

Феноменология против символического искусственного интеллекта: философия научения Хьюберта Дрейфуса

СЕРГЕЙ АСТАХОВ

Преподаватель, Школа философии, факультет гуманитарных наук,
Национальный исследовательский университет «Высшая школа экономики»
(НИУ ВШЭ). Адрес: 105066, Москва, ул. Старая Басманная, 21/4. E-mail:
sastakhov@hse.ru.

Ключевые слова: искусственный интеллект; феноменология; научение; ментальные репрезентации; моторная интенциональность.

В 1960-е годы между исследователями искусственного интеллекта (ИИ) и феноменологом Хьюбертом Дрейфусом завязался конфликт, который продлится до 2000-х годов. Создатели первых программ ИИ считали, что научение представляет собой решение проблем с помощью особых ментальных репрезентаций, эвристик. Дрейфус же доказывал, что эвристики не требуются для научения, поскольку сознание и тело позволяют гибко реагировать на проблемные ситуации без всяких ментальных репрезентаций. Автор статьи показывает, как из критики символического ИИ возникла феноменология человеческого навыка, освещая предысторию конфликта и анализируя основные противоречия между двумя концепциями научения. Выступая против ассоцианизма в работах Г. Саймона, А. Ньюэлла и Э. Фейгенбаума, Дрейфус предложил свое понимание взаимоотношений между сознанием и телом. С его точки зрения, человек обладает периферийным сознанием, инсайтом и терпимостью к неоднозначности, у него есть осо-

бая структура тела и потребности, что позволяет ему различать релевантное и нерелевантное в окружающей среде и стремиться к ее максимальному постижению.

Автор статьи демонстрирует, как теории научения, разработанные в рамках символического ИИ, повлияли на пятиэтапную модель навыка самого Дрейфуса. С одной стороны, эта модель объясняла, почему программы Саймона и его коллег сначала достигли успеха; с другой стороны, она четко обозначила пределы их развития. Чтобы прояснить телеологию научения, Дрейфус исследовал связи между идеей моторной рациональности Мориса Мерло-Понти и нейросетевым моделированием. Представлены два взгляда исследователей на роль Дрейфуса в истории ИИ, наряду с причинами, по которым его философия почти не повлияла на сообщество ИИ, хотя оказалась очень популярной в социогуманитарных дисциплинах. Наконец, описаны вызовы, которые стоят перед феноменологией научения сегодня.

Введение

Два любителя борются друг с другом, и обоим озаряет, у каждого возникает коварный план: заманить противника в ловушку, обычно хода за два. Либо он попадетсЯ, либо нет. Дрейфус серьезно проигрывал, и тогда он нашел ход, которым мог съесть вражеского ферзя. Для противника выход был один: своим ферзем ставить Дрейфусу шах до того момента, пока его король и ферзь не попадут в вилку — тогда можно будет сделать размен. Программа так и поступила. Как только размен произошел, стратегия Дрейфуса развалилась на части, и программа поставила ему мат прямо в центре доски. Так что игра вовсе не была механической; это была типичная партия между двумя людьми, с моментами драмы и провала, которые и бывают в таких играх. Это было замечательно¹.

ТАК ГЕРБЕРТ САЙМОН, один из отцов-основателей искусственного интеллекта и будущий лауреат Нобелевской премии по экономике, описывает партию между американским философом Хьюбертом Дрейфусом и программой *MacHack VI* в 1967 году². Дрейфус не ожидал, что проиграет: до этого он читал только о примитивных программах, которые не справлялись даже с начинающими игроками. Так, в 1960 году шахматная программа Герберта Саймона, Аллена Ньюэлла и Клиффа Шоу проиграла десятилетнему ребенку за 35 ходов³.

Я благодарю Татьяну Ермолаеву и Сергея Захарова за помощь при подготовке текста.

1. *McCorduck P. Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*. Natick, MA: A.K. Peters, 2004. P. 231–232.
2. Подробную запись партии можно найти здесь: The Chess Game “Hubert Dreyfus vs. Mac Hack VI” (1967) Annotated // Ingram Braun. 25.04.2017. URL: <https://ingram-braun.net/public/research/parlour-games/article/computer-chess-richard-greenblatt-match-mit-philosophy-artificial-intelligence-history>.
3. *Dreyfus H. L. Alchemy and Artificial Intelligence*. RAND papers, 1965. P. 6. URL: <https://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf>.

Теперь проиграл он, и его поражение смаковало все сообщество исследователей искусственного интеллекта⁴. В специальном журнале вышла статья с заголовком «Дрейфус: десятилетка может побить машину. Но машина может побить Дрейфуса». Мы взяли цитату в начале статьи из интервью, которое Саймон дал писательнице и историку ИИ Памеле Маккордак. Из отрывка видно, насколько он был рад, что *MacHack VI* победила. Для него поражение Дрейфуса оказалось символическим событием.

Это была не просто игра, это было столкновение двух концепций. Как Дрейфуса, так и Ричарда Гринблатта, создателя *MacHack VI*, волновала природа человеческого мастерства. Гринблатт, вслед за Саймоном, Ньюэллом и Шоу, считал, что эксперты обладают особыми ментальными репрезентациями, эвристиками, которые помогают им выбирать хорошие ходы и отбраковывать неудачные. Эвристики были особыми правилами, которые не обладали универсальной применимостью, но часто помогали в решении задач. Гринблатт встроил в *MacHack VI* пятьдесят эвристик, опираясь на свое знание шахмат. Дрейфус же полагал, будто эксперты не пользуются никакими эвристиками, им не нужны ментальные репрезентации для игры в шахматы. Скорее, особое устройство сознания и тела позволяет человеку накапливать опыт в разных ситуациях, из-за чего он все лучше и лучше ухватывает важные для практики аспекты реальности. Раз у компьютера нет сознания и тела, то он не сможет достичь уровня человека-эксперта.

Поражение Дрейфуса не означало, будто его теория провалилась. Он был в лучшем случае любителем⁵, и машина победила просто потому, что у него было совсем мало игрового опыта. Успех *MacHack VI*, однако, на самом простом уровне подтверждал связь научения и эвристик. Впоследствии *MacHack VI* станет первой программой, которая будет играть против людей на официальных соревнованиях и даже получит шахматный рейтинг.

К 1967 году не существовало единой теории человеческого мастерства. Научение (*skill acquisition*) было глубоко проблематичным феноменом: ученые искали его базовые механизмы и пыта-

4. Далее — ИИ. Термин «искусственный интеллект» имеет два значения: с одной стороны, он отсылает к особой дисциплине, с другой — к программам, которые разрабатывают специалисты в этой области.

5. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. N.Y.: The Free Press, 1986. P. 112.*

лись объяснить, как они обеспечивают развитие навыков до экспертного уровня. Эта проблема научения, как мы будем ее дальше называть, с трудом поддавалась анализу, так как сами эксперты не могли четко определить, что делает их экспертами. И шахматная партия была лишь небольшим эпизодом из истории конфликта вокруг природы навыка. С середины 1960-х годов проблема научения превратилась в поле битвы, на котором феноменолог Дрейфус пытался доказать, что компьютерные программы Саймона, Ньюэлла, Эдварда Фейгенбаума и многих других не смогут достичь человеческого мастерства.

В этой статье мы расскажем, как из критики символического ИИ развился феноменологический подход, который переопределил проблему научения в терминах телесного сознания и ситуативности. Сначала опишем, как зародился конфликт между Дрейфусом и сообществом ИИ и в чем заключались их содержательные расхождения. Затем мы покажем, какие подходы к научению предложили Саймон, Ньюэлл и Фейгенбаум. Реагируя на их тезисы, Хьюберт и его брат Стюарт создадут знаменитую пятиэтапную модель развития навыка. Эта теория появится в 1980-е годы, и Дрейфус будет вносить в нее изменения, даже когда перестанет выпускать книги об ИИ. В конце мы расскажем о неоднозначном успехе феноменологии научения и о тех вызовах, с которыми она сталкивается сейчас.

Травля в MIT

Хьюберт Дрейфус не сразу стал врагом сообщества ИИ, и вначале проблема научения не входила в круг его интересов. В 1950-е годы он прошел несколько стажировок в европейских университетах, посетил Фрайбург, Лёвен, где работал в архиве Эдмунда Гуссерля, и Париж. Он прослушал курс Карла Ясперса, встречался с Мартином Хайдеггером и Жаном-Полем Сартром, исследовал работы Мориса Мерло-Понти. По возвращении в США он подготовил и защитил диссертацию на тему «Феноменология восприятия Гуссерля». А пока он изучал континентальную философию, в США разворачивалась масштабная когнитивная революция⁶, затронувшая антропологию, нейрофизиологию, лингвистику, психологию. Дисциплина ИИ возникла в рамках нового движения, и Дрейфус

6. Gardner H. The Mind's New Science: A History of the Cognitive Revolution. N.Y.: Basic Books, 1985.

столкнется с ее представителями, когда будет работать в компании *RAND*⁷.

Феноменологу было совсем не просто попасть в учреждение, занимавшееся государственными оборонными контрактами. Стюарт Дрейфус работал в *RAND* на компьютере *JOHNNIAC*⁸. В это время Пол Армер, глава отдела компьютерных наук в *RAND*, искал сторонних экспертов для оценки проекта *Cognitive Simulation*, созданного Саймоном, Ньюэллом и Шоу. Стюарт посоветовал своего брата как консультанта, Армер проверил рекомендации Хьюберта и нанял его в 1964 году⁹. Дрейфус должен был дополнить исследования по когнитивной симуляции философским анализом. Его отчет не понравился Армеру, который в интервью Маккордак говорит о «ленивой философии», «плохой философии» Дрейфуса¹⁰. Он не хотел публиковать отчет в *RAND*, и вполне возможно, что его поддерживали сторонники Саймона. Но внутри компании было и другое крыло: Стюарт Дрейфус, известный математик Ричард Беллман и даже ассистент Армера психолог Роберт Райанстет настаивали, что текст надо опубликовать. Армер задержал публикацию на девять месяцев¹¹, но в итоге выпустил текст в печать как служебный отчет (*memo*)¹².

Текст назывался «Алхимия и искусственный интеллект»¹³. Дрейфус обрушивался на ИИ, сравнивая всю новую многообещающую область с алхимией. Он доказывал: Саймон и Ньюэлл просто сами не понимают, что предлагают. Среди прочего они заявляли, что к 1967 году цифровой компьютер станет шахматным чемпионом, докажет важную математическую теорему и создаст музыку, ценность которой признают критики¹⁴. Дрейфус писал, что эти предсказания не сбудутся.

7. Сокращение от *Research and Development*. Корпорация *RAND* была создана как некоммерческий аналитический центр, консультирующий Министерство обороны США по стратегическим проблемам. Среди прочего консультанты из *RAND* занимались вопросами, связанными с космической «гонкой», ядерным вооружением, ИИ и национальным здравоохранением.
8. *JOHNNIAC* был назван в честь Джона фон Неймана и представлял собой ламповый компьютер, который использовался для научных и инженерных расчетов.
9. *McCorduck P. Machines Who Think. P. 226.*
10. *Ibid. P. 227.*
11. По воспоминаниям Дрейфуса, на год. См.: *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 8.*
12. Это был самый низкий статус публикации в *RAND*.
13. *Dreyfus H. L. Alchemy and Artificial Intelligence.*
14. *Ibid. P. 3.*

В это время Хьюберт Дрейфус преподавал философию в Массачусетском технологическом институте (MIT), одном из центров развития ИИ, так что после публикации статьи с ним перестали общаться. Академические проявления вежливости стали просто неуместны; как вспоминает Дрейфус, «студенты и профессора, работавшие над проектом робота, не смели обедать вместе со мной, чтобы не вызвать недовольства своих руководителей»¹⁵. Стюарт Дрейфус работал на той же системе, что и Саймон, — JOHNNIAC, — и теперь они «едва разговаривали друг с другом»¹⁶.

Многие исследователи сдались бы под таким давлением. Но статья Дрейфуса стала очень популярной, о ней упомянули в *The New Yorker*, в разделе *Talk of the Town*. Он стал часто появляться в медиа как философ — эксперт по ИИ. Во многом поэтому его пригласили сыграть против *MacHack VI*. Когда он проиграл, «это вызвало овации исследователей „компьютерного интеллекта“»¹⁷.

Противостояние поставило под угрозу карьеру Дрейфуса. Как он впоследствии вспоминал, «они хотели помешать мне получить пожизненный контракт. <...> Я получил его через их трупы»¹⁸. Добившись контракта в MIT, в 1968 году он уехал в Калифорнию преподавать в Беркли.

После успеха «Алхимии...» Дрейфусу предложили написать книгу об ИИ. В книге он развернул тезисы, критикующие отцов-основателей ИИ, их амбиции, модели и методы¹⁹. В 1984 году Дрейфуса пригласили на передачу *Computer Chronicles* на телеканале KCSM для беседы об ИИ. Он должен был сбалансировать свидетельства сторонников ИИ — на передачу позвали программистов Джона Маккарти²⁰ и Майкла Генесерета. Однако в последний момент Маккарти отказался выступать на одной площадке с Дрейфусом²¹. Съемки перенесли, а Дрейфусу предложили составить вопросы, которые зададут Маккарти в следующий раз. Дрейфус сделал список из шести вопросов, но в студии прозвучали только два, причем их значительно отредактировали²².

15. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 9.*

16. *Ibidem.*

17. *Тоффлер Э. Шок будущего. М.: АСТ, 2002. С. 233–234.*

18. *Hanley J. Hubert Dreyfus Interview // Full-Tilt Boogie. 20.10.2005. URL: <http://full-tilt.blogspot.com/2005/10/hubert-dreyfus-interview.html>.*

19. *Дрейфус Х. Чего не могут вычислительные машины. М.: Прогресс, 1978.*

20. Маккарти придумал термин «искусственный интеллект» и написал язык *Lisp*.

21. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 13.*

22. *Nichols P. The Computer Chronicles — Artificial Intelligence (1984) // YouTube. 09.11.2012. URL: https://youtu.be/_S3moV_ZF_Q.*

После первой книги он написал еще два тома, основанные на обзорах нововведений в сфере ИИ²³. На два десятилетия он стал главным философским критиком ИИ в США. Орудием его критики была континентальная философия, авторы от Гуссерля и Хайдеггера до Мишеля Фуко. Он был далек от аналитической философии, доминировавшей на кафедрах философии в Америке, хотя, например, поддерживал связи с Джоном Сёрлом. Согласно социологам Вадиму Волкову и Олегу Хархордину, именно Дрейфус познакомил американских студентов с французской и немецкой философией, заразил их экзистенциализмом и герменевтикой²⁴.

На протяжении трех книг Дрейфус не менял своего главного аргумента. Проект ИИ не добьется успеха, пока будет развивать допущение, что человеческое мышление — это формальные операции с символами. Человеческое сознание нельзя понять, если рассматривать его как программу и описывать только на языке ментальных репрезентаций — образов, представлений, мыслей, пропозиций, правил, эвристик и т. д., — которые хранятся в его памяти. Из этих идей возникла его критика символического ИИ, а затем и феноменология научения. Прежде чем перейти к проблеме научения, мы рассмотрим самые общие расхождения между Дрейфусом и сообществом ИИ.

Против символического искусственного интеллекта

Дисциплина ИИ развивалась в рамках когнитивной революции 1950-х и 1960-х годов. Среди особенностей когнитивных наук историк и психолог Говард Гарднер указывает два ключевых свойства: постулирование уровня ментальных репрезентаций и «веру, что для понимания человеческого сознания особенно важен электронный компьютер»²⁵. Отцы-основатели ИИ Саймон и Ньюэлл развили эти установки в гипотезе системы физических символов: «Физическая символьная система имеет необходимые и достаточные средства для произведения основных интеллектуальных операций»²⁶. Гипотеза делала когнитивную симуляцию главной зада-

23. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine; Dreyfus H. L. What Computers Still Can't Do: A Critique of Artificial Reason. Cambridge, MA: MIT Press, 1992.*

24. *Волков В. В., Хархордин О. В. Теория практик. СПб.: ЕУСПб, 2008. С. 48.*

25. *Gardner H. The Mind's New Science. P. 6.*

26. *Newell A., Simon H. A. Computer Science as Empirical Inquiry: Symbols and Search // Communications of the ACM. 1976. Vol. 19. № 3. P. 116.*

чей инженеров и программистов: они должны были создать такую программу, чтобы она могла производить ключевые интеллектуальные операции.

В истории ИИ эту модель часто называют символическим ИИ. Она предполагала, что операции с ментальными репрезентациями, которые проделывает человек, можно понять через манипуляции с формальными символами.

Новая парадигма противоречила всему, что Дрейфус узнал о человеческом мышлении из текстов Хайдеггера, Мерло-Понти и позднего Витгенштейна. Сначала в «Алхимии...», а потом и в своих книгах он атакует Саймона и Ньюэлла, доказывая, что сознание не работает как компьютерная программа, а мышление не всегда нуждается в ментальных репрезентациях. Здесь проблема научения еще не выходит на передний план, но Дрейфус закладывает основы своей позиции, которые предопределяют его понимание навыка.

Ассоцианизм

Когда «Алхимия и искусственный интеллект» вышла в *RAND*, классической формулировки гипотезы Саймона–Ньюэлла еще не было. Вместо этого Дрейфус писал об ассоцианистском допущении: «Мышление должно быть анализируемо через простые определенные операции»²⁷. Если интеллектуальные процессы просты и легко отделяются друг от друга, то их можно повторить через манипуляции символами. Такие процессы не сводятся ни к биохимическим феноменам в мозге, ни к социальным и культурным феноменам. Дрейфус связывал ассоцианизм с конкретной философской традицией, идущей от Тита Лукреция Кара к Рене Декарту и Дэвиду Юму, полагавшим, будто мышление можно разложить на простые элементы. Они, в свою очередь, повлияли на научную психологию XIX века и в конечном счете на идею ИИ.

Дрейфус утверждал, что ассоцианизм нельзя обосновать эмпирически. Исследования, которые проводят в *MIT*, Университете Карнеги–Меллона и *RAND*, методологически ограничены и не доказывают, что мышление — это переработка информации, которую можно легко представить в виде дискретных операций с символами. Дрейфус рассматривал конкретную программу Саймона и Ньюэлла — *General Problem Solver (GPS)*. Они позиционировали ее как способную решать самые разные задачи — из гео-

27. *Dreyfus H. L. Alchemy and Artificial Intelligence. P. 48.*

метрии, логики предикатов, шахмат. Ученые предполагали, что базовое свойство мышления — это способность решать проблемы (*problem-solving*). Поэтому они предлагали экспертам описать вслух свои действия при решении задач, чтобы составить особые протоколы. Далее они находили в этих протоколах эвристики, формализовывали их и передавали программам.

Если эвристики были получены правильно, то программа должна была повторить человеческие действия при решении проблемы. Поэтому они сопоставляли устные протоколы и последовательные шаги программ, чтобы доказать, что машина способна повторить интеллектуальные операции. В одной из статей Саймон и Ньюэлл сообщали о пяти расхождениях между протоколами и действиями программ. И при этом они продолжали утверждать, будто человеческое поведение можно рассматривать как «результат сложного, но конечного и определенного набора законов [предположительно ассоцианистских]»²⁸. Дрейфус правомерно замечает, что научные законы не признают исключений.

Развивая мысль, он указывает, что Саймон и Ньюэлл не корректируют свою гипотезу, а, наоборот, заявляют, что она все больше подтверждается. Но главная проблема заключается в том, что сама гипотеза производит свидетельства, которые ее обосновывают. Программы типа *GPS* или *Logical Theorist* — ассоцианистские теории, предполагающие, что мышление можно представить как формальные операции. И когда эти программы производят правильные решения тех или иных проблем, может показаться, будто они подтверждают ассоцианизм, то есть свое исходное допущение. Очевидно, тут возникает круг в обосновании, и его можно разорвать, только используя независимые факты.

Но не только методологические ошибки беспокоят Дрейфуса. Проект когнитивной симуляции уязвим, поскольку ассоцианистское допущение нельзя обосновать априори. Во-первых, далеко не всякое знание можно легко формализовать, и это подтверждают расхождения между человеческими протоколами и действиями программ. Во-вторых, даже если сознание или бессознательные когнитивные процессы можно описать с помощью дифференциальных уравнений, это совершенно не означает, что при своей работе сознание или бессознательное решает уравнения²⁹.

Если в «Алхимии...» речь шла только об одном ассоцианистском допущении, то в книге «Чего не могут вычислительные ма-

28. *Dreyfus H. L. Alchemy and Artificial Intelligence. P. 51.*

29. *Дрейфус Х. Чего не могут вычислительные машины. С. 118.*

шины» оно разделяется на четыре: биологическое допущение предполагает дискретность при передаче сигналов между нейронами; психологическое допущение представляет мышление как переработку информации через бинарный код; эпистемологическое допущение задает любые знания, все, что может быть понято, в терминах логических отношений; онтологическое допущение интерпретирует происходящее в мире как множество фактов, каждый из которых логически независим от остальных³⁰.

Такое распределение позволяет Дрейфусу сделать критику более точечной, атаковать несколько целей. Он проблематизирует каждый переход: от мира к знанию о мире, от сигналов среды к мозгу, от мозга к сознанию. Теперь становится ясно, как много сомнительных метафор использовали ИИ-специалисты. Они не обосновали, почему метафора переработки информации применима к мозгу и почему эту переработку можно выразить через бинарный код. Но даже если мы допускаем «переработку информации» в сознании человека, мы не должны думать, что можно так же легко снять различие между сигналами из внешней среды и процессами в мышлении. В самом понятии «информация» смешиваются обыденные представления и технический смысл, оно слишком широкое³¹. Кроме того, Саймон, Ньюэлл и их последователи слишком быстро перескакивали от сознания к мозгу³², от синапсов к внутренним содержаниям мышления. В этом смысле биологическое допущение поддерживает психологическое.

Следующая ошибка — полагать, что мозг перерабатывает внешние сигналы как цифровой компьютер. Дрейфус вовсе не против обсудить биохимические особенности мозга, и он подчеркивает, что его можно описывать скорее как аналоговое устройство³³. В любом случае нельзя автоматически предполагать, что нервные ткани работают с бинарным кодом. Дрейфус посвятит множество страниц вопросу о функционировании мозга, некоторое время он будет интересоваться голографической гипотезой нейрофизиолога Карла Прибрама³⁴, потом станет поддерживать модель нейронных сетей³⁵.

30. Дрейфус Х. Чего не могут вычислительные машины. С. 106.

31. Там же. С. 116.

32. Там же. С. 207.

33. Там же. С. 209–212.

34. Dreyfus H. L., Dreyfus S. E. *Mind Over Machine*. P. 59–61; см. также: Прибрам К. Язык мозга. Экспериментальные парадоксы и принципы нейропсихологии. М.: Прогресс, 1975.

35. Dreyfus H. L. *What Computers Still Can't Do*. P. xxxiii.

А главное, символический ИИ не учитывал реальную работу сознания, которую исследуют гештальт-психологи и феноменологи. По Дрейфусу, человек воспринимает паттерны в окружающей среде совершенно особенным образом. Он обладает периферийным сознанием (*fringe consciousness*³⁶), позволяющим ему выхватывать из мира сигналы в фоновом режиме. И хотя таких сигналов может быть неопределенно много, человек оказывается защищен от комбинаторного взрыва, когда решение задачи многократно усложняется с каждым следующим шагом доказательства. Именно с этой проблемой не справился GPS. Далее, человек может различать релевантное и нерелевантное на основе глубокого понимания ситуаций, инсайта (*insight*). Программы же различают важное и неважное согласно правилам, составленным их авторами.

Кроме того, люди чувствительны к контексту и обладают терпимостью к неоднозначности (*ambiguity tolerance*). Они могут понять двусмысленное предложение, они воспринимают иллюзии вроде куба Неккера или эффекта Мюллера-Лайера³⁷. Иными словами, люди владеют экстралингвистической информацией и принимают контекстуальное использование языка.

Вместе периферийное сознание, инсайт и терпимость к неоднозначности опрокидывают простое ассоцианистское описание мышления. По Дрейфусу, их нельзя передать через дискретные операции с символами. При этом он не отстаивает иррациональную природу мышления, а, скорее, призывает специалистов в области ИИ учитывать холистические особенности сознания и принимать ограничения своей области.

Тело и сознание

И критика ассоцианизма, и описание особенностей сознания уже присутствовали в «Алхимии и искусственном интеллекте». Но в статье для RAND Дрейфус практически не писал о телесности человека. Этот шаг был необходим, чтобы из него развернулась философия научения. Дрейфус переходит к идее телесного сознания в своей первой книге. Он опирается на раннего Мерло-Понти³⁸ и диссертационную работу феноменолога Сэмюэла Тодса, посвященную тому, как структура тела должна влиять на со-

36. Термин Дрейфус заимствует у Уильяма Джеймса.

37. *Dreyfus H. L. Alchemy and Artificial Intelligence*. P. 56–58.

38. *Мерло-Понти М. Феноменология восприятия*. СПб.: Ювента; Наука, 1999.

знание³⁹. Именно значение тела упустили программисты и инженеры, когда разрабатывали все новые и новые программы.

Для Дрейфуса фундаментальная связь сознания и тела была аргументом против правилосообразности человеческого поведения. Описывая, как человек улавливает ритм, он отмечал:

Наше тело не есть какое-то правило, локализованное в разуме, — правило, которое можно сформулировать или использовать вне зависимости от реальной деятельности по антиципации ритмических длительностей⁴⁰.

Иными словами, правила не обладают достаточной гибкостью, чтобы работать во всех возможных контекстах.

Если правила внеконтекстуальны, то тело, наоборот, очень чувствительно к контексту или, согласно феноменологии Дрейфуса, к ситуации. Дрейфус взял у Мерло-Понти опорную идею: тело — это «синергетическая система», то, что позволяет человеку переходить от одной сенсорной модальности к другой⁴¹. Именно тело позволяет нам чувствовать запах яблока, видеть цвет яблока, прикидывать вес яблока в руке и соединять эти разные образы в единый предмет. По Дрейфусу, оно делает это не благодаря какому-то списку правил; тело способно напрямую, без промежуточных процессов, реагировать на разные ситуации, при этом запоминая свои реакции. Поэтому человеку не нужно огромное количество расчетов, чтобы воспринимать внешний мир⁴².

Дальше Дрейфус сделал удивительный шаг: он привлек феноменологию Гуссерля, которую обычно относил к враждебной философской традиции. Хотя он и критиковал Гуссерля за трансцендентализм, это не помешало ему связать концепцию внутреннего и внешнего горизонтов с идеями Мерло-Понти и Тодса⁴³. Внешний горизонт отвечает за фон, за информацию в восприятии, которая не исключена до конца. В терминологии Дрейфуса внешний горизонт — это «общее чувство ситуации в целом»⁴⁴. В «Алхимии и искусственном интеллекте» он назывался периферийным сознанием. Внутренний горизонт определяет, что у лю-

39. Todes S. *Body and World*. Cambridge, MA: MIT Press, 2001.

40. Дрейфус Х. Чего не могут вычислительные машины. С. 213.

41. Мерло-Понти М. Феноменология восприятия. С. 300–301.

42. Дрейфус Х. Указ. соч. С. 218.

43. Там же. С. 213–214.

44. Там же. С. 206.

бого объекта больше аспектов, чем мы наблюдаем в данный момент. Он включает в себя прошлый опыт общения с объектом, то, как человеку раскрывались его детали⁴⁵. Именно поэтому внутренний горизонт задает наши ожидания от объекта, предвосхищения. Кроме того, внутренний горизонт зависит от тела, потому что именно оно связывает друг с другом разные аспекты объекта, в том числе сенсорные модальности. Таким образом, восприятие объектов, их стабильность, способность видеть аспекты, предвосхищать поведение объектов — это все телесные возможности. И они гораздо более гибкие, чем основанные на правилах программы: внутренний горизонт позволяет человеку предвосхищать частично неопределенные данные⁴⁶, причем это предвосхищение невозможно описать однозначно. Предвосхищение так же подвижно, как и ситуации, с которыми оно стыкуется.

У символического ИИ нет горизонтов⁴⁷, у него нет воспринимающего и активного в среде тела. А еще у него нет потребностей, которые задавали бы, что для него релевантно, а что нерелевантно, тем самым ограничивая потенциальную бесконечность информации из мира. Эту мысль Дрейфус взял у Тодса, подкрепляя ее цитатами из Хайдеггера и позднего Витгенштейна⁴⁸. Причем человек не обладает «изначальным генетическим реестром потребностей и ценностей»⁴⁹, потребности не предзаданы, и потому их нельзя рассматривать как цели, к удовлетворению которых человек рационально стремится. Потребности структурируют ситуации, но и сами от них зависят, так что их удовлетворение напоминает творческий процесс.

Из этой теории вырисовывается ранняя концепция научения Дрейфуса. Благодаря своему сознанию и телу человек может ухватывать определенные аспекты объектов в разнообразных ситуациях, запоминать их и на основе накопленного опыта все успешнее реагировать на их изменения. За счет горизонтов тело обладает гибкостью: если человек ошибается, восприятие ситуации и объекта меняется, и он пробует новую реакцию на ситуацию. Причем и тут ментальные репрезентации не нужны — человек просто «схватывает объект на основе предвидения»⁵⁰. Речь

45. Дрейфус Х. Чего не могут вычислительные машины. С. 205–206.

46. Там же. С. 220.

47. Там же. С. 205.

48. Там же. С. 228–229.

49. Там же. С. 247.

50. Там же. С. 215.

не идет о мистическом прорицании, предвидение возникает за счет экстраполяции прошлого опыта. Схватывание же объекта стремится к максимальному постижению сути, к наиболее четкому ухватыванию всех аспектов, и именно поэтому человек способен развивать навыки. Важно, что максимальное постижение невозможно описать вне контекста: поскольку оно связано с изменчивыми ситуациями и человеческими целями, то может принимать самые разные формы. То есть в разных практиках успех выглядит по-разному.

Стратегия Дрейфуса в 1970-е годы была достаточно четкой: во-первых, показать, что ИИ покоится на целом множестве сомнительных допущений, и во-вторых, прояснить различия между людьми и компьютерами так, чтобы ограничения самых продвинутых программ стали очевидны. Поскольку у компьютера нет человеческого мозга, нет сознания и сложно организованного тела, нет потребностей, он не сможет повторить человеческих достижений. Программы не живут в человеческом мире.

Однако у этой стратегии были свои проблемы. Во-первых, многие программы успешно справлялись со своими задачами. Они доказывали теоремы, переводили простые предложения и обыгрывали философов в шахматы. Как мы отмечали, поражение не опрокидывало его феноменологию. У Дрейфуса было мало опыта, и к тому же он никогда не заявлял, будто компьютеры не смогут играть на любительском уровне⁵¹. Но проблема лежала глубже. Дрейфус проиграл, потому что программы становились сильнее, развивались. И его феноменология не могла установить предел развитию символического ИИ. Не было никаких философских гарантий, что через какое-то время машина не побьет чемпиона мира.

Во-вторых, Дрейфус построил свою антропологию на критике ментальных репрезентаций. Однако он не мог полностью отказать людям в ментальных репрезентациях: тогда не получилось бы объяснить, как человек способен мыслить правила и следовать им. Ему пришлось бы доказывать, что математика и логика невозможны, что рациональная философия от Платона до Гуссерля просто не могла появиться. А он посвятил множество страниц описанию и критике именно этой традиции. При этом он сам иногда использовал термин «цель» для описания научения⁵². Чтобы

51. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 112.*

52. *Дрейфус Х. Чего не могут вычислительные машины. С. 215.*

избежать этих противоречий, нужно было прояснить роль ментальных репрезентаций при развитии навыка, соотнести правила и ситуации внутри некоторой новой схемы.

Научение у Саймона, Ньюэлла и Фейгенбаума

Дрейфус развивал феноменологию научения, реагируя на самые радикальные тезисы в сообществе ИИ. Саймон и Ньюэлл в 1950-е и 1960-е годы искали универсальную рациональную основу научения и потому задавали экспертам задачи по математике, логике и шахматам. Ученые полагали, что у экспертов было больше правил и сами эти правила были качественнее, чем у новичков. По Дрейфусу, эта идеология себя не оправдала. Как мы уже отмечали, между протоколами и пошаговыми действиями машин возникали расхождения, которые поставили под вопрос возможности когнитивной симуляции. Кроме того, программа *GPS* столкнулась с проблемой комбинаторного взрыва. И хотя ее дизайн повлиял на историю программирования, специалисты на некоторое время перестали конструировать общие модели логического мышления.

С точки зрения феноменолога, проект Саймона и Ньюэлла и не мог достичь серьезных результатов, потому что эксперты не занимались бессознательными вычислениями с использованием скрытых эвристик. Скорее, их сознания и тела позволяли так отбирать реакции на проблемные ситуации, что они накапливали уникальный опыт решения задач.

Концепция эвристик была не единственной влиятельной гипотезой Саймона. Вместе с психологом Уильямом Чейзом он разработал концепцию чанкинга⁵³. На этот раз ученые анализировали движения глаз шахматистов, чтобы понять, как они принимают решения. Опираясь на оценку объема краткосрочной памяти Джорджа Миллера (7 ± 2 элемента)⁵⁴, Саймон и Чейз предположили, что опытные игроки могут запоминать в рамках тех же 7 ± 2 элементов краткосрочной памяти больше позиций, чем неопытные. Это означает, что их элементы, или порции (*chunks*), включали в себя больше связей и единиц. Если экстраполировать эту гипотезу, — что сделали многие последователи Саймона и Чей-

53. Chase W. G., Simon H. A. Perception in Chess // Cognitive Psychology. 1973. Vol. 4. № 1. P. 55–81.

54. Gardner H. The Mind's New Science. P. 89–90.

за⁵⁵, — то секрет мастерства в разных областях заключается в том, насколько больше информации человек может кодировать с помощью тех же 7 ± 2 элементов.

Другое направление было связано с Ньюэллом и его коллегами Джоном Лэйрдом и Полом Розенблумом, которые развили принципы *Logic Theorist* и *GPS* в когнитивной архитектуре⁵⁶ *SOAR*⁵⁷. *SOAR* также рассматривала научение как решение проблем, учитывала чанкинг и, кроме того, определяла развитие навыка через ускорение⁵⁸. Она оказалась большим достижением как в когнитивной теории, так и в исследовательской практике: специалисты стали использовать ее для самых разных задач, начиная с тренировочных программ для авиационных пилотов и заканчивая компьютерными играми и мобильными приложениями⁵⁹. Хотя *SOAR* была создана в 1983 году, она до сих пор поддерживается и обновляется исследовательской группой Лэйрда в Мичиганском университете.

Дрейфус же продолжал критиковать новые ветви символического ИИ. Он видел в чанкинге ассоцианистское допущение и утверждал, что эксперты не перебирают порции информации в кратковременной памяти, когда анализируют позицию в шахматах⁶⁰. Более того, порции не имеют смысла, если не встраивать их в общую ситуацию на доске, чего Саймон и Чейз не учитывали. В примечаниях доставалось и особенностям *SOAR*: ситуации не описываются с помощью списка объективных признаков, они изменчивее; чтобы запустить правило действия, нужно идеальное совпадение между ситуацией для навыка и тем описанием, которое хранится у агента. В феноменологии Дрейфуса для экс-

55. Например, чанкинг — ключевая тема одного из самых популярных онлайн-курсов в мире: *Learning How to Learn: Powerful Mental Tools to Help You Master Tough Subjects*// Coursera.org. URL: <https://ru.coursera.org/learn/learning-how-to-learn#instructors>. Создательница курса Барбара Оаки понимает чанкинг как универсальный механизм научения.

56. «Термин „когнитивная архитектура“ отсылает как к абстрактным когнитивным моделям для естественных и искусственных [интеллектуальных] агентов, так и к программам, разработанным на основе данных моделей, которые затем применяются в области ИИ» (*Lieto A. et al. The Role of Cognitive Architectures in General Artificial Intelligence*// *Cognitive Systems Research*. 2018. Vol. 48. P. 1).

57. *Speelman C. P., Kirsner K. Beyond the Learning Curve: The Construction of Mind*. Oxford; N.Y., 2005. P. 54–55.

58. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine*. P. 209–210.

59. В частности, *SOAR* используется для приложения *Liar's Dice* и таких игр, как *Starcraft* и *Minecraft*.

60. *Ibid.* P. 34.

пертного действия было достаточно общего сходства между новой и старой ситуацией. В результате Ньюэлл понимал научение как ускоряющееся решение проблем, а Дрейфус видел в навыке «прерывное движение, в котором ситуативные проблемы вызывают гибкий и беглый ответ»⁶¹.

В исследованиях ИИ критику Дрейфуса вызвала еще одна концепция научения — теория Фейгенбаума, ученика и частого соавтора Саймона. Сначала Дрейфус рассматривал Фейгенбаума как союзника, противопоставляя его «инженерии знаний» (*knowledge engineering*) и символический ИИ. Он думал, что база знаний в таких программах была похожа на игру или микромир в программе *SHRDLU*, то есть обладала четкими и стабильными критериями релевантности. В открытом мире таких критериев нет, но в закрытой информационной среде они вполне могли быть. Дрейфус считал, будто «инженерия знаний» работает только с конкретными задачами в очень ограниченной сфере.

Хотя речь шла о специальных областях, Фейгенбаум и его сторонники все же претендовали на большее: они хотели формализовать мастерство на высших уровнях навыка. Поэтому уже в работе 1986 года Дрейфус обрушивается на его концепцию⁶². Возможно, он передумал из-за книги Фейгенбаума и Маккордак, посвященной экспертным системам⁶³.

Фейгенбаум, по Дрейфусу, продолжал дело Саймона и Ньюэлла: он искал ментальные репрезентации, которые делают человека экспертом. Саймон и Ньюэлл разработали некоторые общие стратегии, их *GPS* был программой, моделирующей интеллект вообще. Фейгенбаум же сразу дистанцировался от моделирования «общего решателя задач», будучи заинтересован в отдельных сферах экспертизы. Он определил экспертные системы следующим образом:

Экспертные системы — это компьютерные программы, работающие на уровне людей-экспертов в различных профессиональных областях⁶⁴.

Экспертные системы работали с тремя видами знания. Во-первых, они использовали общие правила и стратегии, например поиск решения в базе данных или обратный вывод. Это был необходи-

61. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 210.*

62. *Ibid. P. 101–121.*

63. *Feigenbaum E. A., McCorduck P. The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World. L.: Pan, 1984.*

64. Цит. по: *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 101.*

мый минимум, который разработали Саймон и Ньюэлл. Во-вторых, они работали со множеством специальных фактов, которые обычно можно найти в «учебниках и статьях из этой области»⁶⁵. Наконец, системы использовали эвристическое знание, построенное на опыте экспертов и их «искусстве верно угадывать» решение проблемы. На простейшем уровне экспертная система состояла из двух частей: «базы знаний», в которой находятся экспертные данные, и «машины вывода» (*inferential engine*); машина вывода вела поиск по базе знаний, чтобы произвести те или иные суждения о конкретном случае⁶⁶.

Из этого описания становится ясно, что Фейгенбаум в целом разделял установки символического ИИ. Например, он верил в эвристики, более того, рассматривал профессиональный опыт как выводное знание (*inferential knowledge*), работающее с символами и основанное на опыте⁶⁷. Принципиальное отличие экспертных систем от *GPS* и *Logic Theorist* заключалось в том, что они работали в намеренно ограниченных областях, то есть не сталкивались в полной мере с проблемой релевантности/нерелевантности и имели дело с задачами с закрытой, однозначной структурой⁶⁸. Фейгенбаум подчеркивал: эксперты не всегда понимают, что именно делает их экспертами, так что задача специалистов по ИИ — найти эти скрытые правила верного угадывания, возникшие путем проб и ошибок. Специалистам по «инженерии знания» приходилось опрашивать экспертов, выуживать из них эти скрытые эвристики⁶⁹.

Дрейфус снова выступил против экспертных эвристик: они представляли собой дискретные ментальные репрезентации, а значит, основывались на том же ассоцианистском допущении, то есть в конечном счете на неправильном описании работы сознания и тела. Но он не полностью отрицал ментальные репрезентации. Правила и эвристики были нужны новичкам, чтобы понять базовые принципы практики. Однако на высоких уровнях навыка работа шла с тысячами конкретных случаев, для которых не могло быть конечного набора эвристик. Эксперты были хороши не эвристиками, а опытом, умением реагировать на мельчайшие аспекты профессиональных ситуаций. Именно поэтому

65. Dreyfus H. L., Dreyfus S. E. *Mind Over Machine*. P. 104.

66. Ibid. P. 101.

67. Ibid. P. 104–105.

68. Дрейфус Х. Чего не могут вычислительные машины. С. 223.

69. Dreyfus H. L., Dreyfus S. E. *Mind Over Machine*. P. 106.

было так сложно найти скрытые правила — эксперты, по Дрейфусу, не пользовались эвристиками, их знание невозможно было формализовать.

Фейгенбаум и его последователи создали несколько экспертных систем: *MACSYMA* для арифметических и алгебраических задач; *CONGEN*, которая генерировала модели молекулярных структур по определенным данным; *R1* для сборки компьютеров системы *VAX*; геологическую программу *PROSPECTOR*; диагностические программы *INTERNIST-1*, *MYCIN* и *PUFF*. Образцом «инженерии знаний» и прообразом всех экспертных систем была программа *DENDRAL*. И Дрейфус раскритиковал ее, желая показать, что сама идея экспертной системы, которая использует эвристики, не работает. *DENDRAL* была написана в 1960-е годы на языке *Lisp* при участии Фейгенбаума и программиста Брюса Бьюкенена; она должна была помочь химикам определять структуру органического вещества на основе данных спектроскопии. Дрейфус связался с Бьюкененом и узнал, что программа работала с ограниченным набором соединений и, действительно, распознавала их состав успешнее, чем профессиональные химики⁷⁰. Но загвоздка была в том, что сама область спектроскопии была выбрана для экспертных систем именно потому, что в ней человеческое умение распознавать сложные паттерны было не очень важно — гораздо важнее были расчеты и системы выводов. Дрейфус писал:

Поскольку развитый навык в спектроскопическом анализе требует расчетов, успех программы *Heuristic Dendral*⁷¹ не опровергает нашу гипотезу⁷².

Для него расчеты и логические выводы, работавшие с дискретными операциями, не отображали особенностей человеческой экспертизы и поэтому не могли рассматриваться как ее формализация и/или когнитивная симуляция. Кроме того, Дрейфус выяснил, что *heuristic DENDRAL* устарела и не используется — программу просто никто не обновлял. Конечно, благодаря *heuristic DENDRAL* возникла популярная система *CONGEN*, но она при работе уже

70. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 111.*

71. *DENDRAL* — название не одной программы, а целого проекта, в рамках которого были разработаны несколько программ. *Heuristic DENDRAL* была одной из таких программ, которая действительно опиралась на эвристики при анализе данных.

72. *Ibidem.*

не опиралась на эвристики и в основном занималась сложными вычислениями. Получается, что гипотеза скрытых эвристик не получила поддержки специалистов. По Дрейфусу, успех идей Фейгенбаума был связан не с тем, что ему удалось ухватить профессиональный опыт экспертов, а с ростом вычислительных мощностей, который не имели отношения к загадке человеческого мастерства.

Пять этапов при развитии навыков

В 1970-е годы у Дрейфуса уже был набросок феноменологии научения. Но, критикуя Саймона, Ньюэлла и Фейгенбаума, он столкнулся с двумя проблемами: 1) как объяснить частичный успех их программ; 2) как показать, что, хотя люди могут работать с ментальными репрезентациями, например с правилами, их деятельность гораздо больше зависит от телесного взаимодействия с объектами в разнообразных ситуациях. Для этого нужно было найти связь между правилами и ситуациями в человеческом поведении.

Уже после публикации первой книги к Хьюберту Дрейфусу обратился психолог Джон Торп, работавший на военно-воздушные силы США⁷³. В это время мастерство пилотов было принято анализировать в терминах правил, так что пилоты проходили разные тесты, доказывая, что они хорошо помнят самые сложные инструкции. Торп же считал, что экспертное владение навыком нельзя свести к выполнению правил, и искал поддержку в академической среде. Дрейфус согласился помочь и привлек к работе своего брата Стюарта. В 1980 году они опубликовали доклад для ВВС США, в котором предложили знаменитую модель пяти этапов⁷⁴. Окончательно модель оформилась к 1986 году⁷⁵, но Хьюберт и дальше будет менять некоторые ее положения.

В платонической традиции, к которой Дрейфус⁷⁶ причислил Саймона, Ньюэлла и Фейгенбаума, новичок начинает с конкрет-

73. Conversations With History: Hubert Dreyfus // Conversations With History, Institute of International Studies, UC Berkeley. 01.11.2005. URL: https://conversations.berkeley.edu/dreyfus_2005.

74. Dreyfus H. L., Dreyfus S. E. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Berkeley: University of California (Berkeley), Operations Research Center, 1980.

75. *Idem*. Mind Over Machine.

76. Хьюберт всегда подчеркивал, что феноменологию научения он придумал вместе с братом Стюартом. Тем не менее в книге 1986 года, которую они

ных примеров и постепенно добирается до самых абстрактных правил. Дрейфус переворачивает схему: сначала у обучающегося есть абстрактные правила, потом он переходит ко все более и более специфичным ситуациям. Дрейфус разработал следующую последовательность:

- Новичок (*novice*). Он ничего не знает, не понимает контекста, не ухватывает свойства среды. Чтобы отделить важное от неважного, он опирается на правила, которые сами не всегда работают. Новичок похож на компьютер, у него аналитическая установка.
- Продвинутый новичок (*advanced beginner*) уже видит дополнительные аспекты ситуации. Совершенствуясь, он начинает распознавать новые аспекты, зависящие от контекста. Постепенно у него накапливается опыт. Продвинутый новичок все еще похож на компьютер, у него аналитическая установка, он не вовлечен эмоционально.
- Компетентный (*competent*) обучающийся воспринимает все больше деталей. Он не может описать все вариации, правила не исчерпывают все ситуации. Его ошеломляет множество аспектов, как внеконтекстуальных, так и зависящих от ситуации. Ему нужна перспектива, ограничивающая и облегчающая принятие решений. Кроме того, он эмоционально вовлекается в процесс. Дрейфус отмечает, что эмоциональная вовлеченность — это не только желание победить. Это ответственность за действия, это постоянное переосмысление успешных и неуспешных приемов⁷⁷. В процессе решения профессионал может откатываться к предыдущим стадиям, но эмоциональная вовлеченность это усложняет.
- Опытный (*proficiency*). Опытному практику не приходится выбирать, как реагировать. Реакции автоматизируются. Он отлично видит цели и аспекты ситуации, но не всегда чувствует правильное решение. Ему доступно множество решений, но у него еще мало опыта, недостаточно перспектив, чтобы быстро и правильно принимать решения.
- Эксперт (*expertise*). У него большой репертуар перспектив, много опыта и очень подробные различия. Ему

подготовили в соавторстве и при поддержке писателя Тома Атанасиу, они решили ради удобства писать от лица Хьюберта. См.: Ibid. P. 1. Для удобства изложения мы последуем их примеру.

77. Возможно, Дрейфус имеет в виду руминации.

вообще не нужны ментальные репрезентации. Он действует автоматически. Изредка, когда он не уверен в перспективе, он может замедлить принятие решения и задуматься о возможных вариантах. Но даже тогда, уверен Дрейфус, это не холодный математический расчет, не калькулятивная рациональность, а критический анализ интуиций, отличающий «делиберативную», «обдумывающую» рациональность⁷⁸.

Эта схема решала сразу обе проблемы. Программы Саймона, Ньюэлла и Фейгенбаума потому добивались успеха, что могли построить свои действия на правилах — как новички, продвинутые новички и даже компетентные обучающиеся. То есть люди не просто работают с ментальными репрезентациями, «овнутряя» правила, — без этих внеконтекстуальных правил человек не может начать развиваться в той или иной практике. Таким образом, теории ИИ повлияли на Дрейфуса настолько, что он поместил правилосообразное поведение в основу своей пятиэтапной модели.

Но на последних двух уровнях люди действуют совершенно иначе, чем программы. Они не подбирают правила и не пытаются вывести из них новые действия. Правила остаются те же, что и раньше, — они не становятся лучше и их не становится больше, как считали многие специалисты ИИ. Скорее благодаря накопленному на предыдущих стадиях опыту становится возможным чисто интуитивное принятие решений, без напряженного целеполагания. Происходит переход от «логической переработки атомарных фактов к распознаванию [ситуаций] без отсылки к обособляемым элементам»⁷⁹. Возникают беглость, гибкость и уверенность, и их источник Дрейфус видит в интуиции. Он понимает интуицию не как нечто мистическое, а как человеческую способность без усилий видеть связь новой ситуации со старой. Эта способность, или «холистическое распознавание сходств», предполагает, что часто люди могут видеть паттерны, не раскладывая их на элементы⁸⁰. В свою очередь, интуиция зависит от согласованной работы внешнего и внутреннего горизонтов, то есть от телесной основы человеческого мышления.

Все этапы описывались через несколько повторяющихся примеров: вождение автомобиля, обучение авиационных пилотов, тренировки шахматистов. Но нельзя сказать, что авторы провели эмпири-

78. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 31–32.*

79. *Ibid. P. 66.*

80. *Ibid. P. 28.*

ческое исследование; скорее, они построили модель по результатам философского, феноменологического анализа.

В схеме Дрейфуса произошел постепенный переход от новичка с аналитическими установками к эмоционально вовлеченному профессионалу, опирающемуся на опыт и интуицию. Поскольку символический ИИ не обладает телом и не способен к интуитивному познанию, то уровни опытности и экспертизы оказывались закрыты для когнитивной симуляции.

Ядро теории научения

Концепция пяти этапов была важным расширением феноменологии научения, однако, как мы показали, Дрейфус описал базовый механизм образования навыка еще в 1970-е годы. У любого человека есть сознание и тело, благодаря которым мы можем детализировать объекты вокруг нас и адаптироваться к динамичным ситуациям. Как и когнитивисты, Дрейфус придавал огромное значение памяти, но при этом не описывал ее в терминах ментальных репрезентаций. В феноменологии научения память трактовалась как прошлый опыт; без всяких дополнительных шагов опыт меняет восприятие настоящей ситуации, и именно он задает длящуюся перспективу изменяющейся релевантности⁸¹. Прошлый опыт поддерживает способность различать аспекты объектов. Когда человек учится — с помощью инструкций, или учителя, или методом проб и ошибок, — его сознание стремится к максимальному постижению всех релевантных аспектов. Именно поэтому он может развиваться.

Дрейфус объяснял максимальное постижение, ссылаясь на изменчивые цели обучающегося и подвижные ситуации, в которых он оказывается. Это было плохое объяснение, поскольку оно не затрагивало телеологию навыка, то есть не объясняло, почему повторение действия приводит к его улучшению. Поэтому Дрейфус работал над ядром теории научения и в последующие годы.

В 2000-е годы он конкретизировал центральный механизм научения, снова привлекая Мерло-Понти и дополняя его идеи новейшими научными данными⁸². Как и раньше, обучающийся по-

81. *Dreyfus H. L., Dreyfus S. E. Mind Over Machine. P. 88–89.*

82. *Dreyfus H. L. A Phenomenology of Skill Acquisition as the Basis for a Merleau-Pontian Non-Representationalist Cognitive Science. Berkeley: University of California, 2002; Idem. Merleau-Ponty and Recent Cognitive Science // Skillful Coping: Essays on the Phenomenology of Everyday Perception and Action. Oxford, UK: Oxford University Press, 2014.*

лучает знание не в форме репрезентаций в сознании, но как все более и более подробное видение ситуаций. Если ситуация не вызывает успешного действия, обучающийся может и дальше прояснять свои различия ситуации. Чем больше оттенков ситуации он будет видеть, тем больше «очищенных» действий появится. Все действия связаны с телом, поэтому Дрейфус писал об особом, воплощенном агенте и интенциональности двигающегося тела. Мерло-Понти называл эту интенциональность «двигательной»⁸³, или моторной, а обратную связь между воплощенным агентом и воспринимаемым миром обозначал как *интенциональную дугу*⁸⁴.

Максимальное постижение, или захват реальности, оказывается встроенным в интенциональную дугу. Его телеологию Дрейфус описывает метафорически: научение стремится к равновесию или оптимальному гештальту. Он приводит в пример картину: у каждой картины есть оптимальное расстояние, с которого ее лучше всего созерцать; у микроскопа есть лучшая точка фокуса; и каждый навык обладает особым равновесием. Такое решение было проблематичным, поскольку равновесие или гештальт очень сильно напоминали цели обучающегося. Дрейфус уже ввел ментальные репрезентации в описание трех стадий совершенствования, и, если бы оказалось, что все направление совершенствования задается каким-то оптимумом, вся схема попала бы под удар символического ИИ и когнитивистики.

Феноменологии научения нужен был такой язык описания, чтобы он сохранял телеологию, но не вводил ментальные репрезентации. Как ни странно, Дрейфус обратился к идеям, которые были ближе к области ИИ, чем к феноменологии, — к теории нейронных сетей.

Сначала он экспериментировал с работами нейробиолога Уолтера Джексона Фримана III⁸⁵, представлявшего живой организм как сенсомоторную систему. В его теории сенсомоторная система является энергетическим ландшафтом, который структурируется возможностями успешно справиться с той или иной ситуацией, что, в свою очередь, создает новые нейронные связи в мозге. Когда к организму приходит новый чувственный сигнал, он приводит сенсомоторную систему в окрестность конкретного аттрактора. И организм начинает двигаться таким образом, чтобы состояние

83. Мерло-Понти М. Феноменология восприятия. С. 152.

84. Dreyfus H. L. Merleau-Ponty and Recent Cognitive Science. P. 234.

85. *Idem*. Phenomenology of Skill Acquisition as the Basis for a Merleau-Pontian Non-Representationalist Cognitive Science. P. 14–16.

его системы было на самом дне «бассейна притяжения» к аттрактору⁸⁶. Дрейфус описывает это стремление через игру в «холодно/горячо», но уточняет, что энергетический ландшафт дает больше информации. Когда сенсомоторная система чувствует притяжение какого-то конкретного аттрактора, ее можно сравнить с теннисистом: он просто чувствует, как нужно двигаться, чтобы лучше играть. Даже если он не знает, как именно выглядит идеальная позиция, он чувствует верное направление.

Дрейфус выложил исходную версию статьи только на сайте Калифорнийского университета Беркли. Потом, перед официальной публикацией, он внес в текст серьезные изменения. По неизвестной причине он дистанцировался от проекта Фримана: в тексте осталось всего одно упоминание о его аттракторах, все подробности были опущены⁸⁷. Теперь Дрейфус ссылался на работы по обучению сетей с подкреплением⁸⁸. В нейронных сетях одни связи между узлами могут «усиливаться», другие «ослабляться». Сила синаптической связи между нейронами, которые симулируются в такой сети, называется весом. В ходе обучения благодаря специальным алгоритмам веса перераспределяются так, чтобы сеть могла лучше справляться с той или иной задачей. И этот принцип позволяет избежать репрезентации прошлого опыта сети через «прецеденты или правила, определяющие дальнейшие действия»⁸⁹. Для Дрейфуса это означало, что такие модели способны распознавать паттерны и менять под них свое поведение без аналогов ментальных репрезентаций. Модели с подкреплением вдобавок могли без какой бы то ни было эксплицитной цели оценивать долгосрочные последствия своих действий⁹⁰. В этом смысле они отлично подходили для имитации обучения на поздних стадиях.

Дрейфус понимал, что Мерло-Понти не принял бы никакого редуccionистского объяснения для человеческого поведения⁹¹. Поэтому он оставил в своей феноменологии научения пространство, свободное от нейрофизиологических моделей. Он указал на серьезное ограничение нейронных сетей на пути когнитивной симуляции — их «бестелесность». Дрейфус снова обратился к идеям Тодса: у нейронных сетей нет верха и низа, нет предпочтитель-

86. *Dreyfus H. L. Merleau-Ponty and Recent Cognitive Science*. P. 15.

87. *Idem*. P. 247, footnote 29.

88. См.: *Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction, Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press, 1998.

89. *Dreyfus H. L. Merleau-Ponty and Recent Cognitive Science*. P. 236.

90. *Idem*. P. 242–244.

91. *Idem*. P. 246.

ного способа передвижения, нет ориентации вперед, эмоциональной реакции на успехи и провалы⁹². Такие сети не смогут так же распознавать паттерны, так же обобщать старые и новые ситуации, как это делает человек, а значит, они не смогут повторить их траектории научения.

Таким образом, с 1970-х годов по 2000-е изменялись представления Дрейфуса о базовом механизме научения. Критикуя ИИ, он постоянно отслеживал инновации Саймона, Ньюэлла, Фейгенбаума, многих других исследователей, и это заметно повлияло на его представления о навыках. Во-первых, он включил правилосообразное поведение в свою модель, тем самым признав его значение. Во-вторых, выступив против символического ИИ, Дрейфус поддержал альтернативные проекты, возникавшие в рамках нейросетевого моделирования, — подход Фримана и обучение с подкреплением. И хотя ему не удалось до конца примирить феноменологию Мерло-Понти и нейронные сети, он проложил дорогу для подобных исследований.

Роль Дрейфуса в истории искусственного интеллекта

Наша статья может показаться однобокой, так как описывает только феноменологию Дрейфуса и не реконструирует возражения со стороны сообщества ИИ. Дело в том, что таковых было очень мало, сообщество не восприняло его философию всерьез. В истории науки столкновения часто изображаются как яростные обмены критическими репликами, но это не тот случай. Диалога между сторонами не произошло. Да, Хьюберт Дрейфус общался с программистами Терри Виноградом и Джозефом Вейценбаумом, но к тому моменту они сами превратились в критиков ИИ и уже не имели такого влияния на большинство специалистов. Поэтому нельзя сказать, что Дрейфус вдохновил программистов и робототехников отказаться от символического ИИ Саймона и Ньюэлла в пользу нейронных сетей или «воплощенного ИИ» (*embodied AI*).

Маккордак приводит разные причины того, почему даже продуктивные с ее точки зрения идеи Дрейфуса не получили развития. Саймон и Ньюэлл решили не вступать в полемику и никак на него не ссылаться, чтобы его идеи не распространялись дальше. Сеймур Пейперт, организатор партии с *MacHack VI*, готовил критический отзыв на «Алхимию и искусственный интеллект», но так

92. *Dreyfus H. L. Merleau-Ponty and Recent Cognitive Science*. P. 238.

его и не закончил. Многие инженеры и программисты в принципе не принимали критику от того, кто не был специалистом⁹³ и занимался континентальной философией. Их позицию замечательно выразил Фейгенбаум: «И что он предложил нам взамен? Феноменологию! Эту чепуху! Эту пустышку!»⁹⁴

Как соавтор Фейгенбаума, Маккордак сама была предвзята. Даже проведя с Дрейфусом интервью, она не стала относиться к нему с пониманием: «Он так провоцировал своими насмешками, что отпугивал любого, кого мог бы просветить»⁹⁵.

Несмотря на безразличие сообщества ИИ, последнюю книгу на эту тему Дрейфус издал в 1992 году. В ней он провозгласил поражение символического ИИ благодаря трем направлениям: концепции Филипа Агре и Дэвида Чэпмена, «обучению через подкрепление» и нейронным сетям⁹⁶. Дрейфус заканчивает свой обзор там, где начинается история «нового ИИ» (*nouvelle AI*).

Он предчувствовал крах символического ИИ. И действительно, новые тренды в робототехнике⁹⁷ и нейросетевом моделировании⁹⁸ поставили гипотезу Саймона–Ньюэлла под вопрос. Но едва ли феноменология научения повлияла на сообщество ИИ напрямую. Отсутствие диалога не помешало Дрейфусу заявить о своей победе. Торжество наполняет фильм Тао Располи «Бытие в мире», построенный на основе интервью с Дрейфусом и его самыми известными учениками⁹⁹. Третья часть фильма посвящена дебатам вокруг ИИ; философы Шон Дорранс Келли, Марк Рэтхолл, Тейлор Кармэн и Джон Хогеланд¹⁰⁰ рассказывают историю Давида (Дрейфуса) и Голиафа (компьютерщиков из MIT), подчеркивая, что символический ИИ проиграл схватку с философией. Хогеланд придумал для устаревшей традиции специальный термин — «ста-

93. McCorduck P. *Machines Who Think*. P. 234.

94. Ibid. P. 230.

95. Ibid. P. 236.

96. Dreyfus H. L. *What Computers Still Can't Do*. P. xxx.

97. Brooks R. A. *Elephants Don't Play Chess*// *Robotics and Autonomous Systems*. 1990. Vol. 6. № 1–2. P. 3–15.

98. Sutton R. S., Barto A. G. *Reinforcement Learning*.

99. Ruspoli T. *Being in the World*, documentary, 2010.

100. Все последователи Дрейфуса так или иначе занимаются феноменологией и философией сознания. Келли работает в Гарварде, Рэтхолл и Карман — известные специалисты по Хайдеггеру и континентальной мысли — преподают в Оксфорде и Колумбийском университете соответственно. Хогеланд работал в Питтсбургском и Чикагском университетах; он внес большой вклад в философию сознания и критику ИИ, прежде чем умер в 2010 году.

рый добрый искусственный интеллект» (*Good-Old-Fashioned Artificial Intelligence, GOFAI*)¹⁰¹.

Таким образом, конфликт в 1960-е годы не привел к междисциплинарному общению. В результате возникло два параллельных взгляда на критические аргументы Дрейфуса. В одной версии он оказался критиком, слишком яростным и слишком поверхностным, чтобы серьезно повлиять на историю ИИ. В другой версии Дрейфус с помощью экзистенциальной феноменологии предсказал закат символического ИИ и победил в противостоянии с его адептами. Оба нарратива настолько предвзяты, что выбирать между ними не имеет смысла.

Что можно утверждать однозначно, так это успех пятиэтапной модели научения в социогуманитарных науках.

Вызовы для феноменологии научения

Популярность схемы Дрейфуса можно сравнить разве что с успехом пирамиды Маслоу. Глория Дэлл'Альба и Йорген Сандберг в обзоре существующих моделей научения называют его концепцию «одной из наиболее разработанных и влиятельных», причем влияние это распространилось на самые разные эмпирические исследования в менеджменте, преподавании, социальной работе и даже в программировании¹⁰². Отдельно нужно выделить рецепцию феноменологии научения в работах Патрисии Беннер, которая применила идеи Дрейфуса в области сестринского дела¹⁰³.

Сложно сказать, почему модель пяти этапов получила такое распространение. Возможно, главной причиной была ее интуитивная ясность: какие бы сложные аргументы ни лежали в ее основе, в конечном счете она утверждала — человек накапливает знания и развивается от уровня новичка до эксперта. Этот тезис соответствует нашему повседневному опыту, более того, он хорошо согласуется с кривыми обучения, которые психологи исследовали со времен Германа Эббингауза¹⁰⁴.

Возможно, сыграл роль талант Дрейфуса как популяризатора. Именно он открыл Хайдеггера американской академической пуб-

101. *Haugeland J.* Artificial Intelligence: The Very Idea. Cambridge, MA: MIT Press, 1985.

102. *Dall'Alba G., Sandberg J.* Unveiling Professional Development: A Critical Review of Stage Models // *Review of Educational Research*. 2006. Vol. 76. № 3. P. 386.

103. *Benner P. E.* From Novice to Expert: Excellence and Power in Clinical Nursing Practice. Upper Saddle River, NJ: Prentice Hall, 2001.

104. *Speelman C. P., Kirsner K.* Beyond the Learning Curve. P. 31.

лике, причем его интерпретация была настолько необычной, что Сёрл назвал ее «Драйдеггером»¹⁰⁵. Именно он стал одним из главных критиков ИИ в США. Дрейфус приложил много усилий, чтобы сделать свою модель известной, причем не только идею пяти этапов, но и феноменологию Мерло-Понти. Он был среди авторов, заложивших основы теории «воплощенного познания» (*embodied cognition*).

Феноменология научения Дрейфуса развивалась как критика символического ИИ, с самого начала она находилась в четкой оппозиции к работам Саймона, Ньюэлла и Фейгенбаума. Так что каждый раз, когда их идеи обсуждались, модель Дрейфуса получала свою долю внимания читающей публики. Поскольку символический ИИ был частью менталистского мейнстрима в когнитивных науках, все исследователи, выступавшие против всевластия ментальных репрезентаций, изучали работы Дрейфуса. Кроме того, Дрейфус очень удачно выбрал союзников за пределами философии, когда связал свою концепцию с нейросетевым моделированием. Его статьи об интенциональной дуге и обучении с подкреплением появились в 2000-е и 2010-е годы, когда нейронные сети переживали бум.

Благодаря своей известности феноменология научения стала альтернативой для когнитивных архитектур наподобие SOAR, выросших из идей символического ИИ. Однако из-за своей популярности концепция Дрейфуса оказалась открыта для критики с самых разных сторон. В последнее время феноменология научения сталкивается с вызовами, на которые ей только предстоит дать ответ.

Прежде всего, ее раскритиковали сторонники ментальных репрезентаций¹⁰⁶. Саймон и Ньюэлл не вступали в прямую полемику с Дрейфусом, чего нельзя сказать об их последователях в психологии и других дисциплинах. Их главное возражение касалось исследовательских методов: модель Дрейфуса не была эмпирически обоснована. Как возникла схема, какие процедуры использовали братья, когда готовили отчет для ВВС, как они тестировали свои гипотезы?

105. Woessner M. V. Heidegger in America. Cambridge; N.Y.: Cambridge University Press, 2011. P. 208.

106. Gobet F., Chassy P. Towards an Alternative to Benner's Theory of Expert Intuition in Nursing: A Discussion Paper // International Journal of Nursing Studies. 2008. Vol. 45. № 1. P. 129–139; Ungureanu C., Rotaru I. Philosophy of Skillful Coping. Motor Intentionality vs. Representations for Action // Procedia — Social and Behavioral Sciences. 2014. Vol. 163. P. 220–229.

Феноменологи критиковали Дрейфуса за жесткое противопоставление правилосообразных рефлексивных действий и ситуативных автоматических действий¹⁰⁷. Действительно, градация пяти этапов была слишком линейна и сводилась к постепенному замещению правил чувствительностью к ситуации. К тому же, если рассматривать схему Дрейфуса как результат феноменологического анализа, многим оставалось непонятно, в чем состоят особенности его метода. Он постоянно атаковал Гуссерля, и возникал вопрос: что именно он берет из его работ, а что отвергает?

Некоторые проблемы Дрейфус как будто специально оставил для своих учеников. Выделяя три сценария обучения — метод проб и ошибок, подражание и обучение через инструкции, — он подробно описал только последний вариант. Именно под него выстроена модель пяти этапов. Возможно, что для подражания и метода проб и ошибок модель нужно будет изменить.

Далее, Дрейфус больше всего концентрировался на обучении одного человека, за что его критиковал социолог науки Гарри Коллинз¹⁰⁸. Он избегал коллективных навыков (командные виды спорта, киносъемки, постановка пьесы, строительство, работа в лаборатории и т. п.); вероятно, считал, что обучение абстрактного индивида с помощью инструкций — это парадигма, из которой потом можно развить и анализ сложных социальных форм.

Отдельным вызовом стал вопрос об антропоцентризме. Дрейфус критиковал проекты ИИ не только и не столько за провалы — часто программы справлялись с поставленными задачами. Он громил их за то, что они не решали задачи как люди. Иными словами, феноменология научения рассматривала ИИ как масштабный проект когнитивной симуляции, а значит, ее аргументы не затрагивали такие концепции ИИ, которые отходят от имитации человеческого мышления. Этот антропоцентризм проникает в самое ядро теории научения. Каждый из примеров Дрейфуса, будь то вождение автомобиля, или теннис, или наблюдение живописи, включает манипуляции с неодушевленными объектами. И в то же время он почти не ставит вопрос о том, как эти нечеловеческие сущности влияют на человека в ходе его совершенствования или как человек меняет их. Конечно, в феномено-

107. *Hoffding S.* What Is Skilled Coping? Experts on Expertise // *Journal of Consciousness Studies*. 2014. Vol. 21. № 9–10. P. 49–73.

108. *Collins H.* Four Kinds of Knowledge, Two (or Maybe Three) Kinds of Embodiment, and the Question of Artificial Intelligence // *Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus*, Vol. 2. Cambridge, MA: MIT Press, 2000.

логии научения есть тезис о детализации¹⁰⁹: чем более развиты навыки человека, тем больше аспектов объектов он воспринимает. Но по текстам сложно понять, чем отличаются аспекты объектов от аспектов ситуаций, а поскольку ситуации играют у Дрейфуса важную роль, они как будто «заслоняют» объекты в концептуальном плане.

Самая большая проблема связана с тем, как философия Дрейфуса рассматривает структуру научения. Во-первых, его концепция не была приспособлена для описания креативности эксперта¹¹⁰. Как сказано выше, схема Дрейфуса начинается с рационального следования инструкциям у новичка и заканчивается автоматическим принятием решений у эксперта. Именно так объясняется скорость мастера — он не думает, ему не нужны ментальные репрезентации, его опыт сам собой трансформирует восприятие ситуации, он просто мгновенно реагирует. Получается, что такой эксперт в меньшей степени демонстрирует рациональность, он гораздо больше похож на гениальное животное. Конечно, в противовес такому толкованию Дрейфус ввел понятие делиберативной рациональности, когда эксперт замедляет беглость практики, чтобы оценить свои интуиции. Однако делиберативная рациональность сохраняет эксперта на высшем уровне, он просто выбирает из готовых перспектив. Такая рациональность совершенно не ухватывает способность эксперта полностью поменять правила игры и/или создать новую практику.

Во-вторых, хотя люди способны к максимальному захвату, к оптимуму практики, возникает вопрос: как они поддерживают его, если ситуации постоянно меняются? И в таком случае научение внутри практики — это достижение единственного оптимума или скорее переход от одного оптимума к другому?¹¹¹ Если в практике больше одного оптимума, то как оптимумы соотносятся друг с другом? Могут ли они, например, составлять иерархию?

Наконец, есть вызов, опасный как для концепции Дрейфуса, так и для любой современной теории научения. Он так и не объяснил, почему что-то в принципе позволяет людям совершенство-

109. Похожую идею также развивал Чарльз Гудвин. См.: *Goodwin Ch. Professional Vision // American Anthropologist*. 1994. Vol. 96. № 3. P. 606–633.

110. Возможно, такие сложности объясняются тем, что Дрейфус подробно описал только научение через следование инструкциям. Если бы в феноменологии научения был сценарий для метода проб и ошибок, креативность не была бы так ограничена.

111. *Dreyfus H. L. A Phenomenology of Skill Acquisition as the Basis for a Merleau-Pontian Non-Representationalist Cognitive Science*. P. 18.

ваться в практике. Связь сознания и ситуаций постоянно развивается, оптимумы уточняются. Но что толкает человека к уточнению оптимума? И с другой стороны, почему в какой-то момент люди достигают оптимума и перестают развиваться, то есть почему у мастерства есть пределы? Почему люди не совершенствуются постоянно? Возможно, сама идея оптимума блокирует такое развитие. Но тогда откуда берутся оптимумы, кто или что их задает? Человеческое сознание, внешняя реальность, структура практики — Дрейфус не дал точного ответа.

Таким образом, за несколько десятилетий Дрейфус выстроил альтернативный подход к проблеме научения. Если для сторонников символического ИИ оно было решением задач с помощью экспертных эвристик, то в феноменологии Дрейфуса научение стало процессом, в котором сознание и тело, обладая моторной интенциональностью, гибко решают проблемные ситуации и стремятся ко все большему захвату реальности. И хотя проект символического ИИ не реализовал амбиций своих отцов-основателей, едва ли можно сказать, что философия Дрейфуса разгадала все парадоксы совершенствования.

* * *

В 1997 году, через тридцать лет после партии Дрейфуса и *MacHack VI*, суперкомпьютер *Deep Blue* победил чемпиона мира Гарри Каспарова. Казалось бы, машина преодолела барьер «компетентного обучающегося» и достигла уровня эксперта, а значит, модель Дрейфуса была ошибочна. Он прокомментировал успех программы в небольшой сноске: «Эта программа, которая победила чемпиона мира по шахматам, не является экспертной системой, она не работает с правилами и прецедентами¹¹². Скорее, *Deep Blue* использовала метод полного перебора, просматривая миллиард ходов в секунду, чтобы видеть все комбинации на семь ходов вперед. Не считая грубой оценочной функции, которая выбирала, какой начальный шаг приведет к лучшему раскладу через семь ходов, *Deep Blue* не умнее калькулятора»¹¹³. Несмотря на сокручитель-

112. В оригинале — *cases*; Дрейфус имеет в виду термин *case-based learning*, который на русский принято переводить как «обучение на основе прецедентов» по аналогии с «рассуждениями на основе прецедентов» (см.: Варшавский П., Еремеев А. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений // Искусственный интеллект и принятие решений. 2009. Т. 6. № 2. С. 45–57).

113. *Dreyfus H. L. Merleau-Ponty and Recent Cognitive Science*. P. 239.

ные факты, Дрейфус продолжал отстаивать различие между человеческим мастерством и калькулирующей рациональностью.

Библиография

- Варшавский П., Еремеев А. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений // Искусственный интеллект и принятие решений. 2009. Т. 6. № 2. С. 45–57.
- Волков В. В., Хархордин О. В. Теория практик. СПб.: ЕУСПб, 2008.
- Дрейфус Х. Чего не могут вычислительные машины. М.: Прогресс, 1978.
- Мерло-Понти М. Феноменология восприятия. СПб.: Ювента; Наука, 1999.
- Прибрам К. Языки мозга. Экспериментальные парадоксы и принципы нейропсихологии. М.: Прогресс, 1975.
- Тоффлер Э. Шок будущего. М.: АСТ, 2002.
- Benner P. E. From Novice to Expert: Excellence and Power in Clinical Nursing Practice. Upper Saddle River, NJ: Prentice Hall, 2001.
- Brooks R. A. Elephants Don't Play Chess // Robotics and Autonomous Systems. 1990. Vol. 6. № 1–2. P. 3–15.
- Chase W. G., Simon H. A. Perception in Chess // Cognitive Psychology. 1973. Vol. 4. № 1. P. 55–81.
- Collins H. Four Kinds of Knowledge, Two (or Maybe Three) Kinds of Embodiment, and the Question of Artificial Intelligence // Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus, Vol. 2. Cambridge, MA: MIT Press, 2000. P. 179–195.
- Conversations With History: Hubert Dreyfus // Conversations With History, Institute of International Studies, UC Berkeley. 01.11.2005. URL: http://conversations.berkeley.edu/dreyfus_2005.
- Dall'Alba G., Sandberg J. Unveiling Professional Development: A Critical Review of Stage Models // Review of Educational Research. 2006. Vol. 76. № 3. P. 383–412.
- Dreyfus H. L. A Phenomenology of Skill Acquisition as the Basis for a Merleau-Pontian Non-Representationalist Cognitive Science. Berkeley: University of California, 2002.
- Dreyfus H. L. Alchemy and Artificial Intelligence. RAND papers, 1965. URL: <http://rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf>.
- Dreyfus H. L. Merleau-Ponty and Recent Cognitive Science // Skillful Coping: Essays on the Phenomenology of Everyday Perception and Action. Oxford, UK: Oxford University Press, 2014. P. 129–150.
- Dreyfus H. L. What Computers Still Can't Do: A Critique of Artificial Reason. Cambridge, MA: MIT Press, 1992.
- Dreyfus H. L., Dreyfus S. E. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Berkeley: University of California (Berkeley), Operations Research Center, 1980.
- Dreyfus H. L., Dreyfus S. E. Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. N.Y.: The Free Press, 1986.
- Feigenbaum E. A., McCorduck P. The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World. L.: Pan, 1984.
- Gardner H. The Mind's New Science: A History of the Cognitive Revolution. N.Y.: Basic Books, 1985.

- Gobet F., Chassy P. Towards an Alternative to Benner's Theory of Expert Intuition in Nursing: A Discussion Paper // *International Journal of Nursing Studies*. 2008. Vol. 45. № 1. P. 129–139.
- Goodwin Ch. Professional Vision // *American Anthropologist*. 1994. Vol. 96. № 3. P. 606–633.
- Hanley J. Hubert Dreyfus Interview // Full-Tilt Boogie. 20.10.2005. URL: <http://full-tilt.blogspot.com/2005/10/hubert-dreyfus-interview.html>.
- Haugeland J. *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press, 1985.
- Hoffding S. What Is Skilled Coping? Experts on Expertise // *Journal of Consciousness Studies*. 2014. Vol. 21. № 9–10. P. 49–73.
- Learning How to Learn: Powerful Mental Tools to Help You Master Tough Subjects // *Coursera.org*. URL: <https://ru.coursera.org/learn/learning-how-to-learn#instructors>.
- Lieto A., Bhatt M., Oltramari A., Vernon D. The Role of Cognitive Architectures in General Artificial Intelligence // *Cognitive Systems Research*. 2018. Vol. 48. P. 1–3.
- McCorduck P. *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*. Natick, MA: A.K. Peters, 2004.
- Newell A., Simon H. A. *Computer Science as Empirical Inquiry: Symbols and Search* // *Communications of the ACM*. 1976. Vol. 19. № 3. P. 113–126.
- Nichols P. *The Computer Chronicles — Artificial Intelligence (1984)* // *YouTube*. 09.11.2012. URL: http://youtu.be/_S3moV_ZF_Q.
- Ruspoli T. *Being in the World*, documentary, 2010.
- Speelman C. P., Kirsner K. *Beyond the Learning Curve: The Construction of Mind*. Oxford; N.Y.: Oxford University Press, 2005.
- Sutton R. S., Barto A. G. *Reinforcement Learning: An Introduction, Adaptive Computation and Machine Learning*. Cambridge, MA: MIT Press, 1998.
- The Chess Game “Hubert Dreyfus vs. Mac Hack VI” (1967) Annotated // *Ingram Braun*. 25.04.2017. URL: <http://ingram-braun.net/public/research/parlour-games/article/computer-chess-richard-greenblatt-match-mit-philosophy-artificial-intelligence-history>.
- Todes S. *Body and World*. Cambridge, MA: MIT Press, 2001.
- Ungureanu C., Rotaru I. Philosophy of Skillful Coping. Motor Intentionality vs. Representations for Action // *Procedia — Social and Behavioral Sciences*. 2014. Vol. 163. P. 220–229.
- Woessner M. V. *Heidegger in America*. Cambridge; N.Y.: Cambridge University Press, 2011.

PHENOMENOLOGY VS SYMBOLIC AI: HUBERT DREYFUS'S PHILOSOPHY OF SKILL ACQUISITION

SERGEY ASTAKHOV. Lecturer, School of Philosophy, Faculty of Humanities, sastakhov@hse.ru.

National Research University Higher School of Economics (HSE), 21/4 Staraya Basmannaya St., 105066 Moscow, Russia.

Keywords: artificial intelligence; phenomenology; skill acquisition; mental representations; motor intentionality.

A conflict between artificial intelligence (AI) researchers and phenomenologist Hubert Dreyfus arose in the 1960s and continued until the 2000s. The creators of the first AI programs believed that skill acquisition is a matter of solving problems by using particular mental representations, or heuristics. Dreyfus set out to prove that heuristics are not needed for skill acquisition because the human mind and body are capable of reacting to problematic situations in a flexible way without any mental representations. By clarifying the backstory of the conflict and analyzing the fundamental contradictions between the two theories of skill, the article shows how the phenomenology of skill acquisition originated from a critique of symbolic AI. Dreyfus developed his understanding of interconnections between mind and body in opposition to the associationism in the theories of Herbert Simon, Allen Newell and Edward Feigenbaum. He maintained that human beings have fringe consciousness, insight and tolerance of ambiguity and that they have a specific body structure and needs which make it possible to discriminate between relevant and irrelevant features in the environment and get a maximum grip of it.

The author analyzes how theories of learning created within symbolic AI influenced Dreyfus's five-stage model of skill acquisition. That model explained why programs by Simon and his colleagues achieved initial success, but it also exposed their limitations. To clarify the teleology of skill, Dreyfus explored how the idea of motor intentionality is connected with neural network modeling. Two perspectives on the role of Dreyfus in the history of AI are outlined together with the reasons why his philosophy had almost no effect on the AI community even though it was influential in the social sciences and humanities. Finally, current challenges facing the phenomenology of skill acquisition are explored.

DOI: 10.22394/0869-5377-2020-2-157-190

References

- Benner P. E. *From Novice to Expert: Excellence and Power in Clinical Nursing Practice*, Upper Saddle River, NJ, Prentice Hall, 2001.
- Brooks R. A. Elephants Don't Play Chess. *Robotics and Autonomous Systems*, 1990, vol. 6, no. 1–2, pp. 3–15.
- Chase W. G., Simon H. A. Perception in Chess. *Cognitive Psychology*, 1973, vol. 4, no. 1, pp. 55–81.
- Collins H. Four Kinds of Knowledge, Two (or Maybe Three) Kinds of Embodiment, and the Question of Artificial Intelligence. *Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert L. Dreyfus*, Vol. 2, Cambridge, MA, MIT Press, 2000, pp. 179–195.

- Conversations With History: Hubert Dreyfus. *Conversations With History, Institute of International Studies, UC Berkeley*, November 1, 2005. Available at: http://conversations.berkeley.edu/dreyfus_2005.
- Dall'Alba G., Sandberg J. Unveiling Professional Development: A Critical Review of Stage Models. *Review of Educational Research*, 2006, vol. 76, no. 3, pp. 383–412.
- Dreyfus H. L. *A Phenomenology of Skill Acquisition as the Basis for a Merleau-Pontian Non-Representationalist Cognitive Science*, Berkeley, University of California, 2002.
- Dreyfus H. L. Alchemy and Artificial Intelligence. *RAND papers*, 1965. Available at: <http://rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf>.
- Dreyfus H. L. *Chego ne mogu vychislitel'nye mashiny* [What Computers Can't Do], Moscow, Progress, 1978.
- Dreyfus H. L. Merleau-Ponty and Recent Cognitive Science. *Skillful Coping: Essays on the Phenomenology of Everyday Perception and Action*, Oxford, UK, Oxford University Press, 2014, pp. 129–150.
- Dreyfus H. L. *What Computers Still Can't Do: A Critique of Artificial Reason*, Cambridge, MA, MIT Press, 1992.
- Dreyfus H. L., Dreyfus S. E. *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition*, Berkeley, University of California (Berkeley), Operations Research Center, 1980.
- Dreyfus H. L., Dreyfus S. E. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, New York, The Free Press, 1986.
- Feigenbaum E. A., McCorduck P. *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*, London, Pan, 1984.
- Gardner H. *The Mind's New Science: A History of the Cognitive Revolution*, New York, Basic Books, 1985.
- Gobet F., Chassy P. Towards an Alternative to Benner's Theory of Expert Intuition in Nursing: A Discussion Paper. *International Journal of Nursing Studies*, 2008, vol. 45, no. 1, pp. 129–139.
- Goodwin Ch. Professional Vision. *American Anthropologist*, 1994, vol. 96, no. 3, pp. 606–633.
- Hanley J. Hubert Dreyfus Interview. *Full-Tilt Boogie*, October 20, 2005. Available at: <http://full-tilt.blogspot.com/2005/10/hubert-dreyfus-interview.html>.
- Haugeland J. *Artificial Intelligence: The Very Idea*, Cambridge, MA, MIT Press, 1985.
- Hoffding S. What Is Skilled Coping? Experts on Expertise. *Journal of Consciousness Studies*, 2014, vol. 21, no. 9–10, pp. 49–73.
- Learning How to Learn: Powerful Mental Tools to Help You Master Tough Subjects. *Coursera.org*. Available at: <https://ru.coursera.org/learn/learning-how-to-learn#instructors>.
- Lieto A., Bhatt M., Oltramari A., Vernon D. The Role of Cognitive Architectures in General Artificial Intelligence. *Cognitive Systems Research*, 2018, vol. 48, pp. 1–3.
- McCorduck P. *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*, Natick, MA, A. K. Peters, 2004.
- Merleau-Ponty M. *Fenomenologiya vospriyatii* [Phénoménologie de la perception], Saint Petersburg, Iuventa, Nauka, 1999.
- Newell A., Simon H. A. Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM*, 1976, vol. 19, no. 3, pp. 113–126.

- Nichols P. The Computer Chronicles — Artificial Intelligence (1984). *YouTube*, November 9, 2012. Available at: http://youtu.be/_S3moV_ZF_Q.
- Pribram K. *Iazyki mozga. Eksperimental'nye paradoksy i printsipy neiropsikhologii* [Languages of the Brain: Experimental Paradoxes and Principles in Neuropsychology], Moscow, Progress, 1975.
- Ruspoli T. Being in the World, documentary, 2010.
- Speelman C. P., Kirsner K. *Beyond the Learning Curve: The Construction of Mind*, Oxford, New York, Oxford University Press, 2005.
- Sutton R. S., Barto A. G. *Reinforcement Learning: An Introduction, Adaptive Computation and Machine Learning*, Cambridge, MA, MIT Press, 1998.
- The Chess Game “Hubert Dreyfus vs. Mac Hack VI” (1967) Annotated. *Ingram Braun*, April 25, 2017. Available at: <http://ingram-braun.net/public/research/parlour-games/article/computer-chess-richard-greenblatt-match-mit-philosophy-artificial-intelligence-history>.
- Todes S. *Body and World*, Cambridge, MA, MIT Press, 2001.
- Toffler A. *Shok budushchego* [Future Shock], Moscow, AST, 2002.
- Ungureanu C., Rotaru I. Philosophy of Skillful Coping. Motor Intentionality vs. Representations for Action. *Procedia — Social and Behavioral Sciences*, 2014, vol. 163, pp. 220–229.
- Varshavskii P., Ereemeev A. Modelirovanie rassuzhdenii na osnove pretsedentov v intellektual'nykh sistemakh podderzhki priniatiia reshenii [Modeling of Case-Based Reasoning in Intelligent Decision Support Systems]. *Iskustvennyi intellekt i priniatie reshenii* [Artificial intelligence and Decision Making], 2009, vol. 6, no. 2, pp. 45–57.
- Volkov V., Kharkhordin O. *Teoriia praktik* [Theory of Practices], Saint Petersburg, EUSPb, 2008.
- Woessner M. V. *Heidegger in America*, Cambridge, New York, Cambridge University Press, 2011.