Министерство образования Российской Федерации
Томский государственный университет

В.А. ЛАДОВ

ФИЛОСОФСКИЕ ПРОБЛЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Учебно-методическое пособие для специализации
«ГУМАНИТАРНАЯ ИНФОРМАТИКА»

Томск 2005

Рассмотрено и утверждено на заседании методической комиссии
философского факультета Томского государственного университета

Протокол № «____» от «____» _____ 2005 г.

Председатель
методической комиссии ФсФ _____ С.С. Аванесов

Целью данного курса является изложение ряда философских проблем, связанных с развитием систем искусственного интеллекта. Анализируется соотношение понятий естественного и искусственного интеллекта, обсуждаются различные парадигмы понимания разумной деятельности, раскрываются предпосылки возникновения и история развития систем искусственного интеллекта.

Для студентов, обучающихся по специализации «Гуманитарная информатика».

Составитель: к.филос.н., доцент В.А. Ладов

ТЕМАТИЧЕСКИЙ ПЛАН

Тема 1. Основные проблемы философии искусственного интеллекта. 2 ч.

Тема 2. Предпосылки возникновения систем искусственного интеллекта. 2 ч.

Тема 3. Парадигма "интеллект как исчисление понятий". 2 ч.

Тема 4. Парадигма "интеллект как восприятие". 2 ч.

Тема 5. Парадигма "интеллект как рефлексия".
Парадигма "интеллект как самоидентичность". Тест Тьюринга. 2 ч.

Тема 6. Понятие интенциональности. Парадигма
"интеллект как интенциональность". Аргумент "китайская комната" 2 ч.

Тема 7. Понятие производной интенциональности.
Операциональная деятельность. 2 ч.

Тема 8.Синтаксис и семантика языка систем искусственного интеллекта.
Проблема гомункулуса. 2 ч.

Основная литература

1. Винер Н. Творец и робот. М., 1966.
2. Лорьер Ж.-Л. Системы искусственного интеллекта М., 1991.
3. Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики. М., 2003.
4. Ракитов А.И. Философия компьютерной революции. М., 1991.
5. Сергеев В.М. Искусственный интеллект: Опыт философского осмысления //Будущее искусственного интеллекта. М., 1991.
6. Серл Д. Мозг, сознание и программы // Аналитическая философия: становление и развитие (антология). М., 1998. с. 376 - 400.
7. Урсул А.Д. Информатизация общества. – М., 1990.
8. Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17-19 января 2005г. М.: ИФ РАН, 2005.
9. Хофштадтер Д., Деннет Д. Глаз разума. Самара, 2003.
10. Шанже Ж.П., Конн А. Материя и мышление. М., 2004.

Дополнительная литература

1. Анисимов А.В. ЭВМ и понимание математических доказательств //Вопросы философии, 1987, N 3.
2. Арбиб М. Метафорический мозг. М., 2004.
3. Арнольдов А.И. Информация - глобальная ценность XXI века. М., 1997.
4. Грачев М.Н. Кибернетический подход и система философских взглядов Норберта Винера: Автореферат диссертации на соискание ученой степени кандидата философских наук. М., 1994.
5. Дубровский Д.И. Искусственный интеллект и проблема сознания // Философия искусственного интеллекта. М., 2005. с. 26-32.
6. Кочергин А.Н. Искусственный интеллект и мышление // Философия искусственного интеллекта. М., 2005. с. 37-39.
7. Ладов В.А. Интенциональность как основание различия человеческого сознания и искусственного интеллекта // Философия искусственного интеллекта. М., 2005. с. 39-43.

8. Макарчук М.М. Об основном отличии искусственного и естественного интеллекта // Философия искусственного интеллекта. М., 2005. с. 50-52.
9. Маслов С.Ю. Теория дедуктивных систем и ее применения. - М., 1986.
10. Моисеев В.И. Интервал Тьюринга и имитация интеллекта // Философия искусственного интеллекта. М., 2005. с. 307-310.
11. Пенроуз Р., Шимони А., Картрайт Н., Хокинг С. Большое, малое и человеческий разум. М., 2004.
12. Ракитов А.И. Философия компьютерной революции. - М., 1991.
13. Саймон Г. Науки об искусственном. М., 2004.
14. Сергеев В.М. Искусственный интеллект как метод исследования сложных систем // Системные исследования: методологические проблемы (ежегодник) М., 1984.
15. Сергеев В.М. Искусственный интеллект: Опыт философского осмысления // Будущее искусственного интеллекта М., 1991.
16. Юлина Н.С. Обретение разумности: «сократический диалог» и интеракция с компьютером // Философия искусственного интеллекта. М., 2005. с. 82-84.
17. Anderson, D. Is the Chinese room the real thing? *Philosophy* 62, 1987.
18. Block, N. Troubles with functionalism. // Perception and cognition: Minnesota studies in the philosophy of science, vol. 9, ed. W. Savage. Minneapolis: University of Minnesota Press. 1978.
19. Brentano F. Psychology from an Empirical Standpoint. London: Routledge and Kegan Paul. 1973.
20. Dennet, D. Haugeland, J. Intentionality // The Oxford Companion to the Mind, in R. L. Gregory, ed., Oxford University Press 1987.
21. Dennett, D. Evolution, Error and Intentionaly // Sourcebook on the Foundations of Artificial Intelligence, in Y. Wilks and D. Partridge, eds, New Mexico University Press, 1988.
22. Dretske, F. Machines and the mental. // Proceedings and Addresses of the American Philosophical Association 59(3, November), 1985.
23. Dreyfus, H. What computers can't do. New York: Harper Colophon, 1979.
24. Feigenbaum, E.A. and Feldman, J. ed. Computers and Thought. McGraw-Hill Company, New York and San Francisco, 1963.
25. Fodor J. Representations. Cambridge, MA: MIT Press, Bradford Books, 1981.
26. Fodor, J. The Language of Thought. Thomas Y. Crowell, New York, 1975.
27. Grice P. Studies in the Way of Words. Cambridge, Mass: Harvard University Press, 1989.
28. Hauser, L. Why Isn't My Pocket Calculator a Thinking Thing? // Minds and Machines, Vol. 3, No. 1 (February, 1993).

29. Kripke S.A. Wittgenstein on Rules and Private Language. Oxford: Basil Blackwell, 1982.
30. MacCarthy, J. What Is Artificial Intelligence? // Stanford University Bulletin. April 4, 2000.
31. Malcolm, N. Knowledge of other minds. Journal of Philosophy LV, 1958.
32. Minsky, M. Steps Toward Artificial Intelligence // www.consciousness.arizona.edu
33. Minsky, M. Will Robots Inherit the Earth? // Scientific American. October 1994.
34. Searle J. Intentionality: An Essay in the Philosophy of Mind. Cambridge: Cambridge University Press, 1983.
35. Searle J. Minds, Brains, and Programs // The Philosophy of Artificial Intelligence, in M. Boden, ed., Oxford University Press, 1990.
36. Searle, J.R. Is the Brain a Digital Computer? // www.consciousness.arizona.edu
37. Turing, A. Computing machinery and intelligence. *Mind* LIX, 1950.
38. Winograd, Terry and Flores, Fernando. Understanding Computers and Cognition: A New Foundation for Design. Addison-Wesley, 1987.

ТЕМА 1

Основные проблемы философии искусственного интеллекта.

Аннотация

Актуальность философского рассмотрения феномена искусственного интеллекта. Формулировка проблем философии искусственного интеллекта. Основные эпистемологические проблемы. Социальные, психологические и общекультурные проблемы. Обзор литературы.

Конспект лекции

Научно-техническая революция, инспирированная развитием точных, естественных и технических наук в европейских странах, начиная с XVII в., внесла коренные изменения в мир. Бурное развитие техники позволило человеку Нового времени осуществить мощное наступление на противостоящую ему окружающую среду с целью ее подчинения запросам человечества, пытающегося обеспечить себя необходимыми условиями жизнедеятельности. По мнению многих ученых именно этот процесс окончательно выделил человека из всего круга сущего. Человечество совершило мощный эволюционный прорыв, оставив далеко позади другие биологические формы жизни. Движимый развитием техники процесс освоения природной среды, сложность социальной жизни человека, наполненной искусственными техническими изобретениями, достиг своего апогея в XX веке.

В этой ситуации представляется вполне естественным, что философия, которая всегда старалась быть неким универсальным мыслительным предприятием, рефлексирующим по поводу самых общих, глобальных проблем, стоящих перед человечеством в ту или иную эпоху его развития, не могла не отреагировать на происходящие в мире науки и техники процессы. Философские проблемы развития техники, философские проблемы освоения окружающей среды (environmentalism), философские проблемы виртуальной реальности – вот названия тех направлений, которые получили широкое распространение в западной философии XX века.

В определенный момент процесса изобретения и внедрения различных технических устройств, предназначенных для средств освоения и подчинения окружающего пространства, для средств коммуникации и расчетов своих действий, человек продуцировал очень необычный, неизвестный ранее феномен – искусственный интеллект. Ранее развитие техники сосредотачивалось на конструировании устройств, имитирующих с гораздо более высокой производительностью, нежели в их естественном проявлении, внешние органы

чувств и органы действий человека: вместо естественного зрения – микроскоп или бинокль, вместо руки – экскаватор, вместо естественного слуха – радиосвязь, вместо ног – автомобиль и т.д. И вот появилось устройство, призванное имитировать и замещать, казалось бы, самое главное в человеке – то, что с давних времен признавалось его самым существенным признаком – разумность. Системы искусственного интеллекта были призваны воспроизвести и, возможно, в перспективе заменить на более высоком качественном уровне процесс мышления человека, его способность к рациональным интеллектуальным действиям.

Конечно же, философия не могла не откликнуться и на этот, возможно, самый интригующий эпизод в развитии техники. Впервые в человеческой истории рядом с человеком на Земле появился разумный сосед – система искусственного интеллекта, что позволило в новой форме и с новой силой поставить классические философские вопросы о сущности разума, о сущности самого человека. Разработкой и решением данных вопросов, в частности, занимается и такое сравнительно новое направление в западной философской традиции, как философия искусственного интеллекта.

Когда мы заговариваем об искусственном интеллекте, о принципах его построения, о его функционировании, о возможностях его работы мы всегда сравниваем его с естественным разумом человека. При этом, как правило, в научно-популярной литературе, понимание самого человеческого разума уже предзадано. Формулировка понятия искусственного интеллекта дается на основе уже существующего предпонимания естественного разума, которое мыслится, чаще всего, как нечто само собой разумеющееся. Такой подход представляется неверным.

Своеобразие проблем определения понятия искусственного интеллекта и прояснения вопросов о возможностях его работы состоит как раз в том, что сами технические устройства, которые мы называем интеллектуальными, нам хорошо известны. Человек сам построил их. Он может досконально проследить весь процесс создания и функционирования этих систем. Главная сложность заключается в том, что сейчас, в начале XXI века, у нас по-прежнему нет четких и однозначных ответов о естественном разуме, о его сущности и принципах деятельности. На настоящий момент существует достаточное количество конкурирующих теорий, повествующих о том, что следует считать существенными признаками разумности. Отсюда следует, что при выборе той или иной теории естественного разума у нас будет изменяться взгляд и на то, что мы называем искусственным интеллектом. Будет изменяться наша оценка его возможностей.

Все это заставляет нас констатировать тот факт, что, хотели бы того или нет пропоненты ИИ, философия искусственного интеллекта должна обсуждать вопросы

естественного и искусственного разума неразрывно, обращаясь, в том числе, и к достаточно богатой традиции понимания естественного разума в истории философии.

Исходя из вышесказанного, можно сформулировать главные проблемы, которые будут обсуждаться в дальнейшем:

1. Каково соотношение человеческого разума и искусственного интеллекта?
2. Мыслит ли компьютер?
3. Каковы существенные признаки естественного разума?
4. Способен ли искусственный интеллект удовлетворять существенным признакам естественного разума или даже превосходить их?

Кроме этих, определяющих все дальнейшее изложение эпистемологических проблем, конечно же, можно выделить и другую группу вопросов, которые мы будем считать дополнительными. Тем не менее, достаточно большой массив литературы в данной области знания признает эти темы так же чрезвычайно значимыми и актуальными. Это такие вопросы, как:

1. Какова цель создания систем искусственного интеллекта?
2. Существует ли опасность выхода системы ИИ из под контроля человека?
3. Каковы новые психологические факторы поведения человека в коммуникации с системами ИИ?
4. Можно ли предполагать возникновение этических проблем при коммуникации человека с системами ИИ?
5. В чем состоят принципиальные изменения в социальной жизни человека при введении в эту сферу систем ИИ?

Кратко прокомментируем каждый из этих вопросов.

С какой целью вообще создаются искусственные интеллектуальные системы? Если цель состоит в том, чтобы помочь, а в перспективе вообще освободить человека от интеллектуальной деятельности, тогда можно поставить вопрос – зачем это человеку? Не потеряет ли он своего главного своеобразия – способности мыслить? Не деградирует ли в этом состоянии интеллектуальной лени и все возрастающего комфорта жизни?

Очень важной оказывается проблема безопасности. При дальнейшем развитии самоорганизующихся интеллектуальных систем вполне можно прогнозировать их постепенный выход из-под контроля человеческого управления. Как поведут себя машины? Что будет с человеком?

Стремительно ворвавшись в нашу жизнь, системы ИИ привнесли с собой новый психологический климат в повседневную жизнь человека. Какие положительные и

отрицательные факторы, воздействующие на психику человека мы можем зафиксировать здесь?

Вопрос об этике – это, конечно, вопрос на перспективу. Можно прогнозировать построение высоко интеллектуальных систем-роботов, способных, возможно, даже к определенным эмоциональным реакциям в коммуникации с людьми. Могут ли здесь возникнуть новые этические проблемы для человека, связанные, например, с ликвидацией интеллектуальной системы? Те проблемы, которые уже сейчас оказываются актуальными для биоэтики, обсуждающей результаты экспериментов генной инженерии – проблемы клонирования.

Уже сейчас становится ясно, что с дальнейшим внедрением в нашу жизнь систем ИИ человеческий социум должен будет подвергнуться значительной модификации принципов своего устройства. Какими должны быть эти изменения? Что нового они привнесут в организацию жизни человеческого сообщества?

В 50-е – 60-е годы XX века в западной философии сложилась достаточно устойчивая традиция осмысления проблем, связанных с понятием искусственного интеллекта. Философским проблемам ИИ посвятили свои работы такие известные философы, математики и лингвисты, как Алан Тьюринг, Норберт Винер (основатель кибернетики), Марвин Мински, Хьюберт Дрейфус, Уоррен Маккалох, Фрэнк Розенблант и др.

В настоящее время на западе существуют специализированные научно-философские сообщества, обращающиеся к проблемам искусственного интеллекта. Эти сообщества организуют конференции, издают журналы и сборники работ, посвященные данной тематике.

Что касается главных вопросов настоящего курса – о соотношении искусственного интеллекта и естественного разума человека, о прояснении специфики человеческого мышления, о возможности уподобления искусственного разума естественному – автор опирается, прежде всего, на работы таких известных современных представителей философии сознания, языка и искусственного интеллекта, как Дениэл Деннет, Джон Серл, Теодор Дрецке, Ларри Хаусер, Джерри Фодор, Сол Крипке и др.

Литература

1. Арнольдов А.И. Информация - глобальная ценность XXI века. М., 1997.
2. Винер Н. Творец и робот. М., 1966.

3. Дубровский Д.И. Искусственный интеллект и проблема сознания // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 26-32.
4. Лорьер Ж.-Л. Системы искусственного интеллекта М.: Мир, 1991.
5. Урсул А.Д. Информатизация общества. – М.: 1990.

Контрольные вопросы

1. В чем состоит актуальность философского рассмотрения феномена искусственного интеллекта?
2. Какие философские проблемы возникают перед человечеством в связи с развитием систем искусственного интеллекта?
3. Почему проблему о соотношении естественного разума и искусственного интеллекта следует считать важнейшей?

ТЕМА 2
Предпосылки возникновения систем искусственного интеллекта

Аннотация

Идея имитации разумного поведения: историческая справка. Предпосылки возникновения систем искусственного интеллекта. Определение мышления в философских системах Нового времени. Развитие математической логики. Формализация процессов рассуждения, использование алгебраического языка. Логистическая интерпретация электроники. Электронная имитация процесса логического рассуждения.

Конспект лекции

Идея создания человекоподобных мыслящих машин имеет давнее происхождение в человеческой истории. Еще в Древнем Египте простой люд испытывал страх и благоговение перед особыми культовыми фигурами, способными совершать движения и изрекать пророчества. Естественно, что все эти устройства управлялись непосредственно самим человеком наподобие того, как артист-кукольник управляет марионеткой.

В Средневековье, известном своей «подпольной» (неодобряемой церковью) страстью к естественным и техническим дисциплинам и экспериментам, были созданы более сложные

человекоподобные конструкции, приводимые в движение за счет взаимодействия механизмов.

В Новое время, благодаря бурному развитию техники, подобные идеи стали еще более популярными. Французский изобретатель Жак де Вокансон изготовил механического флейтиста в человеческий рост, который исполнял двенадцать мелодий, перебирая пальцами отверстия и дуя в мундштук, как настоящий музыкант. В середине 1750-х годов австриец Фридрих фон Кнаус, служивший при дворе Франциска I, сконструировал серию машин, которые умели держать перо и могли писать довольно длинные тексты. Другой мастер, Пьер Жак-Дроз из Швейцарии, построил пару изумительных по сложности механических кукол размером с ребенка: мальчика, пишущего письма и девушку, играющую на клавесине. В 1830-х годах английский математик Чарльз Бэббидж задумал, правда, так и не завершив, сложный цифровой калькулятор, который он назвал аналитической машиной; как утверждал Бэббидж, его машина в принципе могла бы рассчитывать шахматные ходы. Позднее, в 1914 г., директор одного из испанских технических институтов Леонардо Торрес-и-Кеведо действительно изготовил электромеханическое устройство, способное разыгрывать простейшие шахматные эндшпили почти также хорошо, как и человек. Компактный механический арифмометр, сконструированный в начале XX века стал очень популярным в среде инженерных работников, экономистов, учителей и продержался в эксплуатации вплоть до последней четверти XX века.

Однако несомненным прорывом в области построения искусственного интеллекта, конечно же, следует считать разработку и конструирование электронных устройств. Пожалуй, только с 40-50 годов XX века техника поднялась до того уровня, когда возможное моделирование не только телесных движений, но и умственных действий, стало почти реальностью.

Тем не менее, следует помнить, что для того, чтобы данные проекты оказались технически реализуемы, должна была быть проведена сугубо теоретическая работа по интерпретации сущности человеческого мышления и его возможной репрезентации. Главными этапами в этой теоретической работе были следующие:

А) Сущностью мышления, как оно понималось, в частности, Декартом и Лейбницем (хотя истоки такой интерпретации можно обнаружить еще у Аристотеля), была аналитико-синтетическая деятельность, способность к сочленению или разъединению понятий, образование на этой основе суждений и умозаключений, позволяющих проводить дискурсивное рассуждение.

В) Формализация естественного языка, начало которой также положено Аристотелем при формулировке фигур возможных умозаключений. В XIX веке на основе алгебры Дж.

Буля, которая представляла собой формализацию арифметических действий, было предложено воспользоваться алгебраическим языком и для формализации логического процесса рассуждения. Были введены символы для всех возможных логических констант, характеризующих формальные элементы в суждениях – логические союзы. Отрицание «-», дизьюнкция «V», коньюнкция «&», импликация «⊃», тождество «≡».

C) Формализация главных логических референтов «истина» и «ложь» в виде арифметических символов 1 и 0 (Дж. Буль). Задание референтов для каждого логического союза, проведенное Фреге. Позднее, введенные, в частности, Витгенштейном, таблицы истинности, завершили процесс формализации мышления.

D) В 30-е годы XX в. пионеры информатики, прежде всего, американский ученый Клод Шеннон, поняли, что двоичные единица и нуль вполне соответствуют двум состояниям электрической цепи (включено-выключено), поэтому двоичная система идеально подходит для электронно-вычислительных устройств. Была осуществлена аналогия представления логических референтов суждений при помощи электрических сигналов: 1 – наличие сигнала; 0 – отсутствие сигнала.

E) Построение ламповых и транзисторных электрических схем. Создание логических микросхем.

Примеры.

Рассуждение 1.

Закон транзитивности

$((A \supset B) \,\&\, (B \supset C)) \supset (A \supset C)$

Рассуждение 2.

В преступлении подозреваются трое: Иванов, Петров, Сидоров. Они дали следующие показания. Иванов утверждал, что если преступление совершил Сидоров, то Петров его не совершал. Петров настаивал на том, что если Иванов совершил преступление, то Сидоров тоже в этом участвовал; но если Иванов не причем, то это сделали либо он сам, либо Сидоров. Сидоров отрицал свою вину, но настаивал на виновности либо Петрова, либо Иванова. При условии, что все говорили правду, кто виновен?

$((C \supset \neg\, П) \,\&\, ((И \supset C) \,\&\, ((\neg И \supset (П \vee C))) \,\&\, (\neg\, C \,\&\, (П \vee И))) \supset И$

$((C \supset \neg\, П) \,\&\, ((И \supset C) \,\&\, ((\neg И \supset (П \vee C))) \,\&\, (\neg\, C \,\&\, (П \vee И))) \supset П$

$((C \supset \neg\, П) \,\&\, ((И \supset C) \,\&\, ((\neg И \supset (П \vee C))) \,\&\, (\neg\, C \,\&\, (П \vee И))) \supset C$

Рассуждение 3.

Кто-то (Иван, Петр, Алексей, Николай или Борис) съел банку варенья. Известно, что если съел Борис, то вместе с ним ели Иван и Николай. Если же съел Иван, то вместе с Петром. Петр и Алексей не могли есть варенье вместе, это мог быть только один из них.

Алексей мог есть варенье только вместе с Николаем. По крайней мере, Николай или Борис съели варенье. Кто съел варенье?

((Б ⊃ (И & Н)) & (И ⊃ П) & (П V А) & (А ≡ Н) & (Н V Б)) ⊃ Б

((Б ⊃ (И & Н)) & (И ⊃ П) & (П V А) & (А ≡ Н) & (Н V Б)) ⊃ И

((Б ⊃ (И & Н)) & (И ⊃ П) & (П V А) & (А ≡ Н) & (Н V Б)) ⊃ П

((Б ⊃ (И & Н)) & (И ⊃ П) & (П V А) & (А ≡ Н) & (Н V Б)) ⊃ Н

Эти рассуждения могут быть смоделированы с помощью электрических схем.

Данный процесс электронного калькулирования событий имеет поистине могущественные последствия – все события мира могут быть формализованы и подвергнуты электронной обработке. Правда, для этого, по словам Х. Дрейфуса, необходимо сначала задать соответствующую онтологию мира. По мысли данного американского мыслителя, зеленый свет могущественному развитию цифровых компьютеров дала как раз онтология логического атомизма – развитая в начале XX века в аналитической философии (Фреге, Рассел, Витгенштейн). Если весь мир представить как совокупность простых объектов, которые в определенных сочленениях образуют факты, а логическую структуру языка отождествить с логической структурой мира, то технологии цифрового «рассуждения», т.е. обработки информации охватят собой не только область мышления человека, но и сущее в целом. Дрейфус считает, что цифровые технологии стали столь популярными именно исходя из господства данной онтологической парадигмы в XX веке. При введении другой онтологии цифровой компьютер может уже не выглядеть столь могущественным.

Литература

1. Винер, Н. Творец и робот. М., 1966.
2. Грачев *М.Н.* Кибернетический подход и система философских взглядов Норберта Винера: Автореферат диссертации на соискание ученой степени кандидата философских наук. М.: Российская академия управления, 1994.
3. Лорьер Ж.-Л. Системы искусственного интеллекта М.: Мир, 1991.
4. Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики. М.: Едиториал УРСС, 2003.

Контрольные вопросы

1. Когда и где возникла идея имитации разумной деятельности?
2. Что в новоевропейской философии способствовало развитию идеи искусственного интеллекта?
3. Какова роль математической логики в создании систем искусственного интеллекта?

4. В чем состоит специфика аналитико-синтетической деятельности рассудка?
5. Каковы основные характеристики математического исчисления?
6. Каковы основные характеристики логического исчисления?

ТЕМА 3

Парадигма "интеллект как исчисление понятий"

Аннотация

Аналитико-синтетическая деятельность рассудка по формированию суждений и умозаключений как существенная черта интеллекта. Логическое исчисление. Критика парадигмы «интеллект как исчисление понятий».

Конспект лекции

Так или иначе, если закрыть глаза на негативную критику Х. Дрейфуса (указывая на онтологию логического атомизма как только на одну из возможных и апеллируя к М. Хайдеггеру, Дрейфус все же не смог представить какую-либо внятную альтернативу критикуемой позиции), то мы сталкиваемся со следующей ситуацией. При принятии парадигмы «интеллект как исчисление понятий» (т. е. как аналитико-синтетическая деятельность по сочленению или разъединению понятий и суждений, как последовательное продуцирование умозаключений с оценкой их истинности) мы вынуждены признать: система искусственного интеллекта мыслит. Истинность этого тезиса подтверждается следующим умозаключением:

1. Исчисление понятий есть мышление
2. Искусственный интеллект исчисляет понятия

---

3. Искусственный интеллект мыслит

Данный вывод заставил тех, кто не желал смириться с признанием возникновения рядом с человеком подлинного конкурента по разуму, искать переформулировку существенных признаков мышления, таких, которые бы отсутствовали у систем ИИ. Так возникла новая парадигма: «интеллект как восприятие». В соответствии с этой позицией, утверждалось, что аналитико-синтетическая деятельность ума представляет собой только лишь малую часть возможностей разумного поведения человека. Выдающимся проявлением

разумной жизни признавалась возможность к восприятию объектов в мире, к их осмыслению, подведению под различные родовые понятия, возможность разнообразных волевых действий, направленных на эти объекты, возможность обучения ориентации в окружающей среде.

Вместе с тем нашлись и те, кто принял вызов этой новой парадигмы интеллекта и утверждал возможность построения систем ИИ, соответствующего ей. Среди наиболее известных представителей этого направления можно назвать Нортона Винера, который занимался теоретическим обоснованием возможности построения искусственных систем, соответствующих биологической природе человека. Винер проводил исследования по моделированию чувственного восприятия человека, его ориентации в окружающей среде, его способности к обучению методом проб и ошибок, используя достижения различных областей знаний (математика, нейрофизиология, физика, электроника).

Литература

1. Анисимов А.В. ЭВМ и понимание математических доказательств //Вопросы философии, 1987, N 3.
2. Маслов С.Ю. Теория дедуктивных систем и ее применения. - М., Советское радио, 1986.
3. Сергеев В.М. Искусственный интеллект как метод исследования сложных систем // Системные исследования: методологические проблемы (ежегодник) М.: Наука, 1984.
4. Сергеев В.М. Искусственный интеллект: Опыт философского осмысления // Будущее искусственного интеллекта М.: Наука, 1991.

Контрольные вопросы

1. Каковы основные тезисы парадигмы «интеллект как исчисление понятий»?
2. Почему при принятии данной парадигмы интеллекта необходимо признать, что компьютер мыслит?
3. Каковы критические аргументы по отношению к парадигме "интеллект как исчисление понятий"?

ТЕМА 4

Парадигма "интеллект как восприятие".

Аннотация

Фиксация и манипулирование с объектами мира как специфическая характеристика интеллекта. Перцептрон. Затруднения парадигмы "интеллект как восприятие". Попытки различения деятельности животного, человека и системы искусственного интеллекта. Невозможность фиксации уникального способа деятельности, характеризующего человека.

Конспект лекции

Из тех, кто на практике пытался реализовать парадигму «интеллект как восприятие» при построении систем ИИ следует выделить, прежде всего, Фрэнка Розенбланта. В 1958 г. им была предложена модель электронного устройства, названного им перцептроном, которое должно было бы имитировать процессы чувственного восприятия. Перцептрон должен был передавать сигналы от "глаза", составленного из фотоэлементов, в блоки электромеханических ячеек памяти, которые оценивали относительную величину электрических сигналов. Эти ячейки соединялись между собой случайным образом в соответствии с господствующей тогда теорией, согласно которой мозг воспринимает новую информацию и реагирует на нее через систему случайных связей между нейронами. Два года спустя была продемонстрирована первая действующая машина "Марк-1", которая могла научиться распознавать некоторые из букв, написанных на карточках, которые подносили к его "глазам", напоминающим кинокамеры. Чтобы научить перцептрон способности строить догадки на основе исходных предпосылок, в нем предусматривалась некая элементарная разновидность автономной работы или "самопрограммирования". При распознании той или иной буквы одни ее элементы или группы элементов оказываются гораздо более существеными, чем другие. Перцептрон мог научаться выделять такие характерные особенности буквы полуавтоматически, своего рода методом проб и ошибок, напоминающим процесс обучения.

Однако возможности перцептрона были ограниченными: машина не могла надежно распознавать частично закрытые буквы, а также буквы иного размера или рисунка, нежели те, которые использовались на этапе ее обучения. Вообще, идея, предрекающая перцептрону большое будущее, как оказалось, выдавала желаемое за действительное. До сих пор конструирование на практике систем ИИ, способных ориентироваться в окружающей среде, распознавать объекты и классифицировать их, продвигается очень медленно. Один из авторитетных исследователей в области ИИ – Марвин Мински – продемонстрировал крайний пессимизм по этому поводу, сказав, что, не говоря о роли подвижных роботов или машин, способных читать, слушать и понимать прочитанное или услышанное, перцептроны

никогда не обретут даже умения распознавать предмет частично заслоненный другим. Глядя на торчащий из-за кресла кошачий хвост, подобная машина никогда не сможет понять, что она видит.

Для тех же, кто с радостью воспринял поражение перцептрона в соревновании с разумом человека, пришлось столкнуться с весьма неприятным теоретическим курьезом парадигмы «интеллект как восприятие», отмеченным Ларри Хаусером – одним из исследователей в области ИИ. Дело в том, что когда Декарт и Лейбниц предлагали парадигму «интеллект как исчисление понятий», они руководствовались желанием провести кардинальное различие между человеком и животным. Способность к последовательному рассуждению признавалась главным специфическим признаком человеческого разума в виду того, что провести данное различение по другим пунктам представлялось маловероятным – ведь животные также как и люди наделены перцептуальным аппаратом, позволяющим различать предметы и должным образом, исходя из биологических потребностей, ориентироваться в окружающей среде. Кроме того, животные способны (что в настоящее время получает все большее подтверждение) и на продуцирование эмоциональных состояний.

Принятие парадигмы «интеллект как исчисление понятий» привело, как мы установили выше, к признанию разумности искусственной интеллектуальной системы. ИИ оказался полностью подобным человеку в его существенных признаках. Те, кто не желал смириться с этим положением дел и пытался спасти уникальность человека, предложили парадигму «интеллект как восприятие», настаивая на том, что сущностные черты человеческого разума находятся за пределами аналитико-синтетической деятельности. В такой парадигме человека, действительно, удается отличить от ИИ. Но при этом вышеупомянутые мыслители не заметили того, что эта новая парадигма снова низводит человека до уровня животного, не позволяет провести здесь какого-то кардинального различия, т.е. того, к которому как раз и стремились Декарт и Лейбниц, формулируя свое понимание разума. В итоге, курьезная ситуация борьбы этих двух парадигм состоит в следующем: человеку никак не удается найти уникальное место, он оказывается подобным либо системе искусственного интеллекта (при выборе парадигмы «интеллект как исчисление понятий»), либо обычному биологическому организму животного (при выборе парадигмы «интеллект как восприятие»).

Литература

1. Кочергин А.Н. Искусственный интеллект и мышление // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 37-39.

2. Макарчук М.М. Об основном отличии искусственного и естественного интеллекта // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 50-52.
3. Моисеев В.И. Интервал Тьюринга и имитация интеллекта // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 307-310.
4. Пенроуз Р., Шимони А., Картрайт Н., Хокинг С. Большое, малое и человеческий разум. М.: Мир, 2004.

Контрольные вопросы

1. Каковы основные тезисы парадигмы "интеллект как восприятие"?
2. В чем состоит различие между человеческой деятельностью и поведением животного?
3. В чем заключался принцип работы перцептрона?
4. В чем состоят курьезы парадигмы «интеллект как восприятие»?

ТЕМА 5
Парадигма "интеллект как рефлексия". Парадигма "интеллект как самоидентичность". Тест Тьюринга.

Аннотация

Основные тезисы парадигмы "интеллект как рефлексия". М. Мински о специфике человеческой деятельности: мышление о мышлении, продуцирование методологии. Возможности моделирования рефлексии в системах искусственного интеллекта. Основные тезисы парадигмы "интеллект как самоидентичность". «Внутренний» мир субъекта как основная характеристика разумной жизни. Критика парадигмы "интеллект как самоидентичность". Тест Тьюринга. Эвристическая функция теста Тьюринга.

Конспект лекции

Наиболее отчетливо парадигму «интеллект как рефлексия» можно зафиксировать в исследованиях М. Мински. Мински утверждает, что об интеллекте можно говорить тогда, когда система не просто способна совершать операциональные действия, но и делать предметом внимания сам способ оперирования с объектами, т.е. рефлексировать по поводу

своей собственной деятельности.

Если поставить вопрос о реализации данной парадигмы в индустрии систем искусственного интеллекта, то мы можем констатировать, что даже одни из самых первых технических устройств, претендовавших на статус интеллектуальных, уже обладали вышеуказанной характеристикой. В пример можно привести компьютерные программы для игры в шахматы, которые получили свое развитие еще в 60-х годах XX века. Эти программы устроены так, что они способны к самосовершенствованию, которое достигается за счет того, что предметом информационной обработки становятся сами операциональные действия этой же программы.

Мыслители, разделяющие парадигму «интеллект как самоидентичность», полагают, что сущностью разума не следует считать ни способность к рассуждению, ни осмысленное восприятие объектов. Основной отличительной чертой разумного поведения признается способность к самоидентификации. Самоидентификация – это осознанность своих действий, способность в любой произвольно взятый момент отдать себе отчет о своих внутренних состояниях, способность подвести все свои психические переживания под единое основание сознания, которое в языке обозначается словом «Я».

Выработав данное определение разума, представители этой парадигмы тут же поспешили заявить, что именно здесь проходит подлинная демаркация человеческого разума и искусственного интеллекта. Системы искусственного интеллекта могут быть способны на моделирование рассуждений и даже на моделирование восприятия, но ни одна из таких систем не способна на осуществление актов самоидентификации. Проще говоря, ни один компьютер не способен сказать себе «Я».

Однако такая позиция также оказалась уязвимой для критики. Главный критический аргумент здесь состоял в фиксации препятствия для интерпретации интерсубъективности. Допустим, перед вами находятся двое – компьютер на столе и человек, стоящий рядом. Вы задаете вопрос человеку, и он вам отвечает. Посредством соответствующего интерфейса вы задаете вопрос компьютеру, и он точно так же, как и человек, отвечает на вопрос. Например, вы спрашиваете человека-энциклопедиста: «Как называется столица государства Непал?». Он вам отвечает: «Катманду». Очевидно, что точно такой же вопрос можно поставить перед справочной системой электронной энциклопедии и получить точно такой же ответ. В таком случае, можно спросить, на каком основании находящемуся перед вами человеку вы приписываете свойство самоидентификации, а стоящему рядом компьютеру – нет? Строго говоря, свойство самоидентификации с полной очевидностью мы можем обнаружить только в нас самих. Другим субъектам мы приписываем это свойство только по аналогии со своим собственным. Мы не можем проникнуть в их внутренний мир, в их сознание. Мы можем

лишь наблюдать подобие между собой и другими субъектами во внешнем поведении. И только на основании этого подобия мы совершаем так называемый «аналогизирующий перенос», т.е. наделяем противостоящего субъекта самосознанием, предполагаем, что и его внутренний мир также подобен нашему. Однако, если аналогия становится возможной только на основании внешнего поведения, то почему мы не осуществляем этого аналогизирующего переноса на компьютер? Его внешнее поведение также оказывается вполне подобным нашему. Ему задают вопрос – он отвечает. Нет ничего невозможного даже в том, чтобы научить компьютер на вопрос: «Это ты?» отвечать: «Я», или на вопрос: «Ты понимаешь, что это ты?» отвечать: «Да, я понимаю, что это Я». На основании этих данных система ИИ вполне заслуживает того, чтобы ей, также как и человеку, была приписана способность к самоидентификации.

В данном случае, главная проблема состоит даже не в том, почему мы не приписываем свойство самосознания компьютеру, а в том, почему, на каком основании мы приписываем свойство самосознания другим существам, наделенным естественным разумом? Пока не найдется ответа на этот вопрос, мы не сможем провести различие между человеческим разумом и системой искусственного интеллекта по заданному основанию.

Учитывая вышеизложенные проблемы с фиксацией существенных признаков естественного разума, американский исследователь Алан Тьюринг сформулировал свой тест на определение разумности искусственного интеллекта, который впоследствии стал широко употребим в виду своей простоты. Тьюринг исходил из бихевиористских позиций, он предложил рассматривать человеческое сознание в качестве так называемого «черного ящика». Мы знаем, что подается на «вход» этого «устройства», знаем, что имеем на «выходе», но просто перестаем задавать вопрос о том, что происходит внутри него. Тем самым, все сложные вопросы с определением того, как именно на физическом и на психическом уровнях функционирует человеческий разум оказываются вне рассмотрения. Мы обсуждаем только внешнее проявление человека. Мы задаем ему вопросы и фиксируем его реакцию – т.е. слушаем ответы, наблюдаем за его действиями.

Точно так же Тьюринг предложил оценивать и работу системы искусственного интеллекта. Главный тезис его теста гласит следующее: если по внешним признакам своей деятельности машина демонстрирует свое подобие деятельности человека, то, вне зависимости от того, что происходит «внутри» машины, будем считать ее поведение разумным. Проще говоря, если при заданных обстоятельствах, при заданных условиях поиска решения задачи машина ведет себя так же, как бы в этих обстоятельствах повел себя человек, то будем считать ее разумной. Например, если на вопрос «Сколько будет 2+2?» машина выдает ответ 4, будем считать это разумным ответом.

Тест Тьюринга оказывается крайне либеральным по отношению к многочисленным системам ИИ. С этой точки зрения, например, мой карманный микрокалькулятор следует признать мыслящей сущностью.

Литература

1. Лорьер Ж.-Л. Системы искусственного интеллекта М., Мир, 1991.
2. Моисеев В.И. Интервал Тьюринга и имитация интеллекта // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 307-310.
3. Пенроуз Р., Шимони А., Картрайт Н., Хокинг С. Большое, малое и человеческий разум. М.: Мир, 2004.
4. Хофштадтер Д., Деннет Д. Глаз разума. Самара: Издательский дом «Бахарм-М», 2003.

Контрольные вопросы

1. Каковы основные тезисы парадигмы "интеллект как рефлексия"?
2. Каковы основные тезисы парадигмы "интеллект как самоидентичность"?
3. Каковы критические аргументы по отношению к парадигме "интеллект как самоидентичность"?
4. В чем заключается тест Тьюринга?
5. Какова эвристическая функция теста Тьюринга?

ТЕМА 6

Понятие интенциональности. Парадигма "интеллект как интенциональность". Аргумент "китайская комната".

Аннотация

Существенные признаки содержания понятия "интенциональность". Различие физических и психических объектов. Онтологическая несамостоятельность психического объекта. Основные тезисы парадигмы "интеллект как интенциональность". Интенциональность как субъективная осознанность деятельности, понимание происходящего, приписывание объекту деятельности смысловой характеристики.

Содержание аргумента "китайская комната". Система искусственного интеллекта как неинтенциональная система. Различие естественного и искусственного интеллектов.

Конспект лекции

Intendo – указываю на; имею в виду. Избрав, вслед за средневековыми схоластами, этот латинский глагол для образования базового понятия своего учения, Ф. Брентано в значительной степени определил тезаурус философской мысли XX века. Интенциональность – одно из немногих понятий, которое с высокой степенью интенсивности одновременно использовалось в очень разных, порой даже противоположных друг другу, философских школах современности.

Брентано вводит понятие интенциональности для проведения решающего различия между физическими и психическими явлениями. Любая физическая вещь автономна. Это значит, что она способна существовать без какой-либо поддержки извне. Конечно, всякая физическая вещь также не мыслима без тех отношений, в которые она вступает с другими вещами. Тем не менее, анализируя суждение «Книга лежит на столе», мы не сомневаемся в том, что данное отношение является внешним для каждого из объектов, представленных здесь. И книга, и стол без какого-либо ущерба для своего существования могут быть выведены за пределы этого отношения и поставлены в связь с другими объектами. Эти новые связи могут быть манифестированы в суждениях «Книга лежит в портфеле» и «Стол находится в комнате». Такой способ отношений как раз и характеризует специфику натуралистических связей вещей объективного мира.

Совсем иная ситуация предстает перед нами в качестве результатов анализа так называемых интенциональных связей. Здесь, как утверждает Брентано, связываемые вещи находятся во внутреннем отношении, они не мыслимы друг без друга за пределами этой связи. В интенциональное отношение вступают также два элемента. Однако, в отличие от натуралистической связи здесь взаимодействуют не две физические вещи, а определенное психическое состояние и тот объект, в отношении к которому оно находится. Этот способ отношения может быть манифестирован в таких суждениях, как «Я вижу дерево за окном» или «Я хочу стакан воды».

Именно способность человеческой психики быть направленной на объект познания, принимать его во внимание в качестве некоторого идеального смыслового единства признается в данной теории самой существенной характеристикой разума. Человек не просто бездумно оперирует с объектами мира, он наделяет их смыслом, он понимает происходящее.

Включившись в дискуссию об AI, Д. Серл представил свой, ставший широко известным, «аргумент китайской комнаты», суть которого сводится к следующему.

Допустим, человека, владеющего только английским языком, помещают в изолированную от внешнего мира комнату и предоставляют ему для чтения текст на китайском. Естественно, в виду того, что он не имеет ни малейшего представления о значении китайских иероглифов, текст оказывается для него набором чернильных закорючек на листе бумаги – человек ничего не понимает. Затем ему дают еще один лист бумаги, исписанный по-китайски, и, в придачу к этому, определенную инструкцию на родном для него английском о том, как можно было бы сравнить два китайских текста. Эта инструкция научает выявлению тождественных символов и определению закономерности их вхождения в более общий контекст. Когда приносят третий китайский текст, к нему прилагают вторую английскую инструкцию о сравнении последнего с двумя предыдущими и т. д. В итоге, после продолжительных упражнений испытуемому приносят чистый лист бумаги и просят что-нибудь написать по-китайски. К этому времени человек из китайской комнаты настолько хорошо освоил формальные символические закономерности, что, на удивление, действительно оказался способным написать вполне связный и понятный любому грамотному китайцу текст. И наконец, чтобы произвести должный эффект, человека выводят из комнаты на обозрение широкой публике и представляют как англичанина, изучившего китайский, что сам виновник презентации не замедлит подтвердить своим безукоризненным знанием иероглифического письма.

Так понимает ли наш испытуемый китайский? Серл дает категорически отрицательный ответ на этот вопрос. Понимание должно сопровождаться актами первичной интенциональности, в которых сознание, еще до всякого обращения к каким-либо материальным носителям, т. е. к речи или письму, способно концентрироваться на внутренних интенциональных содержаниях, как нередуцируемых ни к чему другому фактах автономной психической жизни. Интенциональность языка производна, она возникает при намеренном наделении изначально пустых знаков значением, посредством замещения внутреннего интенционального содержания пропозициональным содержанием синтаксически организованных структур.

Для общественности, которая оценивала результаты обучения человека из китайской комнаты, возникла иллюзия того, что экзаменуемый действительно овладел китайским. Причина этой иллюзии кроется в той привычке, в соответствии с которой люди предположили за пропозициональными содержаниями продуцированных человеком синтаксических форм его внутренние интенциональные содержания, явившиеся основой первых. Но на деле обучение в китайской комнате принесло прямо противоположные результаты. Человек научился формальным операциям со знаковой системой без какого-либо собственного «интенционального участия» в этом предприятии. Пропозициональные

содержания представленного на обозрение китайского письма имели смысл только для тех, кто действительно мог подкрепить их более фундаментальными интенциональными содержаниями своей психики. Человек из китайской комнаты сам не понял ничего из того, что написал.

По мысли Серла действия испытуемого англичанина полностью аналогичны работе AI. Искусственный интеллект, несмотря ни на какие интенсификации в сфере технологий, никогда не сможет достичь уровня человеческого сознания именно из-за невозможности преодолеть фундаментальный разрыв между первичной и производной интенциональностями. С помощью специальных программ, настраивающих на формальное оперирование символическими образованиями, AI может создавать иллюзии мощнейшей мыслительной активности, многократно превышающей способности человеческого сознания. Результаты такой деятельности AI оказываются, в самом деле, чрезвычайно полезными для человека. Тем не менее, у нас нет никаких оснований тешить себя иллюзией существования «братьев по разуму». AI не мыслит. Всю работу по содержательному наполнению пустых символических структур берет на себя человек, «прикрепляя» последние к внутренним интенциональным содержаниям – подлинным элементам разумной жизни.

Литература

1. Дубровский Д.И. Искусственный интеллект и проблема сознания // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 26-32.
2. Кочергин А.Н. Искусственный интеллект и мышление // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 37-39.
3. Ладов В.А. Интенциональность как основание различия человеческого сознания и искусственного интеллекта // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 39-43.
4. Серл Д. Мозг, сознание и программы // Аналитическая философия: становление и развитие (антология). М., 1998. с. 376 - 400.

Контрольные вопросы

1. Каковы существенные признаки понятия "интенциональность"?
2. Каковы основные тезисы парадигмы "интеллект как интенциональность"?
3. В чем смысл аргумента "китайская комната"?
4. К каким следствиям приводит рассмотрение данного аргумента?

ТЕМА 7

Понятие производной интенциональности. Операциональная деятельность.

Аннотация

Существенные признаки понятий "первичная интенциональность" и "производная интенциональность". Критика теории первичной интенциональности. Понятие первичной интенциональности как фундаментальная предпосылка новоевропейской философии в осмыслении проблемы разумности. Фиксация иллюзорности первичной интенциональности естественного интеллекта. Механизм приписывания производной интенциональности. Операциональная деятельность человеческого сознания. Операциональная деятельность системы искусственного интеллекта. Фундаментальная идентичность работы систем естественного и искусственного интеллектов.

Конспект лекции

Обсуждая работу автомата по продаже Пепси-Колы, еще один американский философ Д. Деннет без колебаний принимает серлевские аргументы. Что значат наши фразы: «Автомат понял, что я поместил в него настоящую американскую монету, и выдал мне банку с напитком» или «Я его обманул: вместо монеты я опустил в приемник подходящий кусок металла, он ошибся и снова угостил меня баночкой Пепси»? Только то, что мы используем исторически сложившуюся, привычную для нас форму речи. Мы антропоморфизируем автомат, приписывая ему знакомые нам самим интенциональные состояния.

Аргументируя в защиту данного тезиса, Деннет приводит в пример комичный случай из истории торговой индустрии. В пятидесятых годах американские автоматы по продаже напитков появились в Панаме, будучи оборудованы специальным детектором для приема панамских монет. Однако панамские и американские четвертаки оказались на то время подобными друг другу по форме, весу и даже по тому материалу, из которого они изготовлены. Автомат все чаще начал ошибаться, выдавая банку Пепси в том случае, когда вместо панамской монеты, в него помещали американскую. В результате эти первые автоматы по продаже напитков быстро исчезли из Панамы – их использование было невыгодно для панамского правительства.

Что же мы имеем в виду, когда говорим здесь об ошибках устройства? Вполне можно представить себе ситуацию (правда, теперь не реальную, а только воображаемую), что панамское правительство благосклонно отнеслось к равноправному хождению панамских и американских денег на территории своей страны. Будет ли тогда действие автомата, в том

случае, если он принимает американскую монету, считаться ошибкой? Очевидно, нет. Значит одному и тому же физическому действию устройства могут быть приписаны различные интенциональные характеристики, различные модусы интенции. В данном случае - правильное восприятие и ошибочное восприятие. У нас нет сомнения в том, что само устройство не обладает «внутренним чувством» того, что оно ошибается в данный момент. Его физическое состояние всегда одно и то же, оно совершенно индифферентно по отношению к каким-либо интенциональным оценкам. Мы приписываем действию автомата производные интенциональные состояния только потому, что сами обладаем внутренней интенциональностью. Интенциональность наших действий во внешнем мире также производна, как и интенциональность действий автомата, но у нас есть что-то еще: первичная интенция, абсолютно недоступная обсуждаемому устройству.

Отметим еще один момент. Конструкторы детектора по приему монет могут проявлять чудеса инженерной мысли, научая автомат различать не только вес, форму и материал монет, но и отчеканенные на ее поверхностях знаки и рисунки так, что устройство окажется способным отличать друг от друга равные по весу, форме и материалу американские и панамские четвертаки. Тем не менее, даже в этом случае мы, по-прежнему, не сможем утверждать, что автомат понимает, что перед ним американская монета. Отчеканенные знаки и рисунки на американском четвертаке что-то «значат» для автомата только в отношении внешнего сравнения этих неровностей и закорючек с неровностями на панамской монете. Взятые сами по себе, без какого-либо внешнего материального отношения, эти неровности и закорючки не значат ничего. В процессе работы автомата они не отсылают ни к какому внутреннему интенциональному содержанию. Детектор в принципе не может понять, что такое американская монета.

Если бы исследование Деннета заканчивалось только этим тезисом, то сложно было бы отыскать какой-то особый смысл в том, чтобы обсуждать его теорию отдельно. На самом деле, позиция этого американского философа оказывается гораздо более оригинальной и, в конце концов, радикально отличной от интенционализма Д. Серла.

Деннет, как мы только что увидели, полностью соглашается с Серлем в том, что ИИ не обладает первичной интенциональностью, а довольствуется лишь ее производными формами, навязанными ему извне человеческим сообществом. Но в отличие от Серла он утверждает следующее: не только ИИ, но и человек не обладает первичной интенциональностью. Миф о первичной интенциональности – один из самых глубоких предрассудков классической философской традиции Запада. ИИ оказывается действительно подобным человеческому сознанию, но не в том, что он, как и человек, обладает первичной интенциональностью, а, наоборот, в том, что человек, как и ИИ, ею не обладает. Не ИИ

похож на человека, а человек на ИИ. Деннет снова пытается презентировать свою позицию с помощью конкретных примеров. Последуем за ним.

Некто Джонс, отправившись в космическое путешествие, прибывает на планету Земля-Двойник (ЗД). Все здесь оказывается Джонсу знакомо: люди, дома, деревья, небо – все как на Земле. Пообедав в ресторане, пообщавшись с местными жителями и неспешно прогуливаясь по городу, Джонс наткнулся на рекламный проспект, сообщавший об очередном туре скачек на лошадях на местном ипподроме. Джонс был очень возбужден этим обстоятельством и немедленно отправился на ипподром. А возбуждение его было связано с тем, что на Земле он был предупрежден об одной странности фауны той планеты, на которую он улетал. ЗД есть точная копия Земли с одним исключением. Там, на скачках, кроме лошадей можно встретить особых животных – смошадей. Смошади ни по виду, ни по повадкам совершенно не отличаются от лошадей. Тем не менее, смошади не есть лошади.

Так как Джонс имел интерес к познанию и был склонен к самонаблюдению, то его очень волновал вопрос о том, что с ним будет происходить, когда он увидит на ипподроме животных, как он будет пытаться отличить лошадь от смошади. При этом он знал, что данная эпистемологическая ситуация радикализируется тем фактом, что местные жители на ЗД для именования и смошадей, и лошадей используют одно и то же слово – «лошадь», так что выяснить у них с помощью вопроса то, с чем он имеет дело в своем восприятии, не представляется возможным.

Так вот, попав на ипподром и тщательно сосредоточившись на своих внутренних состояниях, наш герой с очевидностью обнаружил, что не имеет в данный момент ничего, что можно было бы назвать первичным интенциональным содержанием. Глядя на проносившихся мимо него животных, он не знал как себя вести, о чем думать: о том, что он имеет действительное восприятие лошади; о том, что он имеет восприятие лошади, но ошибается, так как перед ним на самом деле смошадь; о том, что он имеет действительное восприятие смошади; или о том, что имеет восприятие смошади и ошибается, так как перед ним на самом деле лошадь?

Суть проблемы в том, что восприятие как определенное психическое переживание, действительно, имеет место так же, как имеет место физическое состояние автомата Пепси-Колы в тот момент, когда в него опускают монету, но вот само интенциональное содержание в качестве смысловой интерпретации воспринимаемого объекта равным образом отсутствует в обоих случаях.

Как же тогда возникает определенная смысловая интерпретация? Она возникает из фона, окружения, из определенных, но, в конечном счете, произвольных правил приписывания интенциональных содержаний тем или иным состояниям. Если окружающие

меня люди соглашаются признать в созерцаемых животных смошадей, то эти животные становятся смошадьми. Окружающие начинают и моему восприятию приписывать определенное интенциональное содержание и говорят: «Сейчас он видит смошадь». В конце концов, я совершаю самый изощренный психический пируэт. Я сам на свое полое переживание налагаю производное интенциональное содержание, принятое мной из сообщества, и убеждаю себя в том, что, в самом деле, с очевидностью, вижу смошадь.

Нет сомнения, что сколь бы фантастическим ни выглядел пример Деннета, он, в качестве универсального эпистемологического аргумента, вполне может быть распространен на любое проявление познавательной активности субъекта, на все сферы опыта вообще. Чтобы увидеть здесь проблему, не нужно отправляться в далекое космическое путешествие – разве на Земле нет смошадей? Мы уверены в этом?

Если появление любого интенционального содержания в сознании человека зависит от согласованных правил операций с объектами (более строго – с символами объектов, хотя Деннет не заостряет внимание на лингвистической стороне вопроса), то ИИ думает и понимает ничуть не меньше человека, точнее, человек понимает ничуть не больше, чем ИИ. Система программ возможного ИИ может охватить собой весь мир так, что при взаимной согласованности правил обхождения с объектами своей деятельности каждый элемент ИИ будет демонстрировать понимание (в прямом и единственном смысле этого слова) происходящего, этот мир будет также полон смысла, как и человеческий мир.

Литература

1. Арбиб М. Метафорический мозг. М.: Едиториал УРСС, 2004.
2. Ладов В.А. Интенциональность как основание различия человеческого сознания и искусственного интеллекта // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 39-43.
3. Макарчук М.М. Об основном отличии искусственного и естественного интеллекта // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 50-52.
4. Сергеев В.М. Искусственный интеллект: Опыт философского осмысления // Будущее искусственного интеллекта М.: Наука, 1991.

Контрольные вопросы

1. Каковы существенные признаки понятий "первичная интенциональность" и "производная интенциональность"?

2. Каковы критические аргументы по отношению к теории первичной интенциональности?
3. Можно ли зафиксировать отличия в операциональной деятельности человеческого сознания и искусственного интеллекта?

ТЕМА 10

Синтаксис и семантика систем искусственного интеллекта. Проблема гомункулуса.

Аннотация

Специфика синтаксиса и семантики языка систем искусственного интеллекта. Использование языка человеком и машиной: сходства и различия. Вопрос о возможности задания синтаксической интерпретации языка системы искусственного интеллекта. Проблема гомункулуса. Уровни постановки проблемы гомункулуса.

Конспект лекции

С позиции лингвистического подхода в ИИ-философии серлевский аргумент «китайская комната» утверждает то, что язык искусственных систем не имеет семантики. Вся работа в системе интерфейса человек-компьютер со стороны машины происходит исключительно на синтаксическом уровне. Компьютер «обучен» определенным программам-алгоритмам связи символических элементов знаковой системы так, что возникает впечатление относительно их семантической нагруженности.

Возьмем в качестве примера работу географической электронной энциклопедии. Система ИИ запрограммирована так, чтобы, получив от человека запрос: «Как называется столица Непала?», выдавать ответ: «Катманду». При этом очевидно, что компьютер не понимает, что собственно стоит за теми знаками языка, которые использованы в данном запросе, семантически они для него пусты. Просто в соответствующей программе дана директива: «при запросе, представляющим собой один синтаксический комплекс, выдавать в качестве ответа другой».

Машина может действовать как формальный логик. Отвлекаясь от какого-либо содержательного наполнения, она способна к оценке истинности сложных высказываний на основании анализа истинностных функций составляющих их простых атрибутивных суждений. Она способна оценить истинность вывода как в случае содержательно

наполненного высказывания «Если на улице идет дождь, то асфальт мокрый», так и в случае высказывания, продуцированного на «тарабарском» языке: «Если жунсы губеют, то брунсы тернеют». Путем объединения логических электронных микросхем, принцип работы которых будет соответствовать истинностным функциям для логических союзов, в общую схему, мы можем построить электрическую цепь, моделирующую весьма сложные дедуктивные рассуждения. Однако эта «дедуктивная ловкость» машины будет только подтверждать тезис Серла - в работе компьютера нас завораживает именно эта невероятная синтаксическая мощь, скорость оперирования знаками. Тем не менее, какими бы головокружительными ни были операции, связанные с синтаксисом формальной знаковой системы, компьютер никогда не сможет самостоятельно задать им какую-либо семантическую интерпретацию.

На основании информации о морфологии языка машина может даже имитировать продуцирование связанного текста. Известно, что те или иные части речи подчиняются специфическим законам словообразования и потому данный процесс также можно формализовать. Например, читая фразу: «Глока куздра бодланула бокра и бодрячит бокренка» мы вполне можем предположить здесь наличие частей речи – существительных, глаголов и прилагательных, - структура которых формально будет напоминать обычные, т.е. семантически нагруженные термины нашего естественного языка. Данные морфологические элементы, опять же в соответствии с определенными синтаксическими правилами, будут занимать соответствующие места в предложениях – места подлежащего, сказуемого, определений и т.д. Если машину запрограммировать на соответствие всем этим правилам, то ее успехи также могут быть очень впечатляющими. Однако в данном случае система ИИ будет себя вести в точном подобии с упоминаемым выше персонажем из китайской комнаты, который научился лишь филигранному оперированию синтаксическими элементами знаковой системы по определенным правилам без какой-либо семантической интерпретации.

Итак, на основании серлевского аргумента китайской комнаты мы можем утверждать, что с точки зрения лингвистического подхода в ИИ-философии существенным признаком разумной деятельности будет считаться способность к семантической интерпретации знаковой системы. И именно этим признаком не обладают системы искусственного интеллекта.

Аргумент китайской комнаты вызвал бурные обсуждения в рамках традиции ИИ-философии. Здесь нашлись как его приверженцы, так и оппоненты. Одни утверждали, что системе ИИ, действительно, никак нельзя приписать способность к семантической интерпретации, другие настаивали на том, что в определенном смысле эту способность

нельзя приписать даже и человеку. При этом все молчаливо согласились тезисами относительно синтаксиса.

Интересный момент в исследовании данной проблемы, на который хотелось бы обратить внимание в этой теме, заключается в том, что спустя десятилетие тот же Д. Серл весьма оригинальный образом пересмотрел свой собственный аргумент. На этот раз вопрос был поставлен о синтаксисе. А можем ли мы, в самом деле, утверждать – так как мы это делали ранее - что машина способна на выполнение синтаксических процедур в рамках заданной знаковой системы? Теперь американский философ дал отрицательный ответ и на этот вопрос.

Для того, чтобы прояснить смысл серлевской аргументации, вновь вспомним, для начала, англичанина, изучающего китайский. Критический аргумент относительно семантики начинался с того, что человек не понимает значений написанных на бумаге символов. Осваивая формальные правила операций с данными символами, он овладевает определенной синтаксической техникой, которая создает иллюзию семантической осведомленности. Однако не пропустили ли мы здесь, при описании данного лингвистического действия, одного важного момента, на который нам следовало бы обратить внимание еще до начала формулировки критического аргумента относительно семантики? Что именно может увидеть человек на предоставленных в его распоряжение листах бумаги? Строго говоря, на физическом уровне на листе бумаги виден лишь хаотический набор чернильных пятен различной формы. Получается, что прежде чем констатировать свою неосведомленность относительно семантики языка, человек из китайской комнаты уже должен задать определенную синтаксическую интерпретацию! Он должен понять вот эти чернильные пятна на листе бумаги именно как знаки, которые, возможно, объединены какой-либо системой правил функционирования, составляя при этом единое целое – язык. Он должен понять, что вот эти чернильные пятна в принципе могут что-то обозначать. При рассмотрении того, как тот или иной субъект – неважно, человек или машина - овладевают и пользуются языком, синтаксис не должен возникать по принципу Deus ex machina. На физическом уровне, в среде материальных носителей языковых структур нет никакого синтаксиса. Для того, чтобы тот или иной материальный объект оказался знаком, ему следует задать не только семантическую интерпретацию, которая покажет, что, собственно, этот знак обозначает, но и, прежде всего, интерпретацию синтаксическую, которая покажет, что данный материальный объект в принципе может что-то обозначать, т. е. является знаком.

Известно, что в основание информатики была положена гениальная в своей простоте идея, которую продуцировали американские математики и техники, в частности, Клод

Шеннон – объединение логики и электричества. К тому времени – 30 гг. XX века – в логике уже прочно зарекомендовал себя новый подход, основанный на слиянии формально-логической символики и языка математики. Сначала на основе алгебры Дж. Буля, которая представляла собой формализацию арифметических действий, было предложено воспользоваться алгебраическим языком и для формализации логического процесса рассуждения. Были введены символы для всех возможных логических констант, характеризующих формальные элементы в суждениях – логических союзов. Отрицание «-«, дизьюнкция «V», коньюнкция «&», импликация «⊃», тождество «≡». Затем было предложено заменить логические референты «истина» и «ложь» арифметическими символами 1 и 0. Далее Г. Фреге и другие философы аналитической традиции построили систему референций для каждого логического союза – это были так называемые таблицы истинности. Таким образом, появилась возможность оценивать истинность какого-либо сложного высказывания на основе анализа составляющих его простых высказываний и их истинностных функций. Все это, в свою очередь, привело к возможности формального контроля над системой дискурсивного рассуждения вообще, т.е. к оценке логической последовательности и необходимости выводов из посылок какой угодно степени сложности.

Так вот новизна идеи информатики заключалась лишь в том, что было предложено интерпретировать наличие или отсутствие напряжения в электрической цепи как знаки арифметических символов 1 и 0 соответственно. Так и возник цифровой компьютер. Теперь все логические наработки по анализу рассуждений можно было бы смоделировать на электрическом уровне путем создания соответствующих элементов цепи – сначала электрических лам, затем транзисторов, и, в конце концов, электронных микросхем, закодированных на выполнение-имитацию истинностной функции какого-либо логического союза.

Еще раз обратим внимание на ключевые элементы данного процесса интерпретации. Сначала цифры 1 и 0 были поняты как знаки логических референтов «истина» и «ложь». Арифметика в данном случае оказывалась синтаксисом для логической семантики. Но составляет ли этот синтаксис «онтологию машины»? Т. е. действительно ли hardware компьютера – это нагромождение нулей и единиц? Если бы это было так, то идея информатики не представляла бы какой-либо ценности. В том-то и дело, что физика не имеет синтаксиса вообще. Наличие напряжения в электрической цепи – это еще не единица. Это лишь наличие напряжения в электрической цепи. Представим себе по ходу ситуацию, что мы вкручиваем лампочку в патрон настольной лампы с плохим качеством контакта проводников тока. Лампочка то загорается, то гаснет – будем ли мы считать данные события физического уровня передачей, скажем, какого-либо зашифрованного кода? В данном случае, конечно же,

нет. В этом как раз и состоит первичный интерпретативный шаг – понять определенный уровень электрического напряжения как знак. И пока не важно знак чего: логического референта или, скажем, предупреждения об опасности пожара. Физический уровень в качестве материального носителя языковых выражений прежде семантической должен вначале получить синтаксическую интерпретацию, которая задается внешним образом, через пользователя данной знаковой системой.

Если при использовании электронной энциклопедии я задаю вопрос о столице Непала и получаю надлежащий ответ, то машина не только не понимает значений символов, с которыми она оперирует в соответствии с определенным алгоритмом, она не представляет собой даже и формальную синтаксическую систему. На физическом уровне вслед за одним случаем высокого уровня напряжения в определенном участке цепи, возникает другой случай – только и всего. Для того, что бы эти факты высокого напряжения понять как знаки, которые могут подчиняться определенным операциональным правилам их сочетания, необходимо задать первичную синтаксическую интерпретацию, на которую в дальнейшем и будет опираться программист при формулировке соответствующего алгоритма операций.

Интересно, что в связи с появлением данного аргумента в отношении синтаксиса системы ИИ в когнитивной науке с новой силой вспыхнули дискуссии вокруг так называемой проблемы гомункулуса. Параллельно развитию информатики и ИИ-философии в когнитивной науке, основанной на современных достижениях нейрофизиологии, очень активно стали проявлять себя исследования, основанные на интерпретации мозга как цифрового компьютера. Известно, что на физическом уровне движение нейронов представляет собой, в определенном смысле, движение электрических зарядов, а нейронные цепи можно уподобить цепям электрическим. Если понимать компьютер как скопление информации, закодированной в цифровом виде, то точно также можно было бы и отнестись к работе головного мозга. В таком представлении мозг оказывался подобным машине, где “hardware” представляет собой физическое наличие нейронных связей, на которые накладываются синтаксическая и семантическая интерпретации. Однако если даже синтаксис не свойственен физике, а привносится на материальный уровень внешним образом, то здесь возникает проблема. Кто задает подобные синтаксические интерпретации? С компьютером, который мы покупаем в магазине, все проще – здесь интерпретацию задает пользователь. Но как быть с головным мозгом? Получается, что, кроме того, кто является физическим носителем нейронных связей, должен существовать еще и тот, кто интерпретирует эти нейронные соединения как синтаксические элементы системы. Так возникает проблема гомункулуса – своеобразного «разума в разуме», того, кто

интерпретирует физику. По сути же, с точки зрения материалистической онтологии, в мозге, как и в компьютере, нет никаких нулей и единиц.

Литература

1. Searle, J.R. Is the Brain a Digital Computer? // www.consciousness.arizona.edu
2. Дубровский Д.И. Искусственный интеллект и проблема сознания // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 26-32.
3. Кочергин А.Н. Искусственный интеллект и мышление // Философия искусственного интеллекта. М.: ИФ РАН, 2005. с. 37-39.
4. Хофштадтер Д., Деннет Д. Глаз разума. Самара: Издательский дом «Бахарм-М», 2003.

Контрольные вопросы

1. Какова специфика синтаксиса и семантики языка систем искусственного интеллекта?
2. В чем состоит различие в использовании языка человеком и машиной?
3. В чем состоит проблема гомункулуса?

Глоссарий

Аналитическая философия – одно из ведущих направлений современной западной философии; распространена, прежде всего, в англоязычных странах; рассматривает философские проблемы логики, математики, сознания, языка, искусственного интеллекта.

Интенциональность – понятие, введенное в философский лексикон австрийским мыслителем Ф. Брентано; обозначает свойство сознания быть направленным на предмет.

Исчисление понятий – логическая операция по сочленению понятий в суждения и умозаключения, результатом которой является построение дискурсивного рассуждения.

Логический референт – особый предмет, на который указывает предложение языка; Г. Фреге выделял два логических референта: «истина» и «ложь».

Операциональная деятельность – манипулирование с объектами по определенным, заранее заданным правилам – алгоритмам; характерна для систем искусственного интеллекта.

Перцептрон – техническое устройство, имитирующее естественный процесс восприятия и классификации объектов; разработано в середине XX века американским ученым Ф. Розенблантом.

Проблема гомункулуса – теоретическое затруднение в области когнитивной науки, связанное с необходимостью ответа на вопрос о субъекте, генерирующем семантическую и синтаксическую интерпретацию физических процессов головного мозга.

Рефлексия – способность сознания обращать внимание не только на предметы, но и на сам способ внимания к предметам, т.е. на свою собственную деятельность.

Самоидентичность – способность мыслящего субъекта отдавать себе отчет о своем «я».

Семантика языка – поле значений, смыслов, к которым отсылает знаковая система.

Синтаксис языка – структура знаковой системы, подчиняющаяся определенным законам и правилам построения.

Система искусственного интеллекта – техническое устройство, производящее имитацию или моделирование интеллектуальной деятельности.

Формализация естественного языка – процесс перевода знаков естественного языка в термины логики и математики.

Эпистемология – учение о познании; обсуждает вопросы о структуре и процессах познания окружающей действительности человеком.

Философия искусственного интеллекта – одно из направлений современной философии; обсуждает вопросы, связанные с появлением и развитием искусственных интеллектуальных систем; распространена, главным образом, в англоязычных странах в качестве раздела аналитической философии.

Возможные темы курсовых работ

1. Тест Тьюринга для систем искусственного интеллекта.
2. Аргумент «Китайская комната» в философии искусственного интеллекта.

3. Понятие «производной интенциональности» в философии искусственного интеллекта Д. Деннета.
4. Проблема следования правилу в философии искусственного интеллекта.
5. Предпосылки возникновения систем искусственного интеллекта.
6. История развития систем искусственного интеллекта.
7. Современные направления в философии искусственного интеллекта.
8. Искусственный интеллект: перспективы развития.
9. Перцептрон Ф. Розенблатта и современная робототехника.
10. Парадигма «интеллект как рефлексия» в философии искусственного интеллекта М. Мински.