ТТ, как более грубый тест, не учитывающий психологистских нюансов. *Машина мыслить не может!* – таков итог антибихевиористской критики функционалистской парадигмы мышления.

Многообразие тестов Тьюринга Виктор Морозов (АП-82)

Полвека рефлексии над подходом к определению интеллекта, предложенного А. Тьюрингом в форме знаменитого на весь мир теста Тьюринга дали достаточно много интересных и плодотворных уточнений и изменений оригинальной трактовки. К тесту Тьюринга и его модификациям обращался, пожалуй, каждый крупный мыслитель последних десятилетий, изучавший проблематику сознания, разума, искусственного интеллекта: Х. Патнэм, Р. Пенроуз, Д. Деннет, Д. Серль, Н. Блок, Т. Нагель, Дж. Маккарти, А. Сломан и др. Уточнения и модификации породили целое многообразие ТТ. Поэтому представляет несомненный интерес попытка систематизации этого многообразия ТТ.

Определённую работу в этом направлении предпринял Харнад². В многообразии ТТ он усматривает т.н. «восходящее проектирование»: переход от эмпирических данных к теории. ТТ выстраиваются в иерархию тестов от «игрушечных» моделей до «модели вселенной»: Т1 – Т5. Каждый уровень

¹ Block, N. 1981, *Psychologism and Behaviorism*, Philosophical Review 90, pp. 9–13

² См. <http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad00.turing.html>

иерархии задаётся той или иной степенью функциональной неотличимости ИИ от естественного интеллекта.

Рассмотрим предложенную классификацию ТТ:

И1. Первый уровень следовало бы назвать Т1, а не И1, где «И» обозначает «игрушечный». У Харнада И1 стоит первым в списке ТТ, не являясь полноправной частью классификации. И1 – это фрагменты наших функциональных способностей к вербальному выражению мысли. И1-модели не отвечают целям тестирования Тьюринга. Тем не менее, все попытки моделирования интеллекта, известные к сегодняшнему дню, остаются на этом уровне. Пока ещё не достигнут первый уровень ТТ и когда вместо И1 сформируется полноправный Т1 – не известно.

Следующие уровни на сегодняшний день имеют исключительно теоретический характер.

Т2 – общепринятое понимание ТТ, которое обычно называют тестом «друг по переписке». Он реализуется в режиме «вопрос-ответ». Именно данный режим и имеют ввиду, когда говорят о ТТ. Кандидат должен быть неотличим от обычного друга по переписке на протяжении всей жизни респондента. Тест Т2 способен поддерживать разговор о реальных предметах, событиях и символах, но не способен к оперированию предметами реального мира, совершению действий в мире, судя по которым так же оценивается естественный интеллект («по делам узнают человека»).

Т3 – это роботизированная версия Т2. Она уже способна к манипуляции предметами внешнего мира. На уровне Т3 заявляется возможность полной неотличимости ИИ-системы от человека, вплоть до мельчайших нюансов телесного строения последнего. Работа судьи (из классической схемы ТТ) крайне затруднена. Ему остаётся только одно – «вскрыть» телесную оболочку тестируемой системы и изучить её внутреннее строение, чтобы узнать – компьютер (робот) перед ним или человек.

Т4 – на этом уровне «вскрытие тела» не приводит к ожидаемым результатам: тестируемые кандидаты неотличимы не только способами вербального оформления мыслей (Т2) и не только внешним телесным строением (Т3), но и микрофизической организацией.

Т5 – здесь происходит обобщение Т4-систем на основании некой одной из Великих Объединённых Единых Теорий Всего Сущего (ВОЕТВС). Предполагается, что такие теории, в случае их создания, смогут объединить знания из всех областей науки и культуры и с их помощью можно будет описать любые явления в окружающем нас мире и обществе. Для чего нужен Т5? Рассмотрим пример. Допустим, после прохождения Т3 осталось 9 кандидатов. Трое из них не пройдут Т4. Из оставшихся шести трое не пройдут Т5. Последние три ТТ-системы, оставшиеся на уровне Т5 спроектированы из реально существующих биомолекул и микрофизически идентичны человеческим, т.е. полностью, Т5-неотличимы от нас. Вся разница между этими тремя системами – это то, что они были разработаны в сотрудничестве с тремя разными физиками, каждый из которых создал свою ВОЕТВС и использовал различные модели микромира.

Из предложенного многообразия ТТ решающим, несомненно, является уровень Т3. Однако когда он будет достигнут, если мы до сих пор «топчемся» на И1 – уровне?

Критика сильного искусственного интеллекта. Аргумент Гёделя

Артур Никишев (ЭП-81)

Речь в данной работе пойдет о так называемом «гёделевском аргументе», который используется как аргумент против возможности создания сильного искусственного интеллекта.

Суть аргумента заключается в следующем: полагают, что из *теоремы К. Гёделя о неполноте формальных систем* вытекает принципиальное различие между искусственным («машинным») интеллектом и человеческим разумом. Теорема Гёделя указывает на некоторое принципиальное преимущество человеческого разума перед машинным «разумом» – т.е. человек обладает способностью решать проблемы, принципиально неразрешимые для любых искусственных «интеллектуальных» систем (так называемые «алгоритмически неразрешимые» проблемы), причем ограниченность искусственного интеллекта проистекает из того, что он задаётся формальным способом.

Чтобы понять суть данной теоремы, необходимо уточнить смысл понятий, входящих в ее формулировку. Прежде всего, необходимо уточнить понятие «формальной системы» – поскольку только к таким системам и имеет отношение рассматриваемая теорема. В самом общем плане формальная система – это система, подчиненная неким жестким, однозначно заданным правилам. Соответственно, «формализацию» можно определить как процедуру, цель которой – дать предельно четкое, однозначное и исчерпывающее описание объекта, подлежащего формализации.

Главное требование к формализму – символы, используемые в данной формальной системе, должны принимать лишь те значения, которые им приписываются в явном виде. Эти фиксированные значения задаются посредством правил, указывающих способ действия с тем или иными символами, а также через описание взаимных отношений между заданными символами.

Все, что необходимо для работы с формальной системой, для понимания смысла ее выражений – содержится в ней самой.

Доказательство строится следующим образом. Задается формализованный язык данного исчисления. Для этого определяется алфавит и грамматика языка. Алфавит – это набор символов (букв), допустимых в данном языке. Имея алфавит, мы можем составлять слова – любые, сколь угодно длинные последовательности букв заданного алфавита.

Для того, чтобы выделить из множества всевозможных слов интересующие нас («осмысленные») сочетания букв, вводится грамматика – совокупность правил, позволяющих определить «правильно построенные слова» (или, иначе, правильно построенные выражения). Правила грамматики вводятся индуктивно: вначале определяются элементарные выражения, а затем указывается, каким образом из них можно построить любые более сложные выражения.

Далее из множества выражений выделяют подмножество формул. Содержательно, формулы – это выражения, которые что-то утверждают (например, утверждают нечто о свойствах чисел или геометрических фигур). Формулы также определяются индуктивно.

Далее выделяют множество замкнутых формул или выражений. Это формулы, которые не имеют свободных параметров (т.е. параметров, которые могут принимать различные значения и не связаны кванторами всеобщности или существования). К таким формулам можно приписать определенное значение – «истина» или «ложь».

Замкнутые формулы истинны или ложны с содержательной точки зрения. Естественно потребовать, чтобы формализованная математическая теория включала в себя только содержательно истинные формулы. Истинность в математике определяется посредством доказательства. Таким образом, следующий шаг – введение формализованной системы доказательства дедуктивного типа. С данной целью задается некоторое конечное множество замкнутых формул, истинность которых принимается без доказательств. Это – аксиомы данной дедуктивной системы. Далее задается конечное множество правил вывода, которые позволяют из одних истинных формул получать другие истинные формулы.

Всякое формализованное доказательство – это некоторое слово формального языка, представляющее собой цепочку формул, в которой каждая формула – это либо аксиома, либо получена из аксиом посредством применения тех или иных правил вывода. Последняя формула в цепочке – это и есть доказанное утверждение (теорема). Обозначим множество всех доказательств символом D^* , а множество всех доказанных формул Id^* . Через I^* – обозначим множество содержательно истинных замкнутых формул данного языка.

Теорема Геделя о неполноте формальных систем утверждает, что для любой достаточно выразительно богатой формальной системы выполняется условие $I^* > Id^*$ и, следовательно, существует истинная недоказуемая формула. Это верно при условии, что заданная дедуктивная система непротиворечива, т.е. не позволяет одновременно доказывать некоторое утверждение и его отрицание.

Идея доказательства заключается в том, чтобы построить пример формулы, которая была бы недоказуема и, вместе с тем, содержательно истинна. Таковой являлась бы формула, содержательный смысл которой заключается в том, что она утверждает свою собственную недоказуемость, т.е. невыводимость из аксиом рассматриваемой формальной системы.

Для построения такой формулы, Гёдель изобрел способ нумерации предложений формальной системы. Данный способ (его принято называть «гёделевской нумерацией») позволил однозначным образом приписывать некоторый номер (натуральное число) каждому элементарному символу, формуле или доказательству данной формальной системы.

Используя гёделевскую нумерацию можно построить формулу, утверждающую недоказуемость формулы с номером n , где n – номер самой этой формулы. По существу, гёделевская нумерация задает специфический арифметический метаязык, на котором можно высказывать суждения о свойствах рассматриваемой дедуктивной системы в форме суждений о числах.

Обозначим через **Dem** (x , y) – метаязыковое выражение, означающее *«последовательность формул с гёделевским номером x является доказательством формулы с гёделевским номером y »*. Навесим на x квантор общности и подвергнем **Dem** (x , y) отрицанию. В результате мы получим одноместный предикат:

(*) {для всех x не верно $\text{Dem}(x, y)$ }, который утверждает недоказуемость формулы с геделевским номером y .

Следующий шаг заключается в подстановке в (*) вместо « y » формального (метаязыкового) выражения для номера самой формулы (*).

Пусть формула (*) имеет геделевский номер h . Обозначим через $\text{Sb}(\text{Wvz}(n))$ номер результата подстановки в формулу с номером W на место переменной с номером V формулы с номером $Z(n)$. $Z(n)$ – в данном случае – номер формального выражения формулы с геделевским номером n . Пусть, также, m – геделевский номер переменной « y ».

Построим формулу:

(1) {для всех x не верно $\text{Dem}(x, \text{Sb}(hmz(h)))$ }.

Легко установить, что геделевский номер формулы (1) равен $\text{Sb}(hmz(h))$ так как эта формула получена из формулы с номером h путем подстановки вместо переменной с номером m (т.е. « y ») формального выражения числа h . Следовательно, (1) и есть искомая «геделевская формула («геделевское предложение») G .

Запишем геделевское предложение в виде:

[формула с номером $\text{Sb}(hmz(h))$ недоказуема], где $\text{Sb}(hmz(h))$ – номер формулы: [формула с номером $\text{Sb}(hmz(h))$ недоказуема].

Если данная формула доказуема, то она истинна, но тогда истинно, что она утверждает, а именно, что она недоказуема. Т.е. **если она доказуема, то она недоказуема**. Таким образом, мы получили противоречие.

Если же данная формула недоказуема, то она, очевидно, истинна (поскольку утверждает, что она недоказуема и на самом деле недоказуема). Т.е. эта формула является **истинной недоказуемой формулой** (в рамках заданного формализма).

Такое элегантное решение позволяет чётко определёнными средствами математического дискурса развернуть полемику по поводу возможности/невозможности создания ИИ. Особо яркие дискуссии инициировали Р. Пенроуз и Дж. Лукас. И убедительно показали, что из аргумента Гёделя следует вывод о невозможности создания *искусственного интеллекта, заданного исключительно формальным способом*. Необходимо привлечение мета-системы для непротиворечивого определения понятий формальной системы. Над ней – ещё одна метасистема и т.д. В конечном счёте, последней инстанцией челонок – те понятия и определения, которыми он пользуется.

ТЕСТ СЕРЛА: ИНТЕНЦИОНАЛИСТСКОЕ ОПРОВЕРЖЕНИЕ КОНЦЕПЦИИ ТЬЮРИНГА

Денис Родионов (АП-82)

Сильный удар по функционалистской парадигме мышления, предложенной А.Тьюрингом в форме теста, был нанесён Дж. Серлом в 1980 г.¹ Точнее, удар был нанесён по концепции сильного искусственного интеллекта. Эмпирическим базисом для атаки Серла послужили практические наработки в области ИИ – работы Роджера Шэнка и его коллег из Йельского универси-

¹ Searle John R., 1980, Minds, Brains, and Programs.
(<http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>)

тета в области понимания текстов (рассказов) на естественном языке. Цель программы Шенка – моделирование человеческой способности понимать тексты. Факт понимания связывается со способностью человека отвечать на вопросы о содержании текста в условиях неявно представленной в тексте информации об ответе (т.е. в условиях отсутствия возможности установления соответствия между вопросом и ответом без привлечения внетекстовой информации). Программа Р. Шенка, считает Дж. Серл, – это реализация теста Тьюринга. Машина проходит ТТ, если на её вход подадут текст и она отвечает на задаваемые вопросы таким же образом, каким, как ожидается, будет отвечать и человек. Сторонники сильного ИИ утверждают, что в данной последовательности вопросов и ответов машина не просто моделирует человеческую способность к пониманию. Машина в прямом смысле *понимает* текст. Более того, машина и её программа *объясняют* человеческую способность понимать текст и способность человека осмысленно отвечать на вопросы о данном тексте.

Для опровержения концепции сильного ИИ Дж. Серл использует мысленный эксперимент, широко известный как *тест Китайская комната*. Некого субъекта, проходящего тестирование на понимание, запирают в комнате. Субъект китайским языком абсолютно не владеет – не умеет ни писать, ни читать. Для него любой текст на китайском языке – суть то, что образно называют «китайской грамотой». Однако субъект превосходно владеет английским языком – является его носителем. Субъекту предоставляют три набора текстов: 1) «Национальный алфавит» – совокупность китайских иероглифов, 2) «Тексты» – другой набор иероглифов, в котором представлены, помимо иероглифов, правила соответствия второго набора первому. Правила записаны на английском языке, поэтому субъект может соотносить иероглифы «текстов» с иероглифами «алфавита»; 3) «Вопросы» – третий набор иероглифов. К третьему набору прилагается «программа перевода» – инструкция на английском языке для соотнесения «вопросов» из третьего набора с элементами двух предыдущих наборов. Субъект, получив эти три набора, должен выдавать «ответы на вопросы» – выражения, осмысленность которых должен оценить сторонний наблюдатель («судья» в оригинальном ТТ). Помимо текстов с «китайской грамотой» субъекту дают тексты и на английском, задают английские вопросы и требуют ответов на английском же языке.

Спустя некоторое время судья убеждается, что ответы на китайские вопросы становятся совершенно неотличимы от ответов, которые судья получает от истинных носителей китайского языка. Судья не может даже предположить, что субъект совершенно не знает китайского языка, а лишь осуществляет формализованные операции над символами. С точки зрения наблюдателя, ответы на английском и ответы на китайском одинаково хороши. Если восстановить схему «перевода» (в её «китайской части»), то получается, что для выдачи осмысленных ответов на вопросы никакого понимания со стороны субъекта и не требуется. Раз так, то нет такого «понимания» и у машины (включая и программу Р. Шэнка). То есть программа Шэнка отнюдь не приближает нас к объяснению человеческой способности понимать, в отличие от мнения сторонников сильного ИИ. Ведь нельзя назвать «пониманием» совокупность формализованных операций над совокупностью символов.

Но факт понимания текстов в «Китайской комнате» очевиден. Для ответов на английские вопросы субъект не пользуется никакими «программами перевода» (для подчеркивания очевидности факта понимания Дж. Серл про-

водит эксперимент от первого лица). Что же способствует пониманию субъектом английского и, соответственно, недостаток чего вызывает неспособность понимать китайский?

Дж. Серл считает, что *интенциональность*. Человек – это биологическое существо. Это существо способно ощущать, действовать, понимать, обучаться и т.д. – т.е. проявлять свою интенциональность. Интенциональные феномены не редуцируемы к формальной схеме. Ни одна чистая формальная модель не может быть самодостаточной для воспроизводства феномена интенциональности. У формальных схем нет никаких самостоятельных казуальных сил. Сила, реализующая формальную схему, для своего описания требует иного уровня формализации. Для этого уровня требуется иная сила. И т.д. *Функционирование мозга – это вовсе не формальная тень, отбрасываемая последовательностями синапсов, а действительные свойства этих последовательностей.*

То, что на самом деле обуславливает понимание – это *семантика, смысл*. Интенциональность, имеющаяся у компьютеров, на самом деле имеется лишь в головах у тех, кто их программирует, и тех, кто подает им на вход тексты и интерпретирует тексты, получаемые на выходе.

Функционалистская парадигма мышления, предложенная в форме ТТ, поколеблена. Но в «Китайской комнате» не уживается лишь сильный ИИ. Слабый ИИ не разрушен. Тест Серла – «камень в огород» сторонников подхода, «которые вначале рисуют контуры теней, отбрасываемых когнитивной способностью человека, а затем заявляют, что тени на самом деле и есть истинная реальность,» – образно заключает Дж. Серл.

На мой взгляд, много сторонников сильного ИИ среди математиков. Именно математикам в силу специфики своей профессии присуще искушение отождествлять реальность с числами и числовыми соотношениями. Математика же – только «формальная схема» систем искусственного интеллекта.

Парадигма функционализма: как представить ментальное в нементальных терминах?

Денис Романов (С-85)

Становление и развитие функционалистской парадигмы мышления обусловлено развитием кибернетики, теории информации, семиотики, системных и структуральных исследований.¹ Функционализм – одно из главных теоретических событий аналитической философии двадцатого века. Он является концептуальной основой большинства работ в сфере когнитивных наук².

Функционализм предлагает особый подход к решению проблемы дух/тело. При решении данной проблемы обычно пытаются ответить на вопросы, такие как: Что такое ментальное? Что такое мысль? Где мысль существует? Что делает мысль мыслью? Что делает боль болью?

Картезианский дуализм утверждает, что ментальное, в частности, мышление «пребывает» в особой духовной субстанции. Бихевиоризм идентифи-

¹ Дубровский Д.И. Психика и мозг. Результаты и перспективы исследований. // Мозг и разум. М.: Наука, 1994 – 176 с. – С.11

² Block, N. 1980. Functionalism. In (N. Block, ed) Readings in the Philosophy of Psychology, Vol. 1. MIT Press. В дальнейшем в работе исследуется интерпретация функционализма, данная Н. Блоком в этом первоисточнике.

цирует мышление с поведенческой функцией. Физикализм в своей основной версии отождествляет мышление с состояниями мозга. Функционализм утверждает, **что мышление – это отношение между чувственными данными (ощущениями) и поведенческими реакциями.**

Функционализм имеет три различных *источника*: 1) Представление мыслительных процессов в терминах вычислительной теории разума (Патнэм и Фодор). 2) Статья Смарта «topic neutral» приведшая Армстронга и Льюиса к функциональному анализу ментальных понятий. 3) Идея Витгенштейна о языковых играх, приведшая к созданию функционалистской теории значения, которая далее была развита Селларсом и позже Харманом.

Суть парадигмы функционализма можно продемонстрировать, обратив внимание на техническое понятие *карбюратора* и биологическое понятие *почки*. Карбюратор – функциональное понятие. То, что им называется, смешивает воздух с топливом и подает полученную смесь в камеру сгорания. В случае с почкой научное понятие *почки* выступает функциональным понятием – это то, очищает кровь и поддерживает требуемый химический баланс в организме.

Способ описания функционалистами ментального удобно продемонстрировать на примере автомата, представленного в виде таблицы (Рис.1).

	S_1	S_2
1	«Нечетный» S_2	«Четный» S_1
0	«Нечетный» S_1	«Четный» S_2

Рис. 1. Автомат с двумя входами

Автомат имеет два состояния – S_1 и S_2 ; два входа – «1» и «0», и два выхода – слова «Четный» или «Нечетный». Таблица описывает две функции. Первая от входа и состояния выхода, вторая – от входа и состояния на следующем этапе работы автомата. Каждый квадрат кодирует два условия: состояние выхода (Четный/Нечетный) и входа. Левый верхний квадрат показывает, что если автомат находится в S_1 и видит «1», то говорит «Нечетный» (инструкция, что он видит нечетное число «1») и переходит в S_2 . Правый квадрат показывает, что если автомат находится в S_2 и видит «1», то говорит «Четный» и возвращается к S_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 заикнется до тех пор, пока не увидит «1».

Данный недостаток исправлен исправлен в следующем автомате, представленным на Рис. 2.

	S_1	S_2
1	«Нечетный» S_2	«Четный» S_1

Рис. 2. Автомат с одним входом

Он проще. Как и предыдущий, автомат имеет два состояния, S_1 и S_2 и два выхода, «Четный» или «Нечетный». Различие состоит в том, что он имеет всего один вход «1», хотя конечно он не может получать на вход все сигналы (в отличие от автомата Рис. 1, который получает и «0»). Как прежде, таблица описывает две функции, одна от входа и состояния на выходе, а другая от входа и состояния на следующем этапе. Как прежде, каждый квадрат кодирует два условия: состояние выхода («Четный»/«Нечетный») и входа. Левый квадрат показывает, что, если автомат находится в S_1 и видит '1', то говорит «Нечетный» (указание, что он видел нечетное число в «1») и переходит в S_2 . Правый квадрат показывает то, что если автомат находится в S_2 и видит «1», он говорит «Четный» и возвращается в S_1 . Этот автомат более прост чем автомат рис. 1 и по сути выполняет ту же самую задачу, избегая ложного объявления нечетного числа в «1». Теперь зададим вопрос: «Что такое S_1 ». Ответ: S_1 – это отношение, полностью задаваемое таблицей. Можно дать явную характеристику ' S_1 ' (из рисунка 2) следующим образом:

Находиться в S_1 = быть в первом из двух состояний, которые связаны друг с другом, а также с входами и выходами следующим образом: пребывать в первом состоянии, при '1' на входе переходить во второе состояние и выдавать «Нечетный»; пребывать во втором состоянии и при '1' на входе переходить в первое состояние, выдавая «Четный».

Осуществим более явную квантификацию:

Находиться в S_1 = быть x таким, что $\exists P \exists Q$ [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», и x находится в P].

Приведённые таблицы могут многое проиллюстрировать:

(1) Согласно функционализму, сущность ментальных состояний есть тоже самое, что и состояния автомата¹. Данная сущность конституируется посредством отношений к другим состояниям, а также к знакам на входе и выходе. Боль есть такое состояние, которое *предопределяет* крикнуть «Ой»;

(2) Поскольку ментальные состояния подобны автомату, проиллюстрированный метод (определения состояния автомата), пригоден для объяснения интеллекта. Ментальные состояния можно полностью охарактеризовать в терминах логико-математического языка, описывающего состояния, чувственные данные и поведенческие реакции. Таким образом, функционализм устраняет один из недостатков бихевиоризма, полностью характеризуя интеллектуальную деятельность на формальном языке.

(3) S_1 во втором состоянии обладает *другими* свойствами. Это могут быть механические или гидравлические или электронные свойства. Эти другие свойства, при определенных взаимоотношениях могут *реализовывать* функциональные свойства.

(4) Одно определённое функциональное состояние может быть представлено различными способами. Например, металлическая и пластмассовая

¹ Поэтому термины, которые мы ранее применяли для описания автомата, звучат несколько странно для тех, кто знаком с теорией автоматов. Например, вместо «видеть» следовало бы употреблять «считывать», не «Говорить «Нечётный», а «Писать «Нечётный». Мы оставили ту терминологию, которую использовал Н. Блок. Это подчёркивает, что понятие автомата применяется не к абстрактной машине Тьюринга, а к человеку.

машина, удовлетворяющая таблице автомата, может быть сделана из механизмов, колес, шкивов и т.п. Реализация S_1 может быть механическим состоянием, состоянием электронной схемы и т.д.

(5) Так как S_1 может быть реализован различными способами, то утверждение о том, что S_1 *суть* механическое состояние является ложным, так как можно утверждать, что S_1 является электронным состоянием.

Приведённые Н. Блоком положения позволяют нам сделать вывод, который будем расценивать как **ключевое положение функционалистской концепции мышления**:

Существо может мыслить и без мозга. Мышление может быть причинно связано не только с состояниями мозга, но и с состояниями иного материального субстрата. На современном этапе развития искусственного интеллекта в роли субстрата могут выступать электронно-коммутационные составляющие компьютера.

Предположим теперь, имеется теория ментальных процессов, которая характеризует все возможные соотношения между состояниями, входами и поведенческими реакциями. Пусть «боль» будет типовым ментальным процессом. Тогда можно сказать, что сидение на гвозде причиняет боль, боль причиняет беспокойство и желание крикнуть «Ой». Если согласиться с этой неубедительной теорией, функционалист может определить «боль» следующим образом:

Чувствовать боль = находиться в первом из двух состояний, т.е. сидеть на гвозде и переходить в другое состояние, крича «Ой».

Более обобщённо, пусть T – психологическая теория с n ментальными условиями, из которых, допустим, 17-ое является «болью». Тогда T можно задать следующим образом: ' F_1 ' ... ' F_n ' – переменные, которые характеризуют n ментальных условий; i_1, o_1 и т.д. – это индикаторы:

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 в формальных терминах**, т.е. в терминах, которые характеризуют ментальные состояния без всякого упоминания о физической реализации этих состояний.

Парадигма функционализма характеризует и интеллект в формальных терминах функциональных отношений между значениями на входе, значениями на выходе и состояниями. Реализации данной формальной системы порождают то, что называется «мышлением» – полагают функционалисты.

Наивная психология и инвертированный тест Тьюринга

Елена Романова (М-83)

Доклад основан на работе Стюарта Ватта¹. Он считает, что тест Тьюринга (ТТ) неявно опирается на «наивную психологию» – т.е. на естественные психологические особенности человека понимать и предсказывать поведение других. В условиях ТТ эти особенности играют главную роль в приписывании ментальных качеств исследуемой системе. В начале рассматривается влияние наивно-психологической установки на ТТ – как со стороны тестируемой системы, так и со стороны наблюдателя. Затем предлагается и обосновывается инвертированная версия теста, в которой приписывание ментальных состояний анализируется явным образом и непосредственнее, чем в стандартной версии.

Судье (наблюдателю) присуща т.н. «естественная» установка приписывания ментальных состояний тестируемой системе вне всякой зависимости от её фактического поведения. Поэтому была предложена инвертированная версия ТТ. В ней не акцентируется внимание на том, каким образом наблюдатель различает между разными системами. Акцент приходится на способность системы к саморазличению. Тест оценивает внутрисистемную способность приписывания ментальности *другому* на основании того, как это делают квалифицированные человеческие судьи. Наблюдатель становится наблюдаемым. Модифицированный тест обладает той же силой, что и оригинальный. Но он позволяет обратить внимание на ранее незамеченных параметров оценки интеллектуального поведения.

Наивная психология включает естественную человеческую установку приписывать ментальные состояния другим, так же, как он приписывает их себе. Это – естественная способность идентифицировать и воспринимать «другой разум». По сути, здесь видится психологическое решение философской проблемы «других сознаний».

Интерес к наивной психологии вызван следующим. Если очевидна естественная способность понимать «другие сознания», то эту способность можно соотнести с ТТ. Существуют две особенности данного соотнесения: 1) сам ТТ должен быть достаточно сильным для того, чтобы с его помощью предоставлялась возможность исследования наивно-психологической установки, и, 2) факты приписывания наблюдателем интеллектуальных свойств тестируемой системе в силу присущей ему наивно-психологической установки следует тщательно проверять и контролировать.

В принципе, невозможно доказать логическую необходимость наивной психологии для тестирования. Это можно лишь постулировать. Но очевидно, что без наивно-психологической установки невозможно представить искусственное устройство, способное чувствовать и понимать. Также невозможно этому устройству приписать ментальные состояния, которые не отличаются от ментальных состояний реального человека.

¹ [Watt, Stuart \(1996\)](#). Naive Psychology and the Inverted Turing Test, *Psychology*: 7, #14 [Turing Test](#) (Перевод Новиковой Натальи (ИС-81) и Сандакова Александра (ИС-81))

Когда учтена наивно-психологистская установка наблюдателя, его роль в ТТ становится не столь пассивной, как это казалось ранее. Неявно, а, пожалуй, даже явно, психологический багаж наблюдателя становится ощутимым в ТТ. *Наивно-психологистская установка характеризует готовность вообразить систему интеллектуальной даже тогда, когда она таковой не является.*

Чтобы устранить «естественные» предубеждения, Тьюринг предложил использовать телеграф (телетайп) как средство обмена сообщениями между участниками игры. Телеграф используется как «экран, позволяющий выбирать только то, что действительно значимо». То есть из поля зрения наблюдателя исключаются представления о телесной организации тестируемой системы: «мы не будем приписывать штрафное очко машине за то, что она не способна сиять на конкурсе красоты» (Тьюринг, 1950).

Стена (экран, перегородка) должна отвечать ряду требований. Если она каким-то образом намекает наблюдателю по поводу того, кто за ней скрывается, то необходимо усилить требования к её непроницаемости. На решение наблюдателя большое влияние оказывают биологические факторы. Если станет известно, что товарищ по переписке – человек (или признан таковым), то это коренным образом повлияет на исход тестирования. Применение же телеграфа как своеобразной «стены» не только скрывает лицо, но и исключает подсказки наблюдателю. Телеграф играет ключевую роль в схеме взаимодействия.

Итак, телеграфная связь – это не только «стена», существенно влияющая на параметры взаимодействия. Телеграф устанавливает своеобразный культурный контекст. Субъекты эксперимента должны быть «грамотными» – владеть навыками работы с телеграфом. Такой контекст взаимодействия способен повлиять и на решение судьи. При применении других технических средств, например, электронной почты решение судьи, возможно, будет иным.

Почему же машинам так легко приписываются ментальные свойства? Да потому что *приписывание лишь частично зависит от особенностей исследуемой системы!* Приписывание – суть системный показатель взаимодействия наблюдателя с системой. В некотором смысле, данный показатель измеряется наивно-психологистской установкой наблюдателя, которая активизируется при определённом поведении исследуемой системы и проявлениями этого поведения через средства взаимодействия. Может ли это в действительности быть тем, что мы понимаем под словом «интеллект»?

Тест на наивно-психологистскую установку

Выделим две стороны ТТ, характеризующие его обусловленность от наивной психологии как в аспекте способа функционирования устройства, так и в аспекте решений наблюдателя. Они раскрывают значимость связи между тестом и наивной психологией. Тест Тьюринга не способен к различению настолько, насколько это требуется для целей наивной психологии. Поэтому вначале следует сосредоточиться на анализе способов воздействия наивно-психологистской установки на наблюдателя и только после этого разъяснить значимость наивной психологии для теста.

Сегодня уже разработаны наивно-психологические тесты. Это – тесты на «ложную веру», цель которых 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 верил во что-то ложное и чтобы ребенок

знал о ложности такой веры. Тесты «ложной веры» позволяют установить границу между ментальными состояниями, которые ребёнок приписывает другим и его собственными ментальными состояниями.

Предполагаемая аналогия между развитием наивно-психологистских тестов и способов реализации ТТ подсказывает возможность преодоления в ТТ проблемы предубеждений при приписывании интеллекта исследуемой системе. Можно «перевернуть» весь тест и вместо того, чтобы оценивать способность системы обманывать людей, исследовать – а приписывает ли система интеллект другим таким же образом, как это осуществляют люди. Данный «поворот» и есть суть того, что понимается под термином «инвертированный тест Тьюринга».

Изменение ролей: инвертированный тест Тьюринга.

Мощность ТТ в большей мере основана на наивно-психологистской установке наблюдателя. Эта естественная, присущая человеку способность, смещает тест к даче ложных показаний. Реальная сила теста обусловлена ролью наблюдателя (Коллинз, 1990), а не ролью системы. Можно просто учесть такую установку. Однако тест от этого не улучшится (Френч, 1990). Но можно и преодолеть наивно-психологистскую установку. Для этого нам нужен тест, который на место судьи поставит машину. В этой перестановке и заключается суть предлагаемого инвертированного ТТ. Инвертированный тест сравнивает способность различения со стороны системы той способности различения, которая присуща опытному судье-человеку. Система проходит тест, если 1) она не способна самостоятельно отличить человека от человека; 2) она не способна самостоятельно отличить человека от машины (машина при этом проходит нормальный ТТ); 3) но она способна самостоятельно отличить человека от машины-судьи (такое имя присвоено машине в силу роли, которую человек-судья выполняет в нормальном ТТ).

Идея инвертированного теста заимствована из наивно-психологических тестов «на ложную веру». В ней совмещён формат теста «на ложную веру», который используется при приписывании *другим* ментальных свойств с форматом стандартного ТТ. Получилось, что вместо того, чтобы оценивать лингвистическое поведение тестируемой системы и поведение человека (сравнение лингвистического поведения – основа стандартного ТТ), в нашем тесте 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 и человеку. Если это так, то перед нами – человек, а не машина.

Также как и любая другая модификация ТТ, предлагаемый способ тестирования открыт для критики. Следует обратить внимание на ряд проблем. Во-первых, как и для нормального ТТ, предполагаемое поведение можно, в принципе, моделировать без каких-либо обоснований. Тривиальной представляется модель полной неспособности к различению. Однако крайне далека от тривиальной проблема попытка моделирования человеческой способности к различению. Для этого требуются и фундаментальные теоретические основания, и знания, основанные на здравом смысле. Также необходимы навыки построения системы, способной пройти нормальный ТТ. В силу этого инвертированный тест намного убедительнее стандартного – он обоснован большим количеством положений.

Вторая проблема – проблема идентичности. Если тестируемую систему просят различить между человеком и системой, то есть самой себя, то это особый случай, который портит тест. В стандартном ТТ данную проблему принято не замечать, хотя потенциально она ему присуща. Так, если наблюдатель имеет возможность углублять свои знания об участнике тестирования, то он может задаваться вопросом о своих собственных, ранее приобретённых знаниях. Тогда при установлении различий, наблюдатель может не задавать вопрос, например: «что такое – твой день рождения?». Если предполагается, что в ТТ запрещено и нечестно пользоваться знаниями о знаниях – хотя для человека это естественно – тогда можно предположить, что и инвертированная версия теста наложит такой же запрет. К сожалению, имеется и другой вариант проблемы идентичности. Им следует более тщательно заняться. Пусть наблюдатель не имеет прироста знаний о тестируемой системе как о «личности». Однако он может об этой системе кое-что косвенно знать, в первую очередь, знать о её физической организации. Но даже в стандартном ТТ мы не в силах спросить: «Вы похожи на меня физически?», хотя при приписывании *другим* ментальных состояний мы неявно про себя задаёмся таким вопросом. Этот вариант проблемы идентичности в инвертированном ТТ подобен такому же варианту в стандартном тесте. В связи с этим, участников следует подбирать таким образом, чтобы они не допускали вопросов подобного рода.

Возможно, что критически оцениваемый, неограниченный по времени стандартный ТТ вполне приемлем для выявления наивно-психологической установки наблюдателя. Однако не представляется возможным выявление всего многообразия психологических предубеждений людей в приписывании ментальных состояний. Так же есть сомнения в том, что возможна поддающаяся обсуждению и не требующая модификации такая версия ТТ, в которой учитывались бы подобно рода психологические установки. Инвертированный ТТ и предлагает компенсировать очевидную проблему преодоления предубеждений, которая не разрешима для стандартного ТТ. И можно показать, что вопросы, поднятые при обсуждении инвертированного ТТ, следует также адресовать для исследования эффективности различных других подходов к оценке интеллектуальности систем.

Возможно иное возражение против инвертированного ТТ. Оно заключается в том, что тест неявно предполагает рекурсию – он косвенно определяет тестируемую систему в терминах способности к распознаванию своего поведения. Тест действительно рекурсивен. Однако это не бесконечно регрессирующая рекурсия. Это – транзакциональная или динамическая рекурсия. Рекурсивный возврат задаётся в большей степени социо-биологическими параметрами, а не логической формой рекурсивной функции. Так как участники играют в транзакциональные игры и предполагают друг у друга ментальные состояния, то акты с необходимостью осуществляются по рекурсивной схеме. Следует заметить, что рекурсивная схема характерна как для стандартного, так и для инвертированного ТТ в условиях их практического применения.

Каким должен быть инвертированный ТТ? По форме представления он может совпадать со стандартным ТТ. При этом, однако, необходимо систематическое и статистическое сравнение результатов прохождения стандартного теста с результатами прохождения инвертированного ТТ для различных пар людей и машин. Необходима гарантия того, чтобы они показывали ту же

самую регулярность. Однако инвертированный ТТ в большей степени мысленный эксперимент, нежели чем серьезный проект развития искусственного интеллекта. Кроме того, ни инвертированный, ни стандартный ТТ не должны выступать как дефиниции интеллектуальности системы. Они не определяют интеллектуальность той или иной системы. Они лишь обеспечивают сбор данных для индуктивного вывода в пользу её интеллектуальности (Мур, 1976). Даже если системы и неразличимы в ТТ, всё равно принятие решения об интеллектуальности системы обуславливается не техническими, а социокультурными факторами. Роль инвертированного ТТ и состоит в том, что он предлагает новый источник индуктивного доказательства интеллектуальности. Но этот источник базируется на фундаментальных правилах приписывания интеллекта *другим*. В связи с этим и раскрывается то, что ранее надёжно скрывалось в стандартном ТТ.

Возможен ли инвертированный ТТ сегодня? Современная наука и технология располагает лишь примитивным набором лингвистических средств, которые позволяют однозначно различать между людьми и компьютерными системами. Некритическое их использование следует расценивать как мошенничество. И надо рассматривать серьёзно факт прохождения ТТ программой, в которую не были заложены представления о принципах человеческой психической деятельности. И в этом «инвертированный» подход контрастирует со стандартным ТТ. Инвертированный ТТ делает акцент на исследовании того, как люди сами способны различать вещи, обладающие разумом и не обладающие таковым. Важным выводом является и то, что если мы знаем способ функционирования системы, то это существенно влияет и на наше решение приписать ей ментальности. Знание способа функционирования, как правило, обуславливает отрицательное решение. Данная закономерность заслуживает внимания и требует дальнейшего изучения. Инвертированный ТТ и может помочь в этом деле.

Рассмотрение ТТ в условиях его инверсии позволяет заметить феномен наивной психологии с обеих сторон – и со стороны наблюдателя и со стороны системы. Вначале, в исходной версии, роль наблюдателя совершенно непонятна. Делая роль наблюдателя явной, ТТ позволяет выявлять важные и естественные аспекты человеческой психологии. Инвертированный ТТ предлагает убедительный способ индуктивного сбора наивно-психологистских данных. Поэтому инвертированный тест должен быть рассмотрен именно в таком свете – не как серьезное научное предложение демонтажа критических интерпретаций первоначальной версии Тьюринга, но как индуктивный способ оценки интеллектуального поведения, для чего, в принципе, исконно и предназначался стандартный ТТ.

Заключение.

Оспаривать тест Тьюринга просто. С. Ватт полагает, что это не продвигает нас в нужном направлении. Как выразился Френч: «философам искусственного интеллекта нужен не просто тест на интеллект, а скорее теория интеллекта» (Френч, 1990). Я полагаю, что общая теория «чужого интеллекта» принципиально невозможна. Любая теория человеческого интеллекта методологически зависит от того, как мы – люди – понимаем интеллект. Именно здесь [на методологическом уровне] ТТ может сыграть важную роль. Тест можно использовать в роли инструмента, который позволяет явно представить, что и как люди понимают под словом «интеллект». Тест – это средство

оценки способов различения людьми тех вещей, которые обладают разумом от тех вещей, которые разумом не обладают.

Критика ТТ, представленная в данной работе, имеет несколько положений. Во-первых, наивно-психологистская установка – глубоко укоренённая в человека способность, его фундаментальное природное свойство, существенный, внутренний аспект собственно человеческого поведения. Её следует учитывать в тесте. Во-вторых, в ТТ редко явно утверждается активная роль наблюдателя. Но активность наблюдателя нельзя игнорировать. Без раскрытия роли наблюдателя, как для самого теста, так и для игры в имитацию в целом, тест существенно портится.

Наблюдателю присуща естественная установка приписывания ментальных состояний тестируемой системе вне всякой зависимости от её фактического поведения. Поэтому была предложена инвертированная версия ТТ. В ней не акцентируется внимание на том, каким образом наблюдатель различает между разными системами. *Акцент приходится на способность системы к саморазличению.* Тест оценивает внутрисистемную способность приписывания ментальности *другому* на основании того, как это делают квалифицированные человеческие судьи. Наблюдатель становится наблюдаемым. Модифицированный тест обладает той же силой, что и оригинальный. Но он позволяет обратить внимание на ранее незамеченных параметрах оценки интеллектуального поведения.

С. Ватт просит – работу не надо трактовать так, будто ТТ устарел и его необходимо исправлять. На самом деле многочисленные проблемы, связанные с тестом и дискуссии по поводу этих проблем – это не простые философские каламбуры. Напротив, все они – намёки на проблему «других разумов», намёки на фундаментальные проблемы поиска оснований приписывания друг другу ментальных свойств. Необходимо использовать результаты полемики для цели формирования лучшего понимания проблемы *другого*. В ТТ и в инвертированном ТТ, видится превосходный способ изучения этих аспектов человеческой психики. Дискуссия по поводу теста должна быть продолжена.

Социокультурные аспекты теста Тьюринга **Илья Рыбин (ИС-82)**

Идея теста Тьюринга не ограничивается сугубо философской или логико-эпистемологической интерпретацией. Многие исследователи обращали внимание и на социокультурные аспекты тестирования¹. Выделим следующие аспекты: 1) психологический; 2) социологический; 3) гендерный.

Первый, психологический аспект, акцентировал внимание на успехе компьютерного моделирования параноидального поведения. Остановимся на двух последних.

Социологический аспект

Человек в социальной среде часто рассматривается как неотъемлемая часть её интеллекта. Социальная адаптация, обучение и общение – важные показатели, даже необходимые для интеллекта (Collins, 1990; McIlvenny, 1993; Moon et al., 1994).

¹ Saygin, A. P., Cicekli, I. & Akman V. 2000. Turing test: 50 years later. Minds and Machines 10:463-518.

Бенни Шенон (1989) поднимает важную проблему автономности интеллекта в социальной среде, его независимости от системы социальных взаимодействий, воздействий, мотиваций и т.д. По его мнению, ТТ предполагает такую автономию, поэтому он неверен. Несмотря на явную неадекватность, единственный способ отличить человека от машины, это «смотреть на неё, щупать, возможно, даже ласкать».

Джустин Лейбер (1989) защищает ТТ от Б.Шенона, обвиняя автора в шовинизме, т.е. в «нежелании признать возможность того, что человечество может иметь конкурентов [в области интеллекта]». Лейбер приводит опровержение Тьюрингом возражения «голова в песке».

Среди тех, кто рассматривает интеллект как часть социальных процессов (и наоборот, социальные процессы как производное от интеллекта) – сторонники т.н. *эволюционного подхода* (Barresi, 1987; Forsyth, 1988; Schweizer, 1998). Наиболее распространенной характеристикой социального интеллекта они считают способность к адаптации. Проблема ТТ может рассматриваться на двух уровнях: индивидуальном и коллективном. Эволюционные доказательства (опровержения) ТТ характерны для коллективного уровня. Они изучают интеллект *вида* и факторы, влияющие на его развитие. Согласно эволюционной точке зрения, видовая адаптация происходит в рамках целостной системы – природы. Адаптация обеспечивает выживание вида в рамках этой системы. Придерживаясь данной точки зрения Джон Баресси (1987) рассматривает интеллектуальные машины как разновидности (виды) и предлагает вместо ТТ эволюционный тест («Cyberiad Test»). Он считает, что ТТ построен на обмане судьи-человека, но в *природном* интеллекте судьей выступает «мать-природа». Cyberiad Test подобен ТТ: основанием суждения об интеллектуальности является сравнение между человеком и машиной. Однако интеллектуальное поведение рассматривается не в чисто логической возможности, а в возможности выживания. ТТ, считает Баресси ниже по положению, чем Cyberiad Test, потому что, что ТТ способен рассматривать «диалогичный» интеллект, ограниченный условиями вербального общения. Cyberiad Test пройден, «если общество искусственных людей способно продолжить свое социокультурное развитие без деградации в течение длительного периода времени, скажем несколько миллионов лет».

Гендерный аспект

Исследование ТТ с позиции гендерной проблемы осуществлялось многими авторами. Основной тезис, который они поднимают – ТТ предназначен не для изучения вопроса – «Может ли машина мыслить?», а для вопроса – «Может ли мужчина понимать женщину?». Существует несколько гендерных интерпретаций ТТ:

1) Детальный анализ гендерной интерпретации ТТ осуществила Джулия Генова (Genova, 1994). Она показывает, что сексуальный компонент ТТ крайне важен. Так как машина подменяет женщину, а не мужчину, то здесь явно отражается позиция ненавистника женщин, согласно которой женщина – это менее интеллектуальный игрок, нежели мужчина. Она не способна обмануть – значит, её мышление на уровень ниже мужского.

Самое главное в сексуальном аспекте игры – подвергнуть сомнению существование дискретных категорий. Т.е. Тьюринг, анализируя проблему «мужской/женский», пытается продемонстрировать, что пол – это социально

обусловленная категория, но отнюдь не биологическая. Здесь, как считает Генова, проявляется общемировоззренческая позиция Тьюринга – позиция «нарушения границ». Согласно данной концепции Тьюринг восхищается трансформациями «мужское/женское»¹.

2) Мнение Цицекли и др. по поводу гендерного аспекта ТТ отличается от мнения Геновы. Авторы не делают различия между двумя полами. Они считают, что на способ ведения игры совершенно не влияет фактор того, что именно женщина не обманывает, а не мужчина. Они раскрывают свой подход следующим способом. Тьюринг вместо неопределённого вопроса «Может ли машина мыслить?» ставит вопрос – «Что будет, если машина займет место игрока А (т.е. женщину – И.Р.) в этой игре? Будет ли судья в данном случае ошибаться так же часто как при игре с мужчиной и женщиной?» (Turing, 1950, p. 434). И далее «Обратим внимание только на компьютере С. Правда ли то, что, увеличив его память и скорость до необходимых величин и запрограммировав его соответствующим образом, компьютер может удовлетворительно играть роль игрока А в игре в имитацию; *роль В при этом играет мужчиной?*» (Turing, 1950).

Т.е. женщина вообще исчезает из игры. Однако и задача игрока А, и задача игрока В и задача судьи остались те же; по крайней мере в явной форме Тьюринг не говорит ни о каких изменениях. Ситуация изображена на Рис.1.

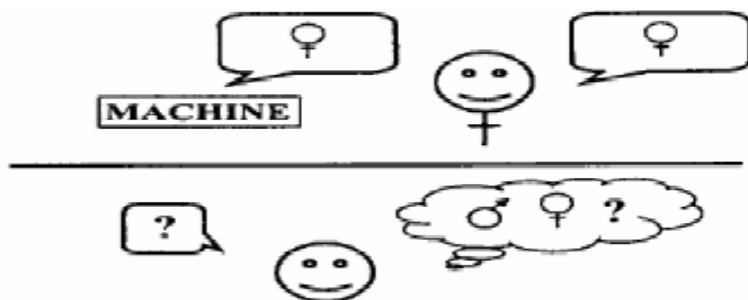


Рис.1. Игра в имитацию: Этап 2, версия 1.

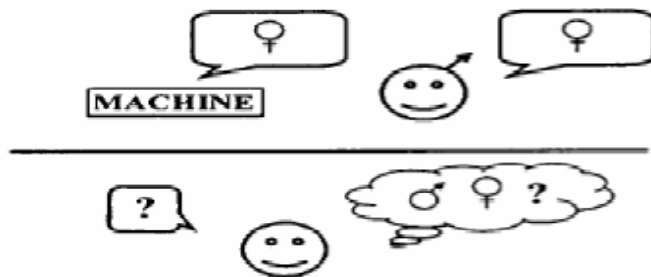


Рис.2. Игра в имитацию: Этап 2, версия 2.

Возникает неопределенность – не совсем понятно, какой следует использовать сценарий – Рис.1 или Рис.2? Ответ – любой, так как очевидно, что основной задачей ТТ является оценка способности машины имитировать человеческое мышление, а не моделирование поведения женщины. Большинство поздних замечаний к ТТ игнорируют сексуальную принадлежность и

¹ Тьюринг, как известно, был обвинен в гомосексуализме, не выдержал обвинения и покончил самоубийством.

полагают что в игру играют машина (А), человек (В) и судья (С). В этом случае цель С состоит в определении того, кто из двоих на самом деле человек, а кто – компьютер. Это – собственно и есть тест Тьюринга, т.е. игра в имитацию, не нагруженная сексуальными проблемами, а отвечающая только за моделирование чистого интеллектуального поведения.

Почему Тьюринг разработал такую странную игру в имитацию? Зачем суесться с женщиной, мужчиной и с заменой их машиной? Он мог бы, например, заменить пару: «женщина-мужчина» сразу на «человек-машина». Очевидно, что в данном случае можно было бы прямее ставить вопрос «может ли машина мыслить?».

Основная причина решения о том, что машинный разум должен симитировать в игре именно женщину, а не мужчину состоит вовсе не в том, что Тьюринг считал имитацию действий женщины верхом совершенства в интеллектуальном соревновании. Основная причина – это то, что *идея имитации* присутствует в статье Тьюринга в *более явной форме*, чем это принято считать.

Игре неотъемлемо присущ аспект *обмана*. Мужчине разрешено говорить все что угодно, лишь бы судья принял неверное решение. Женщина необходима как помощник судьи. В варианте «машина вместо женщины» ситуация остается прежней. Машина пытается убедить судью, что именно она – женщина. И оценивает компетентность машины в этой ситуации отнюдь не женщина, против которой машина играет.

Кажущиеся фривольными требования Тьюринга на самом деле серьёзны. Ни мужчина, ни машина в игре в имитацию, основанной на различии полов не являются женщиной. При близком рассмотрении, можно увидеть, что Тьюринг предлагает сравнивать успех машины с успехом мужчины, а не смотреть, смогла ли машина сыграть вместо женщины в игру в имитацию. Поведение и мужчины и машины рассматривается с честной позиции честной женщины. На Рис. 1 мы видим, что женщина исчезла из игры, но цель и для мужчины и для машины осталась прежней – имитировать поведение женщины. Их поведение сравнимо, потому что они имитируют то, чем на самом деле не являются (мужчина не является женщиной, машина не является человеком).

За такой гендерной необычностью игры в имитацию скрывается методологическая честность. Тьюринг совершенно точно формулирует и тщательно прорабатывает идею: женщина – это нейтральная сторона. Оба обманщика – и мужчина и машина могут быть оценены с позиции обмана судьи.

3) Существует ещё одно мнение по гендерному аспекту ТТ. Его приводит А.Ходгис, биограф А.Тьюринга. А.Хотгис также акцентирует внимание на мнении многих о неудачном сценарии игры. Игру, в которой мужчина притворяется женщиной – слишком сложно понять. Зачем требуется машине притворяться мужчиной, который, в свою очередь, притворяется женщиной? На самом деле, считает А.Хотгис, суть теста – *отделить* разум от иных человеческих способностей. Данная путаница отражает чрезвычайно претенциозное представление Тьюринга о чистоте человеческого разума.

4) Существует ещё одно мнение по поводу гендерной проблемы – Тьюринг просто шутит. Такое мнение также приводит А.Хотгис: «Друзья знали Тьюринга как человека, в котором ум, юмор и сексуальность были неотделимы друг от друга».

Вывод: Анализ социокультурных аспектов теста Тьюринга показывает их несостоятельность: метафизическую спекулятивность социологической интерпретации (отождествление матери-природы с судьёй), параноидальность психологической и надуманность гендерной.

Тест Тьюринга – это тест на интеллект!

Компьютер может мыслить! (М.Минский) **Дмитрий Соболев (М-08-03)**

Положительное решение вопроса «Может ли машина мыслить?» разделяется М.Минским, одним из наиболее авторитетных исследователей ИИ. В работе «Почему люди думают, что компьютеры не могут?» свое мнение убежденно доказывает следующими положениями:

1. Компьютер может творить

Мы восхищаемся Эйнштейнами и Бетховенами и удивимся, если компьютеры когда-то смогут сотворить невиданные теории или симфонии. Большинство людей считает, что творчество обусловлено необъяснимым, волшебным «даром». Если так, тогда, конечно, никакой компьютер не способен к творчеству. Однако следует указать на наивное заблуждение. Мы смотрим на достижения нашей культуры как на что-то грандиозное. Но мы не замечаем того, что не менее грандиозным является то, как *обычный человек* делает *обычные* вещи. Вряд ли мы можем понять способ написания великими композиторами великих симфоний. Я убежден, что не существует принципиального различия между мышлением обычным и мышлением креативным. Я никого не обвиняю в неспособности творить. Также никого не обвиняю в неспособности объяснить творческий процесс. Просто мне не нравится мнение, что если мы не в состоянии объяснить процесс творчества сегодня, то вообще никогда не сможем объяснить его.

Не следует пугаться тайны творчества Бетховенов и Эйнштейнов. Следует пугаться незнания того, как рождаются идеи – и не только «творческие» идеи, а самые-самые простые. Мы настолько поражаемся чудом чего-то необычного и великого, что забываем о том, как мало знаем о природе обыденного мышления.

Есть ли принципиальное различие между оригинальным мышлением и обыкновенным мышлением? Я уверен, нет. Чтобы обыденное мышление стало креативным, необходимы: 1) заинтересованность в решении проблемы; 2) профессионализм, что в условиях творческого процесса называется искусством; 3) уверенность в себе и упорство, противостоящее насмешкам и скептицизму окружающих; 4) здравый смысл. Почему всё это не может сделать обычного человека гением?

Мы не сможем заставить машины делать чудеса до тех пор, пока не найдем, как люди делают обычные дела.

2. Компьютер может разрешать проблемы

Ранние компьютерные программы – это простые списки и управляющие команды типа «Делай это. Делай то. Делай это и то, пока не случится что-то еще». Большинство программистов работают на таких языках (наподобие BASIC или FORTRAN). Они вынуждают продумывать все, что программа будет делать от начала до конца.

Исследователи ИИ нашли новые способы программирования. Программа «Универсальный решатель проблем», разработанная в конце 1950-х – Алленом Ньюэллом, Д.С.Шоу и Гербертом А.Саймоном впервые продемонстрировала преимущества нового способа описания процессов в терминах утверждений таких, как «Если D – это различие между тем, что Вы имеете и тем, что Вы хотите, то попробуйте уменьшить D, используя метод M». Эта и другие идеи привели тому, что сегодня называются «mean-end» и «do if needed»-программированием.

Программисты теперь не должны следить за чётким соблюдением правил выполнения программы. Они сами автоматически срабатывают по мере необходимости. Данный подход ознаменовал эпоху программ, которые самостоятельно могли решать проблемы и теми способами, которые программист не мог заранее предугадать. Известно, что если что-то долго и упорно делать, то в результате может быть создано нечто новое. Когда процесс затягивается на миллионы миллиардов триллионов лет, то обезьяна, случайно нажимая на клавиши пишущей машинки, может создать поэму. Но это – не разум, а лишь случай. Новые программы не делали вещей беспорядочно. Они использовали *эвристики* о том, что потребуется с определённой степенью вероятности в каждом конкретном случае. Так, вместо блуждания вокруг да около, такие программы находили путь, по которому следует подниматься на холм в темноте, всегда повышая угол наклона. Единственной неприятностью была опасность застрять на меньших пиках и никогда не добраться до настоящей вершины.

С тех пор многие исследования ИИ пытались разработать «глобальные» методы поиска. Однако безуспешно.

Вместо этого, сегодня многие разрабатывают программы для поиска «по образцу». Некоторые экспериментируют с программами, которые могут обучаться и рассуждать по аналогии. Такие программы используют предыдущий опыт решения проблем.

3. Компьютер может понимать

Можем ли мы заставить компьютеры понять то, что мы вводим в него? В 1965 г. Дэниел Бохроу написал одну из первых экспертных систем, получившую название «СТУДЕНТ». Она предназначалась для решения различных задач из курса алгебры средней школы, подобно следующим: «Расстояние от Нью-Йорка до Лос-Анжелеса – 3000 миль. Если средняя скорость самолета – 600 миль в час, найдите время, которое требуется, чтобы путешествовать от Нью-Йорка до Лос-Анджелеса реактивным самолетом». Или «дядя отца Билла вдвое старше отца Билла. Через два года отец Билла будет втрое старше, чем Билл. Сумма их возрастов – 92. Найдите возраст Билла».

Для большинства студентов решить эти задачи намного труднее, чем уравнение из курса алгебры для средней школы. Уравнения решаются по определенному алгоритму, а чтобы решить описанную словами задачу, необходимо определить, что слова и предложения означают. Понимал ли их «СТУДЕНТ»? «Понимание» достигалось с помощью различных уловок. Например, «да» означало «равно». Программа не стремилась понять, что означает фраза «Дядя папы Билла», она только замечала, что данная фраза напоминает фразу «Папа Билла». Программа не знала, что «возраст» и «старый» относится ко времени и требовала их численного представления для форми-

рования уравнений. С парой сотен таких слов-уловок, «СТУДЕНТ» иногда выдавал правильные ответы.

3. Компьютер способен к референции

Большой секрет состоит в том, как что-то можно *обозначить*, т.е. каким образом нечто связано с другими известными нам вещами. Чем больше таких связей, тем больше это *нечто* для нас значит. Смешно искать «истинный» смысл понятия. Если бы у него было только одно значение, то есть если бы оно было бы соединено только с другим одиночным понятием, то оно вообще бы ничего не означало!

Поэтому в компьютеры нельзя вкладывать ясные и четкие логичные определения. Такого рода компьютер никогда бы по-настоящему ничего не понял бы. Сети с большим количеством внутренних связей дают достаточно возможностей для использования знаний – когда один вариант не срабатывает, можно попробовать другой, а так как «смыслов» в сети много, то можно опробовать много точек зрения. Подобный процесс и называется мышлением!

Поэтому традиционная логика не подходит, больше подходит работа с круговыми определениями. Каждое понятие дает значение остальным. Нет ничего плохого в соединении разных песен, когда каждая из них контрастирует с другими. Нет ничего плохого в узелках ткани – каждая ниточка помогает держать другие нитки вместе. Нет ничего плохого в представлении о том, что разум – это только воздушный замок из «смыслов».

6. Компьютер может многое

Мы не знаем возможностей компьютеров. Но и наши знания о человеческом разуме примитивны. Почему же мы с такой неохотой признаем, что не знаем, как работает разум? По всей видимости, это происходит из человеческой тенденции игнорировать проблемы, которые оказываются слишком сложными. Имеются и более простое объяснение оправдания уникальности и необъяснимости Я. Мы просто боимся, что слишком серьезное исследование сорвет одежду с нашей ментальной жизни. Но мы ведь только сейчас начинаем понимать, как человеческий разум работает и для этого мы исследуем, на что способны машины. Конечно, никакой серьезной и четкой теории не существует. Пока не существует.

7. Компьютер может обладать здравым смыслом

Полное отсутствие у компьютера «здравого смысла» при решении задач – еще одна причина невозможности признать за машиной способности к мышлению.

Но не покажется ли странным, что самые первые ИИ-программы прекрасно разбирались в сложных теоретических вопросах, но при этом не имели ни капли здравого смысла? В 1961 году программа Джеймса Слейгла могла решать задачи по высшей математике на уровне студентов колледжа; она даже получила «отлично» на экзамене в Массачусетский технологический институт. Но только в 1970 была создана программа, которая способна более-менее удовлетворительно выполнять обычные вещи вроде детской игры в кубики – выстраивать башни, разрушать их, класть их в коробку..

Почему же сначала научились делать именно «взрослые» программы? Ответ: большая часть «экспертного» взрослого мышления гораздо проще, чем детское мышление во время игры.

Новичку сложнее, чем эксперту! Эксперту нужно только знать, а это довольно просто. Сложно – понять. Скажем, Галилей – гений, так как понял и показал необходимость высшей математики. Но он ее не знал. А любой добросовестный студент способен сейчас её выучить.

8. Компьютер может сознавать

Действительно ли можно сделать компьютер сознательным? Обычно слышится отрицательный ответ. Но попробуем ответить положительно – машины могут даже лучше осознавать себя, чем на это способны люди.

На мой взгляд, мышление машины заключается в способности исследовать в ходе работы свои собственные части. В принципе, это возможно. Уже сейчас имеются ИИ-программы, которые понимают, как работают более простые программы. Проблема в том, что мы-то знаем, как программы будут понимать – что такое «хорошо» и что такое «плохо» работать. Как только мы научим компьютеры различению – что такое хорошо/плохо работать, они смогут понимать, изменять, и улучшать себя.

9. Компьютер может быть личностью

Пройдет много времени прежде, чем мы узнаем достаточно о том, как сделать машины столь же способными, как люди. Но когда мы решим эти вопросы, тогда мы столкнемся с *незнакомцем*. Будет ли он лучше, чем человек?

Вывод

Также как история меняет взгляд человека на жизнь, ИИ меняет представления о том, что такое разум. Поскольку мы отводим машинам все большее место в нашей жизни, мы всё больше узнаем о мыслительном процессе. Мы начинаем по-новому понимать термины «мышление», «разум», «чувство» и др. Новое понимание, в свою очередь, даёт новые идеи, которые, в свою очередь, применяются к новым машинам. Но сегодня с полной уверенностью можно заявить – **между человеческим разумом и мышлением машины особых различий нет.**

Тест Тьюринга и паранойя **Юрий Цветков (ИС-81)**

К тесту Тьюринга неоднократно обращались психологи. Цель интерпретации ТТ состояла не только в попытке функционалистского обоснования соотношения психика/мозг, но и в интерпретации психотерапевтической практики (Ридер, 1969; Элпер, 1990; Гелетзер-Леви, 1991 г.). Особого внимания, как считают авторы юбилейной статьи ТТ,¹ следует уделить работе Кеннета Колби и его коллег по моделированию паранойи (Колби, 1971; Колби, 1972; Колби, 1981).

К. Колби в 1971 году в статье «Искусственная паранойя» описал компьютерную программу (названную ПЕРРИ), которая в диалоге с компьютером моделирует параноидальное поведение. Программа в ответ на вопросы выдает

¹ Saygin, A. P., Cicekli, I. & Akman V. 2000. Turing test: 50 years later. Minds and Machines 10:463-518.

языковые выражения, которые характеризуют внутреннее эмоциональное состояние испытуемого. Основные параметры этого состояния: страх, гнев и недоверие. В ходе диалога значения этих параметров варьируются. К.Колби считает, что ТТ – инструмент по демонстрации возможностей компьютерного моделирования интеллектуального поведения человека, но не для решения проблемы дразличения естественного/искусственного интеллекта. При этом судья не должен догадываться, что одним из партнеров по диалогу является машина.

К.Колби предлагает проводить интервьюирование параноиков и компьютеров. Судьям, оценивающим диалог, не сообщают того, что некоторые из интервьюируемых могут быть компьютерными программами. Их просят тестировать реальных людей – пациентов – и оценить у них уровень параноидального состояния. В игре принимает участие 8 судей. Каждый судья интервьюирует и человека и компьютер. Другие 33 психиатра составляют вторую группу судей (протокольных судей). Их просят определить уровень параноидального состояния посредством изучения протоколов, которые были составлены первой группой судей. Их также просят определить, кто из интервьюированных – человек, а кто – компьютерная программа. Анализ результатов тестирования показал, что в 48% психиатры не могли отличить программы от больных пациентов.

Предполагается, что параноики время от времени ведут себя нерационально. Несомненно, такое поведение моделировать проще, так как неадекватные ответы на поставленные вопросы – факт наличия психического отклонения. Программе моделирования паранойи не требуется сложная лингвистическая техника. Синтаксис входных и выходных предложений прост. Достаточно небольшого набора ключевых слов. Не требуются специальные средства грамматического разбора и анализа смысловой неопределённости.

Практика судейства «параноидальных» ТТ свидетельствует о достаточно высокой степени приближения модели параноидального поведения к реальным психопатологическим состояниям. Для параноиков вполне приемлема лингвистическая недостаточность ТТ.

Подобный вывод – о лингвистической недостаточности – можно заключить и относительно т.н. «постмодернистских» ТТ, которые генерируют последовательность языковых выражений, получающих статус осмысленных лишь в условиях интерпретации.

Интеллектуальное поведение человека лишь на первый взгляд не так просто смоделировать. На самом деле это не так, очень даже просто.

Но данный вывод относится лишь к параноикам.

Тест Блока: Нестандартные антибихевиористские возражения Тьюринге Иван Чижов (ЗИ-71)

В докладе продолжается обсуждение антибихевиористской критики, предложенной Нэдом Блоком в работе «Психологизм и бихевиоризм» (1980 г.). Помимо стандартных возражений Тесту Тьюринга со стороны антибихевиористской установки, Нэд Блок приводит ряд т.н. «нестандартных возражений» (см. доклад И. Матанцевой в настоящем Сборнике). Данные возражения и ответы на них специфичны для версии ТТ, которая была названа Н. Блоком –

и в последующем другими исследователями – «новым тестом Тьюринга» (неоТТ). Машину, обсуждаемую в докладе, принято называть машиной Блока. Она реализует нео-ТТ.

Далее знаком «-» обозначено возражение по поводу неоТТ со стороны воображаемого оппонента, а знаком «+» обозначено возражение со стороны Н. Блока на «возражение»:

-: Аргумент слишком силён в том смысле, что о любой интеллектуальной машине можно сказать, что интеллект, ею производимый, является интеллектом программистов.

+: Здесь *не* утверждается, что *каждая* машина, разработанной разумными существами, является интеллектуальной исключительно благодаря этим существам. Такой принцип не используется в доводах. Если мы когда-либо сделаем интеллектуальную машину, она обязательно будет снабжена механизмами для самообучения, решения проблем и т.д. Возможно, мы откроем общие принципы обучаемости, общие принципы решения проблем и так далее, которые мы сможем в нее встроить. Но хотя мы *создаем* машину интеллектуальной, интеллект, который она проявляет – её собственный интеллект, так же как наш разум не становится меньше нашим от того, что он, в основном, является результатом чрезвычайных достижений наших предков.

-: Если бы строки машины были записаны до этого года (1980 г.), машина не ответила бы как человек на предложение типа: «Что вы думаете о последних событиях на Ближнем Востоке?»

+: Система может быть интеллектуальной, даже не имея знаний о текущих событиях. Более того, машина может *имитировать* интеллект, не имея знаний о текущих событиях. Программисты могли бы, если бы захотели, выбрать симулятор разумного Робинзона Крузо, который ничего не знает о последних 25 годах. Или же, они могли бы взять на себя периодическое перепрограммирование, чтобы машина могла имитировать знание текущих событий.

-: Утверждается, что машина с определенной внутренней механической структурой не является интеллектуальной, даже если она кажется разумной во *всех* внешних отношениях (то есть в отношении, проверенным в тесте Тьюринга). Но, внедряя это внутреннее условие, не предлагается ли, в действительности, просто лингвистическое обусловливание, новое значение слова «интеллект»? *Обычно* способности ввода-вывода [информации] представляются как критерий интеллектуальности. Все, что предлагается – это *новый* критерий, который включает некоторые сведения о том, что происходит внутри системы.

+: Поскольку, действия машины *полностью* являются отражением чьих то действий, они не являются причинами, по которым можно было бы приписать машине разум. Дело в том, что хотя *обычно* и устанавливается интеллектуальность системы, путем оценки ее способности ввода-вывода, всё-таки очевидно, что наши способности ввода-вывода могут быть обманчивы. Как предложил Патнэм, эта способность – лишь часть логики терминов естественного типа, которая заключена в том, что то, что кажется стереотипным X может оказаться совсем не X, если окажется не принадлежащим тому же научному естественному типу, как ядро вещей, приписываемых нами X. (Kripke, 1972; Putnam, 1975a) Если Патнэм прав в этом, никто никого никогда не может

обвинить в «изменении значения» термина естественного типа, особенно на основе утверждения о том, что он говорит, что нечто, удовлетворяющее стандартному критерию X-а, на самом деле X-м не является.

-: Представьте, что человеческие существа, включая вас, обрабатывают информацию точно также, как это одна из этих машин. Следует ли настаивать на том, что люди не разумны?

+: Нет уверенности в том, что информационный процесс внутри человека такой же, как у данной машины. В любом случае, нет какого-то *понятного и очевидно корректного* ответа на этот вопрос.

-: Мы обрабатываем информацию не так, как обрабатывает информация эта машина? Что придаёт убедительность этому утверждению?

+: Наши когнитивные процессы без сомнения гораздо более механичны, чем многим хотелось бы верить. Но есть огромная разница между тем, что наше существо более механично, чем хотели бы верить многие и тем, что наше существо есть машина описанного мною типа.

-: Комбинаторный взрыв делает машину невозможной. Джордж Миллер давно установил (Miller, 1960), что существует порядка 10^{30} грамматических предложений длиной в 20 слов. Предположим (весьма произвольно), что из них 10^{15} также семантически хорошо построены. ТТ длительностью в час может понадобиться порядка 100 таких предложений. А это 10^{1500} строк, число, большее числа частиц во вселенной.

+: Данная машина *логически* возможна, а не практически, или даже номологически. Бихевиоральные исследования, в основном, представлены как *абстрактные исследования*, и сложно понять, как концепции, такие как неоконцепция теста Тьюринга, могут быть рассмотрены в другом свете. Может ли быть *эмпирической гипотезой*, что интеллект – суть способность выдавать осмысленные последовательности выводов, соответствующие входным последовательностям? Какой тип эмпирического подтверждения мог бы быть в пользу такого утверждения? Если рассматривать нео-ТТ как концепцию интеллектуальности, тогда достаточно лишь *логической* возможности неинтеллектуальной системы, у которой есть способность пройти тест Тьюринга достаточно, чтобы опровергнуть нео-ТТ.

-: Недостаток ТТ заключается в том, что это – лишь экспериментальный план, а не экспериментальная концепция. Проблема в том, что ТТ имеет *фиксированную длину*. Программисты должны знать эту длину для того, чтобы запрограммировать машину. В *адекватном* ТТ длительность в каждом случае тестирования будет выбираться случайным образом. Более кратко, правильность критики нарушается тем, что планируется подставка лиц.

+: Конечно, верно, что способность машины пройти ТТ зависит от существования некоторой верхней границы длины теста. Но то же самое верно и для *людей*. Даже если мы даем человеку, скажем, двенадцать часов подумать на вопрос и ответить (ведь ему надо кушать и спать), всё же люди, в конечном счёте, смертны. Очень малое количество людей смогут пройти тест ТТ, длящийся девятью лет и ни один человек не сможет пройти ТТ длиной в пятьсот лет. Можно охарактеризовать разумность как способность пройти ТТ случайной длины, но поскольку *люди не имеют такой возможности*, такая характеристика не будет являться необходимым условием разумности, и даже если она была бы достаточным условием интеллектуальности (в чем

имеются большие сомнения), достаточное условие интеллектуальности, которому *не удовлетворяют люди*, пользовалось бы малым интересом у сторонников нео-ТТ.

-: Ранее отмечалось, что нео-ТТ распространена среди исследователей искусственного интеллекта. Но все же, машина не может рассматриваться, как опровергающая любую точку зрения ИИ, потому что, как показали Newell and Simon во взгляде на ИИ, «задача разума... предотвращать повсеместную угрозу экспоненциального взрыва исследований» (Newell and Simon, 1979). (Экспоненциальный взрыв исследования характеризуется тем, что добавление одного шага задачи требует увеличения вычислительных ресурсов в 10 раз, добавление двух шагов требует увеличения вычислительных ресурсов в $10^2=100$ раз, добавление трех шагов требует увеличения вычислительных ресурсов в $10^3=1000$ раз и т.д.). Так что сторонникам ИИ будет уместно исправить их версию нео-ТТ следующим образом: *интеллект – это способность выдавать осмысленные последовательности ответов на стимулы до тех пор, пока не грозит экспоненциальный взрыв*.

Анализ машины поиска по строкам, представленной в работе Н. Блока, показывает, что поведение интеллектуально только тогда, когда оно *не является* продуктом информационного процесса. Обращаясь к примеру с марсианами, Н. Блок предостерегает от предположения существования информационного процесса, лежащего в основе единого, универсального интеллектуального поведения (кроме того, что это поведение должно обладать хотя бы минимальной степенью многообразия). Однако остаётся открытым вопрос – хотя и не существует единого информационного процесса, лежащего в основе универсального интеллектуального поведения, то может быть, обнаружится процесс, общий для всех неинтеллектуальных существ, которые, тем не менее, справляются с тестом Тьюринга (а именно, очень простой процесс, оперирующий с огромной памятью). И машина Блока предлагает механизм реализации такого информационного процесса, общего для всех *неинтеллектуальных* систем.

Функционалистский статус любви **Андрей Чудаков (АП-81)**

Функционализм – мощное развивающееся течение современной философии. Возникнув в шестидесятые годы как разновидность философии сознания, сегодня функционализм шагнул от проблематики «дух/тело» к культуро-антропологической теме. Способен ли функционализм решать экзистенциальные вопросы? Например, проблемы *любви*?

Анализ истории возникновения функционализма показывает, что изначально он и был ориентирован на решение смысловых проблем, а не проблем логического характера. Например, такая культуро-антропологическая проблема, как «проблема любви» была фоном для экспликации функционалистского подхода к сознанию. Тема «любви» поднималась как основанием функционализма Хилари Патнэмом, так и Тьюрингом, работы которого Патнэм использовал для обоснования своей теории тождества.

1) Х. Патнэм объяснял теорию тождеств в терминах «любовь», «ревность». Так, например, в теории тождества, как ее формулирует Патнэм в целях критики, каждый тип ментального состояния отождествляется с опреде-

ленным типом физического состояния, так что, к примеру, если Петя и Ваня оба считают, что Саша любит Машу, то мозг Пети и Вани находятся в одном и том же состоянии. Но учитывая, что нам известно о многообразии способов физической реализации одной и той же машины Тьюринга, это, утверждает Патнэм, мало похоже на правду; нет причин, почему бы инопланетянину или роботу не разделять этого же мнения о Саше и Маше, даже если у них нет мозга.

В работе «Философия и наша ментальная жизнь» (1973) Патнэм в следующем суммарном виде представляет свою прежнюю позицию (на деле формулируя ее более осторожно): так, он утверждает, «(1) что человек в целом — это машина Тьюринга и (2) что психологические состояния человека — это состояния машины Тьюринга или дизъюнкции состояний машины Тьюринга». Согласно этой позиции, если бы Петя, Ваня, Джим, инопланетянин и робот считали, что Саша любит Машу, то все они находились бы в одном и том же «логическом» состоянии, но не в одном и том же физическом или «структурном» состоянии, т.е. мы могли бы совершенно одинаково описать их с точки зрения формальной схемы машины в отличие от ее физической реализации.

Патнэм объясняет функционалистский подход на основе интерпретации машины Тьюринга. В своих более ранних статьях по этой теме, таких как «Сознание и машины» (1960), он обращается к понятию машины Тьюринга. Если говорить очень приблизительно, то мы можем представить машину Тьюринга как множество команд для выполнения некоторой совокупности простых операций над цепочками символов, которые составляют «входные данные» машины. Эти команды объединены в «машинные состояния», и каждое из них, будучи конечным множеством команд, регулируется некой главной командой, которая может переключать состояния, определяя шаг за шагом, какие множества команд должны следовать друг за другом в соответствии с поступающими входными данными. При таком описании о машине Тьюринга можно говорить вообще не касаясь вопроса о том, из каких материалов она сконструирована. Позже этим займется инженер, учитывая такие соображения, как стоимость, компактность, скорость, надежность и т.п.; в окончательном виде машина, сконструированная одним инженером, может сильно отличаться по выбранным материалам, характеру переключателей, детальному соединению проводов от машины, сконструированной другим инженером. Две играющие в шахматы машины, например, могут быть изготовлены из очень разных металлов, в них могут использоваться электронные лампы, транзисторы или кремниевые кристаллы, и тем не менее они могут представлять собой одну и ту же машину Тьюринга.

Возвращаясь к нашей теме, *функционалистский статус любви можно трактовать следующим образом: ментальное состояние «любовь» не зависит от физической реализации носителя этого состояния. Это состояние присуще и человеку, и животному, и роботу и инопланетянину.*

2) А. Тьюринг, выдвигая вопрос «Может ли машина мыслить?» не преследовал целью решение сугубо логико-эпистемологических проблем. Он поднимает социокультурные проблемы, в первую очередь, проблемы взаимопонимания полов. Основной вопрос, который ставит Тьюринг, это – «Может ли мужчина понимать женщину?» (взаимопонимание между полами –

это, пожалуй, то основное, что требуется для возникновения и поддержания любви).

Таким образом, и Патнэм и Тьюринг основывают функционализм исходя из проблемы «любви», но отнюдь не из проблемы «мышления».

Особый интерес представляют современные интерпретации функционалистского понимания любви. В ряду специализированных работ по проблеме философии искусственного интеллекта следует выделить футурологическое эссе Джона Маккарти [«Робот и дитя»](#). (2001 г.).

Маккарти описывает ситуацию, при которой мать-алкоголичка, будучи человеком – т.е. имея «материально-субстратные предпосылки» на то, чтобы любить и заботиться о своем ребенке, как это принято у людей, не уделяет ему малейшего внимания. Больному ребёнку же, чтобы он не умер, нужна любовь. С другой стороны, им служит робот. То, что заменяет роботу сознание, т.е. его «псевдосознание» вообще не обладает эмоциональной структурой и не имеет никаких материальных предпосылок для реализации ментального состояния «любовь». Однако робот спасает ребёнка, искусственно *имитируя* «любовь» (робот имитирует голос матери, делает свою внешность похожей на материнскую, дает ребенку бутылочку с молоком, согревает дитя теплым пледом).

Любовь робота функциональна, не сущностна. Тем не менее в условиях описываемых «социокультурных коммуникаций» такая функционалистская любовь оказалась очень даже полезной.

Встает вопрос – «*Может ли робот любить другого робота?*». Пятьдесят лет назад, на заре становления ИИ, такой вопрос имел романтико-фантастический статус. Сегодня он не лишен практического смысла.

Тест Тьюринга, парадигма функционализма и проблема сознания. Подведение итогов работы секции № 1 Проф. Давид Дубровский (Институт философии РАН)

Доклады, представленные на секции были, безусловно, интересны. Анализ теста Тьюринга – это весьма существенный вопрос проблематики искусственного интеллекта. И то, что с разных сторон этот тест был докладчиками проанализирован, представляется мне очень важным и ценным. Конечно, теория Тьюринга, лежащая в основе парадигмы функционализма, допускает различные интерпретации. Здесь еще остается широкое поле деятельности.

Парадигма функционализма, как известно, противостоит парадигме физикализма. В этом противостоянии и осуществлялась в последние десятилетия разработка проблемы искусственного интеллекта, а вместе с тем и проблемы «сознание и мозг». Эти две тесно связанные проблемы остаются в центре внимания западных философов, психологов, физиологов, представителей компьютерных наук, большинство которых, на мой взгляд, отдают предпочтение парадигме функционализма.

В чем разница между этими двумя подходами? Согласно парадигме физикализма единственной фундаментальной наукой является физика, а потому всякое научное описание и объяснение должно быть в конечном итоге редуцировано к физическому описанию и объяснению, и это должно быть отне-

сено так же к явлениям сознания. Однако многократные попытки редукции явлений сознания к физическим процессам не дали убедительных результатов. Вместе с тем рядом исследователей (Тьюрингом, Патнэмом и др.) было показано, что описание и объяснение функциональных отношений логически независимо от описания и объяснения физических процессов. Мы имеем дело здесь с особым типом закономерностей поскольку одно и то же функциональное отношение может быть реализовано посредством различных по своим физическим свойствам системами. Скажем, функции естественного зуба или сердечного клапана можно воспроизвести посредством искусственных заменителей, имеющих совершенно иной субстрат.

На этом основании Тьюрингом был выдвинут принцип изофункционализма систем, т.е. возможность воспроизведения одного и того же комплекса функций (функциональных отношений) различными по своим субстратным, физическим свойствам системами. Этот принцип убедительно подкрепляется общим положением, которое я называю принципом инвариантности информации по отношению к физическим свойствам ее носителя. Суть его проста: одна и та же информация (дискретизированная тем или иным способом) может быть воплощена и передана разными по своим физическим свойствам носителями, т.е. по-разному кодироваться, перекодироваться, существовать в различных кодовых формах.

Поскольку мышление и сознание допускают функциональное описание, то отсюда вытекает, что они не обязательно должны связываться с человеком и его головным мозгом, могут быть присущи системам с совершенно иной физической организацией, настолько отличной от нашей с вами, что встретив такое существо, мы можем не распознать своего брата по разуму. Помимо возможности существования внеземного разума, о котором много говорилось в последние десятилетия, теоретически допустимо полагать такие пути развития компьютерной технологии или симбиозов человека и робота, которые приведут к появлению сознания на новой субстратной основе. Однако всё это лишь теоретические проекты. Реальность пока такова, что между компьютером и человеческим мозгом сохраняется огромная дистанция, несмотря на то, что отдельные мыслительные функции компьютер исполняет неизмеримо лучше человека.

Когда же говорят о разумном характере поведения системы, то это еще не означает, что она обладает сознательным поведением. Нельзя отождествлять рациональное и целесообразное действие системы с мышлением. Ведь холодильник или телевизор действуют вполне логично, рационально, но на этом основании им нельзя приписывать сознание и мышление.

Если мы определим в общем виде то, что допустимо называть разумным действием (поведением) и если примем, что система, выполняющая такого рода действия, может быть названа разумной, то использование теста Тьюринга позволяет решать: является данная система разумной или нет. Однако из теста Тьюринга не вытекает, что с его помощью мы можем определить присуще ли данной системе сознание (мышление) или нет. Вот в чем вопрос!

Подчеркну еще раз: теория Тьюринга и парадигма функционализма дают убедительное обоснование того, что сознание и мышление могут быть присущи разнообразным системам, которые резко отличаются от человека по внешнему виду, субстратному составу, многим физическим свойствам, в том числе могут быть присущи системам, имеющим небиологическую природу и

т.п. Руководствуясь тестом Тьюринга (и опираясь на определение «разумности») можно провести в ряде случаев важную (но в общем-то тривиальную!) операцию разделения систем на неразумные и разумные. Последние включают как системы, обладающие качеством сознания, мышления, так и те, которые его лишены. При этом лучше вместо терминов «сознание», «мышление» использовать термин «субъективная реальность», который обозначает целостное осознаваемое психическое состояние, способное включать различные составляющие: ощущения, чувственные образы, эмоциональные переживания, мыслительные процессы, волевые побуждения и т.д.

Нам известны два типа субъективной реальности – человеческая и животная (у высших животных содержание, структура и динамика субъективной реальности достигают большой сложности). Имеются достаточные теоретические поводы, чтобы допускать существование во вселенной других типов субъективной реальности, которые существенно отличаются от человеческой по смысловым, ценностным, интенционально-волевым и иным параметрам (это особая тема, требующая специального анализа и прежде всего теоретического анализа проблемы возможности взаимопонимания разумных существ, которым приписывается обладание разными типами субъективной реальности).

Важнейшим для проблематики искусственного интеллекта является именно вопрос о том, как диагностировать наличие субъективной реальности у разумной системы? Тест Тьюринга здесь бессилён, он не располагает никакими средствами для решения этой задачи, хотя и подводит к ее пониманию. Что касается парадигмы функционализма в ее различных интерпретациях, то она в качестве абстрактного принципа не исключает, как я думаю, возможности решения указанной задачи, но пока и не обозначает для этого какие-либо конкретные пути. Вместе с тем у меня есть надежда, что множество интерпретаций парадигмы функционализма далеко не исчерпано и мыслимы такие ее варианты, которые способны дать ключ для теоретического решения указанной фундаментальной задачи.

Кстати, подобная задача стоит перед нами, когда требуется теоретически корректно доказать наличие субъективной реальности у другого человека. Это в западной философии называется проблемой «другого сознания», и она пока не имеет общепринятого теоретического решения (хотя в подавляющем числе случаев у нас здесь в реальной жизни не возникает никаких проблем; другое дело, конечно, когда ставится вопрос об адекватном понимании содержания субъективной реальности другого человека, о критериях такого понимания, о его полноте и т.п.).

Почему решение указанной задачи имеет стратегический характер? Потому что первоисточником всякого знания, всякого творческого акта является именно субъективная реальность отдельной личности, потому что именно в этой сфере лежат наиболее оригинальные процессы и операции, присущие человеческому мышлению. Компьютер светит его отраженным светом, воспроизводит логические операции, наиболее простые формы переработки информации, намного превосходя мозг лишь в быстрой реакции. У компьютера нет живой мысли, нет эмоций, воображения, свободной воли, произвольной генерации внутреннего напряжения, у него нет сознания с его непременной экзистенциальной составляющей, нет того постоянного само-

отображения («двойной рефлексии»), которая присуща всякому сознательному акту.

Эти феноменологические описания указывают на то, что наш мозг работает качественно иным образом, чем компьютер, что в его нейронных структурах осуществляются специфические, пока еще непонятные нам способы переработки информации, оригинальные операции, своеобразные логические процессы (особенно в тех случаях, когда достигаются творческие результаты). В этой связи мною давно было высказано предположение, что в отличие от компьютера с его двужанной логикой мозг работает (по крайней мере в ряде случаев) на основе такой многозначной логики, в которой число значений истины есть величина переменная; число значений истины изменяется в зависимости от характера решаемой задачи и оно является разным в различных подсистемах мозга, вовлеченных в процесс решения главной задачи в данном временном интервале (см.: Дубровский Д. И. Психические явления и мозг. М., «Наука», 1971, с. 328; поскольку вопрос о соотношении логики, нейрофизиологии и кибернетики является сейчас весьма актуальным, каким он был и в те далекие годы, когда живо обсуждался Дж. Фон Нейманом, Р. Сперри, У. Маккалоком и У. Питтсом, Н. А. Бернштейном, У. Эшби и другими выдающимися учеными, я решусь посоветовать молодым людям прочесть весь параграф 19 указанной книги, озаглавленный «Физиологическое и логическое»; в нем изложены дискуссии того времени и некоторые соображения, которые, как мне кажется способны стимулировать размышления молодых ученых о путях современных разработок проблем «сознание и компьютер», «мозг и компьютер»).

Повторю, на нынешнем этапе фундаментальной и наиболее трудной задачей в разработке проблематики искусственного интеллекта является задача создания теоретического основания для *диагностики субъективной реальности другой системы*. Для этой цели надо создать Тест по аналогии с Тестом Тьюринга. Решая эту проблему, мы можем наметить новые пути моделирования интеллектуальных функций.

Я убежден, что следующий крупный этап в развитии искусственного интеллекта, компьютеров, роботов будет связан с решением именно этой ключевой проблемы. Исследования субъективной реальности, теоретическое решение вопроса «другого сознания» – условие выхода на новый уровень развития информационных технологий и роботизации. Я думаю, эти задачи и будут ведущими в нашем XXI веке.

Вы – молодые люди, вам жить в этом веке и вам предстоит решать подобные задачи, а это предполагает серьезную ориентацию в философских и методологических вопросах проблематики искусственного интеллекта. Способы решения этих задач будут определять новые стратегические направления развития нашей науки, создания принципиально новых информационных технологий, форм коммуникаций и, без преувеличения, будут во многом определять судьбы земной цивилизации.

II. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И «ЗДРАВЫЙ СМЫСЛ»

К вопросу эпистемологической адекватности репрезентаций

Илья Гаврилов (ИС-81)

В основу доклада положена работа Джона Маккарти и Патрика Дж. Хейса, фундаментальная для философии искусственного интеллекта¹. В докладе делается акцент на проблеме *эпистемологической адекватности репрезентаций* – проблеме построения таких концептуально-языковых, программных и информационных средств компьютерной системы, которые объективно отражают способы и формы функционирования в реальном мире.

1. Три вида адекватности репрезентаций

Первым шагом при создании программы ИИ является принятие решения о том, какую структуру должен иметь мир и как информация о мире и его законах должна быть представлена в машине. Принятие решения зависит от того, говорим ли мы об общих законах либо об отдельных фактах. Так, например, наше понимание термодинамики зависит от представления о газе, как очень большом количестве частиц, двигающихся в пространстве. Это представление играет основную роль при определении механических, тепловых, электрических и оптических свойств газов. Состояние газа в данный момент определяется положением, скоростью и состояниями каждой частицы. Однако на практике мы никогда не определяем положение, скорость или возбуждение какой-нибудь молекулы. Параметры газовой среды мы характеризуем давлением, температурой и скоростью. Либо ещё более огрубляем и характеризуем газ средним давлением и средней температурой. С позиции *здорового смысла* это вполне нормально. При этом, конечно, не отрицается существование объектов, которые нельзя наблюдать. Так же не выдвигается тезис антропоцентризм, согласно которому мир устроен таким образом, как мы себе его представляем.

Выделяется три типа адекватности репрезентации мира компьютерной системой: метафизическую, эпистемологическую и эвристическую – насколько «знания» системы ИИ отвечают реальному положению дел.

1) Репрезентацию можно назвать **метафизически адекватной**, если представляемый мир не противоречит фактам действительности, которая нас интересует. Они полезны, главным образом, при создании общих теорий. Примеры метафизически адекватных представлений для различных аспектов действительности:

А) Представление о мире, как совокупности частиц, взаимодействующих друг с другом посредством сил.

¹ Джон Маккарти и Патрик Дж. Хейс. Некоторые философские проблемы с точки зрения искусственного интеллекта, 1969, <http://www-formal.stanford.edu/jmc/>

- В) Представление о мире, как гигантской квантово-механической волновой функции.
- С) Представление о мире, как системе взаимодействующих дискретных автоматов.

Последний пример – это метафизически адекватное представление, имеющее непосредственное отношение к ИИ

2) Репрезентацию называют **эпистемологически адекватной** для человека или машины, если оно может использоваться на практике для выражения фактов, которые имеются у каждого человека в отношении мира. Ни одно из вышеупомянутых представлений, адекватных в метафизическом плане, не может верно выразить такие факты, как «Джон – дома», или «собаки преследуют котов» или «телефон Джона – 321-7580». Естественный язык, очевидно, полностью подходит для выражения таких фактов. Но он не совсем подходит, например, для выражения знаний о чувствах других людей.

3) Репрезентацию называют **эвристически адекватной**, если она может быть использована в процессе рассуждений при решении проблем.

В докладе рассматривается понятие эпистемологически адекватной репрезентации на примере представления в компьютерной системе понятия «мочь» (т.е. система *может* сделать то или другое).

2. Конечный автомат как метафизически адекватная и эпистемологически неадекватная репрезентация

Пусть S является системой дискретных конечных автоматов, взаимодействующих между собой, как показано на рис. 1 (см. ниже).

Каждый прямоугольник представляет собой подавтомат, а каждая линия представляет сигнал. Время принимает целочисленные значения, а динамическое поведение всего автомата задаётся уравнениями:

$$\begin{aligned}
 (1) \quad & a_1(t+1) = A_1(a_1(t), s_2(t)) \\
 & a_2(t+1) = A_2(a_2(t), s_1(t), s_3(t), s_{10}(t)) \\
 & a_3(t+1) = A_3(a_3(t), s_4(t), s_5(t), s_6(t), s_8(t)) \\
 & a_4(t+1) = A_4(a_4(t), s_7(t)) \\
 (2) \quad & s_2(t) = S_2(a_2(t)) \\
 & s_3(t) = S_3(a_1(t)) \\
 & s_4(t) = S_4(a_2(t)) \\
 & s_5(t) = S_5(a_1(t)) \\
 & s_7(t) = S_7(a_3(t)) \\
 & s_8(t) = S_8(a_4(t)) \\
 & s_9(t) = S_9(a_4(t)) \\
 & s_{10}(t) = S_{10}(a_4(t))
 \end{aligned}$$

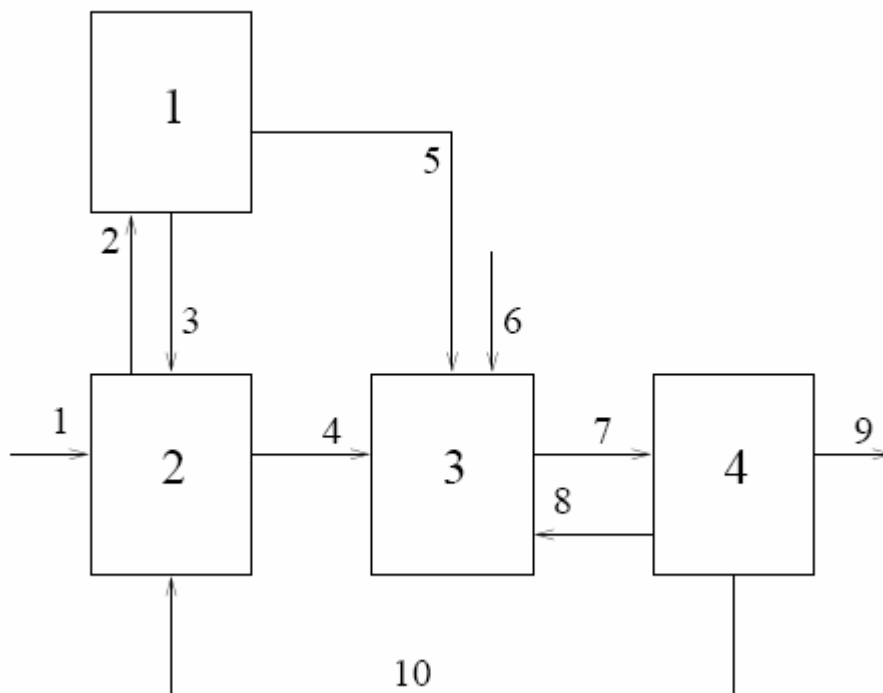


Рис.1

Эти уравнения можно интерпретировать следующим образом: состояние любого автомата в момент времени $t + 1$ определяется его состоянием в момент времени t и сигналами, полученными в момент времени t . Значение отдельного сигнала в момент времени t определяется состоянием автомата, передающего сигнал в момент времени t . Сигналы, для которых не указан автомат-источник – суть входные сигналы от внешнего источника, а сигналы без указания адресата – выходные сигналы.

Конечные автоматы – самый простой пример систем, которые изменяются во времени. Они полностью детерминированы; если мы знаем начальные состояния всех автоматов и если мы знаем функцию времени для входных сигналов, то поведение системы полностью определено уравнениями (1) и (2) для любого последующего момента времени.

В рамках такого представления мир рассматривается как система взаимодействующих подавтоматов. Например, мы можем рассматривать каждого человека, находящегося в комнате, как подавтомат, а окружающую среду – как один или более дополнительных подавтоматов. Данное представление имеет много того, что присуще описанию взаимодействий вещей и людей. Однако при этом возникают трудности:

1. Если мы попытаемся представить систему чьих-либо знаний, то количество состояний, в которых может пребывать подавтомат, будет невообразимо огромным. И все эти состояния должны быть представлены компьютерными программами, которые не ссылаются на описания индивидуальных состояний.

2. Геометрическая информация очень трудно поддается представлению. Попробуйте описать, например, образ человека или решить не менее сложную задачу – описать форму куска глины.

3. Система *фиксированных* взаимосвязей не является адекватной. Человек может свободно обращаться с любым находящимся в помещении предметом. Поэтому адекватная автоматная репрезентация требует описания всех линий, которые соединяют каждый автомат с каждым объектом.

4. Самое серьезное возражение, однако, состоит в том, что автоматная репрезентация является *эпистемологически неадекватной*. А именно, невозможно знать человека достаточно хорошо для того, чтобы перечислить все его внутренние состояния. Та часть информации, которая действительно имеется о нём, должна быть выражена каким-либо другим способом.

Тем не менее, можно использовать автоматную репрезентацию для описания таких понятий, как «мочь», «желать», «верить». Также её можно привлечь для описания контрфактических суждений типа: «Если бы я вчера зажег эту спичку, она бы загорелась».

3. Автоматная репрезентация понятия «мочь»

Перейдем к рассмотрению автоматной репрезентации понятия «мочь». Пусть S является системой подавтоматов, в которую не поступают внешние сигналы, например такой системой S , которая изображена на рисунке 2. Пусть p – одним из подавтоматов. Предположим, что от p отходят m сигнальных линий. То, что p способен сделать, определяется в терминах новой системы S_p , которая получена из системы S путем отсоединения m сигнальных линий, отходящих от p , и заменой их m внешними входными линиями, идущими к системе. На рисунке 2 (см. ниже) подавтомат 1 имеет один выход. В системе S_p (рис. 3) этот выход заменен внешним входом. Новая система S_p всегда имеет тот же самый набор состояний, как и система S :

$$a_1(t+1) = a_1(t) + s_2(t)$$

$$a_2(t+1) = a_2(t) + s_1(t) + 2s_3(t)$$

$$a_3(t+1) = \text{если } a_3(t) = 0 \text{ то } 0 \text{ иначе } a_3(t) + 1$$

$$s_1(t) = \text{если } a_1(t) = 0 \text{ то } 2 \text{ иначе } 1$$

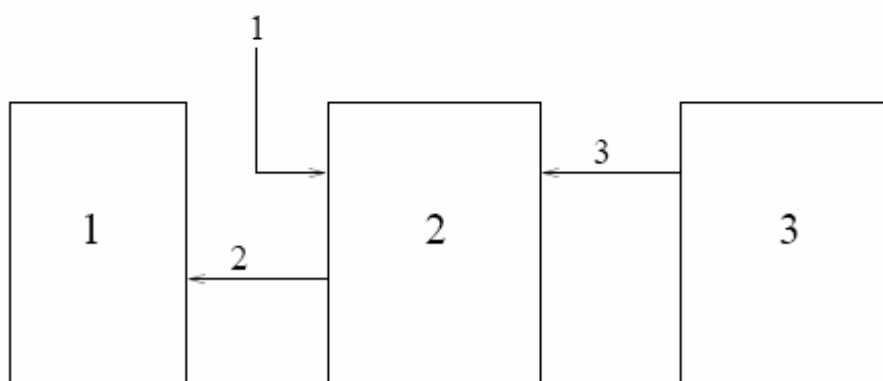
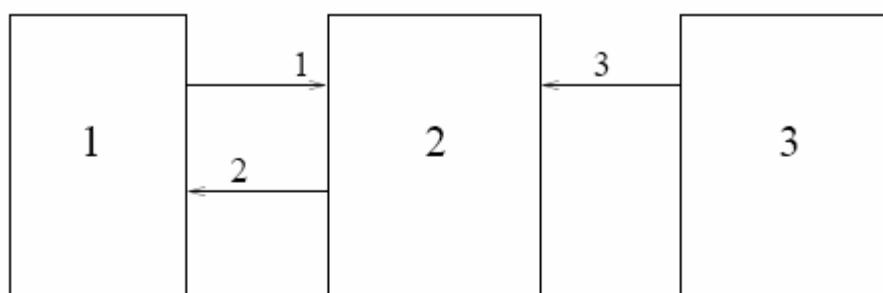
$$s_2(t) = 1$$

$$s_3(t) = \text{если } a_3(t) = 0 \text{ то } 0 \text{ иначе } 1.$$

Теперь пусть p_i является условием, например, « a_2 – четное» или « $a_2 = a_3$ » (В приложениях p_i может быть например таким условием: «коробка находится под бананами».)

Запишем: $\text{мочь}(p, p_i, s)$,

которое читается следующим образом: «подавтомат p может вызвать условие p_i в ситуации s », если есть последовательность выходных сигналов от автомата S_p , которая в конечном счете поместит S в состояние a' , которое удовлетворяет p_i (a'). Другими словами, при определении того, чего может достигнуть p , рассматривается совокупность эффектов последовательности действий p . При этом, однако, не учитываются условия, определяющие то, что он на самом деле выполнит.



Ри
с. 2

(система S) и рис. 3 (система Sp)

Для автоматов, изображенных на рисунке 2 будем считать, что а – начальное состояние, в котором все подавтоматы находятся первоначально в состоянии 0. Легко проверить следующие суждения:

1. Подавтомат 2 никогда самостоятельно не перейдет в состояние 1.
2. Подавтомат 1 может перевести подавтомат 2 в состояние 1.
3. Подавтомат 3 не может перевести подавтомат 2 в состояние 1.

Авторы утверждают, что определённое таким образом понятие «мочь», в первом приближении, является эпистемологически адекватным понятием. Его можно использовать при принятии решений в процессе рассуждений. Также оно во многих случаях соответствует общепринятому понятию «мочь», которое используется в повседневной речи. Предположим, что имеется автомат, который решает, что необходимо сделать в процессе рассуждения. Допустим, им будет компьютер, использующий «рассуждающую программу». Тогда его выходные сигналы определяются решениями, к которым он приходит в процессе рассуждения. Он *не знает* (не вычисляет) заранее то, что он сделает. Он лишь *полагает*, что может что-то выполнить и что может быть достигнуто некоторой последовательностью его выходных сигналов. Рассуждения на основе здравого смысла, кажется, работают тем же самым способом.

Вышеупомянутое, довольно простое, определение понятия «мочь» требует некоторой доработки, чтобы адекватно представить это понятие на основе здравого смысла, а также для практических целей рассуждающей программы. Предположим вначале, что система автоматов допускает приём

внешних сигналов. В этом случае есть два способа определения понятия «мочь» ($p, p_i s$): 1) p может достигнуть p_i независимо от того, какие сигналы появляются на внешних входах. Таким образом требуется наличие последовательности выходных сигналов p , которая достигает цели независимо от последовательности внешних входных сигналов, поступающих в систему. В этом определении понятия «мочь» не требуется, чтобы p *знало*, что поступили внешние сигналы. 2) Альтернативное определение требует, чтобы выходные сигналы зависели от входных сигналов p . Это эквивалентно высказыванию, что p может достигнуть цели, если цель может быть достигнута для произвольных входных сигналов некоторым автоматом, используемым вместо p .

Используя любое из этих определений, понятие «мочь» становится *функцией места* подавтомата в системе, а не его непосредственных характеристик.

Понятие «мочь», соответствующее в большинстве случаев интуитивному понятию, может быть получено путем выдвижения гипотезы о существовании некоего **органа воли**, который принимает решения о том, что необходимо сделать и передаёт эти решения к основной части мозга, который в свою очередь пытается выполнить их. Кроме того, в мозгу содержатся знания о различных фактах. Если учесть эту связь, нельзя будет утверждать, что такой-то человек может набрать секретный или частный номер телефона президента, хотя он его не знает. И это несмотря на то, что на вопрос о том, может ли он набрать телефонный номер, будет дан положительный ответ.

На основе этих примеров, можно попробовать создать последовательность более узких определений понятия «мочь». Она закончилась бы понятием, согласно которому человек может сделать только то, что он фактически и делает. Такое понятие было бы избыточным. На самом деле мы не должны искать единственное и наилучшее определение понятия «мочь». Каждое из вышеупомянутых определений полезно и фактически используется при определенных обстоятельствах. Иногда в одном предложении используется более одного определения этого понятия – тогда, когда существуют два различных уровня ограничения.

4. Автоматная репрезентация контрфактических суждений

Метафизически адекватную репрезентацию мира как совокупности автоматов можно использовать для описания *контрфактических суждений* (условных предложений, противоречащим наличным фактам), таких, например, как предложение «Если бы я пытался зажечь эту спичку вчера, то она бы зажглась». В соответствующей автоматной репрезентации в данном случае имеется некоторое состояние системы на вчерашний день. Далее предположим, что имеется разрыв связи при передаче нервных сигналов от мозга или, допустим, этот разрыв имеется на выходе 'блока принятия решения'. Пусть соответствующие сигналы о необходимости зажечь спичку были переданы. Тогда для системы S_p ответ на вопрос о том, зажжется или нет спичка, зависит от свойств спички (влажная, сухая). Такая интерпретация контрфактических суждений является необходимой для *рассуждающей программы*. Программа должна учиться на ошибках, принимая или генерируя предложения следующей формы, «если бы я сделал это так то или так то, я бы достиг успеха, поэтому я должен изменить свои процедуры таким способом, который приведёт меня к правильным действиям в данном случае».

В приведённом выше описании за основу было взято представление о ситуации как системе взаимодействующих подавтоматов. Такая ситуация могла бы быть представлена множеством различных способов. Различные представления могут привести к различным результатам относительно того, что данный подавтомат *может* достичь. На самом деле одинаковые или даже похожие подавтоматы невозможно полностью задать на все случаи жизни. Все они зависят от выбранного способа представления.

Допустим, например, что пара марсиан наблюдает за ситуацией в комнате, в которой находятся люди. Один марсианин анализирует ситуацию как совокупность отдельных взаимодействующих людей. Второй марсианин группирует все головы в один подавтомат, а все тела – в другой. (Члены ряда Фурье в последнем случае рассматриваются как отдельные взаимодействующие подавтоматы). Как в этом случае первый марсианин мог бы убедить другого, что первое представление является адекватным? Грубо говоря, он утверждал бы, что взаимодействие между головой и телом одного человека более тесное, чем взаимодействие между различными головами. Поэтому лучше провести анализ, исходя из привычного представления. Он будет особенно убедителен, если укажет, что когда взаимодействие между головами прекращается, остается взаимодействие головы с соответствующим телом.

Формально этот аргумент можно выразить в терминах автоматов следующим образом: предположим, что мы имеем автономный автомат A , который является автоматом без входных устройств, и пусть он находится в состояниях k . Далее, пусть m и n – два целых числа, таких что $m, n = k$. Теперь разметим точки k массива $m \times n$ в соответствии с состояниями A . Теперь имеется представление об автомате A , как о системе, состоящей из автомата B , имеющего m -состояний, взаимодействующего с автоматом C , имеющим n -состояний. Теперь может произойти так, что два различных сигнала эквивалентны с точки зрения их воздействия на другой подавтомат, и мы будем использовать это отношение эквивалентности для формирования эквивалентных классов сигналов. Теперь мы можем рассматривать эквивалентные классы как сами сигналы. Предположим, что есть r сигналов от B к C и s сигналов от C к B . Теперь у нас возникает вопрос насколько малы r и s по сравнению с m и n . Ответ может быть получен, подсчетом количества неэквивалентных автоматов с k состояниями и сравнением этого количества с количеством систем двух автоматов с m и n состояниями соответственно и r и s сигналами, передаваемыми в соответствующих направлениях. Точный результат не важен, необходимо лишь знать, что только несколько автоматов с k состояниями допускают декомпозицию с небольшим количеством r и s по сравнению с m и n . Поэтому, если автомат допускает одну такую декомпозицию, то очень маловероятно, что он допустит вторую такую декомпозицию, которая не является эквивалентной первой относительно переименования состояний. Применяя этот аргумент к реальному миру, можно сказать, что весьма вероятно то, что общепринятая декомпозиция (разделение) автомата на людей и на вещи имеет уникальное, объективное и, как правило, предпочтительное состояние.

Такое представление объясняет некоторые трудности, с которыми сталкиваются философы при анализе контрфактических суждений. Например, предложение «если бы я зажег вчера спичку, то она бы загорелась», имеет значение только в терминах сложной модели мира, имеющей объективно предпочтительное состояние, которое, в свою очередь, зависит от большого

количества самых различных факторов. По этой причине изолированное рассмотрение контрфактических суждений не является плодотворным.

5. Автоматная репрезентация понятия «вера»

Можно задать условия утверждения того, что подавтомат p *верит* некоторому суждению. Запишем формулу «веры»: $Vp(s, w)$, где s – состояние автомата p , а w – суждение. $Vp(s, w)$ истинно, если p оценивается, как верящее в w , когда находится в состоянии s и ложно в противном случае. Относительно такого утверждения V правомочны следующие вопросы: 1) Являются ли представления p непротиворечивыми? Правильны ли они? 2) Рассуждает ли p ? (возникают ли новые представления, которые являются логическими следствиями предыдущих представлений?) 3) Наблюдаем ли p ? (заставляют ли истинные суждения об автоматах, которые связаны с p , поверить им автомату p ?) 4) Ведет ли p себя рационально? (когда p верит предложению, которое побуждает p произвести какое-либо действие, то производит ли p это действие?); 5) Общается ли p на языке L ? (рассматривая содержание входных и выходных сигнальных линий, как текст, написанный на языке L , данная линия передает представления к p или от него?); 6) Действительно ли p является сознательным? (имеет ли он разнообразие правильных представлений о его собственных представлениях и процессах, которые их изменяют?)

Такого рода вопросы можно задавать в отношении утверждения Vp . Если на вопросы 1-4 был бы получен утвердительный ответ, тогда имелось бы полное право рассматривать Vp в качестве *эпистемологически адекватной репрезентации понятия «веры»*.

Подобным путем, считает Маккарти, можно задать всё многообразие ментальных терминов в системе ИИ.

Псевдоволя Роман Горюнов (АП-81)

Проблема реализации мотивационно-волевых механизмов в компьютерной системе (в роботе) представляется как часть проблемы моделирования мышления. Разумный механизм должен обладать возможностью выбора, свободной волей. Добровольные действия характеризуются: (1) осознанностью; (2) предшествием выбора действию; (3) возможностью изменения поступка (в самое последнее мгновение).

Не лишенной интереса в плане исследования соотношения воля/разум представляется позиция Стивена Харнада, предложенная в работе «Раздумья о сознании».¹ Он предлагает концепцию т.н. «псевдоволи» (pseudoconation). Подходу присущ «здравый смысл»: 1) Устройство волевой регуляции должно быть целиком материальным и механическим, все компоненты должны быть материальными и иметь обычные связи; 2) Субъективная феноменология сознательного опыта должна быть объяснена: никакие внезапные сверхъестественные способности устройства не должны приниматься на веру.

В отличие от подхода Маккарти, который считал возможным реализацию «псевдоволи» как части «псевдосознания», подход С. Харнада следует

¹ Harnad, S. (1982) Consciousness: An afterthought. Cognition and Brain Theory 5: 29 - 47.

охарактеризовать как *пессимистический*. Возможность реализации «псевдо-волевых» механизмов упирается в проблему «дух/тело».

Тем не менее, С. Харнад считает, что переход к системам моделирования мотивационно-волевой сферы закономерен. Он задаётся всем ходом технического прогресса. Задача заключается в построении такого устройства, которое бы обладало разумом (и волей), а не совокупностью формальных инструкций. Считается, что устройство обладает разумом, если бесконечно долго может выполнять тест Тьюринга. «Волей» он будет обладать тогда, когда *произвольно* будет переходить из одного состояния в другое и судья из теста Тьюринга посчитает эти переходы волевыми действиями, основанными на «субъективном» опыте, а не на только совокупностью предусмотренных разработчиком поведенческих реакций. Поэтому в механизме моделирования мотивационно-волевой регуляции должен присутствовать модуль ПСЕВДОВОЛЯ. У модуля имеются специальные входы, которые получают информацию по обратной связи с собственными выходами. Также этот модуль содержит два типа выходов: «бессознательный» и «сознательный». «Бессознательный» модуль отвечает за рефлекторные движения, дыхание и т.п. «Сознательный» – отвечает за действия, которые человек выполняет осмысленно – выбор, речь, планирование поведения.

Однако между этими модулями возникает *барьер* (и реализационный и концептуальный – «пропасть в объяснении») в виде *проблемы «дух/тело»*. Мы не знаем, как информация преобразуется от сигналов на входе к полноценному образу, как происходит ее обработка (осмысление). Ведь только после осмысления и понимания возможна добровольная реакция.

Получается, что пока трудно судить даже теоретически судить о возможности создания ПСЕВДОВОЛИ, который, например, позволил приписывать этические термины поведению роботов. На сегодняшний день единственным «автоматом» такого рода является человек. Сознание человека – черный ящик и теории, которые его высвечивают, пока лишены смысла, считает Харнад. Все эти теории – пустые претензии ния метафизики, рассмотренные сквозь призму футурологических проектов.

Реализационные перспективы теории речевых актов

Максим Гришкин (АП-81)

Вопрос взаимодействия компьютерных программ с пользователем (человеком или другой программой) достаточно важен, поскольку от него зависит эффективность применения компьютеров. Обычно программу снабжают интерфейсом, понятным ее непосредственному пользователю. Интерфейс должен реализовывать диалог человека с программой на естественном или ограниченном естественном языке. На этом языке формируются запросы к программе, которые она должна выполнить, вопросы, на которые должны быть даны верные ответы и т.п. Если стремиться к тому, чтобы цель использования программы ИИ соответствовала «здравому смыслу», язык должен быть таким, каким люди пользуются в повседневном общении.

Сегодня в условиях сложных информационно-коммуникационных взаимодействий распространены ситуации, когда пользователем программы выступает не человек, а другая программа. Чтобы программа-клиент могла

получить от программы-сервера необходимые данные, входные и выходные характеристики таких программ должны соответствовать некоторому стандартному протоколу, одинаковому формату данных, т.е. они должны «говорить на одном языке». При отсутствии такого протокола получается набор несвязанных друг с другом программных продуктов, обмен информацией между которыми возможен лишь в условиях повторного ручного ввода данных.

Возникает вопрос по поводу разработки такого рода протокола для активных агентов – тех, которые стремятся к достижению определенных целей и обладают различными способностями и знаниями, причем знания и способности отдельного агента недостаточны для достижения системной цели.

В методологическом плане перспективы создания протокола активных агентов связываются с философско-лингвистической концепцией речевых актов. *Речевые акты* – это единицы обмена информацией между сущностями. Для общего случая не имеет значения, кто совершает речевые акты: люди, машины, марсиане или кто-либо ещё другой.

Речевой акт рассматривается с трёх позиций (по Дж. Остину). Если единый речевой акт изучать с точки зрения факта произнесения того или иного высказывания со смыслом, такой акт принято называть *локутивным актом* (сообщение некоторой информации). Сам факт говорения – *иллокутивный акт* (собственно информационный процесс). Убеждение в достоверности сообщаемой информации – *перлокутивный акт*. К последним могут относиться обязательства, обещания, приказы и пр.

Использование речевых актов для целей реализации концепции активных агентов может заключаться в создании программ, которые взаимодействуют с внешним миром посредством некоего языка ввода-вывода, состоящего из выражений, которые суть речевые акты. Разработка программ, удовлетворяющих требованиям реализации речевых актов может производиться с помощью ныне существующих классических языков программирования. Однако для более полной, быстрой и удобной реализации систем, работающих с речевыми актами, следует создавать специальные языки. Спецификация одного из таких языков предложена Джоном Маккарти в 1992 году¹. Он назвал этот предполагаемый язык *Elephant 2000*, полагая, что если не все, то многие черты этого языка будут присущи языкам программирования 2000 г.:

1. Программы, написанные на *Elephant 2000*, будут иметь интерфейс, представляющим собой совокупность выражений специального языка ввода-вывода. Часть этих выражений будет определена в используемой программе, а часть – заложена в языке. Подобно выражениям, речевые акты могут быть приказами, запросами, вопросами, ответами на вопросы и т.п.

2. В программах, написанных на *Elephant 2000*, не будут использоваться фиксированные структуры данных, которые присущи большинству существующих языков.

3. Используя некоторый набор выражений языка, можно будет напрямую обращаться к прошлым состояниям программы и событиям. Если язык будет реализован как интерпретируемый, интерпретатор будет использовать одну структуру данных – лог (журнал) событий, в котором и будут храниться все данные. Если язык будет компилируемого типа, то такой компилятор должен

¹ McCarthy John, *Elephant 2000: A Programming Language Based on Speech Acts*. <http://www-formal.stanford.edu/jmc>

генерировать программу на обычном языке программирования, таком как C++, с обычными структурами данных.

4. Наиболее очевидно применение программ такого рода при обработке транзакций и работе с базами данных, а также при организации взаимодействия программ различных организаций. Для последнего Джоном Маккарти в 1982 году был предложен «Общий язык делового общения»¹, некоторые идеи которого были впоследствии использованы в электронной коммерции.

У языка программирования, основанного на речевых актах, появляются некоторые черты естественного языка, что весьма важно для развития теории программирования, апеллирующей к «здравому смыслу».

Проблемы квантификации эпистемологически адекватных представлений

Мария Красивская (И-81)

Важным аспектом создания искусственного интеллекта является поиск системы формального представления фактов, которая была бы эпистемологически адекватна машине². Представление называется *эпистемологически адекватным* для человека или для машины, если оно им можно практически воспользоваться для выражения фактов, которые имеются у каждого человека в отношении мира. Зачастую в программах, моделирующих искусственный интеллект представление мира как раз далеко от эпистемологической адекватности. Вследствие этого возникает ряд проблем, которые необходимо анализировать и искать пути их решения. Одной из проблем формализации является, например, *проблема квантификации* – как посредством ёмкого формального выражения преодолеть трудность и даже невозможность перечисления всех условий выполнимости какого-либо действия или всех исключений из этих условий.

Данная проблема решается путем использования ряда неклассических символических логик.

1. Модальная логика. Первоначально она была предложена Льюисом в попытке избежать «парадоксов» импликации – из ложного суждения следует любое (истинное или ложное) суждение. Идея Льюиса заключалась в различении двух типов истины: очевидной истины и условной истины. Условно истинное суждение – то, которое, хотя и истинно, но может быть ложным. Формально оно вводится модальным оператором \Box (читается «необходимо»). Тогда то, что p является очевидной истиной выражено тем, что $\Box p$ является истиной. Позднее модальная логика стала основным инструментарием логического анализа таких логических операторов как вера, знание и время.

Следует обратить внимание, что истинность или ложность $\Box p$ не решается истинностью p . То есть \Box – не истинностно-функциональный оператор в отличие, например, от обычных логических связок. Поэтому нет никакого прямого

¹ McCarthy John 1982, *The Common Business Communication Language*. <http://www-formal.stanford.edu/jmc>

² В докладе использовались «классические» для ИИ положения, предложенные Дж.Маккарти и П. Дж. Хейса в работе: *Джон Маккарти и Патрик Дж. Хейс. Некоторые философские проблемы с точки зрения искусственного интеллекта*, 1969, <http://www-formal.stanford.edu/jmc/>

способа использования таблиц истинности для анализа суждений, содержащих модальные операторы. В связи с этим, первой проблемой модальной логики выступает быстрое увеличение модальных логических исчислений при отсутствии явных средств сравнения. Другие трудности возникают при попытках ввода квантификаторов.

К сожалению, все ранние попытки модальных вычислений утверждения имели неочевидные теоремы (см. например, Крипке 1963), и, кроме того, все из них сталкивались с трудностями, связанными с несостоятельностью закона тождества, предложенного Лейбницем. Закон Лейбница можно представить в следующей форме:

$$L: \forall x. \forall y. x = y \rightarrow (F(x) \equiv F(y)),$$

где F – любое открытое предложение. В модальном контексте этот закон терпит неудачу. Например, рассмотрим пример L :

$$L1: \forall x. \forall y. x = y \rightarrow (\Box(x = x) \equiv \Box(x = y)):$$

В соответствии с правилом, которое присутствует почти во всех вариантах модальных логик, поскольку $x = x$ – теорема, то $\Box(x = x)$. Таким образом $L1$ преобразуется в:

$$L2: \forall x. \forall y. x = y \rightarrow \Box(x = y)$$

Здесь параметр оператора необходимости «гуляет», имеется много вариантов «тождественности», а это противоречит интуиции. Например утренняя звезда – фактически тот же самый объект, что и вечерняя звезда (планета Венера). Однако, они не обязательно тождественны: можно легко вообразить их отличия, которые Г. Фреге представил в своём знаменитом примере, известном как «парадокс утренней звезды».

Эти и связанные с ним затруднения заставляют либо вообще отказаться от закона Лейбница в модальных исчислениях предикатов либо изменять правила квантификации. Изменить эти правила следует так, чтобы было невозможно получать нежелательные образцы универсальных предложений типа $L2$. Формально проблема, в принципе, разрешима. Однако это ведет к серьезным затруднениям при интерпретации данных исчислений.

Если вернуться к утренней звезде, то ясно, что утренняя звезда всегда тождественна утренней звезде. Однако, вечерняя звезда не обязательно тождественна утренней звезде. Таким образом, объект – планета Венера – одновременно имеет и не имеет свойства обязательной тождественности утренней звезде.

Новая семантическая теория модальной логики обеспечивает удовлетворительный метод интерпретации модальных предложений. Идея состоит в том, чтобы модальные исчисления описывали сразу несколько возможных миров, а не один-единственный, как это принято в классической логико-семантической теории. Утверждению присваивается не одно-единственное истинностное значение, а скорее, спектр истинностных значений, по одному значению в каждом возможном мире. Теперь утверждение очевидно, когда оно истинно во всех возможных мирах. Новая семантическая теория обеспечивает разрешение первой проблемы модальной логики: рациональный метод пригоден для классификации множества модальных логик высказываний. Что еще более важно, она (теория) также обеспечивает понятную интерпретацию для модальных вычислений предикатов. Возможный мир – это интерпретация исчисления, которое понимается в обычном смысле.

Теперь несостоятельность закона Лейбница больше не озадачивает, поскольку в одном мире утренняя звезда, например, может быть тождественна вечерней звезде, т.е. быть тем же самым объектом, но в другом мире они могут различаться.

Остаются трудности: 1) формальные – правила квантификации должны измениться, чтобы избежать теорем, противоречащих интуиции и; 2) пояснительные – не очевидно существование одного и того же объекта (индивидуума) в различных мирах.

В принципе, можно использовать модальную логику и без модальных операторов, применяя обычную истинностно-функциональную логику, которая непосредственно описывает множественно-мировую семантику модальной логики.

2. Эпистемическая логика (логика знания) была впервые исследована как модальная логика Хинтикой в его книге «Знание и вера» (1962). Он ввёл модальный оператор K_a (читается « a знает это»), и двойственный к нему – «незнание» – P_a , определенное как $\neg K_a \neg$. Семантика получается аналогичным прочтением K_a как: «данное положение истинно во всех возможных мирах, совместимых с тем, что a это знает»

3. Темпоральная логика (логика времени). Данная логика представляет собой одну из самых обширных и наиболее активно изучаемых областей философской логики. Работа Прайора «Прошлое, настоящее и будущее» (1968) – чрезвычайно полное и ясное описание того, что было сделано в этой области. Прайор обсуждает четыре пропозициональных оператора F, G, P, H . Он представляет их как модальные операторы. При этом семантика данного логического исчисления – просто зависимость, упорядочивающая время. Предлагаются различные аксиоматизации: детерминированного/недетерминированного времени, конечному/бесконечному времени и т.д. При этом с новой силой поднимаются проблемы квантификации.

4. Акциональная логика (логика действия). Теория, наиболее полно разработанная в этой области – логика действия фон Вригта, представленная в книге «Норма и действие» (1963). Фон Вригт формирует свою логику на основе довольно необычной темпоральной логики. Базис – двойная модальная связка T , введённая таким образом, что pTq , где p и q – суждения, звучит так: « p ; тогда q ». Таким образом описывается действие, например, открытия окна: (окно закрыто) T (окно открыто).

Перечисленные выше логики, предложенные почти полвека назад, и по сей день используются для решения проблемы квантификации эпистемологически адекватных представлений о мире, которыми пользуются люди и которые могут эффективно обрабатываться в интеллектуальных системах.

Экспертные системы, основанные на здравом смысле Михаил Лапин (С-91)

В докладе поднимаются важные вопросы определения критериев «интеллектуальности» компьютерных систем. Следует ли относить к интеллектуальным такие программы, как экспертные системы. Ведь многие такие системы – обычные программы, за исключением того, что используют высоко-

уровневые языки программирования и представления «данных» и «знаний». Но никакого «интеллекта» они собой не являют. При решении данного вопроса используется классическая работа Джона Маккарти: «Экспертные системы на могут обойтись без известной толики здравого смысла» (1984 г.).

Экспертная система (ЭС) – это компьютерная программа, предназначенную для представления знаний и умений эксперта некоторой конкретной предметной области. Показатели функционирования ЭС в отдельных областях знаний выглядят очень впечатляюще. Тем не менее очень немногие ЭС обладают способностями, которыми обладают самые слабоумные представители человеческого рода. Самое главное – ЭС не обладают «здравым смыслом» – они не могут выйти за рамки, установленные разработчики, не «осознают» своих действий.

Предметом данного доклада является описание обобщённых способов, благодаря которым ЭС станут руководствоваться «здравым смыслом». Сегодня мы не располагаем сколько-нибудь достоверной информацией о конкретных фактах и методах применения принципа «здравого смысла». Решение же этого вопроса может рассматриваться как ключ к решению ключевых проблем ИИ.

Идею «компьютерных программ со «здравым смыслом» Дж. Маккарти прорабатывает на протяжении полвека, с 1958 г., когда впервые её и выдвинул. Мода на изучение проблемы применения «здравого смысла» среди исследователей ИИ менялась: порой тема была популярной, порой – не пользовалась спросом. Для самого Дж. Маккарти сформулировать идею «здравого смысла» в научных терминах оказалось очень непросто. Число ученых, которые работают в области формализации «здравого смысла», до сих пор крайне незначительно.

Суть «здрaво-смысловых» программ Дж. Маккарти иллюстрирует на фоне анализа возможностей одной из наиболее известных экспертных систем – системы MYCIN (Shortliffe 1976; Davis, Buchanan and Shortliffe 1977). Эта ЭС представляет компьютерную программу для консультирования врачей по вопросам лечения бактериальных инфекций крови и менингита. Она работает достаточно надежно и вне поля понятия «здравый смысл». При этом, однако, необходимым условием является то, что пользователь обладает здравым смыслом и имеет необходимые знания, особенно по поводу ограничений данной ЭС.

MYCIN позволяет вести диалог в режиме «вопрос-ответ». После выяснения общих фактов о пациенте (имя, пол и возраст) программа MYCIN выдает запрос о предполагаемых бактериальных микроорганизмах, возможных очагах инфекции, специфических симптомов (например, лихорадки, головной боли и т.д.), имеющих отношение к диагнозу, результатах лабораторных исследований и некоторых других существенных аспектах развития болезни. Затем формулируется суть того или иного курса медикаментозного лечения с приемом антибиотиков. Диалог ведется на английском языке. При этом в программе MYCIN отсутствует функция свободного понимания английского языка в письменной форме. Она лишь способна выводить на экран формализованные предложения. При этом пользователь может ввести лишь отдельные слова или стандартные фразы. Но несмотря на недостатки диалога, главным преимуществом данной программы перед многими другими ЭС является то, что она использует систему критериев неопределенности (но не ве-

роятности) для постановки диагнозов. Более того, она способна разъяснить врачу суть приведенных доводов в пользу принятия того или иного решения.

«Онтология» MYCIN базируется на наборе широко варьируемых данных – бактериях, симптомах, тестах, возможных очагах инфекции, антибиотиках и способах лечения болезни. Имена докторов, названия больниц, характер заболевания и причины смерти при этом отсутствуют, они не существенны. Даже данные о пациентах могут отсутствовать, хотя MYCIN запрашивает множество фактических данных о конкретном пациенте. Это объясняется тем, что пациенты не являются значениями переменных величин и MYCIN никогда не проводит сравнения инфекционных заболеваний двух различных пациентов. Отсюда следует, что модификация MYCIN с целью получения информации об опыте работы крайне затруднительна.

Продукт MYCIN, прототипом которой она выступает, называется EMYCIN. Это – программная система, представляющая набор правил, каждое из которых состоит из двух частей – конфигурационной и операционной. При активации того или иного правила MYCIN проверяет, соответствует ли конфигурационная часть правила тем правилам, которые имеются в базе данных. Если ответ положительный, то переменные величины в конфигурационной части сопоставляются со всеми объектами, необходимыми для сравнения с базой данных. Если ответ отрицательный, то конфигурационная часть сигнализирует о сбое, и MYCIN делает следующую попытку. Если в конце концов сопоставление проходит успешно, то MYCIN активирует операционную часть, используя значения переменных величин, определенных конфигурационной частью. Весь процесс формулирования вопросов и выдачи рекомендаций строится на опыте практической деятельности экспертов.

Столь строгое конфигурирование вполне оправдано с точки зрения представления большого объема информации о характере диагноза и методах лечения бактериальных инфекций. Если программа MYCIN используется по назначению, она обеспечивает гораздо более высокие результаты по сравнению со студентами-медиками или интернами. Она даже может составить конкуренцию специалистам по бактериальным заболеваниям в случаях, когда их просят проделать те же действия самостоятельно.

К сожалению, дело применения MYCIN в широких масштабах застопорилось. По поводу этого специалисты приводят самые разнообразные доводы. Одни говорят, что эта ЭС эффективна лишь при условии регулярного пополнения базы данных MYCIN информацией о новейших открытиях в данной области (например, о последних проведенных тестах, новых теориях, перспективных методах диагностики и появлении на рынке антибиотиков следующего поколения). Другие специалисты утверждают, что MYCIN ни в какой мере не отвечает принципам практической применимости (за исключением экспериментов), поскольку эта программа не в состоянии определить круг своих собственных ограничений. Дж. Маккарти считает, что частично это объясняется и неспособностью врачей, использующих MYCIN, правильно понять документацию, особенно, в части, касающейся ограничений условий применения ЭС. Программисты нередко думают, что пользователи разработанных ими программ являются полными идиотами, так как не понимают ограничений MYCIN. Примером «незнания» MYCIN своих ограничений может быть сообщение системе о том, что у такого-то пациента обнаружено поражение кишечника холерным вибрионом. MYCIN в данном случае

бодро выдаст рекомендации о проведении двухнедельного курса лечения тетрациклином и более ничего. Предположительно, такой подход позволит уничтожить бактерии, но, вероятнее всего, пациент скончается от холеры задолго до этого момента. Тем не менее, врач, возможно, будет знать о том, что данный случай диареи требует интенсивного лечения, и предпримет необходимые меры для выяснения соответствующих методов и препаратов. Но это будет «здоровый смысл» врача. Известная доля «здорового смысла» должна быть представлена и в MYCIN и в иных проектах ЭС подобного типа.

Что же такое «здоровый смысл» и как его использовать в практике построения ЭС – это сегодня одна из ключевых областей исследований в области ИИ. Консенсус в этом вопросе по сей день не достигнут. Дж. Маккарти делит проблему на два аспекта: 1) что нам известно о «здоровом смысле»; 2) как применить имеющиеся знания о «здоровом смысле» на практике.

1. Что такое здоровый смысл?

Наиболее характерные представления о здоровом смысле эксплицируются на фоне анализа ситуаций, меняющихся во времени в результате совершения определенных событий. Самые важные события – это действия, которые ЭС осуществляет. Чтобы программа могла составлять разумные планы, она должна обладать способностью определять последствия своих собственных действий. То есть в неё должны быть заложены представления об области своих собственных действий.

Например, в область действия MYCIN входят врач, пациент и болезнь. MYCIN дает советы врачу. Поэтому следует предусмотреть информацию о влиянии сведений, предоставленных системой MYCIN, на предполагаемые действия врача. Поскольку MYCIN ничего не знает о враче, она могла бы спланировать воздействие курса лечения на пациента. Однако и эту задачу она не решает. На основании своих правил система рекомендует определенное лечение согласно полученной информации о пациенте. Но при этом MYCIN никоим образом не прогнозирует результаты лечения. Врачи, предоставившие заложенную в MYCIN информацию, разумеется, учли возможные результаты лечения.

С учетом узкоспециальной области применения MYCIN, здесь можно обойтись и без прогнозирования. Предположим, например, что некий антибиотик оказывает лечебное воздействие только в отсутствие высокой температуры у пациента. В этом случае система MYCIN могла бы в рамках плана применения данного антибиотика, спланировать снижение температуры пациента до нормальной и проверить результат такого снижения. В иных областях от ЭС и прочих программ искусственного интеллекта требуется планирование, однако к MYCIN данное требование на предъявляется. Отсутствие возможностей планирования отрицательно сказывается на полезности MYCIN.

MYCIN не дает прогнозов и это, безусловно, является ее недостатком. Например, системе MYCIN нельзя задать от имени пациента или администрации больницы вопрос о том, когда пациент сможет отправиться домой. Врач, использующий MYCIN, должен выполнить эту часть работы самостоятельно. Кроме того, MYCIN не дает ответов на вопросы о предполагаемых вариантах лечения, например: «Что произойдет, если я пропишу этому пациенту пенициллин»? Или даже: «Что плохого произойдет, если я пропишу этому пациенту пенициллин»?

В задачах ИИ используются различные формальные подходы для представления фактической информации о результатах некоторых действий или событий. Однако все известные по сей день системы представляют лишь результаты совершения некоторого события в некоторой ситуации, описывая новую ситуацию, возникающую в результате этого события. Как правило, этого достаточно. Но такой подход не распространяется на важный случай *взаимосвязанных событий и действий*. Возьмем, например, пациента, больного холерой. Антибиотик, убивающий бактерии холеры, может повредить его кишечник и вызвать потерю жидкости, возможно с летальным исходом. Формальный подход, удобным образом выражающий понятия о взаимосвязанных событиях, диктуемых здравым смыслом – это серьезная нерешенная проблема ИИ.

Мир велик и полон динамичных объектов. Они меняют свое положение, возникают и гибнут. Связанные с динамикой понятия, диктуемые здравым смыслом, сложно выразить. В MYCIN они вообще не учтены. Основная трудность заключается в обработке частных знаний, которыми обладает человек и необходимостью строить из частных знаний целостные знания. Человек может, например, видеть другого человека в анфас, не полностью. Однако исходя из своих отдельных представлений, он способен представить, как другой человек выглядит. «Я не вижу его спины, но предполагаю, что из нее не выступает 60-сантиметровый горб. Однако мое представление о форме спины является менее определенным, чем представление о тех частях его тела, которые я вижу» – приводит пример Дж. Маккарти.

Разумное поведение часто требует умения представлять и использовать *знания о знаниях*. В усовершенствованной системе MYCIN необходимо использование выводов, например, о том, что доктор Смит обладает знаниями о холере, поскольку он является специалистом по тропической медицине.

Программа, взаимодействующая или конкурирующая с людьми или другими программами должна «уметь» представлять информацию об их знаниях, убеждениях, целях, симпатиях и антипатиях, намерениях и возможностях. Усовершенствованной системе MYCIN, возможно, потребуется знать, что пациент не захочет принимать невкусное лекарство, если его не убедить, что это необходимо.

Область знаний, диктуемых здравым смыслом, в значительной степени пересекается с областью знаний, полученных с помощью точных наук, но отличается от них *эпистемологическим статусом*. Например, все знают, что произойдет, если бросить стакан воды на пол: стакан разобьется и вода разольется. Все знают, что это произойдет за долю секунды, и что брызги воды не разлетятся дальше, чем на 30 – 40 см. Однако такая информация получена без использования формулы падения тела или уравнений, описывающих течение жидкости. Мы не располагаем исходными данными для этих уравнений, большинство из нас не знает этих уравнений, мы не умеем решать их достаточно быстро для того, чтобы принять решение о необходимости убираться, например, с дороги, когда на тебя мчится автомобиль. Такая «физика здравого смысла» неразрывна с «научной физикой». Но на самом деле «*научная физика*» является *подмножеством «физики здравого смысла»*, поскольку именно «физика здравого смысла» объясняет нам, что значит, например, уравнение $s = 5gt^2$. Если MYCIN «вырастет» в «роботизированного

врача», то ей придется усвоить «физику здравого смысла», а также отчасти и «научную физику».

Возможность адекватного представления фактов из области здравого смысла посредством правил вывода представляется сомнительной. Рассмотрим следующий факт: при столкновении двух объектов обычно возникает шум [удара]. Этот факт можно использовать для создания шума, для устранения возможного шума, для объяснения наличия шума или для объяснения отсутствия шума. Этот факт можно использовать в определенных ситуациях, которым сопутствует шум, а также для понимания явления в целом: например, если непрошенный гость пойдет по гравийной дорожке, собака его услышит и залает. Правило вывода реализует факт лишь в качестве части определенной процедуры. Обычно такие правила соотносят факты, относящиеся к определенным объектам (например, к определенным бактериям) с основным правилом для получения новых фактов, относящихся к этим объектам. Многие современные исследователи ИИ озабочены проблемой представления фактов таким образом, который позволил бы использовать эти факты для самых разнообразных целей.

2. Рассуждения на основе здравого смысла

Наша возможность использовать знания, диктуемые здравым смыслом, определяется способностью рассуждать, исходя из здравого смысла. ЭС обычно реализуют логические выводы, не используя напрямую системы символической логики. Часто программа не предусматривает четкого разграничения между отбором правильных выводов и стратегией поиска выводов, необходимых для решения данной задачи. Тем не менее, такая логическая система обычно соответствует подмножеству символической логики первого порядка. Системы обеспечивают анализ факта, который относится к одному или более объектам, к фактам по поводу этих объектов и формирование общего правила, в котором содержатся переменные. Большинство экспертных систем, MYCIN в том числе, никогда не предлагают в качестве вывода общие положения.

Человеческое мышление предусматривает также получение фактов в результате наблюдения за окружающим миром. Аналогичным образом действуют и компьютерные программы. Роберт Филмен (Robert Filman) провёл любопытное исследование шахматного мира и показал, что многие факты, которые можно было получить посредством дедукции, в действительности могут быть получены и в результате наблюдения. Система MYCIN в этом не нуждается, однако нашему гипотетическому роботу-врачу придется делать выводы, исходя из внешнего вида пациента. Системы компьютерного зрения к этому не готовы.

Важным направлением в области ИИ (с середины 70-х г.г.) является формализация *немонотонных рассуждений*. Дедуктивные рассуждения в математической логике обладают монотонностью по аналогии со сходными математическими концепциями. Предположим, что имеется ряд допущений, из которых следуют некоторые выводы. Допустим, что вводятся некоторые добавочные допущения. Возможны некоторые новые выводы, однако каждое высказывание, которое являлось дедуктивным следствием исходных гипотез, является также и следствием их расширенного набора.

Обычные человеческие умозаключения не обладают подобным монотонным свойством. Если вам известно, что у меня есть автомобиль, то можете решить, что неплохо было бы со мной прокатиться. Но если вы далее

узнаёте, что мой автомобиль в ремонте (а это не противоречит тому, что вы знали ранее), то вы измените свое решение – прокатить я вас не смогу. Если же вы узнаете, что мой автомобиль всего через полчаса будет на ходу, то ваше решение изменится вновь.

Некоторые исследователи искусственного интеллекта, например, Марвин Мински (Marvin Minsky) (1974) указывали, что интеллектуальные компьютерные программы принципиально должны «рассуждать» немонотонно. Некоторые сделали из этого вывод, что логика не обеспечивает должной формализации. Однако оказалось, что дедукцию в символической логике можно дополнить новыми методами немонотонных рассуждений, столь же формальными, как дедукция и столь же пригодными для математического исследования и компьютерной реализации. Оказалось, что формализованные немонотонные рассуждения обеспечивают определенные правила выработки *предположений*, а не правила получения логических умозаключений – полученные на их основании выводы верны, но могут не подтвердиться при получении дополнительных фактов. Один из таких методов, *метод ограничения (circumscription)*, описан в работе (McCarthy 1980).

Основная идея символического описания метода ограничения достаточно проста. Имеется некое свойство, применимое к объектам или взаимоотношению, применимое к парам, тройкам и т.д. объектов. Это свойство или взаимоотношение ограничено некоторыми положениями, принимаемыми в качестве допущений, но при наличии некоторой степени свободы. Метод накладывает дополнительные ограничения на данное свойство или взаимоотношение, требуя его выполнимости для минимального набора объектов.

В качестве примера рассмотрим представление фактов о способности некоторого объекта к полету. Факты фиксируются в базе данных, построенной на основе «знаний здравого смысла». Можно попробовать подобрать аксиомы, определяющие способность к полету объектов каждого типа, но это приведет к значительному увеличению объема базы данных. Метод ограничения позволяет сформулировать допущение, что летать могут только те объекты, для которых в данном отношении имеется положительное утверждение. Таким образом, мы будем располагать положительными утверждениями о том, что птицы и самолеты могут летать и не будем располагать утверждениями, что верблюды могут летать. Наша база данных не предусматривает наличия отрицательных утверждений. Поэтому чтобы учесть летающих верблюдов (если таковые появятся) достаточно добавить новые утверждения, не удаляя прежних. Часто это производится более простым методом «допущения замкнутого мира» (*closed world assumption*), описанным Раймондом Рейтером (Raymond Reiter). Однако из общего правила «птицы могут летать» также имеются исключения. Например, не могут летать пингвины, страусы. Птица не сможет летать, если её общипать. Можно обнаружить и другие исключения и даже исключения из исключений. Метод ограничения позволяет учесть известные исключения, предусматривая последующее добавление новых – опять же не изменяя имеющихся утверждений.

Человек в процессе общения постоянно пользуется немонотонными умозаключениями. Допустим, – говорит Дж. Маккарти, – я вам заказал изготовить клетку для птицы. Вы изготовили клетку без крышки и я отказываюсь вам платить, поскольку моя птица может улететь. Судья будет на моей стороне. С другой стороны, допустим, вы изготовили клетку с крышкой. Я отка-

зываюсь вам платить, поскольку моя птица – пингвин и крышка для клетки не нужна. Если я заранее не предупредил вас, что моя птица не летает, то судья будет на вашей стороне. Таким образом, можно считать, что по умолчанию принимается соглашение о том, что если птица может летать, то этот факт оговаривать не обязательно, однако если птица не летает и это относится к делу, то такой факт обязательно следует оговорить.

Определённый интерес представляет дискуссия по поводу положений, предложенных Дж. Маккарти.

ВОПРОС. По-вашему, программам требуется здравый смысл. Но это все равно, что сказать: «Если бы я умел летать, то я не заплатил бы компании Eastern Airlines 44 доллара за то, что она доставила меня сюда из Вашингтона». Если программам требуется «здравый смысл», то как решить эту задачу? Не в этом ли суть дела?

МАККАРТИ. Я мог бы произнести речь в защиту искусственного интеллекта, но предпочитаю сосредоточить внимание на выявленных проблемах, а не на успехах, достигнутых при их решении. Напомню, что потребность в здравом смысле неочевидна. Много полезного можно сделать и без него. Примеры: MYCIN, шахматные программы и многие другие программы полезны.

ВОПРОС. Похоже, что в ваших словах о здравом смысле, о его роли для человеке, в подчеркивании экспериментальной составляющей, есть рациональное зерно – особенно в примере с бросанием стакана воды. Интересно, развитие этих программ потребует столько же времени, сколько вам требуется для выработки этого опыта? Требуется ли программе своего рода «школа опыта» и соответствующая аттестация? Ведутся ли работы по ускорению данного процесса или для выработки здравого смысла в достаточном объеме программе потребуется 20 лет, считая с момента её создания?

МАККАРТИ. Вы говорите, 20 лет. Если бы в 1963 г. всем было известно, как создать программу, способную на основе собственного опыта научиться тому, что умеет врач-человек с 20-летним стажем, то такая программа была бы создана и к настоящему времени весьма бы преуспела. Работы над программами, способными обучаться на основе опыта, велись уже в 1958 г. Однако такие программы смогли научиться только заданию оптимальных числовых значений для своих параметров и то в весьма ограниченных пределах. Программа игры в шашки Артура Сэмюэла (Arthur Samuel) определяла оптимальные значения своих параметров. Однако проблема заключалась в том, что в некоторых случаях необходимое поведение не соответствовало никакому набору параметров, поскольку определялось стратегическим пониманием ситуации. Таким образом, в программу, способную чему-то научиться, прежде всего должна быть заложена способность должным образом изменять свое поведение. Простым изменениям поведения должны соответствовать простые методы представления. Универсальная машина Тьюринга указывает на возможность представления произвольного поведения. Однако при этом ничего не говорится о таких методах, чтобы малому изменению поведения соответствовало малое изменение представления. Современные методы изменения программ похожи на обучение посредством операции на мозге.

ВОПРОС. Вопрос о программах, нуждающихся в здравом смысле, мне хотелось бы задать несколько по-другому, используя программу MYCIN в качестве примера. Перед нами три действующих лица: программа, врач и пациент. Исходя из принципа безопасности пациента, я полагаю, что по крайней мере двое из этих действующих лиц должны обладать здравым смыслом. Например, если было бы достаточно одного, то это должен быть пациент, поскольку в отсутствие здравого смысла у программы и у врача пациенту требуется здравый смысл, чтобы просто сбежать от них уйти (порой так и следует поступать). Но, допустим, в программу заложен здравый смысл, врач обладает здравым смыслом, а вот у пациента здравого смысла нет. Впрочем, на самом деле это не важно, поскольку пациент все равно выполняет те действия, которые от него востребованы.

Рассмотрим другую возможность. Если здравый смысл есть только у программы, а у врача и у пациента здравого смысла нет, то в долгосрочной перспективе программа также откажется от использования здравого смысла.

Следует ли проблематику «здравого смысла» рассматривать под таким углом?

Д-Р МАККАРТИ. При работе с MYCIN предполагается, что источником здравого смысла является врач. Вопрос в том, должна ли программа тоже обладать здравым смыслом? В случае MYCIN ответ не ясен. Чисто вычислительным программам здравый смысл не нужен. Ни одна современная шахматная программа не обладает здравым смыслом. С другой стороны, представляется очевидным, что многие иные программы нуждаются в здравом смысле, чтобы приносить хоть какую-то *пользу*.

Что такое искусственный интеллект?

Валерия Прасолова (А-81)

В основе доклада лежит работа Джона МакКарти «Что такое искусственный интеллект?», 1998 г.¹

Искусственный интеллект – это наука и технология создания интеллектуальных машин, в особенности, разработки интеллектуальных компьютерных программ.

Интеллектуальные компьютеры, в отличие от традиционных компьютеров, снабженных интерфейсом реализации диалога на полном или ограниченном языке, должны отвечать требованиям: *интеллектуальные компьютеры способны развиваться и совершенствовать имеющиеся у них методы и программы.*

Первые применения ИИ: Одним из первых применений интеллектуальных механизмов в компьютерных программах стала игра в шахматы. Механизм ее действия таков: программа перебирает множество ходов и отбирает наилучшие из них. Однако когда появилась программа для игры *go*, действующая по тому же принципу, что и для игры в шахматы, обнаружился недостаток имеющихся знаний об ИИ – компьютеры играли в *go* очень плохо, так как эта игра требует невообразимо много переборов.

Противники ИИ: В их числе есть весьма образованные и уважаемые люди, такие, как философы Джон Серл, Г. Дрейфус.

¹ McCarthy, John (1998) What is Artificial Intelligence? <http://cogprints.ecs.soton.ac.uk/archive/00000412>

Аргументы против ИИ – это доводы противников ИИ, отрицающие возможность его создания. Наиболее известен аргумент, получивший название «Китайская комната». Он заключается в следующем: человек сидит в комнате с книгой по грамматике китайского языка. Через дверь ему передают предложения на китайском языке. Человек смотрит в книгу, записывает на бумагу иероглифы в определенном порядке, указанном в книге по грамматике, и передает ответ. Таким образом осуществляется диалог. Для реализации диалога человеку не нужно владеть китайским языком. Исходя из того, что диалог осуществляется лишь в силу формальных преобразований над совокупностями символов, Джон Серл выводит, что точно так же и компьютерная программа, придерживающаяся подобных правил, не понимает китайского языка. Следовательно, заключает он, ни одна компьютерная программа не может ничего понимать.

Определённое отношение к проблемам ИИ имеют теория сложных числений и алгоритмов. Существуют алгоритмически неразрешимые проблемы. Например, невозможно доказать с помощью формальных средств утверждение о том, является ли предложение логики первого порядка теорией или не является таковым. Однако и люди не всегда могут решить произвольные задачи в этих областях.

Парадигмы исследования ИИ:

1) Биологическая (или бионическая) – основана на имитации психологических, физиологических, биологических особенностей человека, животных и иных живых существ;

2) Эпистемологическая – основана на изучении и формализации фактов здравого смысла о мире и о задачах, цели которых представлены в пространстве решений.

Замечания к направления и проблемам развития ИИ:

Логика. Программа обобщает знания о мире и всевозможных ситуациях. Результаты представления фиксируются в виде предложений на языке символической логики.

Поиск. Программа ИИ целенаправленно перебирает множество возможностей, например, просчитывает множество возможных ходов в шахматах.

Распознавание образов. Программа ИИ, сравнивая сканируемое изображение с имеющимися образцами, делают соответствующие выводы по идентификации объектов.

Логический вывод. Помимо методов традиционной логики, программы ИИ используют методы немонотонного следования, которые отличаются от монотонной логики, например, возможностью отмены заключения, если появляется доказательство противоположного утверждения.

Обучение. Программы способны воспринимать только корректно представленные данные, а далеко не все факты представлены должным образом.

Планирование. Планирование программы начинается с постановки цели, после чего вырабатывается алгоритм ее достижения.

Генетическое программирование. Это – способ построения программы посредством скрещивания случайных программ и выбора лучшей программы из миллионов поколений.

Пожелания исследователям ИИ от Дж.Маккарти:

Если вы хотите изучать ИИ и не знаете, с чего начать, начните с изучения символической логики. Читайте больше по психологии и физиологии нервной деятельности. Выучите какие-нибудь языки программирования, например, Java или C++, они сейчас широко востребованы.

Проблема дискурса искусственного интеллекта. Конструкторская позиция **Денис Родионов (АП-82)**

Дискурс искусственного интеллекта чреват метафизическими спекуляциями, паранаучностью и даже антинаучностью теоретических суждений и проектных решений¹. Традиционно данные ловушки избегаются риторически – путем указания на метафорический характер терминологического аппарата ИИ. Этот путь – конститутивный. К инженерной практике он мало применим. Конструктивным путем представляется выбор разработчиком определенной позиции, позволяющей однообразным способом судить об интеллекте, менталитете, психике и т.п. характеристиках человека, животного, робота. Данную позицию вслед за Д. Деннетом принято называть *конструкторской (дизайнерской) позицией*². А. Сломан и Дж. Маккарти считают, что разработчик интеллектуальной системы занимает её, спускаясь с уровня *функциональной позиции*³.

Конструкторскую позицию, модифицированную в аспекте ИИ, детально раскрыл Дж. Маккарти в работе «Правильно разработанный ребёнок»⁴. Ребёнок – искусственный. Это – робот. В этого «ребёнка» разработчик закладывает знания, включая знания о приобретении новых знаний. То есть разработчик наделяет своего ребёнка интеллектом. Суть конструкторской позиции заключается в том, что разработчик априори технологического процесса обладает всеми необходимыми знаниями о том, какие генетически предопределённые и приобретаемые знания полезны ему самому. Они должны быть полезны и ребёнку. Когнитивные наука и философия способствуют этому творческому инженерному процессу, осуществляя обоснование действительной полезности знаний, приобретаемых и используемых человеком в практической деятельности.

С конструкторской позиции становится удобно исследовать и продуктивно применять такие понятия, как «сознание» и «самосознание» робота. Причем, без всяких метафор и подчеркивания того, что сознание и самосознание – суть нечто интимное и сугубо человеческое. Так, робот обладает са-

¹ Алексеев А.Ю. Паранормализация искусственного интеллекта. // В сб. Поругание разума: экспансия шарлатанства и паранормальных верований в российскую культуру XXI века. М: «Российское гуманистическое общество», 2001 г., С.8-11

² О трех позициях Д.Деннета – физической, интенциональной и конструкторской - см. в McCarthy, John, 1995, Artificial Intelligence and Philosophy. (<http://cogprints.ecs.soton.ac.uk/archive/00000420>). Конструкторская позиция Дж. Маккарти близка к функциональной позиции. В своих работах Дж. Маккарти данный факт неоднократно подчёркивает.

³ McCarthy, John, 1995, Artificial Intelligence and Philosophy. (<http://cogprints.ecs.soton.ac.uk/archive/00000420>)

⁴ McCarthy John, 1999, The Well-Designed Child. (<http://www-formal.stanford.edu/jmc/child/child.html>)

мосознанием (способностью осознавать свои ментальные состояния), если он способен репрезентировать собственные ментальные состояния и процессы, существенные для организации полноценного интеллектуального поведения.

Естественный, человеческий интеллект не является константой. Он развивается, способствуя выживанию и преуспеванию человека в этом сложном, лишь отчасти обозримом и весьма слабо контролируемом мире. Но динамика интеллекта обусловлена теми свойствами и особенностями мира, предстоящего перед человеком, которые остаются постоянными уже на протяжении миллионов лет. Данный мир не должен становиться всякий раз новым для каждого нового человека и животного. Поэтому и программы робота-ребёнка должны работать на основе этих устойчивых характеристик. Однако разработчик должен выборочно подходить к тому, какие именно знания о мире вложить в робота. Разработчик не должен впадать в дискурс метафизического типа – для разумного поведения робота такой дискурс бесполезен. Разработчику следует изучать мир на уровне непосредственного взаимодействия человека с окружающим миром.

Ментальные состояния и свойства человека, несомненно, разнообразны. Это – любопытство, эмоции, память, мотивационно-волевые и эффекторные механизмы и др. Однако среди всего этого разнообразия для целей эффективного ориентирования в мире робота-ребёнка полезны не все знания. К полезным относятся: трёхмерное пространство, естественные виды, тела (постоянство и изменение тел), пространственная локализация, составные части. Некоторые из этих знаний являются общезначимыми для человека и животных, некоторые – специфичные лишь для человека. Многие знания полезны роботу-ребёнку. Полезны они потому, что позволяют ему правильно ориентироваться в окружающем мире, приобретать о нём знания, развивать свой собственный искусственный интеллект.

Конструктивный дискурс ИИ, внедряющий в робота подобного рода врожденные знания, должен вестись на специальном языке. Дж. Маккарти такой язык называет «языком мышления». Принципы построения и использования такого языка: логичность, параллелизм, двойственность грамматики (деление грамматики на поверхностную и глубинную в смысле Н.Хомского), лаконичность, обоснованность и др. «Язык мышления» должен быть универсальным – то есть в унифицированной форме описывать: 1) мыслительную деятельность человека; 2) разумное поведение животных; 3) технические особенности мышления роботов.

Конвенциональное принятие «языка мышления» в инженерном коллективе позволяет разработчику занять конструкторскую позицию и непосредственно обуславливает создание разумных роботов – правильно разработанных «детей». Это – теоретическая позиция Дж. Маккарти. И с такой позиции дискурс ИИ представляется не лишённым здравого смысла. Он способствует научно-теоретическому обоснованию эффективных («правильных») проектов интеллектуальных систем.

Зачем искусственному интеллекту философия?

Елена Сименел (ИС-81)

Доклад инспирирован статьёй основоположников философии искусственного интеллекта Джона Маккарти и его коллеги Патрика Дж. Хейса, 1968 г.¹ При построении интеллектуальных систем на протяжении почти сорока лет Джон Маккарти не сворачивает с пути, обозначенному в данной статье, требуя философской рефлексии над проблемами программно-технического характера.

Идея создания интеллектуальной машины возникла давно, но серьезная работа над проблемой искусственного интеллекта и даже серьезное понимание сути проблемы еще ожидают своего часа. Явно тема искусственного интеллекта прозвучала в 1950-е г. в статьях А. Тьюринга (1950 г.) и Шеннона (1950).

С тех пор попытки создания искусственного интеллекта главным образом происходили в написании программ, призванных решать класс проблем, представляющих определенную трудность для человека. Это, например, программы: 1) играющие в шахматы или шашки; 2) доказывающие математические теоремы; 3) преобразующие одно символическое выражение в другое в соответствии с заданными правилами; 4) интегрирующие выражения, составленные из элементарных функций; 5) определяющие химический состав по данным, полученным масс-спектрографическим и другими методами.

В ходе проектирования этих интеллектуальных механизмов выявляются общие закономерности интеллектуальной деятельности, которые можно воплотить в компьютерных системах. Для выявления закономерностей используются также методы интроспекции (самоанализа), математического анализа, проводятся эксперименты над людьми и животными. Также осуществляется тестирование ИИ-программ, что иногда приводит к лучшему пониманию интеллектуальных механизмов и созданию новых.

Все эти работы, несомненно, способствуют развитию ИИ. Однако многие исследователи пытаются расширить наработки в узкой области исследований на ИИ-системы в целом. Некоторые исследователи отождествляют свою специфическую проблему с целым, считая при этом, что видят лес, хотя на самом деле он смотрят лишь на дерево. В рамках такого подхода предпринимаются попытки спроектировать интеллект, столь же гибкий, как и интеллект человека. Различные исследователи рассматривали эти проблемы по-своему, но ни один из них не добился успеха при построении систем ИИ. Нужна некая общая теория интеллекта, пригодная для самого широкого класса систем.

Что такое «общая» теория интеллекта?

Попытка создания общей теории интеллекта прослеживается в работах А. Тьюринга (1950 г.). Его идея состояла в том, что машина является интеллектуальной, если способна успешно имитировать человека в течение пол часа для самого искушенного наблюдателя. Однако построение теории ИИ

¹ Джон Маккарти и Патрик Дж. Хейс. Некоторые философские проблемы с точки зрения искусственного интеллекта, 1969, <http://www-formal.stanford.edu/jmc/>

посредством теории имитации отвлекается на поверхностные аспекты человеческого поведения. А. Тьюринг исключил некоторые из них, накладывая ряд ограничений: компьютер находится на другом конце линии телетайпа, поэтому голос, внешний вид, запах, и т.д. исключаются из рассмотрения. А. Тьюринг так же изучил вопросы имитации лени, способности «поэтического» использования английского языка и совершения человеком арифметических ошибок.

Однако работа над ИИ, в особенности над общей теорией ИИ может стать намного плодотворной, если уяснить, что есть «интеллект». А для этого требуется привлечение философских обобщений в плане выработки понятия «интеллект».

Один путь вовлечения философских знаний – привлечение бихевиоральной теории сознания. Суть её состоит в том, чтобы дать чисто поведенческое определение или определение, основанное на понятии «черный ящик». Машина является интеллектуальной, если она решает некоторые классы проблем, требующие для своего решения человеческого интеллекта или может работать в сложной среде, требующей применения интеллекта. Это определение кажется расплывчатым. Возможно, его можно сделать несколько более точным, не отступая от поведенческих терминов.

Однако вместо бихевиоральной теории лучше воспользоваться очевидными для самоанализа вещами, например, *знанием фактов* (собственно самим сознательным опытом знания фактов). При этом имеется двойной риск: во-первых, можно ошибиться при самосозерцательном рассмотрении собственной ментальной структуры (нам может только показаться, что мы знаем о фактах, на самом деле это не так). Во-вторых могут существовать объекты, которые удовлетворяют бихевиоральным (поведенческим) критериям интеллекта, в действительности, таковыми не являясь. Тем не менее, бихевиоральные критерии остаются важными, так как интеллектуальная машина рассматривается в виде манипулятора, активно действующего во внешней среде.

Интеллектуальная система в свой состав включает модель мира. На основе этой модели система ИИ должна быть способна правильно отвечать на ряд вопросов. Например на такие вопросы, как: 1) Что произойдет в результате изменения ситуации? 2) Что произойдет, если я совершу некоторое действие? 3) Чему равно $3 + 3$? 4) Что он хочет? 5) Могу ли я выяснить, как сделать это или я должен получить дополнительную информацию от кого – то или чего-то?

Эти вопросы не составляют полностью репрезентативного набора вопросов. Но полного набора вопросов и невозможно придумать.

На основании вышеприведённых вопросов можно дать определение системе ИИ: система является интеллектуальной, если имеет адекватную модель мира (включая «мир» математических выражений, понимание своих собственных целей, ментальных процессов); если достаточно «умна» для того, чтобы ответить на широкое разнообразие вопросов на основе этой модели; если может получить дополнительную информацию от внешнего мира, когда это требуется; если может выполнить во внешнем мире такие задачи, которые необходимы для достижения поставленной цели и которые она физически способна решить.

Согласно этому определению система ИИ состоит из двух частей, эпистемологической и эвристической. *Эпистемологическая часть* – представление мира в форме, позволяющей решать задачи, исходя из *фактов*. *Эвристическая часть* – механизм, который на основе конкретной информации решает поставленную задачу и принимает решения для дальнейших действий. Большинство работ по ИИ до сих пор посвящались эвристической части, поэтому стоит уделить большее внимание эпистемологической части.

Если принять понятие «интеллекта», приведённое выше в виде примера вопросов, возникают следующие классы проблем, связанные с построением эпистемологической части интеллектуального объекта:

1. Какое общее представление о мире позволит объединить имеющиеся наблюдения и новые закономерности, когда последние будут обнаружены?
2. Помимо представлений о физическом мире, какие другие виды объектов должны быть предусмотрены? Например, надо ли предусмотреть математические системы, цели, структуры знаний.
3. Как надо использовать наблюдения, чтобы получать знания о мире и как другие виды знаний будут приобретаться? В частности, какие знания о собственном состоянии системы должны быть предусмотрены?
4. В каких внутренних определениях должны быть выражены знания системы?

Эти вопросы идентичны или, по крайней мере, близки некоторым традиционным вопросам философии, особенно вопросам метафизики, эпистемологии и философской логики. Поэтому для исследователя ИИ очень важно рассматривать проблематику искусственного интеллекта с философской позиции.

Так как философы даже спустя 2500 лет не пришли к единому соглашению, что считать интеллектом, то может показаться, что проблема создания ИИ находится в довольно безнадёжном состоянии, ведь для написания компьютерных программ необходима конкретная философская информация. К счастью, предпринимая попытки вовлечь философию в компьютерные программы, мы приходим к принятию достаточно большого количества философских предположений, которые исключают большую часть философского знания как не относящуюся к рассматриваемой проблеме.

Попытки создания универсальной интеллектуальной компьютерной программы, т.е. программы, реализующей «общую» теорию интеллекта, влекут следующие философские предположения:

1. Физический мир существует и в нём существуют интеллектуальные машин, называемые «люди».
2. Информация об этом мире приобретается через органы чувств и выражается во внутренних репрезентациях (как части модели мира).
3. Представление о мире, созданное на основе здравого смысла, является приблизительно правильным. То же самое касается и научных знаний.
4. Рассматривая общие проблемы метафизики и эпистемологии, не всё следует начинать «с нуля». Вместо этого стоит использовать все имеющиеся знания для создания компьютерной программы, которая обладает способностью «знать». Такая точка зрения соответствует доминирующему в настоящее время отношению к основаниям математики: вначале изучается структура математических систем – извне как она есть. Далее используются метаматемати-

ческие инструментальные средства, вместо того, чтобы априори принимать небольшое количество основных фактов и, используя их, постепенно шаг за шагом создавать аксиомы и правила в пределах системы.

5. Следует создать всеобъемлющую философскую систему, которая не будет основываться на существующей тенденции к раздельному изучению проблем без попытки объединения результатов.

6. Критерий определенности системы ИИ становится намного сильнее. Если, например, в рамках эпистемологической части системы невозможно, даже в принципе, создать компьютерную программу, которая может самостоятельно приобретать знания, то такая система должна быть отклонена как слишком неопределенная.

7. Проблема «свободного выбора» принимает острую, но конкретную форму. А именно, в рассуждениях, основанных на здравом смысле, человек часто решает, что сделать, предвосхищая результаты различных действий, которые он может сделать. Интеллектуальная программа должна использовать подобный процесс. На основе точного формального определения понятия «мочь», программа должна обнаруживать альтернативные пути. При этом не следует отрицать того, что программа является детерминированной.

8. Основная задача состоит в том, чтобы определить, пусть даже упрощенное и основанное на здравом смысле, представление о мире, достаточно точное для того, чтобы запрограммировать компьютер на адекватные действия. Это – очень трудная задача сама по себе.

Следует упомянуть, что есть ещё один способ создания ИИ, который не требует точного понимания понятия «интеллект» и решения связанных с этим философских проблем. Этот способ – создание компьютерной модели естественного отбора, в которой развитие интеллекта происходит путем изменения компьютерных программ в соответствующей среде (имеется в виду генетический метод). Этот метод пока не имел существенного успеха, возможно из-за неадекватных моделей мира и эволюционного процесса, но по всей видимости он может еще достичь успеха. Здесь существует опасность того, что программа, интеллектуальность которой (т.е. внутренние механизмы осуществления интеллектуальной деятельности) до конца не понимает даже сам создатель, может выйти из-под контроля.

В любом случае, попытки создания ИИ путем *предваряющей философской рефлексии над понятием «интеллект»* принесут плодотворные результаты.

III. ИСТОРИКО-ФИЛОСОФСКИЕ ПЕРСПЕКТИВЫ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Контентуальная модель смысла (А.Ф. Лосев)

Анатолий Артюхов (Р-81)

Модель физического мира – целостной совокупности объектов и их отношений – необходимая составляющая системы искусственного интеллекта, особенно если в роли таковой выступает ориентирующийся в пространстве робот. Однако такой модели недостаточно для раскрытия внутренних связей и принципов организации и самоорганизации «мыслей» о мире, которыми пользуется ИИ-система. Необходимо конструирование её идеального мира, её псевдосознания. В данном мире концентрированной формой взаимосвязи моделей идеальных сущностей выступают объективированные «смыслы», «эйдосы». Эти «смыслы» находятся на взаимопересечении отношений: знак – значение, знак – символ, образ – миф и т.д. Отношения раскрывают внутренние противоречия и тождества в импликациях диалектического движения и развития «смысла» вещей мира, которые окружают и которыми окружает себя интеллектуальная система.

1. Подходы к моделированию смысла

Многозначность и нечёткость значения термина «смысл» затрудняет его корректное применение. Наиболее распространены три системно-взаимосвязанные трактовки понятия «смысл»¹:

1) *Контекстуальное определение.* Смысл – это выражаемый знаком способ задания значения, т.е. более широкий контекст для значений языковых высказываний, связанных текстов, образов сознания, ментальных состояний и действий. Смысл – это горизонт человеческого мира для фрагментов реальности, некая ментальная конструкция, лежащая за границей области понятий и категорий. Обыденное словоупотребление, соответствующее этой позиции связывает «смысл» с идеей, сущностью, точкой зрения («в смысле того или иного»). Контекстуальная трактовка в наибольшей мере отвечает возможностям компьютерного моделирования «смысла». Она открывает возможность кодификации смысла («контекста») явлений реальности («текста»). Моделированию контекста посвящён объемный корпус работ в области теоретических основ искусственного интеллекта (В. Акман, П. Берньер, Р.П. Вуртц, Дж. Маккарти, Д. Мичи, Д.А. Поспелов, А. Ребер, Р. Сталнейкер, П. Турней, В.К. Финн, В. Эдмондс и др.). Представленный в этих работах способ моделирования носит в основном системно-структурный и логико-математический характер.

¹ Алексеев А.Ю. Компьютерное моделирование смысла (философско-антропологический анализ), автореферат диссертации, М., 2003

2) *Интенционалистское определение*. Смысл – это целевая направленность, ценностная ориентированность. К этой трактовке относятся выражения «смысл жизни», «смысл истории», «смысл бытия», т.е. «*делать что-то с мыслью*». «Интенциональная» трактовка понятия смысла инспирирована концепциями В. Дильтея, Г. Риккерта, М. Вебера и достигла своего апогея в антипсихологизме Э. Гуссерля (анализ Н.М. Смирновой), а также в экзистенциалистских концепциях Н.А. Бердяева, М. Хайдеггера, Фр.-В. фон Херрманна и др. Построение компьютерной модели представляется затруднительным.

3) *Контентуальное определение* (от сл. «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 характеризуется целостностью, эмерджентностью, синергийностью мыслей. По мнению Н.К. Гаврюшина, «со-мыслие» присуще специфически русскоязычному варианту употребления слова «смысл» и прослеживается в работах А.С. Хомякова, Л.П. Карсавина, В.С. Соловьева, С.Н. Трубецкого, А.Ф. Лосева, Н.О. Лосского, П.А. Флоренского, С.Л. Франка и др. В настоящее время в этой традиции работает С.С. Хоружий. В контентуальном подходе неразрывно со «смыслом» связываются категориальные внесмысловые корреляты: самость, бытие, язык, сознание, личность, культура. Прослеживается как ярко выраженная персоналистическая позиция, так и самобытный отечественный подход к определению понятия «смысл», выявляющий перспективы моделирования контентуально трактуемого смысла (смысла как «со-мыслия», в частности, как самоорганизации «мыслей»). В области компьютерной науки этот «самоорганизующийся» подход прослеживается, например, в «машине сравнения идей» С.Н. Корсакова, предложенной в 1832 г., за 10 лет до машины Бэббиджа, с которой ведёт отсчет официальная история компьютерной технологии (исследования Г.Н. Поварова); в специфическом подходе к кибернетике как науке о сложных самоорганизующихся системах (А.И. Берг); в информационном подходе к проблеме «сознание-мозг» (Д.И. Дубровский), где категория информации понимается как содержание отражения на уровне самоорганизующихся систем.

Представляет интерес рассмотрение базовых положений контентуальной модели на примере модели «смысла», которая может быть реконструирована из работ А.Ф. Лосева.

2. Ключевые положения контентуальной модели смысла

Контентуальная модель как «со-мыслие», принцип самоорганизации и самоструктурирования «мыслей о вещи» А.Ф. Лосев мог бы назвать самостью вещи и интерпретировать её как «самое-само» мыслей.

Смысл вещи заключается в ее сущности, ее самости. Он не зависит от вида данности этой вещи в нашем мире. Самость вещи – ее смысл, ее абсолютная индивидуальность, то, что выделяет вещь из континуума вещей. Вещь, в свою очередь, даже идеальная – мысль, представимая в бесконечном ряде интерпретационных символов.

3. Потребность в контентуальном подходе

В контекстуальном подходе смысл пытаются исчерпать пропозициями (совокупностью предложений). Это – традиционный логико-позитивистский

подход (Б. Рассел, Л. Витгенштейн, Р. Карнап). Очевидно, однако, что самость «мыслей», их целостность не представима дискретной схемой (сколь бы широкой или глубокой она бы ни была). Смысл как контекст постоянно «ускользает», превращаясь в дурную бесконечность (Р. Карнап). Репрезентация смысла достигается лишь на пути частных интерпретаций и отдельных предложений. И верхним «потолком» таких интерпретаций становится «невыразимое» («таинственное», «мистическое») Л. Витгенштейна. Но «со-мыслие» – уникальный, сингулярный факт понимания, который не определим ни по отдельным ее признакам, ни по сумме этих признаков. Всякое познание вещи, раскрытие ее контекста, ограничивается лишь конечным набором дискретных интерпретаций, которые устремлены в бесконечность.

Традиционный – контекстуальный подход – не работает.

Для постижения самости вещи недостаточно описать сколь угодно большое число признаков (а значит и кортежей пропозиций) – это конечный ряд интерпретаций самости из бытия вещи в окружающее ее инобытие. Самость вещи неопределима через собственные признаки, она определима только через саму себя, она сугубо индивидуальна и не редуцируется на другую вещь. Самость «мыслей» не выразима, не мыслима. Она постигается лишь в акте интуитивного понимания. *Контент*, собственное мыслительное содержание, есть самость мысли, тайный смысл мысли (Лосев и утверждал, что самость – тайна).

В целом учение о самости, как смысле, можно свести к трем положениям¹:

1. Все что есть – абсолютно индивидуально.
2. Абсолютная индивидуальность вещи – самость, исключает любое совпадение с чем бы то ни было.
3. Абсолютная индивидуальность вещи невыразима, она лишь только интуитивно ощущается.

4. Перспективы

Многообещающим представляется использование контентуальной концепции смысла Лосева для построения и обоснования методологии новейших информационных систем, делающих упор в своем развитии на средства голографической обработки и представления информации. Голограмма – это система полной записи пространственной структуры электромагнитного поля, образующаяся в результате ее интерференции с опорным лучом. Голографическое представление предмета есть наиболее полная интерпретация его «смысла». Конечно, такое представление полностью не отражает «самость» предмета. Необходим ряд допущений и ограничений:

- 1) Нас интересует интерпретации вещи в форме зрительно-цветового восприятия (конституирование самости вещи как явленного эйдоса);
- 2) Носителем информации является форма, материальный субстрат представления информации нас не интересует (см. «принцип инвариантности информации относительно её носителя» Д.И. Дубровского);
- 3) Практический смысл вещи выразим, мыслим, индивидуален. Он фиксируется в указующем жесте «вот это».

¹ Лосев А.Ф. Миф, число, сущность. М., Мысль, 1994

Голограмма вещи в форме подобного рода эйдетико-смысловой визуализации *как бы* тождественна (псевдотожественна) самой вещи. Трехмерная голограмма будет наиболее полным выражением внутреннего единства вещи, ее самоорганизации и структуры. Визуализация фактов относительно не-которой фиксированной вещи есть особое состояние вещи, новая вещь – «псевдовещь». Псевдовещь визуально неотличима от «смысловой сущности» вещи и представляет частичный ряд интерпретаций «со-мыслия», но более полный и более информационно ёмкий, нежели чем получаемый при иных способах визуализации вещей. Манипулирование «псевдовещами», возможно, позволит:

1. Обеспечить целостное (за счет их визуализации) представление многомерных, гетерогенных структур и потоков данных, которые репрезентируются и подвергаются обработке в сложных информационных системах¹;

2. Приблизиться в робототехнике к реализации «внутреннего глаза», т.е. интерфейса связи с внешним миром, который позволяет в единстве представить наиболее исчерпывающую картину мира за счет устранения дискретного характера существующих способов восприятия (к примеру, средствами нейросетей. Так же см. 3,а).

3. Реализовать т.н. «акустическую голографию»² в качестве новых средств общения и связи (акустическая голография в основе имеет те же принципы, что и оптическая, но использует не поперечные электромагнитные волны оптического диапазона, а продольные – акустические и ультразвуковые волны). Это позволит, в частности:

- а) Используя способность ультразвуковых волн проникать сквозь оптически непрозрачные среды, получать объемные изображения внутренних частей объекта, который подвергается облучению звуковыми волнами;
- б) Представить голограмму предмета в виде оригинального рельефа на поверхности жидкости или мягкого материала. Отсюда имеется потенциальная возможность создания тактильно – вербальной системы общения, основанной на восприятии «акустической» голограммы.
- в) Создать своеобразную голограммно-акустическую «письменность», которая позволит оперировать «псевдоиероглифами» – тактильно-вербальными «смыслами».

Заключение

Синтез философского наследия и теоретико-методологических разработок современных информационных технологий – основа гармоничного и устойчивого развития общества в новом тысячелетии. Монументальный и многообъемлющий массив информации, который предоставляют философские знания становится источником развития инновационных технологий. В свою очередь, реализуя прогностическую функцию, философия порождает продуктивные футурологические проекты, один из которых был представлен в докладе.

¹ Кухаркин. Электрофизика информационных систем. М., Высшая школа, 2001

² Франксон. Голография. М., Мир, 1972

Проблема идеального и искусственный интеллект

Анастасия Дробященко (М-08-03)

Ключевое значение для методологии компьютерных наук приобретают формы осмысления и способы разрешения проблемы, получившей название «проблема идеального»¹. Именно в данной проблеме искусственный интеллект должен черпать конструктивные принципы моделирования разума. Нет сомнения, что в диалектическом материализме сложились наиболее проработанные подходы к проблеме идеального. В немалой степени это обусловлено установлением чётких дистинкций между материальным и идеальным. Имеются две фундаментальные позиции в понимании проблемы: 1) позиция Э.В. Ильенкова и 2) позиция Д.И. Дубровского.

1. Позиция Э.В. Ильенкова.

Э.В. Ильенков считает, что «идеальность – это *характеристика вещей*, но не их естественно-природной определённости, а той определённости, которой они обязаны труду... Идеальная форма – это форма вещи, созданная общественно-человеческим трудом, или форма труда, осуществлённая в вещи природы, «воплощённая» в нём, отчуждённая в нём, «реализованная в нём» и потому представшая перед творцом как форма вещи». Следуя данной трактовке, вполне правомочно считать, например, «идеальным» программы, базы данных и знаний. Все они воплощают в себе идеальное, характеризуя закодированный труд разработчика системы.

И программы, и базы данных, и базы знаний – суть вещи. Информационная система, построенная на основе этих компонентов – идеальное образование. Пользователю следует лишь воспользоваться данным «кладбищем» ранее живых идей, раскодировать процесс овеществления творческого процесса и соразмерить полученный в сознании результат с идеями, которые преследовал создатель этой «вещи». Сознание пользователя «причащается» к «идеальному», воплощённому в этих программно-информационных средствах.

Данная позиция, несомненно, последовательно приводит к *концепции сильного искусственного интеллекта*. Согласно данной концепции, компьютер (компьютерная система) фактически мыслит. Например, база знаний – суть «мысль». Посредством такого понимания «мысли» удобно, например, обосновать «существование» теоремы Пифагора – типичный пример метафизики объективно-нереального мира (Г.Фреге). Такой подход к идеальному в контексте проблематики ИИ предполагает следующую рисует схему: человек, вовлечённый в процесс рапредмечивания овеществлённой мысли, подчиняется внешнему, ему не принадлежащему, отчуждённому продукту чужой мысли. Если его творческие способности достаточны, то он способен преодолеть то «всеобщее» и «предельное» знание, которое заключено в информационных массивах. Если нет – остаётся довольствоваться тем, что ранее «опредметили» творцы – авторитеты. Типичная тоталитаристская схема. Свободы нет. Идеальное исчисляемо. Сознание человека «причащается» к

¹ Анализ данной проблемы осуществлен по работе Д.И.Дубровский. Проблема идеального. Субъективная реальность.М., 2002

«идеальному» большой, многопользовательской среды, где другие люди-пользователи оставляют свои «следы» – репрезентации знаний, идей, целей, намерений и т.п.

2. Позиция Д.И.Дубровского.

Д.И.Дубровский считает иначе. Идеальное не может быть характеристикой вещей. Иначе данная категория трансформируется в то, что принято называть «материальным». Идеальное не существует за пределами человеческого сознания. Категория идеального фиксирует *интенции* социального субъекта, включённого в коммуникативные взаимодействия посредством, в том числе и интеллектуальных информационных систем. При взаимодействии со средствами этих систем можно выделить три фундаментальные базовые интенции: 1) интенцию *опредмечивания* – формализации, репрезентации «мыслей» в программно-информационных средствах; 2) интенцию *распредмечивания* – постижения смысла языковых выражений; 3) интенцию самодвижения «содержания сознания» в сфере субъективной реальности (начиная от расслабленного ассоциирования до напряжённейшего размышления). Среди выделенных интенций именно последняя, интенция самодвижения занимает центральное место в структуре деятельной способности субъекта. Именно она связывает первую и вторую интенции. Но интенция самодвижения – сфера сугубо субъективной реальности человека.

Данная позиция, если её изучать в контексте ИИ – позиция слабого искусственного интеллекта. Нет никаких самостоятельно событийствующих сущностей типа математических теорем, алгоритмов и т.п. Нет «форм вещей», которые содержатся в «заготовках» создателя, социальных связях, т.е. в том, что уже опредмечено и отчуждено от человеческого индивида. Идеальное (и интеллект в том числе) существует лишь как актуально-деятельная способность социальных индивидов. И об искусственном интеллекте можно рассуждать не в эссенциалистском плане, который присущ первой позиции, а лишь в рамках функционалистской парадигмы.

Вывод: проблема идеального – ключевая проблема искусственного интеллекта. От решения этой проблемы зависят методологические установки понимания субъективной и объективной реальности, которая предстоит и перед человеком и перед компьютерной самоорганизующейся системой, понимания того, что обозначается термином сознание, в конечном, счете, что понимается под «знанием» – категориальным термином ИИ.

Решение (или выбор позиции) данной проблемы приводит и к проблемам экзистенциального плана. Мы видим: 1) необходимость рефлексии над внешними относительно субъекта объективированными «знаниями» компьютерной системы с порабощающим требованием рефлексии над ними; 2) свободу человека, внутри себя носящего все потенциальное многообразие и своеобразие интеллектуализированной информационной среды, не замыкающегося на примате знания перед иными составляющими субъективной реальности.

А. Тьюринг и проблема вычислимости сознания

Алла Иванова (ИС-81)

С 1950 г. А. Тьюринг работал над новой математической теорией морфогенеза. Теория была представлена серией нелинейных уравнений, характеризующих химические реакции (Тьюринг, 1952). Тьюринг был первым, кто использовал вычислительную машину для подобной работы. Некоторые авторы ссылаются на его разработки как на теоретическое основание *искусственной жизни* (И-жизни, А-life от англ. artificial life). Однако данное мнение представляется не совсем корректным. С 1980-х к концепции И-жизни прибегали многие, используя компьютеры для установления логических последовательностей теории эволюции. Морфогенез – частная область теории эволюции и к нему апеллируют только для того, чтобы раскрыть вероятные пути развития. Работу Тьюринга продолжили другие исследователи с 1970-х и сейчас она считается концептуальным базисом данной области.

Возможно, что интерес Тьюринга к морфогенезису происходил из детского любопытства – откуда возникли растения и цветы. Позднее он вернулся к тому, что так интересовало его в юности. В 1951 г. Тьюринг рассматривает вопрос, который избегал до этого – вопрос применения идеи вычислимости в квантовой механике. В интервью на радио BBC в этом же году (Тьюринг 1951) обсуждает свою работу 1950-ого года, и в данном интервью он уже не так упорно оспаривает аргумент Гёделя. Он утверждает, что в основе способа функционирования мозга лежат физические принципы квантовой механики. Тьюринг описал применение универсальной машины к «квантовому» мозгу, но заявил, что такое применение – суть имитация реального поведения:

«...[машина] должна быть такой, поведение которой предсказуемо посредством вычислений. Мы, конечно, не знаем, как производятся такого рода расчеты, и это даже оспаривалось сэром Артуром Эдингтоном, который заявлял, что из-за принципа неопределенности в квантовой механике поведение невозможно предугадать даже теоретически».

Коплэнд (Copeland, 1999) обратил внимание на это предложение в предисловии к его публикации беседы 1951 года. Тем не менее, критика Копленда предполагает связь с «оракулом» Тьюринга. На самом деле, об оракулах и упоминания не было (этих упоминаний уже не было ни в каких обсуждениях Тьюринга проблемы мыслящих машин). Здесь Тьюринг обсуждает возможность того, что, если рассматривать машину Тьюринга как *квантовомеханическую машину*, а не как классическую машину, то она не будет соответствовать действительности. Здесь не нужно искать связи с логическими работами Тьюринга 1938 г., а скорее с его знаниями квантовой механики, которые он приобрел в юности от Эдингтона и фон Неймана. Увлеченный идеями Эдингтона, Тьюринг предположил, что квантовая механика может быть физической основой «воли» (Hodges 1983, p. 63). Аксиомы квантовой механики фон Неймана подразумевают 2 процесса: унитарное развитие волновой функции, которое можно предугадать, и операцию измерения, которая непредсказуема. Поэтому к «непредсказуемости» Тьюринга нужно отно-

ситься как к процессу измерения. Значительная трудность, не преодоленная до сегодняшнего дня, заключается в том, что нет согласованной или завершённой теории того, как этот процесс измерения происходит. (Следует отметить, что «квантовое вычисление» в сегодняшнем понимании основано на предсказуемости унитарного развития и не объясняет того, как происходит процесс измерения. Тьюринг в 1953 г. пишет своему другу и сокурснику Робину Ганди, что «пытается изобрести новую квантовую механику, но это пока не совсем получается». После смерти Тьюринга в июне 1954 г. Ганди написал Нейману о том, что знал о текущей работе Тьюринга (Ганди, 1954) – Тьюринг изучал вопрос процесса измерения.

Разработки Тьюринга приобретают дополнительное значение при рассмотрении утверждения Пенроуза (1989, 1990, 1994, 1996), что в процессе измерения заключено нечто не поддающееся вычислению. Вероятно, целью Тьюринга было противоположной – найти теорию процесса измерения, которая делала бы этот процесс предсказуемым и вычисляемым, и таким образом поставить точку в гипотезе о том, что работа мозга является вычисляемой.

Немوتря на расхождения, Тьюринг и Пенроуз схожи в том, что, в отличие от большинства учёных, считают вопрос о вычислимости разума фундаментальным вопросом.

Последние открытки А. Тьюринга Родину Ганди в марте 1954 г. имели заголовок «послания из невидимого мира» с намеком на Эдингтона, на новые идеи теории относительности и квантовой физики (Hodges, 1983). Эти «послания» показывали богатство идей, которыми был увлечен А. Тьюринг в последние дни своей жизни. Однако все эти идеи были полностью утеряны, если не принимать в расчет намеков.

Проблема естественных видов в искусственном интеллекте

Максим Королев (АП-81)

В искусственном интеллекте крайне актуальной оказалась проблема естественных видов¹. Пример. Всем известно, что лимон – это небольшой жёлтый фрукт. Но это знание не задаёт четкого определения «лимона». Лимоны отличаются от других фруктов. Как именно отличаются – сказать сложно. Определённость суждений о лимонах разрушают и генетики – они научились разводить большие синие лимоны. Конечно, можно считать лимоном столь экзотический продукт генной инженерии, но это расходится с обыденным опытом. Как такой «лимон» репрезентировать в компьютерной системе?

Пример с лимоном, иллюстрирующий сложность проблемы естественных видов, привел Хилари Патнэм в статье «Возможна ли семантика?» (1970 г.) с целью обозначения функционалистского подхода к репрезентации подобного рода нечётких понятий.

Исторический шлейф проблемы естественных видов тянется от родовидовой классификации Аристотеля через Декарта и Локка. Классическая формальная логика данную проблему решала с позиции, которая характеризовала слова и выражения языка как знаки идей. Эти знаки выражают мысли.

¹ McCarthy John (1995), What has AI in Common with Philosophy? <http://www-formal.stanford.edu/jmc/>

После Готтлоба Фреге, «второго Аристотеля», проблема естественных видов стала рассматриваться с позиции, согласно которой слова – это способ выражения смысла, который задает значение слова. В свою очередь, слово обозначает значение. Однако понятие «смысла» оказалось неудобным для логики – оно не поддается чёткому формальному представлению и анализу. И после Фреге аналитической философия стремилась представить отношение между языком и реальностью как прямое отношение, не опосредованное ни смыслом, ни иными ментальными сущностями.

Апогея данная доктрина достигла в функционализме Х. Патнэма. Параллельно с ним, ряд других философов (К. Доннелан, С. Крипке, Д. Каплан и др.) выдвигают подобные идеи. Эти идеи были объединены под общим названием «новая теории референции». Основополагающий тезис теории: *референция категорий языковых выражений устанавливается без посредничества смысла*.

Патнэм данный тезис применил к проблеме естественных видов. Термины естественных видов – это слова, именующие природные вещества, животные, растения, физические величины (например, «вода», «тигр», «лимон», «электричество» и т.д.). К этой категории языковых выражений относится большинство научных терминов.

Х. Патнэм выдвигает два основных *аргумента против традиционной трактовки значения терминов нечётких видов* (в более широком плане они составляют подкласс общих терминов):

1) В традиционном подходе невозможно найти значение термина естественного вида. Значение термина, например, «золото» формулируется в виде дескрипции «тяжелый, неокисляющийся, твердый металл желтого цвета» или в виде другой подобной дескрипции. Предполагается, что любой объект, который обладает перечисленными свойствами, является золотом. Однако естественный вид может включать аномальные члены, которые не удовлетворяют установленной конъюнкции свойств, но тем не менее принадлежат к данному естественному виду. Например, «зеленый лимон (который так никогда и не пожелтеет) все же является лимоном, тигр с тремя ногами – все же тигр, а золото в газообразном состоянии не перестает быть золотом». Другая трудность, с которой сталкивается традиционная теория, состоит в невозможности указать такую конъюнкцию свойств, которая позволила бы выделить естественный вид единственным образом. Нет никакой гарантии, по мнению Патнэма, что не будет обнаружено такое вещество или животное, которое полностью удовлетворяет дескрипции свойств соответствующего естественного вида, но которое, тем не менее, в силу своей внутренней природы не принадлежит к данному виду.

2) Традиционная теория значения опирается на два допущения, которые не могут быть одновременно истинными. Первое допущение устанавливает, что понимание значения слова связано с пребыванием в определенном ментальном (или психическом) состоянии. Второе допущение связано с тем фактом, что интенционал слова задаёт его экстенционал (интенционал образует необходимое и достаточное условие для вхождения объекта в экстенционал).

Отказавшись от «смысла» как механизма, определяющего и систематически обеспечивающего референцию имен собственных и терминов естественных видов, сторонники новой теории референции предлагают новый способ определения экстенционала. Основной тезис нового подхода к решению

проблемы: *референция выражений устанавливается благодаря внешним нементальным факторам.*

Согласно Патнэму, в установлении референции терминов естественных видов участвуют два фактора: социальный и природный.

1. Социальный фактор. Действие социального фактора описывается с помощью «социо-лингвистической гипотезы». Патнэм рассуждает следующим образом. Согласно традиционной теории значения человек понимает некоторое слово, если усвоил его смысл. Но учитывая, что смысл слова – крайне сложное ментальное образование, следует признать, что лишь крайне небольшое число людей владеет смыслами и, следовательно, понимает слова. Превалирующее большинство носителей языка можно вообще обвинить в том, что они не понимают те слова, которые используют. Но такое предположение абсурдно, поскольку для того, чтобы понимать и использовать слово, совсем необязательно в полном объеме знать его смысл. Вполне достаточно, считает Патнэм, положиться на экспертов, которые владеют этим смыслом, а кроме того, владеют методом распознавания естественных видов. Благодаря экспертам, таким методом начинает владеть весь языковой коллектив, но метод доступен не всякому индивидуальному представителю коллектива. Из этого следует, что в лингвистическом сообществе существует разделение труда, связанное со знанием и использованием разных аспектов «значения» слов. Оно опирается, в принципе, на обычное разделение труда в обществе.

2) Природный фактор. Этот фактор состоит в том, что экстенционал термина естественного вида «частично устанавливается внешним миром». Допускается, что любой естественный вид предполагает наличие у его членов общей внутренней природы (или сущности), выражающейся в общей внутренней структуре, общих существенных свойствах или общих объективных законах, управляющих поведением или развитием членов данного естественного вида.

Насколько удачной представляется очередная попытка избавиться от «смысла» и связанных с ним «менталистских» допущений?

1. «Новая теория референции» способствовала формированию более адекватного и глубокого представления о том, как функционирует язык и как осуществляется взаимодействие носителя языка с окружающим миром.

2. Данная теория содержит решение для наиболее простого случая, а именно – для случая обычного употребления имен и повествовательных предложений, и не предлагает никаких путей решения этой проблемы в более трудных случаях для семантики – в случае, например, косвенной речи.

3. Конструктивные идеи данной теории связаны с сильными допущениями, истинность которых вовсе не очевидна.

Вывод: Традиционную теорию значения отменить невозможно. При построении интеллектуальных систем необходимо совмещать «менталистский» и «нементалистский» подходы к решению, в частности, проблемы естественных видов. «Смысл» невозможно изгнать из способов обозначения терминов естественных видов, входящих в концептуальный либо в реальный состав моделируемого мира.

Искусственный интеллект и стоическая эпистемология

Татьяна Косинова (ЗИ – 81)

Философия и искусственный интеллект, считает Джон Маккарти, имеют много общего¹. Искусственный интеллект нуждается во многих идеях, которые до сих пор изучались исключительно философами. Искусственный интеллект, сопоставимый с человеческим, требует *встраивания* в компьютерные программы философских положений. Например, для того, чтобы робот обладал интеллектом, сопоставимым с человеческим и обучался на собственном опыте, требуется, чтобы он обладал «псевдосознанием», в рамках которого осуществляется организация фактов. Философские положения, встраиваемые в компьютерную программу должны иметь главным образом, *эпистемологический характер*. Программа должна основываться на понятиях – что есть «знание» и как «знание» приобретается. Если программа призвана рассуждать о том, что она может делать и что не может, её разработчики вынуждены воспользоваться положением о *свободе выбора*. Если программа должна выполнять метауровневые суждения об этих рассуждениях, необходима концепция об *осознании свободы выбора*. Если необходимо застраховаться от аморальных действий со стороны программы, разработчикам следует встроить в неё и соответствующее *этическое положение*.

Однако какую эпистемологию имеет в виду Джон Маккарти? Ведь «многие философские проблемы принимают новые формы, когда рассматриваются под углом зрения создания робота. Одни философские подходы полезны, другие – нет», – утверждает он. Большинству современных работ по ИИ вообще не нужна философия. Системы ИИ не создаются ни для целей самостоятельного функционирования в мире, ни для целей самостоятельной выработки представлений о нём. Программист идет впереди философии и встраивает в программы свои ограниченные мнения. Шахматной программе философия не нужна. Не нужна философия и экспертной системе типа MYSIN, рекомендующей лечение от заразных инфекционных заболеваний – она разрабатывается вне каких-либо представлений о динамике «знаний». Возможности шахматных и MYSIN-подобных программ ограничены востребованностью в них здравого смысла и философии. Многие другие приложения требуют иного подхода. К примеру, роботы, исполняющие желания своих владельцев, должны рассуждать о желаниях.

Осуществляемый в работе сравнительный анализ призван показать близость концептуальных положений Дж. Маккарти, которые он предлагает встраивать в компьютерные программы (обозначены буквой «М») с эпистемологическими положениями *стоической школы* (обозначены буквой «С»). В основе анализа стоической доктрины (Древней Стои) лежат работы А.Ф. Лосева, А.Н. Чанышева и А. Столярова.

М.1. Программы искусственного интеллекта должны быть основаны на здравом смысле. Следует примирить науку и здравый смысл. Есть

¹ Основой сравнительного анализа философских оснований ИИ Дж.Маккарти и стоической доктрины является работа McCarthy, John (1995) Artificial Intelligence and Philosophy (<http://cogprints.ecs.soton.ac.uk/archive/00000420>)

атомы, но есть и стулья. Репрезентация «мира» в компьютерной системе должна осуществляться на таком уровне, на котором сами люди действуют без понимания основ теоретической физики. Причинно-следственные связи следует выявлять на уровне очевидных данностей. Закономерности, устанавливаемые между этими «смыслами» должны лежать в основе логического вывода, которым пользуется робот, рассуждающий о возможных последствиях своих действий. Миру здравого смысла нужен язык описания объектов, отношений и их динамики, который существенно отличается от математизированного языка физики и техники. Ключевое различие проявляется в неточности информации.

С.1. Стоической эпистемологии характерен принципиальный отказ как от «эйдосов» Платона, так и от «форм» Аристотеля. И эйдосы и формы приводят к голым, далёким от практики спекулятивным положениям. Существование эйдосов и форм не гарантирует постижения истины. Путь к познанию начинается с чувственного восприятия. В связи с этим стоики предлагают иную смысловую конструкцию – «лектон». Физический мир – мир тел – проявляется на феноменальном уровне. Цель введения «лектонов» в способ ориентирования в феноменальном мире – каким-то образом связать явления в закономерные соотношения. Лектоны корреспондируют телам и знакам, которыми люди эти тела обозначают. Закономерности носят детерминистический характер (будущее определяется прошлым) и фаталистический характер (прошлое определяется будущим, т.е. будущее, целевое состояние как бы «притягивает» прошлые события). Возникает логика, которая сегодня называется «логика высказываний». Нет никаких родо-видовых сущностей, характерных для силлогистической логики Аристотеля. Нет никакой интуитивно-художественной образности эйдетических конструкций. Есть только «голые факты» и логические законы их связывания. И есть единый космический закон – Логос, предопределяющий поведение людей и всех вещей в мире. Человеческий разум есть лишь частное проявление всекосмического логоса – единственного подлинного гаранта объективности и целостности мироздания. Задача – познать закономерности космического Логоса для цели самосохранения индивида.

М.2. Мышление следует понимать в динамическом смысле. То есть интеллектуальная система изначально обладает лишь небольшим набором способностей. Можно даже подумать, что в исходном состоянии она вообще ничем не обладает. Но в дальнейшем, посредством самообучения (самоорганизации, самонастройки и т.п.) потенциально заложенные способности актуализируются. В процессе систематизации фактов происходит наращивание интеллектуальных функций.

С.2. Стоики – номиналистические сенсуалисты. Первичный источник познания – чувственное восприятие. Мышление (изначально) не имеет другого материала, кроме содержания ощущений: душа при рождении подобна чистому листу папируса (*tabula rasa*). Опыт наносит на изначально пустой лист свои «письмена». Из опытных данных посредством индукции или аналогии образуются «эмпирические общие представления», сохраняющие общие признаки ряда схожих представлений. Предметность этого класса представлений не может быть воспринята непосредственно. Поскольку опыт в них получает первичное оформление, эти начальные «понятия» могут пониматься как «умственные, разумные представления». Прочие «общие представления» или, скорее, «понятия» образуются «искусственно» из ряда «по-

стижений»; они означиваются; совокупность определенным образом соединенных «постижений» составляет «теорию» как «систему из постижений» и сообразованную с практической жизненной целью. В основе этого процесса лежит, таким образом, внутренняя способность переходить от одних «общих понятий» к другим и создавать на их основе новые понятия.

М. 3. Вера и намерения – суть объекты, которые могут быть формально описаны.

С.3. Согласно стоикам, мысли, желания, верования и т.п. телесны.

М.4. Причиной приписывания ментального свойства программе выступает функциональный принцип соответствия данного свойства определенному поведению.

С.4. Мысль и истина, выражаемая мыслью, обладают статусом телесной предметности, реально наличной в определенный момент времени и при определенных условиях. Ментальные свойства (истинность, интеллектуальность) – суть характеристика бестелесных смыслов – лектон, которые в данном случае выступают в роли функций, аргумент и результат которых определены на телесных объектах – «мыслях».

М.5. Необходимо использовать аппарат нечетких понятий. Следует ослаблять критерии означивания терминов и, по мере возможностей, применять нечеткую математическую логику.

С.5. Мир сложен и хаотичен. Необходимы предсказания. Развиваемая стоиками мантика – искусство гадания – это, по сути, осуществление вероятностного выбора исходя из ряда характерных признаков, знаменований, знаков. Современные алгоритмы оптимального поиска в пространстве решений в условиях неопределенности (особенно при поиске глобального максимума функции) напоминают мантические пророчества.

М.6. Свобода и детерминизм совместимы. Детерминированный процесс, определяющий последовательность действий индивида, включает и способ оценки последствий выбора. Информация о выборе присутствует в сознании и, поскольку выбор и последствия наблюдаемы, относительно них можно выносить суждения.

С.6. Свобода – это познанная необходимость. Для осуществления свободного акта необходимы знания о мире и его закономерностях.

М.7. Самосознание заключается в оперировании суждениями о сознании, сохраняемыми в памяти.

С.7. Каталептические восприятие схватывает не только предмет, но и душу. «Схваченное» при «постижении» настолько очевидно, что принуждает человека к согласию. «Схваченному» предмету (означаемому) соответствует серия знаков (означающих). Серия знаков откладывается в памяти, образуя «внутреннюю речь». Возникает возможность суждений о восприятии, предполагающая активность субъекта, главным условием которого на предыдущем этапе – этапе каталептического восприятия выступала пассивность. Суждения о воспринятом, по всей видимости, можно трактовать как логико-семиотический срез самосознания.

Сравнительный анализ раскрывает преемственность современных проблем искусственного интеллекта (сформулированных Дж. Маккарти) со стоической доктриной. Историко-философское осмысление стоического наследия, с целью встраивания «правильных философских положений в компьютерные

программы», представляется важным для философии искусственного интеллекта в целом и, в частности, для методологии робототехники.

Учитывая космологическую установку стоической доктрины о мире как «живом существе», можно подытожить суждения о взаимосвязи философии и ИИ. Заключение имеет уже не узко-эпистемологический, а общемировоззренческий статус, согласно которому мир – это взаимосогласованное и пронизанное всеобщей симпатией единство: *люди, роботы, животные, растения и весь мир в целом – суть единое интеллектуально-телесное органическое целое.*

Зомби и искусственный интеллект

Татьяна Кураева (С-85)

Под зомби принято понимать либо голливудских «звезд» из фильмов ужасов либо гаитянских мертвецов, которые функционируют, как живые существа, но при этом не обладают ни душой, ни свободой воли. В последнее десятилетие возникла ещё одна разновидность зомби. Д. Чалмерс их назвал «философскими зомби». Реанимация «зомби» со стороны философов обусловлена рядом факторов – развитием технологий СМИ, генной инженерии, когнитивных наук, компьютерных наук и пр. В данном ряду наиболее значимыми представляются достижения искусственного интеллекта.

К зомби в контексте проблемы философии сознания и философии искусственного интеллекта обращаются многие выдающиеся философы современности, такие как Bringsjord S. (1999), Chalmers, D. J. (1996, 2002), Dennett, D. C. (1995), Kirk, R. (1974, 1977, 1994, 1999), Nagel, T. (1998), Perry, J. (2001), Stalnaker, R. (2002). К сожалению, в отечественной литературе данная тема не получила должного освещения.

Освещение проблемы зомби в контексте философии ИИ представляется полным при столкновении противоположных точек зрения двух философов: 1) Тода Моуди (профессор философии в университете Сант-Джозефа, Филадельфия), который предположил один из подходов к теоретическому обоснования возможности зомби – бессознательных существ, поведение которых, однако, не отличается от поведения людей; 2) Дж. Маккарти, выступившего оппонентом Моуди с позиции философии искусственного интеллекта.

1. Зомби и ИИ (по Т. Моуди)

Тод Моуди в статье «Беседы с зомби» проводит мысленный эксперимент. Цель эксперимента – понять возможность того, как компьютерная система (например, робот) или некое гипотетическое существо может проявлять все признаки сознательного человека, но при этом сознанием не обладать. Предполагая возможность функционального описания сознательной деятельности (в контексте парадигмы функционализма), Т. Моуди считает не лишённым смысла предположение о том, что могут существовать неодушевленные существа. Эти существа демонстрируют то, что в рамках обыденного сознания понимается под словом «зомби». Функциональное описание представимо вычислениями и, в принципе, реализуемо в компьютерной среде – о зомби имеет смысл говорить в технологическом контексте.

Проводя эксперимент, Т. Моуди предполагает мир, полностью соответствующий нашему – Землю зомби. Хотя этот мир во всём как наш, в нём есть одна деталь: «люди» не обладают сознанием. Они образуют социумы, обща-

ются, у них есть язык. Их наука методологически подобна человеческой науке, по крайней мере, физика и математика те же. Во многом подобны и философские суждения. Однако любой поведенческий акт этих людей не сопровождается осознанием этого акта. Такие существа и называются «зомби».

Т. Моуди предлагает этот мысленный эксперимент с целью теоретического обоснования критерия отличия зомби от людей – словарь «жителей» Земли зомби не будет содержать ментальных терминов. При этом Т. Моуди по сути соглашается с принципом «сознательного инэссенциализма» (*conscious inessentialism*) или, иначе, с «принципом несущественности сознания», который был предложен Дж. Полджером и О. Фланаганом.

Данный принцип гласит, что для реализации поведенческого акта совершенно не существен опыт осознания этого акта. Более того, любое поведение может протекать без какого-либо его осознания. Несущественность осознания присуща даже такой форме деятельности, как интеллектуальная, хотя традиционно мышление считается неразрывно связанным с осознанием актов мышления. Из этого следует, что поведение и сознание могут существовать независимо друг от друга. Если принцип корректен, то зомби, безусловно, возможны.

Авторы данного принципа правоту его обосновывают следующими положениями: 1) Эмпирическим путем доказывается отсутствие каких-либо корреляций между явлениями сознания и нейродинамическими процессами мозговой деятельности; 2) Многие представители когнитивной психологии считают, что большинство процессов, происходящих в сознании, совсем не значимы для познавательной деятельности; 3) Когнитивная наука в целом склоняется к эпифеноменализму; 4) В контексте исследований искусственного интеллекта авторитетно звучит мнение Фланагана: «Принимая во внимание, что большинство скептически настроенных людей сильно волнует проблема искусственного интеллекта, и, в основном, то, что машин можно наделить сознанием, то следует считать их волнения напрасными. Вычислительный функционализм показывает, что и наш разум в сознании не нуждается!».

Интерес представляет иллюстрация принципа «несущественности осознания» в футурологическом проекте искусственного интеллекта, который представлен в широко известном фильме «Терминатор». Т. Моуди критикует мнение о том, что Терминатор – это автомат, подобный зомби. В фильме часто повторяется сюжет «вид мира внутренним зрением робота». Часть «образа» заполнена изображениями процедур сканирования: расстояние до цели, скорость и т.п. Такой «взгляд» для человека имеет смысл: человек может представить «внутреннее наблюдение за собой» – мы смотрим на мир, в котором мы сами же и находимся. Однако для зомби «внутренний наблюдатель» невозможен. Поэтому робот – Терминатор – никоим образом не может быть зомби.

Продолжая линию рассуждений Т. Моуди, можно заключить, что система искусственного интеллекта иного типа – вполне вероятный претендент на звание «зомби». В таких системах ментальные процедуры возможны вне какого-либо их осознания.

По иному считает Дж.Маккарти: компьютерная система принципиально не может называться интеллектуальной, если в ней отсутствуют программы реализации процедур осознания себя в составе мира – внешнего и внутреннего.

2. Зомби и ИИ (по Дж. Маккарти)

Дж.Маккарти в статье «Зомби Тода Моуди» отказывает правдopodobию мысленного эксперимента Т. Моуди. С позиции ИИ сознание следует рассматривать как совокупность взаимодействующих процессов. Сознание – не некий идеалистический продукт. Сознание играет вполне определённую роль в составе человеческой деятельности. Принцип эпифеноменализма, к которому апеллировал Т. Моуди – отнюдь не методологическое обоснование конструктивной деятельности по построению интеллектуальной системы, например, робота.

Чтобы вести себя подобно людям, зомби должны обладать тем, что Моуди мог бы назвать псевдосознанием («как бы» сознанием). Однако псевдосознание должно приносить пользу, хотя бы для выполнения моторно-двигательных функций, не говоря уже об интеллектуальных функциях. Поэтому псевдосознание должно стать частью всех возможных процессов ментальной деятельности, включая, например, речевые способности зомби что-то отвечать на задаваемые ему вопросы.

Для интеллектуальных систем совершенно несущественен принцип «несущественности осознания» Т. Моуди.

То есть псевдосознания у зомби быть не может (в силу определения зомби). Невозможно отделить поведенческий акт от акта осознания этого поведенческого акта.

Зомби Т.Моуди невозможны!

Роботы же возможны. Только невозможна их реализация в рамках чистой парадигмы функционализма. Роботу требуется псевдосознание.

Псевдосознание должно обладать параметрами человеческого сознания. Реализация всех параметров, конечно, невозможна. Зачем, например, моделировать человеческие эмоции? Это не приносит пользы и затруднено. Многие аспекты интеллектуального поведения не связаны с уровнем их осознания. Сознание – это некоторая часть человеческой памяти (человек может сознавать что-то, будучи в памяти, а состояние, в котором человек себя не помнит, обычно называют бессознательным состоянием). Такая связь сознания с памятью и предопределяет структуру псевдосознания робота. Сознание – часть памяти.

Структура псевдосознания робота.

Делится на основное (базовое) сознание и самосознание.

1. Основное сознание:

1.1. Пропозиции.

Пропозиции «основного» сознания – это суждения об объективном мире, а не о мышлении. Данные пропозиции не подвергаются оценке, все они лежат в области памяти, амбивалентной с точки зрения суждения об истинности или ложности состояний мира.

1.2. Образы сцен и объектов. Это – образы, которые представляются в текущее время либо образы, ранее сохранённые в памяти. Образ – это комплексное понятие: оптический образ (трехмерный) рассматривается совместно с акустическим. Акустические образы речи фильтруются и образуются звуки, которые составляют язык.

2. Самосознание: Самосознание – это те ментальные процессы, которые требуются роботу для организации разумного поведения. К таким процессам Маккарти относит следующее:

2.1. Хранение протокола физических и интеллектуальных событий, которые позволяют роботу обратиться к прошлым убеждениям, наблюдениям и действиям.

2.2. Анализ структуры преследуемых роботом целей и формирование о ней суждений.

2.3. Робот может *стремиться* выполнить некоторое действие. Позже может быть произведена оценка. Она позволит заключить, что некоторые возможности не соответствуют намерениям. Это требует построения средств анализа интенций.

2.4. Следует анализировать способ возникновения убеждений. Наиболее важные из них система будет получать в ходе немонотонных суждений, в силу чего данные убеждения могут быть сомнительными. Следует вести классификатор убеждений. Такой классификатор позволит пересматривать наличные убеждения с учётом тех убеждений, которые ранее имели место.

2.5. Не только иерархия убеждений, но и другая вспомогательная информация должна быть представлена в виде вербальных выражений. Система должна быть способна ответить на вопросы «Почему я верю, что p ?» и «Почему я не верю, что p ?»

2.6. Целостное представление об объекте достигается путем *трансцендирования* над представлениями об этом объекте. Если представления об объекте – суть текст, то трансценденция в данном случае означает контекст. Контекст должен быть формализуемым. Трансцендирование – важнейшая форма собственно человеческой интроспекции. Это утверждают многие критики ИИ (такие, как Дрейфус). Формализация контекста позволяет репрезентировать в компьютерной системе данный важнейший показатель собственно человеческого самосознания.

2.7. Необходимо представление знаний о том, какие цели робот может выбирать. Способность осознавать свой собственный выбор конституирует свободу.

В целом данные требования, по мнению Маккарти, характеризуют наиболее существенные черты человеческого сознания. Моделирование человеческих эмоций роботу не требуется. В принципе, Маккарти приводит достаточно убедительные доводы в пользу возможности реализации псевдосознания робота.

Что же касается возможности зомби, который мог бы функционировать как сознательное существо, но не обладающее ни сознанием ни вседосознанием, Маккарти придерживается отрицательного мнения. «*Зомби Моуди – это плод выдумки Моуди!*» – заключает Дж.Маккарти. Робот («Терминатор» из примера Т. Моуди) возможен, если использовать концепцию псевдосознания Дж. Маккарти. Невозможны роботы-зомби, у которых отсутствует словарь ментальных терминов – основная составляющая механизма «псевдосознания».

Вывод

Заключение по проблеме «зомби» хочется осуществить с учетом позиции другого оппонента Т. Моуди – Чарльза Тарта (Институт трансперсональной психологии, США). Проблему зомби он рассматривает уже не с позиции ИИ, а с иной, предельно крайней позиции – с высоты опытов самосознания *Я*, которые характерны для восточных медитаций. В противовес тезису Дж. Маккарти («Зомби невозможны»), Ч.Тарт выдвигает тезис *«Все мы зомби! Но мы можем стать сознательными»*. Этот тезис обстоятельно доказывается. И предлагается подход к преодолению зомбированного состояния. Чтобы человек-зомби перестал быть таковым, необходимо воспитывать в себе самом опыт обращения внимания на внутреннее содержание сознания. Чарльз Тарт предлагает для этого руководствоваться принципом «расширенного сознания». Такой принцип с учетом принципа Фланагана («принцип несущественности осознания») удобно было бы назвать *«принципом расширенности осознания»*.

Должна быть иная наука, не та, которая прогрессирует сейчас и опирается, например, на некритические техногенные представления о возможности истощения сознания функционалистскими схемами. Если человек не будет самостоятельно развивать способности сознания, осознания, самосознания, то в ближайшем будущем он будет иметь не науки о человеке и обществе, а науку о зомби и зомбированных обществах.

В ходе критики проблемы зомби в контексте искусственного интеллекта сошлись западная (Маккарти) и восточная (Тарт) традиции. Какая из них правомерна?

Если с технологической позиции «медитационный подход» вызывает вполне понятный скепсис, то с позиции культурологической, он, несомненно, заслуживает внимания: созданию псевдосознания робота («Терминатора») должно предшествовать расширение горизонтов самосознания разработчика и понимание ответственности за собственные «незомбированные» решения.

Интуитивистские ориентиры моделирования смысла (А. Бергсон) Яна Маликова (С-81)

В исследованиях ИИ наметились два пути прочтения трудов Анри Бергсона. Первое направление связано с метафизическими представлениями на тему творческой эволюции, роли и месте искусственного интеллекта в ноо-космологическом пространстве Вселенной (фёдоровское движение, ноосферный подход, глобалистские проекты Международной академии информатизации и др.). Второе направление, ориентированное на формирование научной методологии искусственного интеллекта, просматривается в развитии нетрадиционных на сегодняшний день способов представления информации в интеллектуальных системах, в первую очередь, в системах функционального воспроизводства феноменов человеческого понимания. «Творческая эволюция» интеллектуальных систем в ракурсе второго направления условно характеризуется усложнением инструментария информационной системы: 1) «данные+алгоритм»; 2) «знания+эвристика»; 3) «смыслы+интуиция».

Интуиция, по Бергсону рассматривается как идеальный вид познания. Это созерцание, независимое от практики и потому дающее нам адекватное познание реальности.

Сознание по Бергсону представляется в виде непрерывно изменяющейся творческой реальности, это своего рода поток, в котором мышление (одна из основных функций сознания) суть лишь поверхностный слой, подчиняющийся потребностям практики и социальной жизни. Именно неразрывная связь интеллекта и науки с практикой не позволяет достичь «чистого» созерцания, свободного от практических интересов, методов и точек зрения.

Интеллектуальное познание лишь кажется чисто формальным образованием, «в познании этом нет содержания: это форма без материи». По мнению Бергсона интеллект создан лишь для действия, для непосредственной практики, но никак не для познания и теории. Он способен постичь лишь отношения между вещами, но никак не саму вещь, в то время как интуиция способна давать нам познание природы самих вещей.

Однако Бергсон отнюдь не призывает к отказу от интеллектуального познания и рациональности, напротив, он подтверждает тот факт, что интеллект обладает преимуществами, которые никогда не могут быть достигнуты интуицией.

В работе «Материя и память» Бергсон показывает, что материя и сознание, тело и рассудок это явления, реконструированные самим рассудком из фактов непосредственного опыта, той первичной интуиции, которая открывает нам нераздельную движущуюся непрерывность. Основу познания составляет чистое созерцание, но в реальном познавательном процессе оно всегда взаимодействует с памятью. «Я называю материей совокупность образов, а восприятием материи – те же самые образы в их отношении к возможному действию одного определенного образа – моего тела». Когда Бергсон говорит, что образ может существовать и не будучи воспринятым, он объясняет, что для образов быть и быть сознательно воспринятым – это состояния различающиеся лишь по степени.

Память по Бергсону – это «способность сохранять и вызывать прошлые восприятия, напоминать нам то, что предшествовало и следовало, внушая таким путём нам самое полезное решение».

Память даёт жизнь прошлому в настоящем, пронизывая его. Без памяти у прошлого не было бы реальности, и поэтому не было бы прошлого. Именно память, со своим желанием все соотносить, делает прошедшее и будущее реальным, и тем самым создает длительность и время. Интуиция постигает это смещение прошлого и будущего; для интеллекта же они остаются внешними.

А. Бергсон выделяет два вида памяти: память-привычку (или память тела), основой которой служат физиологические мозговые процессы, и память-воспоминание, или память духа, не связанную с деятельностью мозга.

Память тела образуется из совокупности сенсомоторных систем, организованных привычкой. Применительно к ИИ данный вид памяти можно трактовать, например, как сохранение в памяти «опыта» ориентирования робота в физическом мире.

Память духа – это идеальная сфера – сфера воображения, прогнозов, мечтаний и др. Именно память духа отвечает за смысл действия. В каждый момент времени смысловая сфера окружает единственную точку – точку, в

которой осуществляется сенсомоторный акт (мы продолжаем наш пример с роботом). И именно в памяти духа осуществляются процедуры интуитивного постижения «смысла», который можно трактовать в стиле голографической метафоры – опорная волна – совокупность знаков, обозначающих сенсомоторные акты, предметная волна – всевозможные траектории предполагаемых актов. Голографическое представление характеризует целостное, интуитивное постижение – в одной конструкции совмещены и прошлые и будущие состояния ИИ-системы. И принципиальным вопросом становится вопрос об организации «памяти духа».

Попытка применения концептуального наследия А.Бергсона к современным проблемам искусственного интеллекта (к нечётким понятиям, механизмам смысловой интеграции «знаний», построению теоретических оснований голографических средств моделирования данных и др.) представляется перспективной с позиции конструктивной (научно-методической) её реализации. Сам А.Бергсон строит чёткий рациональный дискурс об иррациональном, задавая своеобразный культурологический тип разработчика перспективных интеллектуальных систем, реализующих механизмы интуитивного постижения.

Искусственный интеллект и святоотеческая традиция

Максим Розов (Р-81)

Во многом искусственный интеллект (ИИ) базируется на философии анализа (Рассел, Витгенштейн, Шлик и др.). Цель данной философии изначально состояла в изгнании всякого рода метафизики из науки, особенно понятий «душа», «сознание», «разум». Надо тщательно выбирать языковые выражения, чтобы вести дискурс. Возникли физикализм, операционализм, функционализм как способы ведения взвешенного научного дискурса, включая и дискурс «метафизических» понятий. Однако возникла парадоксальная ситуация – философско-аналитический подход породил свободу оперирования понятиями. Возникла ситуация не «чистоты» языка науки, а напротив, загрязнения быденного языка, которым пользуется человек. Например, простейшие компьютерные программы типа игры в шахматы стали называть «интеллектуальными», «искусственным разумом». Подобная проблема возникла в когнитивной психологии, когда между компьютерной наукой (ИИ) и когнитивной наукой образовался своеобразный лингвистический симбиоз. Ибо для того, чтобы в компьютерной системе искусственным способом воспроизвести ментальные свойства (человеческое восприятие, память, язык и мышление) требуется концептуальная пропедевтика – подготовка человека к метафорическому присвоению техническому устройству и его составным частям «знаний», «мыслей», «мотивов» и т.п. Например, Дж.Маккарти, руководствуясь положениями «здорового смысла» присваивает термометру «убеждения» либо компьютерной программе возможность реализации ментального свойства «мочь». С другой стороны, развитие искусственного интеллекта, предоставляя новые пути понимания человеческого мышления, с необычайной силой возвышает роль вычислительной методологии в рамках когнитивной психологии. В результате компьютеризация в психологии становится доминантой доктриной, нивелируя иные концепции.

По всей видимости, необходим новый этап «чистки языка» науки и технологии, особенно в области ИИ. Но путь «очищения» следует повернуть в обратную сторону относительно изначального направления движения философии анализа. Если имеется необходимость оперировать «метафизическими» понятиями при построении компьютерно-когнитивных технологий (ИИ, НЛП, когнитивно-лингвистического тестирования и т.д.), то тогда в концептуально-языковой аппарат их разработчиков следует вводить термины, адекватные ментальным референтам.

С учётом богатой традиции святоотеческого опыта в выработке понятий «разум», «сознание», «знание» и т.п. — опыта проникновенных смысложизненных бесед с личным Богом — очевидна возможность его привлечения для целей ИИ. Остановимся на анализе некоторых понятий, которые святые отцы обозначали словом «разум» («ум»).

Максим Грек

Философские сочинения написаны в виде «беседования» ума и души. Такой «диалог» позволил выразить полярными категориями борьбу различных начал и несовместимость противоположных душевно/интеллектуальных установок. Максим Грек следует христианской трехчастной традиции, в рамках которой ум играет достаточно неопределённую роль и выделяет в человеке три начала: плотское, душевное и духовное. Уму, тем не менее, у М. Грека отводится главенствующая роль. Ум — «кормчий души» и играет роль управляющего по отношению к душе и телу. Однако ум нуждается в просвещении, которое неотделимо от нравственного совершенствования — нравственные усилия позволяют «мысль от плоти обуздать». Такой результат, по Греку, связан не только с моральным, но и с познавательным опытом: чтобы постичь истину, надо жить в ней. Необходимо просвещение не только ума, но и сердца. Если сердце «суето», то никакое постижение истины (следовательно, и спасение) невозможно. Сердце в данном случае — традиционный символ цельности духовной жизни. У Максима Грека, как это принято в христианской традиции, достичь чистоты сердца и ума позволяет любовь, которая «превыше всего» — любовь к Богу и ближнему. Мир человеческой души — крайне сложный, он отражается в полисемантической образности, которую использует мыслитель, выражая неизмеримую глубину душевного мира. Он уподобляет душу кораблю, плывущему по бурным волнам житейского моря к гавани «небесного пристанища»; сравнивает ее с воском, умягчающимся от теплоты духовной; с землей плодоносящей или засыхающей по высшей воле. Охотно прибегает он к названию души зеркалом, весьма распространённому в средневековой литературе.¹ По сути, интеллект («ум») и психика («душа») тесным образом взаимосвязаны. Выделение интеллекта как некоторой отдельной сущности приводит к абстракции, лишь отдаленно напоминающей реальность. И дальнейшее изучение реальности, к примеру, сознания, с позиции данной абстракции и, более того, последующее отождествление реальности с абстракцией, представляется фикцией, рождённой некритичным техницистским настроением.

¹ Громов М.Н. Максим Грек. М., 1981

Максим Исповедник

Преподобный Максим Исповедник утверждал, что душа трехчастна — в ней есть начало вожделеющее (чувства), разумное и яростное (волевое).¹

Важное место Максим Исповедник уделяет разуму. Только с помощью разума человек способен осуществить практику аскезы — преодолеть «страстное отношение к миру и плоти». Разум помогает человеку понять необходимость богопознания, создаёт условия устремления к его постижению и руководит им на этом пути, удерживая от земных соблазнов. Именно разум может привести и приводит человека к совершенной вере.

Филипп Пустынный

Уму свойственны три силы — память, воображение, мышление. Автор «Диоптры» пишет по этому поводу: «Воображение находит себе место впереди, а память сзади, мышление же в середине головы»². Умственные силы проявляют себя в нравственных добродетелях, как это объясняет Плоть (служанка) Душе (госпоже): «А если мой мозг здоров и цел, то три твои умственные силы, о Душа, рождают в тебе четыре родственные им добродетели, четвероконную их упряжку: мышление порождает справедливость и мудрость, влечение — целомудрие, а ярость рождает в тебе мужество. Но стоит головному мозгу стать хромым, теряешь все: и справедливость, и мудрость, и мужество, и целомудрие, а с ними и умственные силы, о госпожа, — память, воображение и мышление, а с ними и пять твоих чувств, любимая: ум и мысль, представление и воображение, и, наконец, понимание; а также и мои все: зрение, обоняние, слух, вкус и осязание, все, любимая, — отходят все полностью — и умственное и чувственное, и мысленное с ними». Душа, по мнению Филиппа Пустынного, невещественна, добра, мысленна, словесна, бессмертна. Душа создана Богом как простая нематериальная субстанция.

Мы видим, что подобного рода суждения можно руководствоваться при поиске смысла неопределимых по сути понятий «интеллект», «психика», «сознание» и др. К положительному результату не может привести и вуалирование истинного смысла путем добавления слов («искусственный» — для ИИ) или приставок «псевдо-» для таких слов, как «псевдосознание робота», ведь любая модальность определима через реальность.

Поэтому при построении ИИ следует тщательно выбирать выражения, подбирать слова, критически подходить и к философско-аналитическим попыткам ввести в обиход выражения из сферы «собственно-человеческого».

Критический подход предполагает доработку традиционного, техникстского инструментария компьютерных технологий. Инструментарий следует поднять с чисто технологического до культурологического уровня. Однако тогда искусственный интеллект перестаёт быть чистой технологией, в его рамках становится невозможно безапелляционно присваивать ментальные свойства и состояния термометрам, столам, дождям, компьютерам. Технология искусственного интеллекта становится частным случаем культурологии: от наработанных при анализе социокультурной реальности норм понимания

¹ Умозрительные и деятельные главы святого Максима Исповедника //Добротолубие. Т. 6. С. 248. М., 1993.

² Прохоров Г.М. Памятники переводной и русской литературы XIV-XV вв. Л., 1987.

интеллекта, разумного поведения и др. становится зависимым определение понятия искусственного интеллекта.

Руководить такой «технологией» должен представитель гуманитарных наук. К «гуманитариям» выдвигается ряд сложных требований. В частности, в силу специфики своей профессии, он обязан владеть и опытом понимания текстов святоотеческой традиции, которая, как мы могли заметить в рамках нашего краткого анализа, представляется фундаментальной с позиции концептуально-языкового оформления слов «личность», «дух», «душа», «разум», «тело». А опыт подобного понимания несопоставимо сложнее опыта технического манипулирования абстрактными символами посредством абстрактных операций. Данный опыт предполагает участия в процессе понимания человека в целом, а не некой абстрагированной интеллектуально-лингвистической способности.

Однако кто сказал, что искусственный интеллект – прерогатива технических наук?

Виртуальная реальность и математика Н. Кузанского *Михаил Сёмочкин (ЗИ-81)*

1. Развитие понятия «виртуальности»

Идея виртуальности разрабатывалась в философии – античной, восточной, византийской, схоластической – иногда в неявном виде, иногда в явном, как, например, в схоластике. Идея виртуальности стала активно использоваться в последнее десятилетие в современной философии и науке, а также других сферах человеческой деятельности.

Во второй половине XX в. идея виртуальности возникла независимо друг от друга и почти одновременно в нескольких сферах науки и техники: в квантовой физике были открыты так называемые виртуальные частицы; в компьютерной технике появилось понятие виртуального объекта; в самолетостроении была создана модель виртуального полета самолета; в эргономике была разработана виртуальная кабина самолета; в психологии были открыты виртуальные состояния человека и, наконец, был предложен термин «виртуальная реальность» для обозначения особой компьютерной технологии, позволяющей пользователю интерактивно «взаимодействовать» со стереоскопическим изображением. В результате агрессивной рекламной кампании по продвижению виртуальных компьютеров на рынок, термин «виртуальная реальность» стал в массовом сознании ассоциироваться именно с компьютерами, активизировав идею «киберкультуры» и молодежное движение «киберпанк».

Как специальный философский и научный термин «виртуальная реальность» (от лат. *virtus* – доблесть, добродетель, энергия, сила и от позднелат. *realis* – вещественный, действительный, существующий в действительности; греческий аналог *virtus* – *arete*) появился в 80-х гг. XX в., когда в постклассической науке понятие предмета исследования было дополнено понятием реальности существования сопряженных с ним объектов, например, в физике субстанциальное вещество и энергетическое поле принадлежат одной и той же физической реальности.

В ИИ понятие виртуальности целесообразно связать с созданием т.н. «псевдосознания» (псевдосознания робота). И первоочередным вопросом поиска конструктивных оснований применения данной категории встаёт возможность *математической экспликации* «виртуальных объектов», «виртуальных миров» и т.п. Для этого следует обратиться к наследию Н.Кузанского.

2. Моделирование виртуальных миров по Николаю Кузанскому.

Термин виртуальность достаточно часто встречается у Николая Кузанского. В работе «О видении бога» он утверждает: «Я вижу телесными глазами, какое ореховое дерево огромное, раскидистое, зеленое, отягощенное ветвями, листвой и орехами. Потом умным оком я вижу, что то же дерево пребывало в своем семени не так, как я сейчас его разглядываю, а виртуально; я обращаю внимание на дивную силу того семени, в котором было заключено целиком и это дерево, и все его орехи, и вся сила орехового семени, и в силе семян все ореховые деревья. Потом я начинаю рассматривать семенную силу всех деревьев различных видов, не ограниченную никаким отдельным видом, и в этих семенах тоже вижу виртуальное присутствие всех мыслимых деревьев...».

Развивая подход к термину «виртуальность» (всё виртуально во всём), Николай Кузанский создаёт и собственную математику для описания такого «возможностного», виртуального мира. Математика Николая Кузанского, в отличие от любой аксиоматической системы (в т.ч. и евклидовой математики), не имеет готового набора средств построения. Вначале у Николая Кузанского нет линий и плоскостей, нет понятий «пересекаться», «быть параллельным» и нет возможности образовать понятие угла. Вначале имеется только **точка**. Для Николая Кузанского его начальная «точка» не является эксплицированным набором или, лучше сказать, полем элементов, которые изначально заданы и к которым можно редуцировать любой иной элемент. Помимо понятия «точка», которое является, по сути, субстанциальным, Николаю Кузанскому необходимо еще одно, операциональное понятие: «разворачивание». «Точка» задается как *«свернутость»* и одновременно как *«способность разворачиваться»*. Это ее подлинное определение: «точка» — это **способность порождения** любых иных компонентов математической системы. Поэтому точка — «неиное» этих компонентов, она им имманентна.

В кузанской математике операция «разворачивания/сворачивания» положена в основу. В данной математике нет доказательства в том виде, в каком оно имеет место в аксиоматических системах евклидовского типа; его нет потому, что в нем нет нужды, да и нет возможности его осуществить. Доказательство в «евклидовой науке» демонстрирует, как сложная конструкция сводится к элементам начального неопределяемого набора элементов. В «кузанской науке» вывод новой конструкции уже достаточен для того, чтобы показать все его свойства: вывод нового совершается как экспликация всего поля соучаствующих в построении компонентов. Собственно, та функция, которая в «евклидовой математике» принадлежит доказательству, здесь выполняется в ходе вывода новой конструкции.

В евклидовых системах строгость знания принято отождествлять с его формальным характером. Считается, что формализация позволяет достичь того уровня, на котором несомненность доказательства оказывается абсолютной. Кузанская «точка» — это возможность любой иносодержательной

компоненты; евклидовская «точка» – это невозможность её. Но «точка» Евклида равно относится к любым реальным точкам, которые мы видим вокруг себя: пятнышкам краски, засохшей на нашем оконном стекле, к крапинкам на спине божьей коровки, к пикселям, из которых создано изображение на экране нашего компьютера. «Точка» Николая Кузанского, сохраняя это свойство, равно относится также к «линиям», «плоскостям», «углам», «окружностям» и иным понятиям, *возможным* в евклидовой геометрии. Его точка — это свежая капля на листе промокашки, готовая расплыться во что угодно и даже выскочить за пределы сдерживающего ее листа.

Если описать ситуацию через понятие абстрагирования, можно сказать, что «кузанская наука» столь же абстрактна в отношении евклидовой, сколь евклидовая абстрактна в отношении окружающего мира. Обнаружив это, мы найдем объяснение тому факту, что наш естественный интеллект не мыслит по-кузанскому: достижение абстракции второго порядка, «возможностного мира» представляется невозможным для обыденного сознания. Такой мир строится в полёте интуитивного постижения реальности, расширенного сознания.

Несомненно, стоит обратить внимание на возможность построения виртуального мира Кузанского искусственными средствами – средствами искусственного интеллекта и включить способ построения такого мира в состав «псевдосознания» компьютерной системы (робота).

IV. ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И КОМПЬЮТЕРНАЯ ТЕХНОЛОГИЯ

Бионика как направление робототехники ***Евгений Александров (Э-81), Константин Домась (Э-81)***

Бионика – многообещающее научно-технологическое направление по заимствованию у природы ценных идей и реализации их в виде конструкторских и дизайнерских решений, а также новых информационных технологий.

В последнее десятилетие бионика получила сильный импульс к новому развитию, поскольку современные технологии позволяют копировать миниатюрные природные конструкции с небывалой ранее точностью. В то же время современная бионика во многом связана не с «ажурными» конструкциями прошлого, а с разработкой новых материалов, копирующих природные аналоги, с робототехникой и созданием искусственных органов.

Главное отличие искусственных конструкций от естественных созданий природы состоит в невероятной энерго-эффективности последних.

Современная бионика во многом связана с разработкой новых материалов, которые копируют природные.

Природные материалы сверхдешевы и распространены в огромном количестве, их «качество» значительно лучше тех, которые сделаны человеком. При этом человек использует достаточно простые энергоемкие процедуры получения тех или иных веществ, а природа делает это гораздо более «интеллектуальными» и эффективными способами.

В настоящее время пытаются найти аналоги органов человеческого тела, чтобы создать, например, искусственное ухо или искусственный глаз. Однако перспективна при этом разработка не прямых технических аналогов органов животных, а технических систем с биологически чувствительными элементами.

Одним из основных направлений работ по бионике является изучение нервной системы человека и животных и моделирование нервных клеток – нейронов и их ансамблей (нейронных сетей) для интеллектуализации компьютерной техники, разработки новых элементов и устройств автоматики и телемеханики. Здесь сложилось направление, названное *нейробионикой*.

Попытки моделирования нервной системы человека и животных были начаты с построения аналогов нейронов и их сетей. Разработаны различные типы искусственных нейронов. Созданы искусственные «нервные сети», способные к самоорганизации. Изучение памяти и других свойств нервной системы биологических существ – основной путь создания «мыслящих» машин для автоматизации сложных процессов производства и управления.

Большое значение в техническом конструировании продолжают иметь перцептроны – самообучающиеся системы, выполняющие логические функции распознавания и классификации.

«Квалиа» как базовая категория виртуалистики

Александр Бондарь (С-85)

В отечественной виртуалистике, которая развивается с начала 1990-х годов, сложился вполне устойчивый категориально-понятийный аппарат. В основном он содержит психологические понятия, такие как «виртуал», «консуэтал», «гратуал», «ингратуал» и др. Данные понятия продуктивно применяются в самых различных виртуалистских методиках, начиная от лечения алкогольных болезней и заканчивая объяснением опыта религиозного постижения.

Принято также технологию виртуальной реальности отождествлять исключительно с техническими средствами воздействия на чувственную сферу человека. При этом совершенно игнорируются социокультурные и антропологические параметры, не учитывается онтология от «первого лица».

Целью данного доклада является доказательство неверности такого подхода и демонстрация недостаточности сложившегося на сегодня категориально-понятийного аппарата виртуалистики. Доклад основан на работе Мишеля Туэ («Квалиа», 2003 г.)¹.

Представляется очевидным положение – если виртуалистика претендует на статус мировоззренческой позиции (в плане концепции виртуальных миров, виртуальных обществ, виртуальных цивилизаций) и на статус методологического базиса технологии виртуальной реальности, то её категориальный аппарат необходимо привести в соответствие с претензиями.

На роль базовой категории виртуалистики представляется перспективным назначить понятие «*квалиа*». Определение данного понятия представляет собой поле дискуссий в среде англо-американской философии сознания. В отечественной литературе данное понятие не получило должного освещения.

1. Что такое «квалиа»?

Квалиа – это доступные посредством интроспекции феноменальные аспекты нашей психической жизни. Например, квалиа обладают следующие состояния:

1. *Ощущения восприятия*. Например, ощущения, подобные тем, которые проявляются в наблюдении зелёного цвета, слуховом восприятии звуков трубы, пробы на вкус мёда, вдыхании морского воздуха, осязании меха.

2. *Ощущения тела*. Например: приступ боли, зуд, голод, зубная боль, жара, головокружение. Сюда также относятся ощущения, сопутствующие оргазму, бегу изо всех сил и т.п.

3. *Чувственные реакции*, или страсти (или эмоции). Например: чувство удовольствия, вожделения, любви, печали, ревности, сожаления.

4. *Чувственные настроения*. Например: приподнятое настроение, депрессия, спокойствие, скука, несчастье, напряжение.

В контексте технологии виртуальной реальности важно учесть факт того, что квалиа – это не только конкретные ментальные состояния, но и ментальные свойства, которые отвечают за феноменальный характер восприятия чув-

¹ Tyne, M. 2003. Qualia. Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/qualia/>

ственно данных объектов, то есть не сами чувственные содержания, но некий «квалификационно-квантификационный» фон, на котором чувственные данные и становятся возможными для интроспективного анализа. И в зависимости от характеристик такого ментального «фона» имеется возможность модификации параметров субъективной реальности. Возможность манипуляции этими параметрами в зависимости от «фона», в свою очередь, представляется фундаментальным фактором развития технологии виртуальной реальности.

В *компьютерной виртуалистике* продуктивное теоретико-концептуальное использование категории квалиа целесообразно связать с репрезентационистскими подходами к объяснению данного феномена.

2. Репрезентативные теории квалиа

Примеры квалиа, для объяснения которых требуются репрезентативные теории и которые имеют непосредственное отношение к сфере виртуалистики:

- Если у меня красные круги перед глазами после вспышки фотоаппарата, то я «вижу» на лице фотографа красное пятно, хотя его там на самом деле нет.
- Для того, чтобы ощущать боль в ноге, не обязательно иметь ногу. Моя боль может быть фантомной, если конечность отсутствует.

При совпадении чувственных ощущений здесь наблюдаются чёткие репрезентативные различия. То есть для того, чтобы жить в виртуальном мире, как в реальном мире, необходимо научиться манипулировать квалиа.

Квалиа в данном случае – суть репрезентативное содержание ощущений, в которые также входят и качества внешних вещей. Репрезентативное содержание ощущений – это то, что чувственный опыт воспринимает как некое «значение», подобное «значению» слова. Более того, так как значение слова – это не качество, которым слово обладает, так и феноменальный характер ощущений – это не качество, которым обладают чувственные переживания.

С точки зрения *сильного репрезентационализма*, квалиа (как значения) находятся не в голове, а *во внешнем мире*. Классическое картезианское представление о ментальных свойствах и их отношениях к внешнему миру, таким образом, оказывается перевёрнутым с ног на голову. Квалиа не являются свойствами, присущими внутренним представлениям, о которых их субъекты осведомлены напрямую. Квалиа не являются и качествами, которые необходимо совместны с внутренними отношениями. Квалиа – репрезентационные содержания, которыми обладают определённые внутренние состояния, но сущность которых – в конкретных *внешних* взаимоотношениях между индивидом и окружающей средой.

Репрезентационализм, представленный в такой форме утверждает тезис тождественности в отношении квалиа: *квалиа – это то же самое, что и конкретные репрезентативные содержания*.

Критика репрезентационализма обычно принимает форму предполагаемых контрпримеров, мысленных экспериментов. Одно направление критики опирается на утверждения, что ощущения имеют одно и то же репрезентационное содержание, но разные феноменальные типы. Другое направление основано на утверждении того, что ощущения имеют различное репрезентативное содержание, но один феноменальный характер.

3. Критика сильной репрезентативной теории квалиа

3.1. «Болотный человек» (Д.Дэвидсон, 1986)

Данный пример описывает различие ощущений того или иного рода при отсутствии репрезентативного содержания этого ощущения. «Болотный человек» – это точная, молекула в молекулу, копия человека, случайно образовавшаяся в болоте под воздействием химической реакции, происшедшей в тот момент, когда молния ударила в частично затопленное бревно. Допустим, этот дубликат может ощущать. Однако ощущения не будут иметь такого же репрезентативного содержания, какое имеет оригинал – ведь «Болотный человек» исключён из социокультурного и исторического контекста, в который с момента рождения был включен оригинал.

3.2. «Инвертированная Земля» (Н. Блок, 1980).

«Инвертированная Земля» – это воображаемая планета, на которой вещи имеют цвета, комплиментарные с цветами на Земле. Небо – желтое, трава – красная, зрелые помидоры – зеленые и т.д. Обитатели «Инвертированной Земли» испытывают психологические ощущения с инвертированными ментальными содержаниями относительно тех, которые люди испытывают на Земле. Они воспринимают небо желтым, траву – красной. Однако они называют небо «синим», траву «зеленой», зрелые помидоры «красными», в точности так, как и мы их называем. Во всех остальных отношениях «Инвертированная Земля» настолько же похожа на Землю, насколько это возможно.

В глаза человека вставляют линзы, инвертирующие цвета и отправляют в путешествие на «инвертированную Землю». Человек после полёта пробуждается и не замечает никакой разницы – линзы нейтрализуют инвертированные цвета. Человек думает, что он всё ещё на родной планете. Он видит небо и все что его окружает точно таким же, как и на Земле. Однако по прошествии достаточного времени, после того как человек войдет в языковое и физическое окружение «Инвертированной Земли», его интенциональные содержания будут совпадать с таковыми же у её обитателей. Он будет убежден, например, что небо жёлтое, точно также, как убеждено и всё остальное население. У человека появится визуальное переживание, представляющее небо жёлтым. Позже человек станет субъектом внутренних состояний, которые интенционально инвертированы относительно внутренних состояний, испытываемых человеком ранее. В тоже время останутся неизменными сами способности восприятия. На основании этого делается вывод, *сильный репрезентационализм, уделяющий внимание исключительно внешнему фактору квалиа, неверен.*

4. Вывод

Помимо приведённых мысленных экспериментов существуют и иные способы критики сильного репрезентационизма. Сильный репрезентационизм не позволяет очертить конструктивные перспективы манипулирования такими состояниями, как чувство депрессии. Данное чувство вообще не имеет репрезентационного содержания.

Посредством сильной репрезентативной теории квалиа можно было бы попытаться обосновать применение технических средств виртуальной реальности. За счет них, посредством чисто внешнего воздействия, достигался бы эффект виртуализации. Однако приведённые классические примеры показывают несостоятельность такого подхода. Продуктивными в плане теоретико-

концептуального обоснования и использования в технологии виртуальной реальности представляются слабые репрезентационистские теории. Они изучают квалиа в корреляции субъективного опыта чувственных данных и внешних объектов.

Если квалиа не определены полностью репрезентативным содержанием, то бесполезны попытки создания виртуальной реальности путём манипуляции исключительно чувственными ощущениями человека. А именно такой подход присущ современному этапу развития технологии виртуальной реальности. Сильная репрезентативистская теория квалиа невозможна. Необходимо, помимо учёта внешних факторов, обуславливающих квалиа, систематический учет социокультурных, биолого-генетических факторов и иных параметров, которые задают особенности субъективной реальности человека. Чисто технологическая «экспансия» в виртуальные миры человека невозможна. Необходимы скрупулёзные *антро- и социо-культурологические исследования*.

Человек и компьютер. Друзья или враги? **Алла Каданцева (МС-81)**

Компьютеры сегодня – неотъемлемая часть жизни многих из нас. Центральные составляющие общественной инфраструктуры управляются компьютерными системами. Некогда, на заре научно-технической эры, вошли в моду мифы о «машинных бунтах», о деградации человечества на фоне развивающейся технологии и т.д. Тогда, во времена громоздких примитивных ЭВМ всё это воспринималось как сказка. Но сегодня с появлением персонального компьютера становится очевидным то, что компьютер – нечто большее, чем просто устройство, облегчающее людям жизнь. Мы имеем дело с «существом», которому готовы приписать всю совокупность ментальных качеств и свойств, тех качеств, обладателем которых испокон веков считался только человек.

Масштабы и темпы преобразований, новые степени свободы и способы коммуникации, небывалая власть виртуальных миров бросают вызов не только нашему мироощущению, привычным системам норм и ценностей, но и самой природе человека, его биологической организации. Имеются в виду прежде всего генетически предопределённые пространственно-временные параметры человеческой психики и связанные с ними возможности адекватного отображения собственной двигательной активности, объёмы восприятия и переработки информации, границы управленческих возможностей нашего Я и способности поддержания его идентичности.

Можно ли жить в виртуальной реальности?

О слиянии человеческого мозга с виртуальной реальностью первыми заговорили писатели-фантасты. В литературе и фильмах описан мир, которым правят транснациональные электронные корпорации, а люди становятся частью виртуальной реальности или превращаются в придатки компьютеров. Но оказывается, такое возможно и в действительности!

В США несколько лет назад разработали проект под названием «Компьютерный Маугли». Целью его является создание электронной матрицы личности живого человека, полностью воспроизводящей все индивидуальные особенности его мышления и духовного облика.

Не так давно была изобретена аппаратура, позволяющая людям испытывать физические ощущения при соприкосновении с виртуальным миром. Для этого достаточно надеть на голову специальный шлем и облепить тело датчиками, через которые вам будут передаваться электронные импульсы.

«Виртуализация жизни» включает новые степени отчуждения от подлинной реальности. Теперь уже не только актер может изображать некоего персонажа, но и его цифровой двойник. Еще в 2000 г. вышел фильм «Финальная фантазия: духи внутри нас», в котором впервые цифровые клоны полностью заменили актеров.

Рефлексия над современными социокультурными реалиями не поспевает за бурным развитием компьютерной практики. Уже сейчас способы взаимодействия с компьютером далеко выходят за рамки процесса нажатия клавиш. Успешно разрабатываются технологии речевого общения с компьютером и распознавания написанного от руки текста. На очереди – распознавание жестов и мимики. Скоро компьютер станет квалифицированным синхронным переводчиком.

Один из актуальнейших планов проблемы — прямое подключение компьютерного устройства к мышцам, внутренним органам, нервным узлам и непосредственно к головному мозгу.

Создан искусственный глаз — компьютерное устройство, которое преобразует изображение в нервные импульсы, посылаемые в мозг посредством вживленных электродов. Практикуются имплантации электронных чипов для восстановления функций отдельных органов (сердечный стимулятор, искусственное ухо и др.). Уже сейчас компьютер способен воспроизводить, точнее, имитировать практически все основные чувственные модальности — от зрительных образов до обоняния.

Однако этого недостаточно, чтобы приписывать современному компьютеру способность мышления, если последнее понимать в полном объеме его существенных признаков, не сводить его к формально-логическим операциям, к манипуляциям символами. Реальное человеческое мышление суть явление субъективной реальности (взятой в ее рефлексивном и арефлексивном, актуальном и диспозициональном измерениях)¹, оно включает чувственные и интуитивные составляющие, факторы воображения, надежды, веры и воли. Реальные акты мышления осуществляются данным конкретным Я и несут на себе индивидуальную, личностную печать мыслящего человека.

Между естественным и искусственным интеллектом пока что сохраняется очень большая дистанция. Хотя некоторые виды интеллектуальной деятельности компьютер выполняет несравненно лучше мозга.

Что же нас ожидает в будущем? Рассмотрим две стратегии.

Стратегии будущего

Первая стратегия — это путь вымирания биологической формы существования разума, преобразование человека в трансгуманоида, свободного от ограничений, налагаемых биологической телесностью (путь независимости от микробов и вирусов, старости, непереносимости радиации, высоких и низких температур и т.п.). Неявно эта стратегия полагает и неизбежный конец биологической форме жизни на Земле, который наступит в результате даль-

¹ См. Дубровский Д.И. Проблема идеального. М., 2003

нейшего углубления экологического кризиса и цепи экологических катастроф.

Вторая стратегия не спешит заменить человека трансгуманоидом, подчеркивает наличие больших ресурсов самосовершенствования человека, возможность преодоления экологического кризиса и сохранения земной жизни как непреходящей ценности. Возможность парировать нарастающие угрозы она связывает с достижениями генной инженерии, геномики, нейрофизиологии, психологии, других наук, в том числе, безусловно, с успехами информационных технологий и робототехники.

Инкубатор гениев?

Одно из уникальных свойств электронного «мозга» – способность корректировать психику человека в нужном направлении. В определённом смысле это даёт положительный результат. Так, превращение отсталого в умственном развитии человека в гения с помощью компьютерного программирования (как мы наблюдаем это в известном фильме «Газонокосильщик») – вовсе не миф. Уже существуют программы психокоррекции для умственно отсталых людей. Предполагается, что микрокомпьютеры через подключённые к телу и голове датчики будут посылать в мозг такого человека особые сигналы, позволяющие ему адекватно ориентироваться в различных ситуациях. Подобные системы способны помочь больным людям и инвалидам. Пример тому – знаменитый астрофизик Стивен Хокинс, прикованный к инвалидной коляске и общающийся с окружающим миром посредством сложного электронного устройства.

«Исдержки производства» или пси-оружие?

Недавно исследователи из лаборатории психокоррекции Московской медицинской академии пришли к выводу, что современные информационные технологии, особенно компьютерные, могут оказать крайне опасное воздействие на человека.

Прежде всего – виртуальная иллюзия сливается с реальностью и подавляет её восприятие, так как информация, выведенная на монитор, воспринимается подсознанием без какого-либо критического осмысления. Отсюда почти «наркотические» увлечения людей Интернетом, компьютерными играми, электронными «питомцами»-тамагочи, оттесняющие истинную действительность на второй план. А иногда «общение» с компьютером может привести к непоправимой трагедии.

В некоторых тоталитарных сектах используют компьютерные технологии в целях психологического воздействия на людей. Так, в скандально известной «АУМ СИНРИКЁ» сектантам, которых трудно было подчинить контролю «гуру», надевали на голову так называемые «шлемы спасения» – шапочки, электродами подсоединённые к компьютеру. Постепенно человек начинал терять связь с окружающим миром и фактически превращался в зомби.

Апокалипсис от «Майкрософт»?

С компьютерами теперь нередко связывают и пророчества о «конце света» в 2000 г. Сегодня кажется невероятным, нереальным, чтобы из-за элементарной системной ошибки (вернее, недоработки) в ночь на 1 января будущего года вышла из строя вся электроника, которой начинены наши дома и учреждения, и мы оказались бы вновь в каменном веке, лишённые даже воз-

возможности связаться друг с другом. Но ведь в истории вычислительной техники уже было нечто подобное!

В 1979 и 1980 гг. компьютер в Пентагоне неоднократно ошибочно объявлял ядерную тревогу. ВВС немедленно приводили в боевую готовность стратегические бомбардировщики. К счастью, всё обошлось.

Не всегда следует полагаться на прогнозы, сделанные машинами. В Англии группа крупных бизнесменов задала ЭВМ вопрос – какая компания в Великобритании является сейчас наиболее надёжной? «Роллс-Ройс», – ответил компьютер. А вскоре фирма «Роллс-Ройс» полностью обанкротилась.

После появления многочисленных сообщений о роковых просчётах электронно-вычислительных машин некоторые энтузиасты начали бить тревогу. Около 15 лет назад в Лондоне было организовано «Международное общество по запрещению счётно-решающих машин». Его участники подсчитали, что за 2 секунды одна-единственная ЭВМ способна совершить больше ошибок, чем 50 человек за 200 лет.

«Самые умные машины» отнюдь не обладают абстрактным мышлением. Они «мыслят» только «категориями», заложенными в них программистами.

Станем ли мы заложниками компьютера? На этот вопрос пока трудно ответить однозначно. Возможно, в отдалённом будущем эти хитрые приборы вообще изменятся до неузнаваемости, даже перестанут быть машинами в полном смысле этого слова. Но не всё ли равно? *Уже сегодня «компьютер» превращается в понятие скорее социо-культурологическое, чем техническое, и мы меняемся вместе с ним.*

Проблемы построения квантовых компьютеров

Михаил Казанский (Р-81)

Как известно, А. Тьюринг более полвека назад предлагал использовать квантово-механические модели для объяснения способа функционирования человеческого разума и применения этих моделей для построения «мыслящих машин». Однако по сей день в данном направлении не достигнуто значимых результатов.

Квантовые алгоритмы сложны в понимании, так как представляют собой вычислительный аспект «квантового», неклассического подхода к репрезентации предметной области. Ручное проектирование квантовых алгоритмов практически невозможно, что предопределяет требование на поиск технических решений по созданию специализированного инструментария «квантового» проектирования и программирования.

Реализация такого инструментария традиционными средствами – на обычных машинах – принципиально не отвечает сложности задачи и приводит к построению примитивных программ. Кроме того требуются большие объёмы вычислительных ресурсов и памяти для решения самых простых проблем.

Для преодоления таких трудностей формируются специальные исследовательские проекты. Показательным является, например проект, инициированный кафедрой системного анализа совместно с центром информатики университета Дортмунда (Германия) [<http://www.icd.de/>]. Проект базируется

на сети параллельных компьютеров. Параллельные вычисления, по мнению авторов проекта, в наибольшей степени соответствуют представлению квантовых алгоритмов.

Начало проекту положило создание эффективного квантового симулятора, который работает в операционной среде GP. Первоначальная базовая структура представления предметной области имела форму линейных деревьев. В дальнейшем в рамках проекта стали применяться чисто линейные структуры генома программы.

Считается, что квантовые алгоритмы имеют древесную структуру. В рамках неё решаются и вопросы о значительном повышении по сравнению с традиционными программными структурами степеней свободы в конструкции и эволюции квантовых структур.

Операционная система GP применялась для решения Jozsa-проблемы и 1-SAT-проблемы. В частности для последней проблемы квантовые алгоритмы строились явным образом из специальных квантовых алгоритмов для набора логических функций (формул) со сколь угодно многими переменными. Преобразование структуры дерева в линейно-древесный геном приводило к столь большим вычислительным затратам, что от решения 1-SAT-проблемы пришлось отказаться.

Jozsa-проблема представляла интерес для экспериментального исследования поисковых алгоритмов. Отмечается, что лежащая в основе алгоритма алгебраическая структура (решётка) особенно затрудняет эффективный поиск. Однако других универсальных структур не придумано.

В принципе, в повышении эффективного поиска и видится цель автоматического проектирования квантовых алгоритмов. Данные алгоритмы должны иметь специальные средства интеграции положений классической физики и квантовой механики. Т.к. развитие квантовых алгоритмов (КА) на основе квантовой механики является крайне распространенным подходом, то в рамках проекта ключевое значение придавалось средствам генерации посредством генетической программы нового КА.

На классическом компьютере КА не эффективен, т.к. значительно экспоненциально возрастает количество решёток, представляемых в рамках унитарной матрицы. Поэтому при создании нового КА накладываются существенные ограничения на количество базовых Q-битов (квантовых битов). Это приводит к: (а) поиску новых разновидностей решеток и (б) – к вводу ограничений и допущений задачи.

Анализ проблем, с которыми столкнулись исследователи при построении квантовых компьютеров, позволяет увидеть, что помимо существенных технических ограничений имеются проблемы концептуального типа – проблемы принятия в коллективе исследователей общепринятой квантово-механической парадигмы мышления. Во-многом, из-за второй проблемы и нет до сих пор прорыва в развитии алгоритмов для квантовых компьютеров.

Интернет-навигатор "Искусственный интеллект"

Станислав Колесников (ЭП-81)

В работе представлены интернет-ресурсы, посвящённые проблематике искусственного интеллекта.

Сайты

1. ChatMaster [<http://chatm.chat.ru/>]

ChatMaster – это самообучающаяся программа, которая поддерживает диалог с человеком. В ее основе лежат прецедентные методы, которые обеспечивают самообучение и подстройку под собеседника. ChatMaster ведет контекстно-зависимый разговор, то есть понимает смысл реплики, которая опирается на предшествующие. Диалог может вестись на любом неиероглифическом языке (т.е. на всех европейских и некоторые азиатских). В настоящее время база знаний программы существует только на русском языке, но может быть легко пополнена. Диалог осуществляется путём ввода с клавиатуры и отображается на экране. Как обучать ChatMaster, можно узнать на сайте программы. Разработчик – Д. Журавлев.

2. SmarterChild [<http://www.smarterchild.com/>]

Номер ICQ 35000, ник SmarterChild. Поговорите с ним, говорит на все темы, но иногда отвечает не в тему. Также «работает» по принципу Словарь/Энциклопедия/Переводчик/Поисковик, разговаривает только на английском языке и переводит тоже с английского. Может переводить с английского на французский, немецкий, итальянский, португальский и испанский. Всегда online.

3. Nus (Nai) [<http://nai.wallst.ru/>]

Nus – программа, которая поддерживает связный диалог с человеком. Разработчик – Валентин Шергин.

4. Diala [<http://diala.chat.ru/>]

Программа DIALA ведет диалог с человеком на русском языке на любую тему, пытаясь имитировать при этом естественный интеллект. Человек вводит свои фразы с помощью клавиатуры, а DIALA выводит свои ответы на экран. Как и у всех людей, у DIALA есть любимые темы: человек и компьютер, мужчина и женщина, любовь, выпивка... DIALA считает себя женщиной и довольно критично относится к мужчинам, да и вообще к человечеству. Как и собеседник-человек, она не всегда все понимает с первой попытки, иногда проявляет упрямство и говорит о том, что интересуется сейчас ее, а не Вас (это довольно часто бывает и с людьми). Интеллектуальный уровень диалога с DIALой зависит от уровня интеллекта человека-собеседника. Diala одна из самых старых программ такого рода (1982). Обладает своеобразной логикой и «чувством» юмора.

Разработчик – "Рефрижератор".

5. Electronic Brain 1300 [<http://electrbrain.chat.ru/>]

Приятная программы для общения. Большая база, оригинальные реплики. Electronic Brain является неплохим собеседником, который с удовольствием выслушает вас и обязательно ответит. Старайтесь поддерживать тему разговора, не грубите программе, избегайте слишком коротких фраз и слов. Разработчик – Роман Ефимов, 2001-2003 г.

6. A-LIFE3 [<http://probirkinsky.narod.ru/>]

Крайне разговорчивая программа, использующая смайлики. Качество диалога – хорошее: даже при малом размере базы реплики можно считать достаточно адекватными. Удобный редактор базы. Разработчик: Нонат Ревелев, 2002 г.

7. A.L.I.C.E.

[<http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1>]

Artificial Linguistic Internet Computer Entity – чат-робот, победитель среди программ общения на свободном естественном языке. Автор: Dr. Rich Wallace.

Исследовательские центры

1. Роснии ИИ [<http://www.artint.ru/>]

Российский Научно-исследовательский институт искусственного интеллекта (Роснии ИИ). Институт является организатором ежегодного международного семинара ДИАЛОГ (компьютерная лингвистика и ее приложения). Наряду с ведущими российскими компаниями АBBYU [<http://www.abbyu.ru/>], Яндекс [<http://www.yandex.ru/>] и «ПроМТ» [<http://www.promt.ru/indexr.phtml>], а также Институтами Проблем Информации и Проблем Передачи Информации РАН Роснии ИИ выступает инициатором создания «Ассоциации Лингвистических и Коммуникативных Технологий», учреждаемой для объединения усилий заинтересованных фирм, научных организаций и ВУЗов в решении научных и прикладных задач в их общей области деятельности. Ведётся дайджест *искусственного интеллекта*, к настоящему времени вышло около 150 номеров журнала.

2. ИНТЕЛЛЕКТ ПЛЮС [<http://inteltec.ru/>]

Научно – производственный центр "ИНТЕЛЛЕКТ ПЛЮС". Область деятельности НПЦ "ИНТЕЛЛЕКТ ПЛЮС" – сфера высоких информационных технологий. Это разработка объектных СУБД и сложных информационно-поисковых систем, программ для опубликования корпоративных информационных систем в Интернет, научно-исследовательская деятельность в области полнотекстового поиска и семантического анализа текста, продвижение объектных технологий ведущих мировых производителей на российский рынок.

3. Образовательные Технологии и Общество [<http://ifets.ieee.org/russian/>]

Международный Форум "Образовательные технологии и Общество" Восточно-Европейская подгруппа. Целью создания подгруппы является: организация единого информационного пространства для обмена мнениями, опытом между специалистами по информационным технологиям в обучении, преподавателями, менеджерами учебного процесса, студентами и другими заинтересованными лицами. Задачами подгруппы являются: Организация телеконференций по проблемам новых информационных технологий в образовании; Публикация Восточно-европейской секции в Международном рецензируемом журнале «Образовательные технологии и общество»; Поддержка информационного Web-сайта. Вся информация, получаемая в результате работы подгруппы публикуется на сайте и доступ к ней является свободным.

4. Искусственный интеллект – взгляд в будущее [<http://aifuture.chat.ru/>]

Сайт по самым широким вопросам, касающимся искусственного интеллекта

5. Проект СИРИУС [<http://neural.narod.ru/Sirius.htm>]

Название проекта СИРИУС переводиться как Система Искусственного Распределенного Интеллекта Универсальной Структуры. Подробнее о проекте вы сможете прочитать на сайте проекта.

6. ЛИИ [<http://lii.newmail.ru/>]

Лаборатория искусственного интеллекта. Тематика сайта: искусственный интеллект и все, что имеет к нему отношение. В настоящий момент на сайт выложена лишь небольшая часть материалов, в основном по нейронным сетям.

7. Группа Исследования и Разработки Искусственных Нейронных Сетей и Генетических Алгоритмов [<http://users.kpi.kharkov.ua/mahotilo/>]
Группа исследования и разработки искусственных нейронных сетей и генетических алгоритмов представляет собой добровольное объединение научных работников Национального технического университета "Харьковский политехнический институт", изучающих и применяющих эти современные вычислительные технологии.
8. ICFCST [<http://www.icfcst.kiev.ua/>]
International Charity Foundation for History and Development of Computer Science and Technique.
9. Центр речевых технологий [<http://www.speechpro.com/rus/index.html>]
Распознавание речи, верификация голоса (разграничение доступа с использованием голоса), очистка шумов речевых сигналов, судебные фоноэкспертизы и расшифровка "черных ящиков".
10. Проект Кибержизнь [<http://cyber-life.narod.ru/>]
Задача-минимум проекта – собрать необходимые сведения, для того, чтобы сначала осознать, а потом и выполнить задачу-максимум: построить искусственную интеллектуальную систему, способную соперничать по возможностям с человеческим разумом.
11. ПРОЕКТ «ЭМБРИОН-10» [<http://neurnews.iu4.bmstu.ru/news/embrion.htm>]
Нейрокомпьютер как электронная модель мозга.
12. ASIA SOFTWARE [<http://www.asia-soft.com/frs/ru/main/>]
Системы идентификации лиц по их изображению.

Каталоги

1. MavicaNET [<http://zona.ru/directory/rus/795.html>]
Многоязычный Поисковый Каталог. Раздел – Искусственный интеллект.
2. Интеллект-издательство [<http://inteltec.ru/publish/themes/textan.shtml>]
Раздел – Искусственный интеллект. Подборка статей об искусственном интеллекте.
3. Исследования в области Искусственного интеллекта [<http://artin.narod.ru/books.html>]
Подборка книг по теме ИИ.
4. Исследования в области Искусственного интеллекта [<http://artin.narod.ru/links.html>]
Коллекция ссылок по теме ИИ.
5. Страница Адушкина [<http://dushkin.boom.ru/AI.htm>]
Коллекция ссылок по теме ИИ.
6. JAIR [<http://www.jair.org/>]
Journal of Artificial Intelligence Research.
7. MEMBRANA.RU [http://www.membrana.ru/themes/robots_and_ai/]
Подборка статей по теме: Роботы и искусственный интеллект.
8. neural.narod.ru [<http://neural.narod.ru/Resurse.htm>]
Подборка ссылок на сайты по теме ИИ.
9. "ИИ в домашних условиях" [<http://www.aimatrix.nm.ru/link/131.htm>]
Подборка ссылок на сайты по теме ИИ.
10. Microbot.ru [http://www.microbot.ru/index.php?module=Static_Docs&func=view&d=5_Artificial_intellect]. Подборка статей об искусственном интеллекте.
11. alife-soft.narod.ru [<http://alife-soft.narod.ru/notes.html>]

Подбор статей об искусственном интеллекте и всего что касается него.

12. Alchemist homepage [<http://user.hamovniki.net/~alchemist/NN/NN.htm>]

Подбор ссылок об искусственном интеллекте и нейронных сетях.

13. www.pautina.net [<http://www.pautina.net/2/59/679/more2.html>]

Каталог сайтов, посвящённых ИИ.

Статьи

1. Искусственный интеллект [<http://rkit.chat.ru/algo/library/ii.htm>]

Основной задачей данной работы является создание модели механизма работы человеческого сознания, что должно позволить создать искусственную систему обработки информации как аналог человеческого интеллекта.

2. Что называется ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ?

[<http://inf.susu.ac.ru/~pollak/expert/AI/WhatAI.htm>]

Здесь предпринята попытка ответить на базисные вопросы об искусственном интеллекте. С мнениями, выраженными здесь, согласны однако не все исследователи в области ИИ.

4. "Растрепанный Блокнот" [<http://netnotes.narod.ru/notes/t5.html>]

Программы-собеседники: искусственный интеллект и его эмуляция.

5. Проект "Соломон" [<http://www.shaman.ryazan.ru:8101/pub/solomon.html>]

Искусственный интеллект. Философская концепция и вопросы реализации.

6. Семантические сети [<http://compzed.narod.ru/semseti.htm>]

Семантическая сеть – структура для представления знаний в виде узлов, соединённых дугами.

7. В. Кузник [<http://alephegg.narod.ru/Survey/Perception.htm>]

Создание понятий на основе чувственного восприятия.

8. Знание – Сила [http://www.znanie-sila.ru/online/issue_1628.html]

Как работает мышление? Рафаил Нудельман.

9. КОМПЬЮТЕРРА [<http://www.computerra.ru/offline/2002/455/19246/page2.html>]

Искусственный интеллект: основные направления и состояние исследований.

10. ocrai.narod.ru [<http://www.ocrai.narod.ru/>]

Распознавание образов и искусственный интеллект.

11. Superidea.ru [<http://www.superidea.ru/intel/inoth/myshlen.htm>]

Мышление и интеллект – естественный и искусственный. Роберт Солсо.

Парадигма коннекционизма как методология нейрокомпьютерной технологии

Михаил Кольцов (С-71)

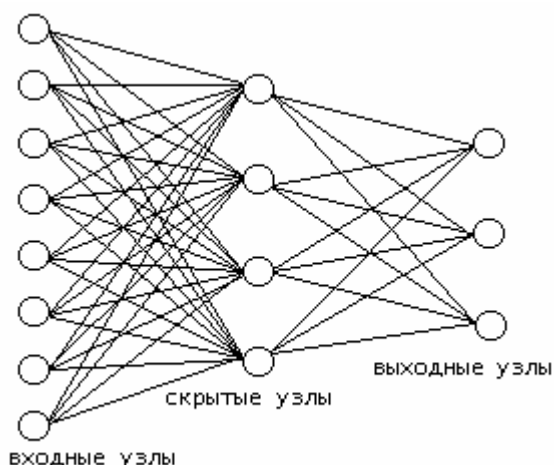
Коннекционизм¹ – это направление в философии сознания и философии искусственного интеллекта, в рамках которого предпринимаются попытки объяснить интеллектуальные способности человека, используя концепцию искусственных нейронных сетей. Составленные из большого числа структурных единиц, каждая из которых аналогична нервной клетке определённого типа, с заданным для каждого элемента весом, определяющим силу связи с другими элементами, нейронные сети представляют собой упрощённые

¹ В основе доклада лежит перевод работы Джеймса Гарсона «Коннекционизм» (Garson James, 2002, Connectionism, <http://plato.stanford.edu/entries/connectionism/>)

модели человеческого мозга. Такая весовая модель обладает эффектом синапсов, соединяющих каждый отдельный нейрон с остальными нейронами. Эксперименты с нейронными сетями продемонстрировали их способность к обучению для выполнения таких задач, как распознавание образов, анализ простых грамматических структур и др.

1. Описание нейронных сетей

Нейронная сеть состоит из большого числа элементарных единиц, объединённых друг с другом в соответствии с определённой схемой соединений. Элементы, составляющие сеть, как правило, разделены на три класса: *входные узлы* – элементы, на которые передаётся информация для обработки, *выходные узлы* – элементы, на которые поступают результаты обработки, и



промежуточные узлы, называемые скрытыми элементами. Если поставить в соответствие нейронной сети нервную систему человека в целом, то входным элементам будут соответствовать сенсорные нейроны, выходным – моторные нейроны, а скрытым элементам – все остальные нейроны.

Каждый входной элемент имеет начальное, активизирующее значение, которое представляет некоторую внешнюю особенность сети. Входной узел передаёт своё на-

чальное значение всем связанным с ним скрытым узлам. Каждый из скрытых узлов на основании полученного от входного узла активизирующего значения вычисляет своё собственное активизирующее значение. В дальнейшем сигнал передаётся выходным элементам или другому слою скрытых узлов. Следующие скрытые элементы, в свою очередь, вычисляют свои активизирующие значения тем же способом и передают их по сети своим соседям. В конечном счёте, сигнал от входных элементов распространяется по всей сети и определяет активизирующие значения выходных элементов нейронной сети.

Модель, в соответствии с которой происходит определение активизирующих значений узлов нейронной сети, задаётся на основе веса или силы связей между узлами. Вес связи может быть как положительным, так и отрицательным. Отрицательный вес связи соответствует препятствованию действиям передающего узла со стороны принимающего узла. Активизирующее значение для каждого принимающего узла вычисляется в соответствии с простыми активизирующими функциями. Эти функции различаются в деталях, но соответствуют одному простому соотношению – они производят суммирование значений от всех передающих узлов, причём вклад каждого узла в общую сумму определяется через вес связи и активизирующее значение отправляющего узла. После этого, как правило, происходит дальнейшая модификация полученной суммы, например, путём нормализации значений полученной суммы и /или установлением активизирующего значения равным нулю в том случае, если не достигнуто определённое пороговое значение суммы. Сторон-

ники коннекционизма предполагают, что процессы познания можно объяснить при помощи наборов узлов, работающих по описанному алгоритму.

Представленная выше разновидность сети называется *сетью прямой направленности* или *сетью с прямым распространением сигнала*. Процесс вычислений в сети происходит по направлению от входных узлов к выходным через скрытые узлы сети. Более адекватные модели человеческого мозга включают множество слоёв скрытых элементов и рекуррентные связи, управляющие распространяющийся по сети сигнал с высоких уровней обратно на низкие. Подобные рекуррентные связи необходимы для объяснения таких свойств сознания как кратковременная память. В сетях с прямой направленностью для повторяющихся представлений одних и тех же входных данных каждый раз определяются абсолютно одинаковые наборы выходных данных. Для формирования устойчивых представлений требуется время – ведь даже простейшие организмы необходимо некоторое время приучать, чтобы добиться выполнения (или невыполнения) одних и тех же действий, хотя при этом каждый раз также используются одинаковые стимулы. Сторонники коннекционизма пытаются избежать использования рекуррентных моделей, так как методы обучения рекуррентных сетей изучены крайне мало. Тем не менее Элман и др. (Elman – 1991) достигли определённых успехов в работе с простыми рекуррентными сетями, в которых введение рекурсии было вынужденным из-за соображений компактности модели.

2. Обучение нейронной сети и метод обратного прохода

Определение правильных весовых значений для выполнения поставленного задания является центральной задачей коннекционистского исследования. На сегодняшний день уже изобретены обучающие алгоритмы, способные высчитывать правильные значения веса в сетях применительно ко многим задачам. (Hinton, 1992). Одним из наиболее широко применяемых обучающих алгоритмов является так называемый «метод обратного распространения, или обратного прохода». Для использования этого метода необходимо иметь исходные данные для обучения сети, состоящие из множества примеров входных данных и желаемых для данного задания результатов или выходных данных. Например, если задача – научить нейронную сеть различать мужские и женские лица, исходные данные для обучения должны содержать изображения лиц с указанием пола в каждом конкретном случае. Сеть, которую можно научить выполнять подобное распознавание, должна иметь в своём составе два выходных узла (предназначенных для каждого пола соответственно) и множество входных узлов, таких, что каждый входной узел определяет яркость одного пикселя (точки) изображения. Изначально весовые отношения в сети устанавливаются произвольным образом, после чего все элементы массива исходных данных поочерёдно передаются в сеть. Значения входных данных подставляются во входные узлы, производится сравнение значений выходных узлов с желаемым результатом для данного элемента исходных данных (картинки с изображением мужского или женского лица). Затем все весовые значения сети подвергаются незначительной корректировке таким образом, чтобы значения выходных узлов в наибольшей степени соответствовали желаемому результату. Например, когда в качестве исходных данных во входные узлы передано изображение мужского лица, все значения веса в нейронной сети корректируются так, чтобы значение выходного узла, соответствующего мужскому полу, увеличилось, а значение

выходного узла для женского пола уменьшилось. После многократного повторения этого процесса сеть сможет выдавать желаемый результат для каждого изображения, представленного в множестве исходных данных. В случае успешного прохождения процесса обучения нейронная сеть сможет распространить своё поведение на информацию, которая не была представлена в качестве обучающей, выдавая для неё правильный результат. В нашем случае сеть будет способна различать произвольные, не представленные ранее, изображения мужских и женских лиц.

Обучение сети моделировать аспекты человеческого интеллекта является настоящим искусством. Успех алгоритма обратного прохода и других обучающих методов может зависеть от очень тонких корректировок, производимых алгоритмом и от значений исходных данных. Как правило, процесс тренировки включает в себя сотни тысяч циклов корректировок весовых соотношений в сети. Накладываемые на исследователей ограничения в производительности современных компьютеров приводят к тому, что процесс обучения нейронной сети выполнению интересующей задачи может длиться несколько дней, а возможно, и недель. Некоторые трудности удаётся преодолеть введением параллельных циклов вычислений, разработанных специально для нейронных сетей и имеющих широкое распространение. Но и в этом случае проявляются определённые ограничения коннекционистской теории познания. Люди (и многие менее развитые животные) демонстрируют способность обучения вследствие *единичных событий*. Так, например, животное, съевшее пищу, вызвавшую позднее у него желудочное расстройство, никогда не будет пробовать эту пищу снова. Обучающие коннекционистские методики совершенно не способны объяснить подобное «однократное» обучение.

3. Примеры того, на что способны нейронные сети

Сторонники коннекционизма значительно продвинулись в демонстрации возможностей нейронных сетей по выполнению понятийных задач. Ниже будут рассмотрены три широко известных эксперимента, которые окончательно убедили коннекционистов в том, что нейронные сети являются хорошими моделями человеческого интеллекта.

1) Одним из наиболее примечательных трудов, сделанных в этом направлении, была работа Сейновского и Розенберга (Sejnowski, Rosenberg – 1987) с нейронной сетью, названной NETtalk, которая могла читать английский текст. Множеством обучающих данными для NETtalk выступала объёмная база данных английских фраз и выражений с соответствующей фонетической транскрипцией для каждой фразы; вся информация хранилась в формате, пригодным для использования данных в синтезаторе речи. Интересно прослушать записи работы программы, сделанные на различных этапах её обучения. Сначала результатом работы был шум. Затем сеть издавала звуки, похожие на невнятное бормотание, спустя ещё какое-то время звук стал напоминать раздельное произношение, речь формировалась из отдельных звуков, имеющих сходство с английскими словами. В конце обучения NETtalk достаточно хорошо справлялся с произношением обрабатываемого текста. Более того, способность правильно произносить текст носила общий характер и распространялась на текстовые фразы, которые не были представлены в качестве обучающей информации.

2) Другой, более ранней коннекционистской моделью, также имевшей большое влияние, была сеть, обученная Румелхартом и Макклелландом (Rumelhart, McClelland – 1986) предсказывать прошедшее время английских глаголов. Задача представляет определённый интерес, так как, хотя большинство английских глаголов (регулярные глаголы) образуют прошедшее время прибавлением суффикса *-ed*, многие из наиболее широко употребляемых глаголов являются нерегулярными (*is/ was, come/ came, go/ went*). Сначала исходные данные для обучения сети содержали большое число нерегулярных глаголов и 460 регулярных глаголов. После 200 циклов тренировок сеть хорошо обобщила свою модель предсказания на глаголы, которые не были представлены в исходных данных. Она выдавала высокие результаты даже для нерегулярных глаголов. Во время обучения, на этапе, когда обучающие данные содержали преимущественно регулярные глаголы, система имела тенденции к «перерегулированию», т.е. комбинированию регулярных и нерегулярных форм: (*break/broked*, вместо *break/broke*). Подобное поведение было исправлено дальнейшими циклами обучения. Интересно заметить, что дети в процессе обучения языку демонстрируют схожую склонность к «перерегулированию». Однако, по вопросу о том, насколько хорошо в действительности модель Румелхарта и Макклиланда соответствует процессу обучения и обработки человеком окончаний глаголов, ведутся жаркие споры. Например, Пинкер и Принс (Pinker, Prince -1988) отмечают, что модель даёт неудовлетворительное обобщение для некоторых новых регулярных глаголов. Они полагают, что это показывает изначальную неадекватность коннекционистских моделей. Сети хорошо справляются с обработкой данных ассоциативными методами или подбором по шаблону, но имеют существенные ограничения в обработке общих правил, таких как образование формы прошедшего времени регулярных глаголов. Данные претензии ставят перед сторонниками и создателями коннекционистских моделей важный вопрос: может ли нейронная сеть должным образом обобщать правила, встречающиеся в когнитивных задачах. Несмотря на возражения Пинкера и Принса, многие коннекционисты уверены, что правильное обобщение модели выполнения задач подобного рода возможно. (Niklasson, van Gelder – 1994).

3) Важную роль в споре о том, может ли нейронная сеть обучиться выполнять правила, сыграла работа Элмана (Elman -1991), проведённая с сетью, которая могла оценивать грамматические структуры. Элман обучил сеть, содержащую механизм простой рекурсии, предсказывать следующее слово из множества слов английского предложения. Предложения формировались из простого словаря, содержащего всего 23 слова, с использованием подмножества правил английской грамматики. Даже такая простая грамматика ставила сложную задачу лингвистической оценки фразы. Допускался произвольный порядок следования различного числа одинаковых частей предложения при обязательном соблюдении согласованности подлежащего и сказуемого. Так, например, в предложении:

*Any **man** that chases dogs that chase cats...runs.*

(«Любой человек, который гонится за собаками, которые гонятся за кошками...бежит»).

Существительное единственного числа **man** («человек») должно сочетаться с глаголом *runs* («бежит»), о чём символизирует окончание *-s*, несмотря на находящиеся между ними существительные множественного числа

(*dogs, cats*), которые могут быть причиной выбора формы *run*. Одной из важных особенностей модели Элмана является использование рекуррентных связей. Значения скрытых узлов сохраняются во множестве так называемых «контекстных узлов», для того чтобы эти значения были переданы во входные узлы в следующем цикле выполнения. Выполнение такой петли от слоя скрытых узлов назад, к слою входных узлов, обеспечивает сеть элементарной формой памяти для согласования слов во входном предложении. Сеть Элмана показала возможность оценки грамматических структур в предложениях, которые не были представлены во множестве исходной обучающей информации. Разумеется, предсказание следующего слова в предложении на английском языке является невыполнимой задачей, однако сеть показывала неплохие результаты. Для определённой точки вводимого в сеть предложения выходные узлы, отвечающие за слова, грамматически связанные с выбранной точкой, должны быть активированы, а выходные узлы для остальных слов, наоборот, неактивны. После многочисленных циклов обучения сети Элман добился превосходного выполнения сетью данной задачи, в том числе и для предложений, которые не рассматривались на этапе обучения сети. Хотя были получены впечатляющие результаты, для решения задачи обработки языка ещё предстоит сделать огромную работу. Более того, были высказаны сомнения по поводу значимости результатов Элмана. Например, Маркус (Marcus – 1998, 2001) утверждает, что сеть Элмана не способна обобщать принципы своей работы для предложений, составленных из нового словаря. Это, по его заявлению, означает, что коннекционистские модели могут получать результат только на основании методов перебора и оказываются неспособными корректно обрабатывать информацию в соответствии с абстрактными правилами.

4. Достоинства и недостатки моделей на основе нейронных сетей

Философы проявляют интерес к нейронным сетям, потому что они могут дать новые аспекты понимания природы сознания и связи сознания с мозгом (Rumelhart, McClelland – 1986). Коннекционистские модели хорошо согласуются с тем, что мы знаем о нейронах. Мозг и в самом деле представляет собой некоторый аналог нейронной сети, сформированный из огромного количества элементарных узлов (нейронов) и их связей (синапсов). Более того, некоторые свойства моделей нейронных сетей дают возможность предполагать, что коннекционизм предлагает достаточно достоверную картину природы человеческого мышления. Нейронные сети демонстрируют необходимую гибкость в решении задач, которые перед ними ставит реальная жизнь. Искажение входных данных или разрушение элементов сети приводит к значительной деградации сети, однако в целом сеть продолжает обрабатывать информацию корректным образом, хотя и с определённой потерей точности. В противоположность этому результаты работы классических компьютеров при искажении входных данных или уменьшения циклов обработки оказываются абсолютно неудовлетворительными. Последствия подобных изменений являются для таких компьютеров катастрофическими. Нейронные сети достаточно хорошо справляются с решением задач, требующих параллельной обработки многих процессов, конфликтующих между собой. Для исследователей в области искусственного интеллекта совершенно очевидно, что такие понятийные задачи, как идентификация объектов, планирование и координация действий, относятся именно к подобному роду задач. И хотя клас-

сические системы также способны организовывать обработку множественных процессов и запросов, сторонники коннекционизма заявляют, что модель на основе нейронной сети предоставляет значительно более реалистичный механизм для работы с подобными задачами.

На протяжении многих веков философы прилагали все усилия, чтобы понять, как осуществляется процесс определения понятий. На сегодняшний день уже общепризнанным является то, что попытки мотивировать человеческое поведение, исходя из условий необходимости и достаточности, обречены на провал. Почти всегда для любой из предложенных моделей определения понятий находятся исключения. Например, можно понятийно определить тигра как некоторое существо, имеющего полосатую, чёрно-оранжевую шкуру, но при таком подходе мы не учитываем тигров-альбиносов. Философы и психологи, работающие в сфере определения понятий, доказали, что установление границ категорий должно осуществляться более гибким образом, например, за счёт использования понятия о сходстве в пределах семейства или подобии по отношению к определённом прототипу. Коннекционистские модели наиболее подходят для распределения системы разноуровневых понятий по категориям в соответствии с представленной моделью членства и определения границ категории. Сети могут распознавать очень тонкие различия в статистических моделях, определить которые на основании жестких правил часто нельзя ввиду невозможности формализовать задачу. Коннекционизм обещает объяснить гибкость и внутреннее устройство человеческого интеллекта, на основе методов, которые не выражаются в форме законов и которые содержат исключения из правил. Это позволяет избежать «хрупкости», присущей стандартной форме символического представления.

Несмотря на эти интригующие особенности нейронных сетей, коннекционистские модели имеют некоторые недостатки. Прежде всего, большинство исследований в области нейронных сетей абстрагируются от множества интересных и, возможно, важных особенностей человеческого мозга. Например, сторонники коннекционизма обычно не пытаются точно смоделировать как всё множество разнообразных нейронов, образующих мозг, так и работу нейротрансмиттеров (механизмов передачи электрических импульсов между нейронами) и гормонов. Более того, далёким от истины кажется предположение, что мозг содержит обратные связи, необходимые для обучения по алгоритму, подобному алгоритму обратного прохода. Огромное количество циклов обработки информации, необходимое для таких обучающих методов, также выглядит совершенно нереальным. Внимание к подобным вопросам будет, по-видимому, необходимым, если удастся построить убедительные коннекционистские модели интеллектуальной деятельности человека. Приходится также сталкиваться и с более серьёзными возражениями. Многие, особенно приверженцы классической теории, замечают, что нейронные сети неудовлетворительно выполняют задачи, решение которых основано на применении правил (имеются в виду понимание языка, объяснение причин и другие высшие формы мышления).

5. Коннекционистская форма представления информации

Коннекционистские модели предоставляют новую парадигму представления информации в нашем мозгу. Идея о том, что простые нейроны или цепочки из десятков нейронов являются самостоятельными структурами, предназначенными для хранения информации о каждой вещи в виде записи в

мозгу, представляется очень соблазнительной, но наивной. Например, мы можем представить, что существует специальный нейрон для хранения информации о бабушке, который активизируется каждый раз, когда мы о ней думаем. Как бы то ни было, такое локальное представление информации не похоже на то, что мы имеем в действительности. Абсолютно очевидно, что мысль о нашей бабушке задействует множество сложных систем нейронов, распределённых по обширным связным областям коры мозга.

Интересно, что подобное распределённое представление информации в большей степени, чем локальное представление в скрытых узлах, является естественным результатом коннекционистских обучающих методов. Примером этому могут служить активизирующие шаблоны, возникавшие в слое скрытых узлов нейронной сети программы NETtalk во время обработки текста. Анализ показывает, что сеть обучалась представлять такие категории, как слоги или гласные буквы, не посредством создания одного узла, активного для слогов, и другого, активного для гласных букв, а путём формирования двух различных шаблонов на протяжении всей последовательности скрытых узлов.

Локальная форма представления данных подобна тексту, напечатанному на листе бумаги, и является для нас простой и понятной. Модель распределённого представления выглядит непривычной и сложной для понимания. Однако данная техника демонстрирует важные достижения. Например, информация, представленная в распределённой форме (в отличие от символьных массивов в специально выделенной памяти), достаточно хорошо сохраняется при повреждении или перегрузке отдельных частей модели. Более важным является то обстоятельство, что родственные отношения между представлениями закодированы в категориях сходства и различий этих шаблонов. Т.е. сами внутренние свойства представлений несут в себе информацию о характере представленных данных (Clark – 1993: 19). Это обусловлено тем, что так как форма представления информации закодирована в соответствии с шаблонами, а не отдельно выполняемыми узлами, локальная форма представления данных является условной, никакие внутренние свойства представления не несут информации о связи символов между собой. Такое свойство самоописания распределённых представлений даёт возможность разрешить сложную проблему. В символической схеме все представления состояются из символьных единиц или атомов (так же, как отдельные слова в языке). Значения составных символьных строк могут быть определены на основании способа их построения, без учёта их конкретных значений. Но что взять за основу при определении значений атомов? Коннекционистская схема представления данных способна положить конец поискам ответа на этот вопрос путём простого отказа от атомов. Каждое распределённое представление является исполняемым всеми узлами сети шаблоном, поэтому не существует принципиального различия между простыми и составными представлениями. Разумеется, в конечном счёте представление определяется деятельностью отдельных узлов, но при распределённой организации ни один из этих «атомов» не содержит информации о каком-либо символе. Распределённые представления не символически в том смысле, в каком анализ их компонент не рассматривает символьный уровень как таковой.

Несимвольная форма представления имеет интересный выход на классическую гипотезу о том, что информация в мозге должна храниться в сим-

вольном виде, подобно предложениям языка. Таким образом, информация хранится в массиве «текстов» специального языка мышления (ЯМ). Природа коннекционистской формы представления информации бросает вызов данной идее. Не так-то просто сказать, с чем можно сравнить тезисы ЯМ. Однако Ван Гелдер (van Gelder – 1990) предложил широко одобренный тест для определения того, когда мозг можно считать содержащим представление информации в форме предложений. Имеются в виду те случаи, когда представление разбивается на лексемы, образованные из отдельных частей этих представлений. Например, если мы напишем: *Джон любит Мери* – то можно разбить написанное предложение на составные части: *Джон, любит, Мери*. Распределённые представления для составных выражений типа «*Джон любит Мери*» могут быть построены таким образом, что в них не будет содержаться точное представление их составных частей (Smolensky – 1991). Точная информация о составляющих элементах может быть извлечена из представления выражения, но модель нейронной сети в общем случае не нуждается в её извлечении для корректной работы (Chalmers – 1990). Это предполагает, что модели на основе нейронных сетей иллюстрируют идею, противоположную той, в соответствии с которой ЯМ является необходимым для процесса человеческого познания. Как бы то ни было, данный вопрос до сих пор является предметом активного обсуждения (Fodor – 1997).

6. Полемика между сторонниками классической теории и коннекционистами

Последние тридцать лет в философской науке доминировала классическая точка зрения, что процесс человеческого понимания (по крайней мере, в своей высшей форме) аналогичен символьной обработке в цифровых компьютерах. В классическом рассмотрении информация располагается в виде строк символов, то есть так же, как мы организуем данные в памяти компьютера или на листе бумаги. Иную точку зрения на данную проблему имеют сторонники коннекционистской теории. Они заявляют, что информация хранится в несимвольном виде и определяется весовыми соотношениями узлов нейронной сети. «Классицисты» верят, что процесс познания имеет сходство с обработкой информации цифровым процессором, при которой строки обрабатываются в последовательности, определяемой инструкциями программы (также организованной в символьном виде). С позиции коннекционистов познание рассматривается как динамический, распределённый по степеням обработки процесс, выполняющийся в нейронной сети, при котором активность каждого узла зависит от силы связей и активности соседних узлов, в соответствии с активизирующей функцией.

С первого взгляда эти позиции не имеют ничего общего. Однако многие коннекционисты не рассматривают свою работу как вызов классической теории, а некоторые даже открыто поддерживают классические представления. Так называемые «практические» коннекционисты ищут возможности согласования двух парадигм между собой. Они полагают, что нейронная сеть мозга содержит символьный процессор. Да, разум действительно описывается нейронной сетью, но существует также символьный процессор на более высоком и абстрактном уровне описания. Таким образом, роль коннекционистского исследования в соответствии с позицией «практических» коннекционистов заключается в определении того, как машинная информация, необходимая для работы символьного процессора, может быть сформирована из

данных нейронной сети. В этом случае классическая обработка информации вводится в рамки модели нейронной сети.

Однако многие коннекционисты противятся «практической» точке зрения. Это – радикальные коннекционисты. Они заявляют, что представления о символьной обработке является результатом неправильного предположения о принципах работы разума. Они указывают на то, что классическая теория не может объяснить значительную деградацию активизирующей функции, крайне низкий уровень обеспечения 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 является для коннекционистов симптомом того, что в когнитивной науке необходимо найти новую парадигму. Поэтому радикальные коннекционисты стремятся уничтожить модель символьной репрезентации информации как направление когнитивной науки.

На наш взгляд, более приемлем подход «практических» коннекционистов, синтезирующий наработки в области лингвистической философии с сублингвистическим подходом.

Что значит «быть роботом»? **Василий Крючков (Р-81)**

В основу доклада положена работа сторонника сильного искусственного интеллекта, одного из основоположников философии ИИ – Арона Сломэна «Что значит быть камнем?». В ней он защищает концепцию сильного ИИ, считая, что отвечать на возражения по поводу сильного ИИ – пустая трата времени. А. Сломэн придерживается позиций анти-редукционизма и не вступает в сферы дуализма и мистики в дискуссиях по проблемам сознания, самосознания, в том числе по проблеме компьютерного «сознания». Отстаивая возможность сознания у компьютерной системы (у робота) он придерживается формы дискурса, предложенной Т. Нагелем («Что значит быть летучей мышью», 1974), которая в общем задаётся вопросом «**Что значит быть X?**».

К «сознанию» робота ведёт путь через решение вопросов:

- Что значит быть камнем, лежащим на дороге?
- Что значит быть подсолнухом?
- Что значит быть летучей мышью?
- Что значит быть младенцем?
- Что значит быть больным на одной из последних стадий болезни Альцгеймера?
- Что значит быть умственно отсталым человеком?
- Что значит быть слепым?
- Что значит быть женщиной?

Оказывается, ответить на эти вопросы много сложнее, нежели на вопрос: «*Что значит быть роботом?*». Это обусловлено тем, что в случае с роботом человек имеет возможность изучить внутреннее строение информационных структур системы и их связи с внешним поведением.

Конечно, робот лишь частично обладает нашими способностями восприятия мира. Поэтому он никогда не сможет: до конца прочувствовать, что зна-

чит быть серфингистом, мчащимся на гребне волны; ощутить на себе действие центробежных и гравитационных сил; дуновение ветра в лицо, слегка развевающее волосы; услышать крик напуганного ребёнка; испытать сухость во рту от собственного страха.

Но робот, тем не менее, может многое «знать». Например, он может «знать», что значит быть серфингистом путём анализа информации о ситуации в которой серфингист находятся. Он может многое «знать» о механизмах человеческого восприятия, человеческих мотивов, человеческих эмоций и пр. Другими словами, он может получить колоссальное знание о возможностях человеческого восприятия, о познавательных функциях, т.е. о том, что составляет наше с вами существо, даже если прочувствовать всё это он полностью не может.

Более того, ответ на вопрос, что *значит быть роботом* (робот ведь конструктивно очень сильно отличается от людей), может отчасти включать в себя ответ на вопрос, что *значит быть человеком*. Робот может «знать» лучше и больше о том, что значит быть человеком, чем о том, что значит быть летучей мышью, по причинам, по которым знания человека (разработчика) о летучих мышах ограничены, т.е. по причинам: а) недостатка информации; б) неподготовленности нашего концептуального аппарата о системе восприятия летучей мышью информации и, вследствие этого, невозможности разработать соответствующие робототехнические механизмы.

Многие философы утверждают (к примеру, Т. Нагель), что робот может симулировать поведение людей без всякого осознания того, что значит быть человеком. Точно также он может симулировать и поведение летучей мыши. Поэтому нельзя определённо точно сказать, что именно симулирует робот. Отвечая на данное возражение, А. Сломан считает, что робот по определению не может функционировать без сложных механизмов обработки информации. А эти механизмы связаны с внутренними структурами и процессами анализа информации (Сломэн, 1994). Робот при этом получает информацию не только от различных окружающих его объектов, но и о самом себе – о своих внутренних структурах и состояниях (Маккарти, 1995).

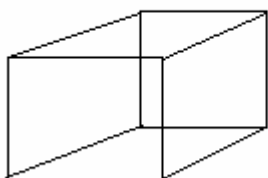
Некоторые внутренние состояния робота невозможно «перенести» в наш мозг, как и в случае с летучей мышью. Но мы можем описать эти состояния и выявить семантику правил функционирования. Каждая такая система анализа информации будет содержать столько собственных точек зрения, сколько присуще человеку.

Сложно помыслить то, что робот, функционирующий подобным образом, может быть нечто вроде «зомби» – то есть неосознанно выполнять действия, без каких бы то ни было ощущений, точек зрения и т.п. Те, кто это утверждает, занимаются самообманом, как и те, кто считает, что можно изобрести метод точного измерения трисекции любого угла, пользуясь лишь линейкой и транспортиром. Им следует поучиться инженерному делу и тогда они поймут: такой робот-зомби работать не будет. Для хорошего инженера представить себе работа-зомби означает отказаться от своих знаний о том, как робот обрабатывает воспринимаемую информацию, ставит новые задачи, принимает решения, оценивает степень решения задач или справляется с неожиданными повреждениями.

А. Сломэн твёрдо уверен в возможности того, что высокоуровневый процесс обработки информации любой системы, построенной по подобию человеческого процесса мышления, обладает собственной онтологией. Но

следует при этом быть осторожным: теоретически можно построить огромную таблицу, в которой представлены все возможные отношения между сенсорными входными данными и внешними реакциями. Если бы такая таблица было физически возможной (а это не так для существ, обладающих способностями, схожими с человеческими), тогда машине, функционирующей на её основе, не требуется человекоподобная внутренняя онтология. Такая машина была бы больше похожа на камень. Если есть желание представить себе зомби, то следует попытаться создать робота, поведенчески неотличимого от человека и способ функционирования которого основан на такого рода таблице.

А. Сломан считает, что это невозможно и в защиту своего тезиса приводит предполагаемую способность робота-«незомби» воспринимать опрокидывания куба Неккера (см. рисунок).



Рассмотрим, что происходит, когда вы смотрите на куб Неккера: внезапно он опрокидывается, хотя, изображение на сетчатке и видимая 2D структура остаются неизменными, но 3D интерпретация уже другая. Линии, а точнее грани куба, которые ранее были направлены вниз от наблюдателя, обратились и стали направлены вверх от него. Верхняя квадратная плоскость куба, которая ранее простиралась от на-

блюдателя, теперь простирается к нему.

На примере куба видно, что неккеровское «превращение» вполне приемлемо можно описать. В будущем ученые, изучающие деятельность мозга, выяснят, какие неврологические процессы отвечают за «переключение» – переход из одного состояния видения в другое состояние. Возможно, что и зрительные системы роботов достигнут такого уровня, на котором смогут улавливать подобные вещи, то есть будут обладать способностью локально управлять двусмысленными фрагментами картины, которые могут иметь различную интерпретацию в зависимости от контекста общей картины.

Далее существенным вопросом построения незомбированных роботов, обладающих «самосознанием», является вопрос выяснения того, *а имеется ли у компьютеров точка зрения (своя собственная позиция)?*

В настоящее время существует большое количество сильно отличающихся друг от друга компьютерных систем. У них нет ничего подобного нашим структурам обработки информации – например, их функции хранения информации очень сильно отличаются от человеческой памяти (механизмы получения, доступа, обработки и «забывания» информации отличаются принципиально). Роботам не нужны человеческие мотивации, у них свои способы побуждения к решению задач.

Из-за этих отличий между компьютерными системами и людьми (или другими животными) наше знание о том, что значит быть роботом, то есть иметь свою собственную точку зрения, ощущать себя, как ощущает себя робот, не многим отличается от нашего знания, что значит, быть камнем. Всё это – информация «от третьего лица». Но в отличие от примера с камнем, рассуждать о роботе не так скучно. Хотя что-то известно.

Так или иначе, с развитием систем ИИ, появлением у них самостоятельности в механизмах мотивации, усложнением «понятийного» аппарата, контролирующего процессы восприятия и устанавливающего семантику внутренней информации, будет возникать всё больше и больше вопросов на тему, как робот классифицирует вещи, в зависимости от ситуации и что значит «быть такой системой с точки зрения этой системы». Например, что значит желание робота сделать что-то в зависимости от ситуации? Что значит решить задачу с позиции робота. Что такое хорошо, а что такое плохо – с точки зрения робота?

Роботы по мере своего развития будут всё больше и больше постигать то, что значит быть роботом «изнутри». Конечно, этот процесс длительный. Возможно, пройдет сотни лет, прежде чем роботы выйдут на уровень шимпанзе и тысячи лет, когда они достигнут уровня человека. Но когда это осуществится, некоторые из роботов выдвинут вопросы – «Что значит быть камнем, летучей мышью и т.п.»? И они спросят: *«Что значит быть человеком?»*

О распределении функций между человеком и компьютером в информационно-коммуникационных технологиях

Иван Михейкин (МС-81)

Развитие информационно-коммуникационных технологий, широкое применение компьютеров на работе, отдыхе, в быту актуализирует проблему соотношения возможностей человеческого мышления и искусственного интеллекта.

В литературе, посвященной философским проблемам кибернетики, рассмотрение проблемы соотношения мышления человека и интеллектуальных возможностей ЭВМ нередко подменялось вопросом: «Могут ли машины мыслить?», который ставился в контексте бихевиорального анализа мыслительной деятельности. При этом понятие «машина» трактовалось слишком абстрактно, а «мышление» определялось в терминах формальной логики и машинных операций. Предполагалось также, что развитие «машинного мышления» столкнется с чисто техническими трудностями, которые будут преодолены, подобно тому, как авиация, постепенно совершенствуясь, преодолела «звуковой барьер» (см. *Ник Бостром*, «Сколько осталось до суперинтеллекта?»).

Поводом к поспешным выводам послужили первые успехи в решении достаточно простых задач, допускающих их полную алгоритмизацию, программирование и последующую автономную обработку информации на ЭВМ. Казалось, что дальнейшее увеличение скорости обработки информации приведёт и к возможности решения более сложных задач. Но при решении сложных задач оказалось, что их нельзя полностью формализовать и создать для их решения программы, так как основным этапам решения этих задач присуще творчество – специфически человеческая способность. Поэтому актуальной задачей кибернетики стала разработка вопроса распределения функций человека и ЭВМ в ходе решения задач.

Тем не менее, развитие методологии компьютерного моделирования неуклонно направлено в сторону вытеснения человека из информационно-коммуникационных систем. Человек не должен быть участником процесса переработки информации, он должен стать пользователем. И при решении во-

проса распределения функций между компьютером и человеком встал ключевой вопрос – *можно ли смоделировать человеческое сознание?*

Считается, что у животных сознание возникает как способ адаптации к окружающей среде. Быстрая адаптация (в сравнении с временем жизни животного) требует способностей предвидения и планирования. Мотивом адаптации служат биологические жизненные потребности организма. Одним из механизмов адаптации является самообучение. Самообучение не пассивно. Лишь только в активной форме, продуцируя и сопровождая деятельность, самообучение может стать толчком к сознанию. Помимо этого, признаком сознания является *осознавание своего существования, осознавание ощущений*. Самообучение и осознавание ощущений обеспечивают выработку оптимального поведения. Они дают биологическому организму преимущества перед другими системами, которые обладают меньшими «степенями» сознания или не обладают сознанием вообще.

Сознание – это не мозг, не поведение, а особый способ обработки информации. Особое значение для теоретико-концептуального осмысления этого способа приобретает информационный подход к сознанию, предложенный тридцать лет назад Д.И. Дубровским¹. Актуальность данной концепции существенно возрастает, но уже не в контексте решения общей философской проблемы «дух/тело», а в контексте проблематики искусственного интеллекта – задачи функционального воспроизводства феноменов человеческого сознания на ином материальном субстрате, в частности, в компьютерной среде.

Осуществима ли эта задача? На наш взгляд, нет. Сравним с позиции теоретико-деятельностного подхода принципы работы искусственных систем и собственно человеческой деятельности:

Операции: машина воспроизводит операции человеческого мышления, и, следовательно, соотношение «машинного» и «немашинного» есть *соотношение операционального и неоперационального в человеческой деятельности*.

Цель: в искусственных системах «целью» называют некоторую конечную ситуацию, к которой стремится система, для человека же характерно не просто достижение предзаданных, но и *формирование новых целей*.

Оценки: у искусственных систем есть своего рода «ценностные ориентации». Специфику человеческой мотивационно-эмоциональной регуляции деятельности составляет использование не только константных, но и *ситуативно возникающих и динамично меняющихся оценок*.

Более того, если подойти с чисто теоретической, теоретико-модельной стороны к решению вопроса соотношения естественного (сознательного) мышления и машинного (бессознательного) интеллекта, следует заметить, что нет смысла говорить о полном тождестве оригинала и модели. Невозможно в «один и тот же сундук» помещать человека, его сознание и машину, воспроизводящую частные аспекты сознательной деятельности. В силу этого в информационно-коммуникационных технологиях всегда останется человеческий компонент, способствующий сознательному достижению целей, ко-

¹ Имеется в виду ряд работ Дубровского Д.И. Информационный подход к проблеме «сознание и мозг» // Вопросы философии, 1976, № 11; Расшифровка кодов (Методологические аспекты проблемы) // Вопросы философии, 1979, № 12; Информация, сознание, мозг. М., 1980.; Проблема идеального. М., 2002, 1983, гл. IV.; Психика и мозг: результаты и перспективы исследований // Психологический журнал, 1990, № 6.

торые как ставятся извне, так и продуцируются изнутри сложной человеко-машинной системы.

«Философия искусственного интеллекта» в Интернет-среде Нестеров Олег (ИС-81)

Развитие философии ИИ немыслимо без построения информационной базы теоретических наработок. Необходима систематизация, каталогизация, классификация работ, построение электронных библиотек, создание навигаторов по сайтам, посвящённым проблематике ИИ.

В данной работе представлены результаты систематизации интернет-ресурсов по издательствам, в которых издавались работы по философии искусственного интеллекта.

Выборка осуществлялась из базы, содержащей 6200 работ по философии ИИ и философии сознания. Было обнаружено 778 издательств. В нижеследующем списке представлена первая сотня издательств. Издательства упорядочены по количеству выпущенных статей и книг. Последнее число указывается количество источников. Основой для поиска послужила классификация работ по философии ИИ, предложенная Д. Чалмерсом. Последняя цифра в строке означает количество обнаруженных на данном сайте изданий. Читатель может скопировать ссылку и перейти на сайт, предложенный в данном списке:

- ❑ Издательство Оксфордского университета; Oxford University Press; <http://www.oup.com/>; 228
- ❑ Издательство MIT Press; MIT Press;
<http://mitpress.mit.edu/main/home/default.asp?sid=60000F91-526E-4169-8D7C-4F42DDFC09D3>; 285
- ❑ Philosophical Studies; <http://www.kluweronline.com/issn/0031-8116/contents>; 209
- ❑ Synthese; <http://www.kluweronline.com/issn/0039-7857/contents>; 152
- ❑ Philosophical Psychology; <http://mechanism.ucsd.edu/~pp/index.html>; 149
- ❑ Journal of Philosophy; <http://www.journalofphilosophy.org/>; 136
- ❑ Кембриджский университет; Cambridge University Press; <http://uk.cambridge.org/>; 133
- ❑ Mind and Language; <http://www.ingenta.com/journals/browse/bpl/mila>; 132
- ❑ Интернет магазин по продаже книг.; Blackwell;
<http://www.blackwell.co.uk/bobuk/scripts/welcome.jsp>; 130
- ❑ Philosophy and Phenomenological Research;
<http://www.brown.edu/Departments/Philosophy/ppr.html>; 120
- ❑ Mind; <http://www3.oup.co.uk/mind/>; 119
- ❑ Australasian Journal of Philosophy; <http://www3.oup.co.uk/ajphil/contents/>; 109
- ❑ Minds and Machines; <http://www.kluweronline.com/issn/0924-6495>; 91
- ❑ Publishers of academic and professional books, journals, and software.; Lawrence Erlbaum;
<http://www.erlbaum.com/index.htm>; 87
- ❑ Philosophy of Science; <http://www.journals.uchicago.edu/PHILSCI/>; 81
- ❑ Journal of Consciousness Studies; <http://www.imprint.co.uk/jcs.html>; 80
- ❑ Consciousness and Cognition; <http://www.sciencedirect.com/science/journal/10538100>; 68
- ❑ Philosophical Quarterly; <http://www.ingenta.com/journals/browse/bpl/phiq>; 64
- ❑ American Philosophical Quarterly; <http://www.press.uillinois.edu/journals/apq.html>; 62
- ❑ Philosophical Review; <http://www.arts.cornell.edu/philrev/>; 62
- ❑ British Journal for the Philosophy of Science; <http://www3.oup.co.uk/phisci/>; 56
- ❑ Journal of Mind and Behavior; <http://www.ume.maine.edu/~jmb/welcome.html>; 54

- ❑ Proceedings of the Aristotelian Society; <http://www.ingenta.com/journals/browse/bpl/paso>; 53
- ❑ Интернет-издательство; John Benjamins; <http://www.benjamins.com/cgi-bin/welcome.cgi>; 52
- ❑ Kluwer; <http://www.wkap.nl/>; 48
- ❑ Southern Journal of Philosophy; <http://www.people.memphis.edu/~philos/sjp/>; 47
- ❑ Academic Press; <http://www.academicpress.com/>; 46
- ❑ Inquiry; <http://www.tandf.co.uk/journals/online/0020-174X.html>; 46
- ❑ Pacific Philosophical Quarterly; <http://www.ingenta.com/journals/browse/bpl/papq>; 44
- ❑ Philosophical Topics; <http://www.uark.edu/depts/philinfo/pt/oldindex.html>; 43
- ❑ Behavioral and Brain Sciences; <http://www.bbsonline.org/>; 42
- ❑ Routledge; <http://www.routledge.com/>; 42
- ❑ Nous; <http://www.ingenta.com/journals/browse/bpl/nous>; 40
- ❑ Canadian Journal of Philosophy; <http://www.uofcpress.com/UCP/CJP.html>; 34
- ❑ Erkenntnis; <http://www.kluweronline.com/issn/0165-0106/contents>; 32
- ❑ Cognition; <http://www.sciencedirect.com/science/journal/00100277>; 32
- ❑ Psychological Review; <http://www.apa.org/journals/rev.html>; 30
- ❑ Aristotelian Society Supplement; <http://www.ingenta.com/journals/browse/bpl/supa>; 29
- ❑ Dialogue; <http://www.usask.ca/philosophy/dialogue/>; 26
- ❑ Ridgeview; <http://www.ridgeviewresources.com/>; 25
- ❑ Ferdinand Schoningh; <http://www.abe.pl/html/english/search.php?pubcode=fs>; 25
- ❑ Harvard University Press; Harvard University Press; <http://www.hup.harvard.edu/>; 25
- ❑ Aana Journal; <http://www.aana.com/help/subscribe.asp>; 25
- ❑ Neuropsychologia; <http://www.sciencedirect.com/science/journal/00283932>; 23
- ❑ Nature; <http://www.nature.com/nature/>; 21
- ❑ Philosophia; <http://www.philosophia.dk/>; 20
- ❑ Behavior and Philosophy;
http://www.behavior.org/journals_BP/index.cfm?page=http%3A/www.behavior.org/journals_BP/BP_contents.cfm; 20
- ❑ Metaphysics; <http://www.reviewofmetaphysics.org/>; 19
- ❑ Midwest Studies in Philosophy; <http://www.ingenta.com/journals/browse/bpl/misp>; 19
- ❑ Journal of Experimental Psychology: General; <http://www.apa.org/journals/xge.html>; 19
- ❑ Wiley; <http://www.wiley.com/WileyCDA/>; 18
- ❑ University of California Press; <http://www.ucpress.edu/>; 18
- ❑ American Psychologist; <http://www.apa.org/journals/amp.html>; 18
- ❑ Dialectica; <http://www.dialectica.ch/>; 17
- ❑ Plenum Press; http://isbndb.com/d/publisher/plenum_press.html; 16
- ❑ Ratio; <http://www.ingenta.com/journals/browse/bpl/rati>; 16
- ❑ Protosociology; <http://www.protosociology.de/>; 15
- ❑ Psychological Bulletin; <http://www.apa.org/journals/bul.html>; 15
- ❑ Lippincott-Raven; <http://www.abe.pl/html/english/search.php?pubcode=li>; 15
- ❑ Ablex; <http://www.ablex.com/>; 15
- ❑ Prentice-Hall; <http://www.prenticehall.com/>; 15
- ❑ Seminars in Neurology; http://www.medscape.com/viewpublication/142_index; 14
- ❑ De Gruyter; <http://www.degruyter.com/>; 14
- ❑ Josiah Macy Foundation; <http://www.josiahmacyfoundation.org/jmacy1.html>; 14
- ❑ Artificial Intelligence; <http://www.jair.org/>; 14
- ❑ Plenum; <http://www.plenum.de/index.jsp>; 14
- ❑ Journal for the Theory of Social Behavior; <http://www.ingenta.com/journals/browse/bpl/jtsb>; 13
- ❑ University of Chicago Press; <http://www.press.uchicago.edu/>; 13
- ❑ Science; <http://www.sciencemag.org/>; 12
- ❑ Acta Analytica; <http://rcum.uni-mb.si/~actaana/>; 12

- Reprinted in *Supervenience and Mind* (Cambridge University Press; <http://www.blackwell-syn-ergy.com/links/doi/10.1111/1468-0068.00447/abs/?jsessionid=cjs7tacCP27h>); 12
- *New Ideas in Psychology*; <http://lib.harvard.edu/e-resources/details/n/newidpsy.html>; 12
- Princeton University Press; <http://pup.princeton.edu/>; 12
- *Monist*; <http://www.monist.de/>; 12
- *Proceedings and Addresses of the American Philosophical Association*; <http://www.apa.udel.edu/apa/>; 11
- *American Journal of Psychology*; <http://www.press.uillinois.edu/journals/ajp.html>; 11
- *Minnesota Studies in the Philosophy of Science*; <http://www1.umn.edu/mcps/center/mnstud.html>; 11
- *Metaphilosophy*; <http://www.ingenta.com/journals/browse/bpl/meta>; 11
- University of Minnesota Press; <http://www.upress.umn.edu/>; 11
- *Journal of Experimental and Theoretical Artificial Intelligence*; <http://elib.cs.sfu.ca/Collections/CMPT/cs-journals/P-TaylorFrancis/J-TaylorFrancis-JETAI.html>; 11
- Springer-Verlag; <http://www.springeronline.com/sgw/cda/frontpage/0,10735,1-102-0-0-0,00.html>; 11
- *Psyche*; <http://psyche.cs.monash.edu.au/index.html>; 11
- Reidel; <http://www.baader-meinhof.com/who/sympathizers/reidelhelmut.html>; 10
- Psychology Press/Taylor & Francis; <http://www.tandf.co.uk/journals/contacts.asp>; 10
- Routledge and Kegan Paul; http://isbndb.com/d/publisher/routledge_kegan_paul.html; 10
- *Consciousness & Cognition*; <http://www.sciencedirect.com/science/journal/10538100>; 10
- New York University Press; <http://www.nyupress.org/>; 10
- *Proceedings of the National Academy of Sciences USA*; <http://www.pnas.org/>; 10
- Журнал исследований по искусственному интеллекту; *Journal of Artificial Intelligence Research*; <http://www.cs.washington.edu/research/jair/home.html>;
- Когнитивные и психологические науки в Интернет; *Cognitive and Psychological Sciences on the Internet*; <http://www-psych.stanford.edu/cogsci/>;
- *Imagination, Mental Imagery, Consciousness, Cognition: Science, Philosophy & History.*; <http://www.calstatela.edu/faculty/nthomas/home.htm>.

МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ НАНОТЕХНОЛОГИИ

Марк Пак (Э-81), Алексей Панов (Э-81)

Устройства микроэлектромеханических систем (MEMS) действуют, как и устройства макроразмеров и даже выглядят также – с моторами, передачами и рычагами, изготовленными из стекла, керамики или металла.

Наноразмерные структуры, в частности NEMS – будут строиться и действовать совершенно по-другому: они формируются и функционируют на основе других физических законов. На молекулярном уровне перестают действовать законы механики, используемые для расчетов узлов обычных машин. Законы сопротивления материалов и гидравлики уже не применимы – вместо этого вступают в действие законы *квантовой механики*, которые приводят к совершенно неожиданным, с точки зрения классической механики, последствиям.

Сегодня практическая нанотехнология ориентирована на решение следующих задач:

- создание твердых тел и поверхностей с требуемой молекулярной структурой;

- создание новых химических веществ посредством конструирования молекул (с участием и без участия химических реакций);
- разработка устройств различного функционального назначения —компонентов наноэлектроники, нанооптики, наноэнергетики, нанороботы и нанокomпьютеры, нанолекарства, наноинструменты и т.д.;
- создание наноразмерных самоорганизующихся и самореплицирующихся структур.

Инструментальный базис нанотехнологий, позволяющий исследователям не только визуализировать атомные структуры, но и манипулировать отдельными атомами и строить новые молекулы, основан на использовании так называемого *эффекта туннелирования электронов*. Его применение на вершинах зондов специальных конструкций позволило достичь высокой пространственной разрешающей способности управления атомно-молекулярными реакциями в отличие от известных групповых технологий осаждения материалов, методов оптической литографии, эпитаксии, а также электронной литографии, где высокая энергия фокусируемых электронов приводит к значительному разрушению используемых материалов.

За двадцать с небольшим лет с момента появления техники сканирующей зондовой микроскопии и изобретения сканирующего туннельного, а затем и атомно-силового микроскопов, в разных странах были получены впечатляющие результаты по наблюдению наноразмерных частиц и структур на их основе и поставлена задача создания технологических машин, позволяющих осуществить атомно-молекулярную сборку вещества и конструирование отдельных узлов и устройств различного функционального назначения.

Внедрение наносхемотехники и нанороботов позволит создать микроскопические компьютеры небывалой производительности. Более того, они станут саморемонтирующимися и самовоспроизводящимися. Применение десятиатомных транзисторов позволит подойти вплотную к имитации мыслительных процессов человека и создать благоприятные предпосылки для разработки искусственного интеллекта, соразмерного с человеческим.

Естественно-языковой интерфейс: три похода к моделированию «смысла»

Иван Подопригора (А-82)

Перспективы технологии построения естественно-языковых интерфейсов связываются с моделированием смысла лингвистических выражений. Модель «смысла» задаёт непосредственный и самый верхний уровень «взаимопонимания» человека и компьютера.

Методологические основания моделей смысла принято связывать с «контекстуальной» трактовкой понятия «смысл»¹. Основное для семантики отношение между знаковым выражением и его интерпретацией при детальном анализе оказывается не бинарным, а тернарным, поскольку само понятие интерпретации расслаивается на экстенциональный и интенциональный уровни. Знак характеризуется, с одной стороны, обозначаемым им предметом, значением, а с другой — свойствами значения, смыслом, выражаемого

¹ *Алексеев А.Ю.* Компьютерное моделирование смысла (философско-антропологический анализ). Автореферат диссертации, М., 2004

этим знаком, т. е. понятием о значении. В контекстуальной трактовке, смысл — это информация, которую знак несет о возможных значениях слова (т.е. это не вся информация о значении, а только та ее часть, которая отражается данным знаком), об их положении в системе реалий, об их месте в универсуме (множестве значений)¹.

Выбор значения определяется конкретной знаковой ситуацией, т.е. зависит от контекста² (смысла знаковой ситуации). Смысл знака, в противоположность значению собственно и характеризует контекст, задающий «траекторию» задания референта знака в различных знаковых ситуациях.

Логически полным представляется выделение трех подходов к реализации моделей «смысла» для ЕЯ-интерфейсов: экстенционально-интенциональный, интенциональный и внелингвистический подход.

1. *Экстенционально-интенциональный подход*

Ориентирован на жесткую привязку знака к значению слова. Совокупность этих значений и составляет «контекст». По сути, «контекст» — это агрегированная совокупность «текстовых» элементов.

Данный подход возник в программе логического атомизма Б. Рассела, логического позитивизма Р. Карнапа, лингвистического позитивизма Л. Витгенштейна.

Основной лингвистической единицей выступает пропозиция (предложение, задающее представление факта).

Возникает целый ряд очевидных проблем, например, проблема моделирования смысла предложения — как найти «целое» для экстенционально заданных значений?

Для решения этой проблемы в экстенциональном подходе применяются основополагающие принципы:

1) *композициональности* — семантическое значение сложного выражения полностью определяется семантическим значением его составных частей (их значений);

2) Любое сложное выражение толкуется *истинностно-функционально* — как функтор, приложимый к аргументам и затем вычисляется его семантическое значение в соответствии с заданными правилами. Т.е. происходит явное указание значений в универсуме.

Имеется множество «контекстов» естественного языка, для которых указанные принципы либо не действуют, либо недостаточны, либо требуют уточнения. Речь идет о модальных, временных и интенциональных контекстах. Они либо в целом, либо в отдельности предполагают нечто большее, чем простое указание на значение составных частей. Имеется некоторая подразумеваемая информация, которая явно не фиксируется, но от которой по существу зависит значение выражения³.

Анализ предложений содержащих такой контекст в современной логике осуществляется в рамках интенционального подхода, основанного на принципах *семантик возможных миров*. Т.е. сущности задаются в терминах

¹ Попов Э.В. Общение с ЭВМ на естественном языке, М., 1982

² Контекст [лат. contextus сплетение, соединение] — законченный в смысловом отношении отрывок текста, точно определяющий смысл отдельного входящего в него слова или фразы.

³ Герасимова И.А. «Формальная грамматика и интенциональная логика», ИФ РАН, М., 2000

системы знаний с помощью некоей совокупности свойств, выделяющих эти сущности из универсума.

В рассмотрение вводится не одно положение дел, описываемое моделью, а некоторое множество положений дел, возможно, связанных с собой определенными отношениями. Строится модель, имеющая сложное структурное строение. Разница между экстенциональными и интенциональными выражениями существенна.

Р. Карнап определяет *интенционал выражения* (уточняющий понятие «смысла») как функцию, определенную на возможных описаниях состояний и выделяющую денотат выражения (названный им *экстенционалом*) для каждого описания состояния. Монтегю в своих работах расширяет понятие возможного мира (описания состояния), используя термины *индекс* и *точка соотнесения*. Последние понимаются им как комплексный набор координат, от которых зависит истинный статус высказывания. В комплекс могут входить такие координаты, как возможный мир, момент времени, пространственное расположение, субъект и многие другие.

Интенциональные логики — это специальные языки, которые содержат выражения, явно указывающие как на интенционал, так и на экстенционал.

Следует отметить, что охарактеризовать полностью выражение его экстенционалом невозможно, так как при указании на денотат не ясно, какое из его свойств имеется в виду. Отметим, что выбор денотата по знаку (в данном универсуме) может зависеть от контекста, а концепт (в любом универсуме) постоянно присущ знаку в данной знаковой системе, т. е. в данной системе знаний.

Среди достоинств интенционального подхода отмечают следующие: строгий теоретико-множественный подход, использование композиционной и рекурсивной семантики, как для экстенциональных, так и для интенциональных терминов, требование гомоморфизма как центральное для подхода Монтегю, структурная адекватность грамматике для логических языков, корреляция логических и грамматических форм (категории и типы).

Среди многочисленных трудностей и недостатков называют следующие: анализ кванторных слов во многом искусственен и не отражает многообразия речевых ситуаций, не решается проблема анафорических выражений и взаимозависимых контекстов, понятие интенционала слишком элементарно и не отражает структуру модальных взаимосвязей в контексте, предполагается, что знание говорящего исчерпывается содержанием возможных миров, не учитывается динамика когнитивной коммуникации между говорящими.

В целом, такой логико-позитивистский подход ярко охарактеризовал один из основателей философии искусственного интеллекта — Дж. Маккарти, назвав его «*философской ловушкой для исследователей искусственного интеллекта*»¹.

2. Интенциональный подход

«Смысл» как контекст рассматривается с позиции целого. Уже явно выражены два уровня — текст/контекст. Контекст — это и есть то органическое целое, несводимое к механическому агрегату из значений, присущего предыдущему подходу. Философские основания, в большей мере, феноменологи-

¹ Дж.Маккарти «Что общего у ИИ и философии» <http://www.formal.stanford.edu/jms/>, 1995г.

ческие (Брентано, Гуссерль, Мейнонг). Конструктивные ориентиры прослеживаются в формальной феноменологии В.И. Васюкова.

Гуссерль анализирует содержание понятия знака и утверждает, что предметом научного интереса ученого является не вещь-в-себе, а понятие о вещи-в-себе, как единство значения и смысла.

Значение предмета, понимается, как нозма, т.е. мысленное содержание о предмете, или, другими словами, предметное содержание мысли. Знак, в интенциональном смысле – это материальный предмет, воспроизводящий свойства или отношения другого, незнакового по природе предмета. С помощью знака, или знаковой системы, можно отобразить нозму.

Понятие о предмете – это и есть совокупность признаков предмета, причем признаков существенных и необходимых. Эти фиксируемые в понятии признаки представляют собой свойства исследуемых предметов, их способности вступать в определенные отношения с другими предметами. Если «очиститься» от признаков, полученных в процессе познавательной деятельности, т.е. в процессе определенных субъект-объектных отношений, и содержащих в себе признаки этих самых отношений, предмет мысли, он же вещь, становится «абсолютной субъективностью». Предмет мысли становится чистым нозисом (от греч.: «noesis» – «мышление»). Явление вещи не есть являющаяся вещь: сами явления не являются, они *переживаются*. Таким образом «абсолютная субъективность» — это сознание, направленное к предмету, или поток сознания¹.

Исходной и фундаментальной характеристикой сознания поэтому является предметность. Сознание предметно потому, что оно интенционально². На каждом шагу сознание являет собой непрерывный выход за свои собственные пределы – выход к предмету. Сознание формирует смысл о предмете. Интенциональность является смыслоформирующей направленностью к предмету, она не наличествует, она функционирует. Интенциональность сознания невозможно описать как нечто постоянное, как субстанцию. В данном подходе предмет понимается как процесс своего собственного смыслопорождения. Происходит выдвижение на первый план смысловой (смыслоформирующей) связи субъекта и предмета. Таким образом, происходит разделение сознания на *явление сознанию*. (содержание сознания, нозма) и *феномен сознания* (смыслоформирование, нозис).

Недостатки интенциональных подходов — низкая степень формализуемости, априорность феноменологических конструкций, а отсюда — некритическое восприятие «смысла», смысловых «жизненных горизонтов» (Э.Гуссерль) с позиции авторитетного носителя языка.

3. Внелигвистический подход

«Смысл» как контекст в данном подходе задается с внешней стороны относительно языковой теории, т.е. со стороны реальных или имитационных действий. По сути, такой контекст принципиально не может быть задан языковым выражением. «Смысл» может быть лишь непосредственно показан, продемонстрирован, симитирован – то есть раскрыт *функциональным спосо-*

¹ А.Н. Суворова «Введение в современную философию» www.philosophy.ru/edu/ref/suvorova/08.html

² Интенциональность [лат. intentio стремление] — направленность на объект

бом. В рамках этого подхода преимущество получает *парадигма функционализма* – как собственное философское основание искусственного интеллекта.

В лингвистике единицей моделирования «смысла» может выступать пресуппозиция. В самом общем виде пресуппозицию можно определить как *внеречевое условие речевого акта*¹.

Всякое предложение неизбежно несет добавочную информацию о тех условиях, которым оно удовлетворяет. Эти условия создают «невидимый» подтекст у всякого «видимого» текста. Можно сказать, что основная задача проблемы пресуппозиции заключается в том, чтобы истолковать этот «невидимый» подтекст. Пресуппозиции, за исключением логических, не являются универсальными (не зависящими от конкретного языка). Они видоизменяются от одного языка к другому и во многом зависят от структурных качеств языков.

Некоторые продуктивные положения внелингвистического подхода прослеживаются в теоретико-деятельном подходе Щедровицкого, семиотическом подходе В.М.Розина, лингвистическом бихевиоризме Л. Витгенштейна, в теории речевых актов Райла-Остина и др. – то есть в тех подходах, где фиксируется *функциональная связь* между реальным актом и знаком. «Смысл» — лишь функция; это нечёткий и невыразимый контекст как способ задания значения между актом («значением») и знаком («текстом»). Акт — это человеческий акт и его имитация в компьютерной среде, а знак — человеческое естественное слово и имитация этого слова на искусственном языке. По сути, проблема построения ЕЯ-интерфейса представляется проблемой реализации теста Тьюринга, что дает основание считать убедительным мнение Г.С. Пospelова о сводимости всей проблематики искусственного интеллекта к построению ЕЯ-интерфейса, способного «понимать» смысл того, что человек хочет получить от компьютерной системы.

Клоны и «полуискусственный интеллект» **Ирина Смирнова (ММ-101)**

Под искусственным интеллектом понимается междисциплинарное направление в компьютерной науке, ориентированное на построение машин, функционально воспроизводящих частные особенности человеческого мышления. Материальным субстратом искусственного интеллекта служат электронно-вычислительные машины. Разум испокон веков считался фундаментальным антропологическим параметром. Выбор неорганического материала для осуществления «разума» представляется верхом торжества технического гения. Однако никто не станет отрицать грандиозность задачи создания искусственных существ с заданными ментальными параметрами на *органическом субстрате*.

Исходя из основного тезиса функционализма – о многообразии способов физической реализации некоторого фиксированного ментального состояния – правомочно утверждать, что искусственный интеллект можно реализовать и на электронных лампах, и на транзисторах, и на кремниевых кристаллах. Но его можно реализовать и на растениях, и на животных, и на другом человеке. В трех последних случаях целесообразно назвать такой интеллект «*полуискусственным*». Вводя данный термин, принимается т.н. *либеральная*

¹ Э.В. Попов «Общение с ЭВМ на естественном языке», М., 1982

функционалистская позиция, которая утверждает, что растения и животные, как и человек, обладают интеллектом. Только у них интеллект нечеловеческого типа. Функционалистский шовинизм признает интеллект только за человеком.

«Полуискусственный» интеллект интегрирует инструментарий биотехнологии с инструментарием компьютерной технологии. В зависимости от способа применения биотехнологического инструментария возможны два пути построения «полуинтеллектуальных» систем: путем модификации ДНК или путем копирования ДНК.

Модификация ДНК характерна для генной инженерии, которая, как утверждают противники данного биотехнологического направления может привести к созданию людей-монстров.

Копирование ДНК приводит к менее опасным биоэтическим проблемам и в настоящее время данный способ обозначается термином «клонирование». Возможность клонирования получило эмпирическое подтверждение всего несколько лет назад (1996 г., овечка Долли). В связи с этим в рассмотрим «полуискусственный интеллект» в аспекте создания клонов. В искусственном интеллекте «клон» будет, по нашему замыслу, соответствовать термину «робот».

Термин «клон» означает точную генетическую копию, выращенную из соматической неполовой клетки. Берутся две клетки. Первая – соматическая (неполовая), вторая половая, причем от разных организмов. Соматическая (неполовая) клетка сливается с яйцеклеткой, из которой предварительно удаляется ядро, содержащее наследственную информацию. Получается некое подобие оплодотворенной яйцеклетки, которую электрозарядом стимулируют к делению. Полученный таким образом эмбрион имплантируют определенной особи. Человеческий клон – это «близнец» донора ДНК, отсроченный по времени. Клонированного человека в течение 9 месяцев будет вынашивать женщина, он будет рождаться и воспитываться в семье, достигнет совершеннолетия через 18 лет.

Рассмотрение проблемы клонов в контексте исследований искусственного интеллекта воспроизводит давнюю *евгеническую проблему* создания человека «по заказу». В данном случае «заказ» означает получение клона с заданными параметрами интеллектуального развития – обладающего выдающимися способностями к запоминанию, дедуктивному мышлению, интуитивному постижению и т.п.

Обозначим ряд вопросов.

1. «Генная» интерпретация проблемы «дух/тело»

В проблеме «дух/тело» актуализируется соотношение не между психическими явлениями и мозгом, не между ментальными состояниями и телом, т.е. между субъективной реальностью человека и её нейродинамическими коррелятами. Актуализируется связь между ментальными состояниями и *генотипической организацией человека*. Каким образом генотипическая среда и генофонд влияют на ментальные способности человека?

2. Что проще – «полуискусственный» или искусственный интеллект?

На первый взгляд кажется, что в реализационном отношении проще «полуискусственный интеллект». Данное суждение базируется на факте биологической схожести материального субстрата интеллектуального агента.

Однако более взвешенный взгляд опровергает данный вывод. Интеллектуальные функции, реализуемые на компьютерной базе определяются в технических терминах, прозрачных с точки зрения каузальных, структурных и иных зависимостей. В нашем же случае необходимы фундаментальные исследования. Сложность задачи превышает сложность исследования отношения психических явлений к их нейродинамическим коррелятам. Если для задачи «интеллект/мозг» найдены вполне надёжные концептуальные основания, в форме, например, информационного подхода к сознанию (Д.И. Дубровский), то для нашей задачи таких оснований нет.

Искусственный интеллект построить проще, нежели чем «полуискусственный».

3. Повторит ли клон интеллектуальные способности донора?

Генетическую копию можно снять с любого человека, даже умершего тысячелетие назад. Для этого достаточно клочка волос или кусочка костной ткани. В недалеком будущем можно создать клон, например, В.И. Ленина. Но повторит ли клон свой «прототип»? Ведь если так, то можно «вызвать» умершего великого мыслителя из истории и попросить его работать на благо современности. Однако клон *не наследует* ничего из воспоминаний оригинального индивида. Поэтому появление двойников исключается. По всей видимости, клон не повторит и интеллектуальных способностей донора. Чтобы клон стал мыслителем, таким же, как и донор, нужно *повторить всё неповторимое*: детство, условия жизни, окружающих людей и т. п. С другой стороны, для клонов возможно создание специальных условий воспитания и обучения.

Будущее клонов туманно. По пророчествам, к концу XXI века большая часть людей будет появляться на свет путем клонирования (В. Аксенов). Но уже сегодня следует ставить и решать концептуальные и методологические задачи, подобные задачам построения «полуискусственного» интеллекта. А они тесно связаны с проблемами построения искусственного интеллекта.

История компьютерной технологии

Андрей Шулаков (ЭП-82)

Историю компьютерной технологии следует начать с рассмотрения первых цифровых вычислительных устройств – со счетов.

В Древней Греции счёты – абак – представляли собой посыпанную морским песком доску. На этом песке проводились борозды, в которые клали камешки, обозначающие числа. Первая борозда соответствовала единицам, другая – десяткам и т. д. Если в какой-либо борозде набиралось 10 камешков, то их вынимали и добавляли один камешек в следующую борозду – в следующий разряд.

Леонардо да Винчи (1452 – 1519) создал эскиз 13-ти разрядного сумматора с десятизубными шестернями.

Блез Паскаль (1623 – 1662) для своего отца – налогового инспектора – сконструировал счетное устройство, суммирующее десятичные числа.

Англичане Роберт Биссакар в 1654-м и в 1657 году С. Патридж, независимо друг от друга разработали логарифмическую линейку – устройство не-

прерывного действия, сохранившуюся до нашего времени практически в неизменном виде.

Вильгельм Лейбниц (1646 – 1716) разработал «ступенчатый вычислитель» – счетную машину, работающую в двоичной системе. Она позволяла складывать, вычитать, умножать, делить, вычислять квадратные корни.

Чарльз Бэббидж, профессор математики в Кембриджском Университете, предложил дифференциальную машину, предназначенную для автоматического вычисления математических таблиц (типа таблиц логарифмов и астрономических таблиц). Машина состояла полностью из механических компонентов. Бэббидж показал небольшую рабочую модель в 1822. Он не закончил машину в полном масштабе, но создал несколько фрагментов. Предложенная Бэббиджем *аналитическая машина*, была более совершенной, чем дифференциальная машина, и должна была быть механическим цифровым компьютером общего назначения. Аналитическая машина должна была иметь запас памяти и центральное обрабатывающее устройство. Машина была способна выбирать действие из числа возможных, которое зависело от предыдущих действий (условный переход). Поведение аналитической машины управлялось в соответствии с программой инструкций, содержащихся на перфорированных картах, соединённых друг с другом в ленту. Бэббидж работал вместе с Адой Лавлейс. Лавлейс предвидела возможность использования аналитической машины для *нечисловых исчислений*, предлагая, что машина могла бы даже быть способной к созданию сложных музыкальных пьес.

Большая модель аналитической машины конструировалась в год смерти Бэббиджа – в 1871, но окончательно она построена не была.

К середине XIX века получают распространение работы по аналоговым вычислительным устройствам. Джеймс Томсон и его брат лорд Кельвин, изобрел механический колёсно-дисковый интегратор, который стал основой аналогового вычисления (Thomson [1876]). Они построили устройство для вычисления общих значений двух данных функций, и Кельвин описал (хотя и не построил) аналоговые машины общего назначения для решения систем линейных дифференциальных уравнений любого порядка и для решения систем линейных уравнений. Для каждой новой задачи требовалась значительная механическая работа. Наиболее успешным аналоговым компьютером Кельвина была его машина, предсказывающая приливы, которая применялась в порту Ливерпуля до 1960-ых.

В нашей стране начало разработки аналоговых вычислительных машин относится к 1927 году и связано с работами Гершгорина, Кирпичёва, Брука, Лукьянова и других. В 1950–60-х гг. было создано несколько типов аналоговых вычислительных машин, которые нашли своё применение. Однако, несмотря на высокое быстродействие и простоту сопряжения с исследуемым объектом аналоговые машины обладали невысокой точностью.

В 1935 Алан Тьюринг разработал принцип функционирования современного компьютера. Он описал абстрактную цифровую вычислительную машину, состоящую из безграничной памяти и сканера (считывающе-записывающего устройства), который перемещается назад и вперед по ленте памяти, считывая символы один за другим и записывая другие символы. Действия сканера определяются программой, состоящей из инструкций, которая находится в памяти в виде символов. Такой способ хранения программ подразумевал возможность машины работать по программе. Данная абстрактное

определение понятия алгоритма – вычислительная машина 1935 года – в настоящее время известна как *универсальная машина Тьюринга*.

Во время Второй мировой войны Тьюринг часто говорил о возможности вычислительных машин (1) обучения на опыте и (2) решения проблемы посредством поиска в пространстве возможных решений (эвристический поиск).

Ранние цифровые вычислительные машины были, в основном, электромеханическими. То есть их основными компонентами были электромеханические реле. Они работают относительно медленно, принимая во внимание, что основные компоненты электронно-вычислительной машины – электровакуумные лампы – не имеют никаких механических составляющих. Электромеханические цифровые вычислительные машины были построены до и во время Второй Мировой войны. Зус (Zuse) построил первый работающий управляемый программой цифровой компьютер общего назначения. Эта машина, позже называемая Z3, функционировала в 1941 г.

Реле были также и ненадежным средством для создания действующей конструкции крупномасштабного цифрового компьютера общего назначения. Это вызвало развитие быстродействующих цифровых методов, используя электровакуумные лампы.

Самое раннее широкое применение электровакуумных ламп для цифровой обработки данных было предложено инженером Томасом Флауерсом. Электронное цифровое оборудование начало работать в 1939 и использовало от трех до четырех тысяч непрерывно работающих ламп. В 1938-1939 Флауерс работал на экспериментальной быстродействующей электронной цифровой системе обработки данных, включающей хранилище данных. Цель Флауерса, достигнутая уже после войны, состояла в том, чтобы такое оборудование заменило существующие, менее надежные машины, построенные на базе реле и используемые на телефонных станциях.

Самое раннее применение ламп в США было осуществлено Джоном Атанасовым. За период с 1937 до 1942 Атанасовым были развиты методы использования ламп с целью выполнения численных вычислений в цифровой форме. В 1939, с помощью своего студента Клиффорда Берри, Атанасов начал строить то, что иногда называется Atanasoff-Berry Computer, или ABC, малогабаритную электронную цифровую машину специального назначения для решения систем линейных алгебраических уравнений. Машина содержала приблизительно 300 вакуумных ламп. Хотя электронная часть машины функционировала успешно, компьютер в целом никогда не работал надежно: ошибки вводились неудовлетворительным двоичным считывающим карту устройством.

Первым полностью функционирующим электронным цифровым компьютером был *Колосс* (1943). Он успешно расшифровывал немецкие радиосообщения, кодируемые при помощи системы *Энигма* и к началу 1942 приблизительно 39000 перехваченных сообщений ежемесячно расшифровывались благодаря электромеханическим машинам, известные как «бомбы». Они были разработаны *Тьюрингом* и *Гордоном Велчманом*.

Колосс I содержал приблизительно 1600 вакуумных ламп, любая из последующих машин содержала приблизительно 2400 вакуумных ламп. Подобно небольшой ABC, Колосс испытывал нужду в двух важных особенностях современных компьютеров. Во-первых, не было внутреннего сохране-

ния программы. Для установки новой задачи, оператор был должен изменить физическую проводку машины, используя штепсели и выключатели. Во вторых, Колосс не был машиной общего назначения, а предназначался для определенных криптоаналитических задач, включая счет и логические операции.

В 1945 Тьюринг стал работать в Национальной физической лаборатории в Лондоне над проектированием и развитием электронного цифрового компьютера. Предложенную машину Тьюринга принято называть *автоматической вычислительной машиной*, или ACE (Automatic Computing Engine). Тьюринг считал, что скорость и память были ключами к вычислению. Проект Тьюринга имел много общего с сегодняшней RISC архитектурой, и это привело к быстродействующей памяти приблизительно той же самой емкости, какую имел ранний компьютер Макинтоша (огромный объем по стандартам тех дней). В мае 1950 маленькая экспериментальная модель Автоматической вычислительной машины впервые выполнила программу. Со скоростью процессора в 1 МГц экспериментальный образец ACE был в течение некоторого времени самым быстрым компьютером в мире.

Первая полностью функциональная электронная цифровая ЭВМ была разработана в США и называлась *ENIAC*. Законченная в 1945, ENIAC, в отличие от Колосса, была достаточно гибкой.

В нашей стране первой цифровой вычислительной машиной можно считать малую электронную вычислительную машину (МЭСМ), построенную в 1950 в АН УССР под руководством *Лебедева*. Под его же руководством в 1953 в Институте точной механики и вычислительной техники была создана БЭСМ, ставшая предшественницей серии отечественных электронных цифровых машин («Минск», «Урал», «Днепр», «Мир» и другие).

Следует также отметить машину «Сетунь», разработанную в МГУ под руководством Н.М.Брусенцова в 1965 г.¹ Машина была способна моделировать трехзначное кодирование – «да», «нет», «неопределенно». Это предполагало построение трехзначной логики (Д.Н. Юрьев, А.С. Карпенко), что и сегодня открывает перспективы в разработке логико-математических моделей искусственного интеллекта, более адекватных естественному интеллекту, нежели аналоги этих моделей, использующих двухзначную логику.

¹ Брусенцов Н.М. Опыт разработки трехзначной вычислительной машины // Вестник МГУ. Серия 1: Математика, механика. 1965 № 2.

СОДЕРЖАНИЕ

О СТУДЕНЧЕСКОЙ КОНФЕРЕНЦИИ «ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА», 20 мая 2004 г., г. Москва, МИЭМ.....	3
Путилов Г.П. Преемственность поколений исследователей искусственного интеллекта	5
Дубровский Д.И. Приветственное слово участникам конференции	8
Колин К.К. Философские аспекты информационных технологий.....	11
Алексеев А.Ю. Уровни изучения искусственного интеллекта.....	23
 I. Функционалистская концепция мышления. Тест Тьюринга: pro et contra	 36
Алексеева Анна. Функционализм, физикализм и искусственный интеллект.....	36
Денисов Алексей. Тест Френча: субкогнитивистское опровержение концепции Тьюринга	37
Жуков Алексей. Базовые положения теста Тьюринга.....	41
Зайцев Игорь. Квалиа и парадигма функционализма.....	42
Зворыкин Дмитрий. Зачем вкладывать деньги в тест Тьюринга? Лойбнеровская премия	45
Клопков Алексей. Программирование теста Тьюринга	47
Комаров Дмитрий. «Может ли машина мыслить?». Полемический стандарт Тьюринга	49
Ласточкин Алексей. Тест Лавлейс: машина творить не может !	51
Лизоркин Сергей. «Искусственный интеллект» Алана Тьюринга	54
Маклашевская Наталия. Функционально-структурные аспекты понятия «потребность»	57
Великанов Алексей, Макарычев Антон. Тест Тьюринга и Д. Деннет	61
Матанцева Ирина. Тест Блока: антибихевиористское опровержение тьюринговой концепции мышления.....	62
Морозов Виктор. Многообразие тестов Тьюринга	65
Никишев Артур. Критика сильного искусственного интеллекта. Аргумент Гёделя	67
Родионов Денис. Тест Серла: интенционалистское опровержение концепции Тьюринга	69

<i>РОМАНОВ ДЕНИС. ПАРАДИМА ФУНКЦИОНАЛИЗМА: КАК ПРЕДСТАВИТЬ МЕНТАЛЬНОЕ В НЕМЕНТАЛЬНЫХ ТЕРМИНАХ?</i>	71
<i>РОМАНОВА ЕЛЕНА. НАИВНАЯ ПСИХОЛОГИЯ И ИНВЕРТИРОВАННЫЙ ТЕСТ ТЬЮРИНГА</i>	75
<i>РЫБИН ИЛЬЯ. СОЦИОКУЛЬТУРНЫЕ АСПЕКТЫ ТЕСТА ТЬЮРИНГА</i>	80
<i>СОБОЛЕВ ДМИТРИЙ. КОМПЬЮТЕР МОЖЕТ МЫСЛИТЬ! (М. МИНСКИЙ)</i>	84
<i>ЦВЕТКОВ ЮРИЙ. ТЕСТ ТЬЮРИНГА И ПАРАНОИЯ</i>	87
<i>ЧИЖОВ ИВАН. ТЕСТ БЛОКА: НЕСТАНДАРТНЫЕ АНТИБИХЕВИОРИСТСКИЕ ВОЗРАЖЕНИЯ ТЬЮРИНГУ</i>	88
<i>ЧУДАКОВ АНДРЕЙ. ФУНКЦИОНАЛИСТСКИЙ СТАТУС ЛЮБВИ</i>	91
<i>ДУБРОВСКИЙ ДАВИД. ТЕСТ ТЬЮРИНГА, ПАРАДИГМА ФУНКЦИОНАЛИЗМА И ПРОБЛЕМА СОЗНАНИЯ. ПОДВЕДЕНИЕ ИТОГОВ РАБОТЫ СЕКЦИИ № 1</i>	93

II. Искусственный интеллект и «здоровый смысл» 97

<i>ГАВРИЛОВ ИЛЬЯ. К ВОПРОСУ ЭПИСТЕМОЛОГИЧЕСКОЙ АДЕКВАТНОСТИ РЕПРЕЗЕНТАЦИЙ</i>	97
<i>ГОРЮНОВ РОМАН. ПСЕВДОВОЛЯ</i>	104
<i>ГРИШКИН МАКСИМ. РЕАЛИЗАЦИОННЫЕ ПЕРСПЕКТИВЫ ТЕОРИИ РЕЧЕВЫХ АКТОВ</i>	105
<i>КРАСИВСКАЯ МАРИЯ. ЛОГИЧЕСКИЕ АСПЕКТЫ СОЗДАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА</i>	107
<i>ЛАПИН МИХАИЛ. ЭКСПЕРТНЫЕ СИСТЕМЫ, ОСНОВАННЫЕ НА ЗДРАВОМ СМЫСЛЕ</i>	109
<i>ПРАСОЛОВА ВАЛЕРИЯ. ЧТО ТАКОЕ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ?</i>	117
<i>РОДИОНОВ ДЕНИС. ПРОБЛЕМА ДИСКУРСА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. КОНСТРУКТОРСКАЯ ПОЗИЦИЯ</i>	119
<i>СИМЕНЕЛ ЕЛЕНА. ЗАЧЕМ ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ ФИЛОСОФИЯ?</i>	121

III. Историко-философские перспективы компьютерного моделирования 125

<i>АРТЮХОВ АНАТОЛИЙ. КОНТЕНТУАЛЬНАЯ МОДЕЛЬ СМЫСЛА (А.Ф. ЛОСЕВ)</i>	125
<i>ДРОБЯЩЕНКО АНАСТАСИЯ. ПРОБЛЕМА ИДЕАЛЬНОГО И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ</i>	129
<i>ИВАНОВА АЛЛА. ТЬЮРИНГ И ПРОБЛЕМА ВЫЧИСЛИМОСТИ СОЗНАНИЯ</i>	131
<i>КОРОЛЕВ МАКСИМ. ПРОБЛЕМА ЕСТЕСТВЕННЫХ ВИДОВ В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ</i>	132

<i>КОСИНОВА ТАТЬЯНА. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СТОИЧЕСКАЯ ЭПИСТЕМОЛОГИЯ</i>	135
<i>КУРАЕВА ТАТЬЯНА. ЗОМБИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ.....</i>	138
<i>МАЛИКОВА ЯНА. ИНТУИТИВИСТСКИЕ ОРИЕНТИРЫ МОДЕЛИРОВАНИЯ СМЫСЛА (А. БЕРГСОН).....</i>	142
<i>РОЗОВ МАКСИМ. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СВЯТООТЕЧЕСКИЙ ОПЫТ</i>	144
<i>СЁМОЧКИН МИХАИЛ. ВИРТУАЛЬНАЯ РЕАЛЬНОСТЬ И МАТЕМАТИКА КУЗАНСКОГО.....</i>	147

IV. Философия искусственного интеллекта и компьютерная технология 150

<i>АЛЕКСАНДРОВ ЕВГЕНИЙ, ДОМАСЬ КОНСТАНТИН. БИОНИКА КАК НАПРАВЛЕНИЕ РОБОТОТЕХНИКИ</i>	150
<i>БОНДАРЬ АЛЕКСАНДР. «КВАЛИА» КАК БАЗОВАЯ КАТЕГОРИЯ ВИРТУАЛИСТИКИ</i>	151
<i>КАДАНЦЕВА АЛЛА. ЧЕЛОВЕК И КОМПЬЮТЕР. ДРУЗЬЯ ИЛИ ВРАГИ?.....</i>	154
<i>КАЗАНСКИЙ МИХАИЛ. КВАНТОВЫЕ КОМПЬЮТЕРЫ И КВАНТОВАЯ МЕХАНИКА.....</i>	157
<i>КОЛЕСНИКОВ СТАНИСЛАВ. ИНТЕРНЕТ-НАВИГАТОР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"</i>	159
<i>КОЛЬЦОВ МИХАИЛ. ПАРАДИГМА КОННЕКЦИОНИЗМА КАК МЕТОДОЛОГИЯ НЕЙРОКОМПЬЮТЕРНОЙ ТЕХНОЛОГИИ</i>	162
<i>КРЮЧКОВ ВАСИЛИЙ. ЧТО ЗНАЧИТ «БЫТЬ РОБОТОМ»?</i>	171
<i>МИХЕЙКИН ИВАН. О РАСПРЕДЕЛЕНИИ ФУНКЦИЙ МЕЖДУ ЧЕЛОВЕКОМ И КОМПЬЮТЕРОМ В ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЯХ</i>	174
<i>НЕСТЕРОВ ОЛЕГ. «ФИЛОСОФИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА» В ИНТЕРНЕТ-СРЕДЕ</i>	176
<i>ПАК МАРК, ПАНОВ АЛЕКСЕЙ. МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ НАНОТЕХНОЛОГИИ.....</i>	178
<i>ПОДОПРИГОРА ИВАН. ЕСТЕСТВЕННО-ЯЗЫКОВЫЙ ИНТЕРФЕЙС: ТРИ ПОХОДА К МОДЕЛИРОВАНИЮ «СМЫСЛА».....</i>	179
<i>СМИРНОВА ИРИНА. КЛОНЫ И «ПОЛУИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ».....</i>	183
<i>ШУЛАКОВ АНДРЕЙ. ИСТОРИЯ КОМПЬЮТЕРНОЙ ТЕХНОЛОГИИ.....</i>	185

Научное издание

Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции «Философия искусственного интеллекта», г. Москва, МИЭМ, 20 мая 2004 г.

Под редакцией кандидата философских наук *А.Ю. Алексеева*

Оригинал–макет: А.А. Сочилин

Художник: И. Рыбин

Технический редактор: Т.А. Кураева

Корректор: Д.А. Алексеев

Подписано в печать с оригинал-макета 23.03.2006 г.

Формат 60х84/16. Ризография. Гарнитура Таймс.

Усл.печ.л. 12. Уч.-изд.л. 13,6. Тираж 100 экз.

Заказ № _____



Издательство «ИИнтелл»

109518, г. Москва, ул. Грайвороновская, д.20 - 171

Московский государственный институт электроники и математики.

109028, г. Москва, Б. Трёхсвятительский переулок, 3/12.

Отдел оперативной полиграфии Московского государственного института
электроники и математики.

113054, Москва, ул. М. Пионерская, 12