

Прагматизм как методологическая рамка: обоснование онтологического статуса искусственного интеллекта

И. К. Ставровский

Институт философии Национальной академии наук Беларуси, г. Минск, Беларусь

tutoriks@gmail.com

Аннотация. Автор рассматривает причину затруднений в решении вопроса об онтологическом статусе искусственного интеллекта, которая обусловлена лакунами в постановке проблемы природы человеческого сознания. Предложено ограничить обсуждение трактовки человеческого сознания методологической рамкой прагматизма, доказана целесообразность такого ограничения. Автор опирается на современный физикализм, а также на информационный подход Д. И. Дубровского, принимая их за основу минимальной теории сознания, и полагает, что прагматистская рамка позволит исключить из дефиниции онтологического статуса искусственного интеллекта ряд спекулятивных суждений.

Ключевые слова: искусственный интеллект, онтологический статус, прагматизм, методология науки, философия сознания, физикализм, информационный подход

Для цитирования: Ставровский И. К. «Прагматизм как методологическая рамка: обоснование онтологического статуса искусственного интеллекта». *Экономические и социально-гуманитарные исследования* 12.2 (2025): 146–154.

<https://doi.org/10.24151/2409-1073-2025-12-2-146-154> EDN: SKXFEEQ.

Pragmatism as a methodological framework: Justifying the ontological status of artificial intelligence

I. K. Stavrovsky

Institute of Philosophy of the National Academy of Sciences of Belarus, Minsk, Belarus

tutoriks@gmail.com

Abstract. The author considers the reason for the difficulties in solving the question of the artificial intelligence ontological status, which is caused by lacunas in the problem statement of the human consciousness nature. It is suggested to limit the discussion of the interpretation of human consciousness by the methodological framework of pragmatism, the feasibility of such a restriction has been demonstrated. The author relies on modern physicalism, as well as on the information approach of D. I. Dubrovsky, taking them as the basis of a minimal theory of consciousness, and believes that the pragmatist framework will allow excluding a number of speculative judgments from the definition of the artificial intelligence ontological status.

Keywords: artificial intelligence, ontological status, pragmatism, scientific methodology, philosophy of consciousness, physicalism, information approach

For citation: Stavrovsky I. K. "Pragmatism as a Methodological Framework: Justifying the Ontological Status of Artificial Intelligence". *Ekonomicheskie i sotsial'no-gumanitarnye issledovaniya* = *Economic and Social Research* 12.2 (2025): 146–154. (In Russian). <https://doi.org/10.24151/2409-1073-2025-12-2-146-154>

Введение

Даже гипотетическая возможность создания мыслящей машины ставит перед исследователями вопрос об онтологическом статусе искусственного интеллекта (ИИ). Сама перспектива появления искусственного интеллекта заставляет переосмыслить такие понятия, как: «разум», «мышление», «сознание», «идея», «чувство» и т. п. Иными словами, всё то, что так или иначе принято относить к категории психического, ментального, когнитивного или, проще говоря, к продуктам деятельности мозга. Если мы можем всерьёз говорить о том, что машина тоже способна мыслить, то ментальное больше не может рассматриваться как исключительно человеческая способность. Более того, нам придется признать, что ментальное во

всех его проявлениях не является свойством только нервной системы биологических существ. Таким образом, в будущем нам придется создать новую психологию и философию сознания, которые будут учитывать нечеловеческую и даже не-биологическую ментальность. А для этого нам потребуется сформулировать минимальную теорию сознания. Постановка цели позволяет нам определить методологическую рамку исследования.

Методологическая рамка исследования

Проблему онтологического статуса ИИ представим как совокупность вопросов, связанных с тем, как следует относиться к тому или иному объекту. Другими словами, под «онтологическим статусом» будем понимать класс, к которому следует отнести тот или иной объект в рамках концептуальной

схемы, т. е. системы категорий, используемой для организации опыта (Дэвидсон, 1993). Имея одни и те же данные, мы можем представить их по-разному. Поэтому нас интересует, какая концептуальная схема предпочтительна при описании ИИ в контексте его сравнения с человеческим сознанием.

При решении проблемы мы будем исходить из прагматистской перспективы. Выбор именно такого подхода обусловлен тем, что вопрос об онтологическом статусе ИИ связан с вопросом об онтологическом статусе человеческого сознания. Данная проблема на данный момент не решена. Но если отказаться от попытки дать окончательный ответ на вопрос «чем является человеческое сознание на самом деле?», тогда вектор нашего интереса сместится на вопрос об успешности предлагаемой концептуальной схемы: «способна ли она описать многообразие фактов при минимуме допущений?». Таким образом, в любой ситуации будет предпочтительно использовать наиболее простую концептуальную схему, усложняя ее в случае необходимости и в той степени, в какой это требуется. Концептуальные схемы, требующие большего количества допущений, исключаются в силу того, что они усложнены более, чем это необходимо на данном этапе исследования. Вместе с тем они не «отбрасываются» раз и навсегда, в случае необходимости к ним всегда можно вернуться.

За основу нашей методологической рамки примем специфический вариант прагматизма, который предлагает У. В. О. Куайн в статье «Две догмы эмпиризма», утверждая, что выбор концептуальной схемы зависит в первую очередь от степени ее эффективности. С этой точки зрения физическая теория сущностно не отличается от мифов, но она оказывается более эффективной моделью для объяснения, систематизации и предсказания чувственного опыта (Куайн, 2010). Любое утверждение о том или ином объекте имеет лишь относительную значимость, так

как может быть пересмотрено при переработке концептуальной схемы. Более того, мы можем поменять концептуальную схему, если этого потребует теоретическая или практическая задача. Если применить данный контекст к проблеме онтологического статуса ИИ, возникает вопрос, насколько сложна должна быть наша концептуальная схема, чтобы описать его адекватно. Чтобы решить этот вопрос, обратимся к работе Д. Деннета «Насосы интуиции», где выделяются три типа установок, которые мы можем использовать для описания поведения объектов в качестве концептуальных схем (Деннет, 2021).

- **Физическая установка** — используется для описания поведения физических тел, которые при этом не являются механизмами. Обычно в рамках физической установки предполагается, что объект будет вести себя однозначно и предсказуемо. Благодаря этому можно точно предсказывать поведение объекта. Отклонение от подобного прогноза рассматривается как аномалия или следствие ошибки в расчетах.
- **Конструктивная установка** — применяется в основном для выявления закономерностей работы различных механизмов. Как и физическая установка, конструктивная позволяет делать достаточно точные прогнозы относительно того, как поведет себя объект. Чтобы предсказать поведение объекта, не обязательно знать, какие физические законы управляют его работой. Достаточно обнаружить закономерности в этой работе и на их основании предположить, как устроен объект. Однако конструктивная установка в большей степени подвержена риску ошибочных предсказаний, так как в основе концептуальной схемы лежит предположение о том, как объект сконструирован, и условие, что в его работе не будет сбоя. Чем сложнее сконструирован объект, тем выше вероятность ошибки.

— **Интенциональная установка** — реализует-ся как стратегия интерпретации поведения людей, животных и иных существей, которые можно рассматривать в качестве агента. Предсказания, сделанные в рамках интенциональной установки, имеют меньшую точность, чем предсказания, опосредованные закономерностями, выявленными в ходе применения конструктивной установки. Мы можем говорить лишь о более или менее вероятном поведении объекта. При этом мы предполагаем, что объект может повести себя неожиданным образом. Иными словами, мы предполагаем наличие свободы воли, т. е. способности выбирать альтернативу. Кроме того, интенциональная установка бесполезна при описании простых объектов. В лучшем случае она ничего не привносит в описание, в худшем — запутывает, создавая ложное впечатление, что поведение объекта сложное и непредсказуемое (усложняет то, что есть на самом деле).

Обоснуем, какие из перечисленных концептуальных схем наиболее соответствуют задаче описания искусственного интеллекта. Из вышеизложенного очевидно, что физическая установка недостаточна для адекватного описания поведения ИИ. Дело не в принципиальной невозможности описать работу машины, редуцировав каждый процесс к физическим закономерностям. Напротив, нет никаких причин сомневаться в том, что это возможно. Однако с прагматической точки зрения в большинстве случаев в подобном описании нет никакого смысла. При таком подходе для описания и прогнозирования каждого действия системы ИИ потребовались бы десятилетия, это усложняет задачу.

Вместе с тем конструктивистская установка значительно упрощает описание и предсказание поведения систем ИИ, если рассматривать сложную машину. Вполне целесообразно опираться на конструк-

тивистскую установку и в отношении простого ИИ, поскольку такой рамки вполне достаточно для описания его работы. Более того, когда необходимо создать или изменить технический объект, для описания работы системы или ее части должно прибегнуть именно к конструктивистской установке.

Интенциональная установка уместна для описания работы системы ИИ в процессах, где ИИ выступает как агент (например, игра в шахматы). При этом реконструировать программу ИИ и устройство невозможно или слишком тяжело. Сложность состоит и в том, что разные системы ИИ могут выполнять схожие или идентичные задачи. Однако если знать, какова «цель» системы, а также предположить, что она «стремится» достигнуть цели наиболее эффективным способом, то появляется возможность предсказывать ее поведение. Таким образом, мы будем говорить о машине как о мыслящей сущности, не считая ее таковой на самом деле.

Из сказанного ясно, что без интенциональной установки мы не можем полноценно описать даже искусственный интеллект, не претендующий на способность эмулировать человеческое сознание. Вопрос заключается в том, требуется ли нам интенциональная установка исключительно для упрощения описания работы ИИ в ряде случаев или же для описания работы ИИ в целом, так как мы принципиально не способны это сделать без ее использования. Кроме того, нас интересует, будет ли описание человеческого сознания иметь концептуальное отличие.

Следует оговориться, что понятия «интеллект», «мышление», «разум», «сознание» для нас являются взаимозаменяемыми. Обозначим две причины этому. Во-первых, если в научной литературе по психологии перечисленные термины разводятся, то в литературе по ИИ они используются примерно в одном и том же смысле: «когнитивные способности, аналогичные человеческим».

Во-вторых, до сих пор нет научного консенсуса по поводу природы человеческого сознания, а попытка дать однозначный ответ по данному вопросу требует отдельного исследования.

Тем не менее нам необходима минимальная теория сознания, т. е. такая концептуальная схема, которая описывала бы природу сознания, опираясь на минимальное число допущений. Необходимо сформулировать ряд посылок, которые позволили бы описать сознание так, чтобы сторонники различных теорий сознания могли принять нашу концептуальную схему за отправную точку.

В современной философии сознания преобладает физикализм, хотя и дуализм полностью не вытеснен (Деннет, 2017). Самым известным современным сторонником дуалистической теории сознания является Д. Чалмерс, с его натуралистическим дуализмом. Данная концепция сравнительно популярна и в какой-то степени влиятельна в рамках современной философии сознания. Мы не можем использовать ее в качестве отправной точки, на это есть две основные причины.

1. Несмотря на специфическую теорию сознания, взгляды Д. Чалмерса (Чалмерс, 2015) на искусственный интеллект практически не отличаются от взглядов таких философов, как Д. Деннет, П. С. Чечленд, Д. И. Дубровский и др. Поэтому рассмотрение проблемы онтологического статуса искусственного интеллекта через призму натуралистического дуализма только усложнит изложение, не добавляя ценного содержания.
2. Собственные взгляды Д. Чалмерса не имеют серьезного обоснования: утверждение, что квалиа существуют как фундаментальные элементы Вселенной, не только не имеет эмпирического подтверждения, но даже не может быть проверено методами современной науки. Иными словами, данная концепция

может быть рассмотрена исключительно как интересное предположение. Однако научные рассуждения, построенные на предположении, должны опираться на надежный теоретический фундамент.

Рассмотрим целесообразность обращения к концепции физикализма. По нашему мнению, ни один из аргументов Д. Чалмерса против физикализма не является достаточно сильным, чтобы отказаться от этой концепции. Максимум мы могли бы говорить о несовершенстве физикализма как концептуальной схемы. Однако в прагматистской перспективе это не является фундаментальной проблемой, так как перед нами не стоит задача найти теорию сознания, которая была бы истинной как таковая. Нас интересует наиболее проработанная на данный момент, непротиворечивая и простая концептуальная схема. Нет никакого смысла отказываться от нее в пользу теории, которая требует принятия большего количества допущений, не предлагая более эффективного описания фактов. Обращение к подобным теориям будет иметь смысл только при условии, если они смогут доказать свое превосходство над физикализмом.

По этой же причине нет смысла обращаться к квантовой теории сознания Р. Пенроуза (Пенроуз, 2015) или ее аналогам. Данная концепция спекулятивная, не имеет эмпирического подтверждения, т. е. является гипотезой, которая не получила научного признания.

Минимальная теория сознания

За основу минимальной теории сознания мы возьмем информационный подход Д. И. Дубровского. Такая теория достаточно проста и систематизирована, потому ее намного проще эксплицировать, а в случае необходимости оппонент сможет легко указать на пункты, с которыми он не согласен. В рамках информационного подхода информацией называется содержание сообщения

или сигнала. Опирается информационный подход на современные научные данные о биологической эволюции и процессах самоорганизации (Дубровский, 2015). Также в основе подхода лежат три базовые посылки:

- 1) информация необходимо воплощена в физическом носителе и вне его существовать не может;
- 2) одна и та же информация может кодироваться по-разному, может быть воплощена и передана с помощью разных по своим физическим свойствам носителей;
- 3) субъективная реальность может рассматриваться как информация, воплощенная в физическом носителе.

Первая посылка выглядит достаточно убедительно, так как мы не знаем примеров существования информации вне материальных носителей. То есть воплощенность в материальном носителе является свойством всей известной нам информации.

Вторая посылка подтверждается эмпирическим опытом, так как мы имеем примеры того, как одна и та же информация может быть записана по-разному. Например, стихотворение может храниться в виде текста на листе бумаги, в виде файла на компьютере, в человеке как воспоминание.

Третья посылка опирается на данные современной нейронауки, которая утверждает, что явления субъективной реальности связаны с процессами в мозговой нейродинамической системе так, что эти два феномена являются одновременными и однопричинными, т. е. находятся в отношениях взаимно-однозначного соответствия. Таким образом, субъективная реальность — кодовое воплощение процессов в мозгу. При этом если субъективная реальность дана нам непосредственно и не требует декодирования, то информация о носителе закрыта и требует расшифровки.

Важно отметить, что отношение между мозгом и субъективной реальностью является случайным, но историчным. Это значит, что нет причин считать мозг необходимо единственным возможным физическим носителем субъективной реальности, мы вправе допустить, что она может существовать и в иных формах. Однако нельзя не заметить, что достоверно известна единственная форма физического носителя субъективной реальности — это мозг, достаточно развитый. Существование субъективной реальности, воплощенной в иных носителях информации, пока не подтверждено. Для нас это значит, что наиболее правдоподобным индикатором наличия субъективной реальности у объекта будет наличие системы, функционально аналогичной человеческому мозгу. Это не является необходимым условием, но всё же дает основание для ожидания: наиболее вероятно, что субъективная реальность должна быть представлена на подобном физическом носителе.

Уже на этом этапе обсуждения очевидно, что наша минимальная теория сознания признает как минимум гипотетическую возможность появления машины как носителя субъективной реальности. Это не только соответствует доминирующим в науке и философии сознания представлениям, но также делает вопрос об онтологическом статусе искусственного интеллекта осмысленным. Если наша минимальная теория сознания исключала бы возможность создания машин с субъективной реальностью, то на данном этапе наше исследование было бы завершено.

На этапе исследования гипотетической возможности создания сильного ИИ требуется ввести важное ограничение, которое позволит в будущем избежать двусмысленности. Часто, когда делается то или иное заявление об искусственном интеллекте, оно оказывается общим; такое характерно для научно-популярных работ (подобные

заявления также нередки в СМИ). Ученые утверждают, что ИИ никогда не сравнится с человеком, в то время как другой ряд ученых уверены, что машины превзойдут своих создателей в самое ближайшее время. Столь общие заявления только актуализируют вопрос. Поэтому мы ограничим дальнейшее обсуждение жесткой рамкой: если машина способна выполнять конкретную задачу на определенном уровне, то мы не вправе предполагать, что она может выполнять эту задачу на более высоком уровне. Данное ограничение является исключительно методологическим и призвано избавить обсуждение от чрезмерного оптимизма в отношении технического прогресса.

Однако мы не отрицаем возможность дальнейшего развития технологии. В работе «Чего не могут вычислительные машины» Х. Дрейфус отмечает, что многие проблемы в области ИИ решаются с помощью приема *ad hoc*¹: машину учат решать только частные задачи, вместо того чтобы давать ей комплексный навык. Например, компьютер можно научить распознавать слово «кошка» при условии, что надпись сделана прописью, а изображение достаточно четкое. Такой «навык» сам по себе практически бесполезен и не кажется чем-то впечатляющим. Однако он подается как первый шаг на пути создания системы, которая безошибочно сможет распознавать любой текст при любом написании. И первое время прогресс в области распознавания новых слов будет быстрым, при обучении системы «вручную».

Проблема в том, что со временем подобный подход становится неэффективным, так как человек должен обучать систему практически бесконечному количеству возможных частных примеров (Дрейфус, 2010). Важно добавить, что даже сегодня задача распознавания текста компьютером все еще строится на *ad hoc* приемах, но гораздо боль-

шей сложности. Например, для разгадывания капчи, т. е. искаженного текста, который позволяет отличить человека от машины по способности разгадывания, потребовалось создать специальную и достаточно сложную систему искусственного интеллекта. При этом ряд исследований показывают, что это не единичный случай в области разработок ИИ. Современные разработчики зачастую даже не пытаются выйти за пределы подобного подхода, потому значительная часть проблем до сих пор решается по принципу выработки всё новых *ad hoc* приемов (Hutson, 2018). Однако разработчики ИИ заинтересованы в том, чтобы их достижения воспринимались как революционные или перспективные. Так создаются условия, в которых выгодно преувеличивать значимость разработок ИИ.

Заключение

Кратко сформулируем подход, по нашему мнению, оптимальный для любого дальнейшего исследования проблемы онтологического статуса ИИ. Мы предлагаем исходить из прагматистской перспективы, подбирая концептуальную схему, которая поможет описать факты, используя минимальное число допущений. В частности, нас будет интересовать, насколько необходима интенциональная установка для описания искусственного интеллекта и обнаружим ли мы концептуальное отличие при описании человеческого сознания. За основу минимальной теории сознания мы предлагаем принять информационный подход Д. И. Дубровского, рассматривающий сознание (субъективную реальность) как информацию, воплощенную в физическом носителе. Для того чтобы избежать необоснованно громких выводов, мы полагаем, что любая разработка в области ИИ является доказательством, что машина может выполнять конкретную задачу

¹ Специально для данного случая.

на определенном уровне. Любые суждения о перспективах технологии в лучшем случае спекулятивны. И хотя предложенный подход имеет методологическое ограничение, он поможет нам не покидать твердую почву фактов.

Список литературы и источников / References

- Деннет Д. К. *Насосы интуиции и другие инструменты мышления*. М.: АСТ; Corpus, 2021. 576 с.
Dennett Daniel C. *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Co., 2014. 512 p.
- Деннет Д. *Сладкие грезы: Чем философия мешает науке о сознании*. М.: URSS; Ленанд, 2017. 304 с.
Dennett Daniel C. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. London: Bradford Books; Cambridge, MA: MIT Press, 2006. 216 p. Jean Nicod Lectures.
- Дрейфус Х. *Чего не могут вычислительные машины: Критика искусственного разума*. Пер. с англ. Н. Родман. М.: ЛИБРОКОМ, 2010. 336 с.
Dreyfus Hubert L. *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper & Row; Toronto: Fitzhenry & Whiteside, 1972. xxxvi, 260 p.
- Дубровский Д. И. *Проблема «сознание и мозг»: теоретическое решение*. М.: Канон+, 2015. 207 с.
Dubrovskiy D. I. *Problem of "Consciousness and Brain": Theoretical Solution*. Moscow: Kanon+, 2015. 207 p. (In Russian).
- Дэвидсон Д. «Об идее концептуальной схемы». Грязнов А. Ф., сост. *Аналитическая философия: Избранные тексты*. М.: Изд-во МГУ, 1993. 144—159.
Davidson Donald. "On the Very Idea of a Conceptual Scheme". *Proceedings and Addresses of the American Philosophical Association* 47 (1973—1974): 5—20. <https://doi.org/10.2307/3129898>
- Куайн У. В. О. *С точки зрения логики. 9 логико-философских очерков* = *From a logical point of view: 9 logico-philosophical essays*. Пер. В. А. Ладова, В. А. Суровцева. М.: Канон+, 2010. 272 с. Б-ка аналитической философии.
Quine Willard Van Orman. *From a Logical Point of View: Nine Logico-Philosophical Essays*. 2nd ed., rev. Cambridge, MA: Harvard Up, 1980. 200 p.
- Пенроуз Р. *Новый ум короля: о компьютерах, мышлении и законах физики*. Пер. с англ. под общ. ред. В. О. Малышенко. 5-е изд. М.: URSS; Ленанд, 2015. 398 с. Синергетика: от прошлого к будущему.
Penrose Roger. *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*. Forew. by M. Gardner. Oxford: Oxford Up, 1999. 602 p.
- Чалмерс Д. *Сознающий ум: в поисках фундаментальной теории*. Пер. с англ. В. В. Васильева. 2-е изд. М.: URSS; Либроком, 2015. 509 с. Философия сознания.
Chalmers David J. *The Conscious Mind: In Search of a Fundamental Theory*. Rev. ed. Oxford: Oxford Up, 1997. 432 p. Philosophy of Mind.
- Hutson Matthew. "Artificial Intelligence Faces Reproducibility Crisis". *Science* 359.6377 (2018): 725—726. <https://doi.org/10.1126/science.359.6377.725>

Информация об авторе

Ставровский Игорь Константинович — младший научный сотрудник Института философии Национальной академии наук Беларуси (Республика Беларусь, 220072, г. Минск, ул. Сурганова, д. 1, корп. 2), tutoriks@gmail.com, SPIN-код: 8095-8273.

Information about the author

Igor K. Stavrovsky — Junior Research Fellow, Institute of Philosophy of the National Academy of Sciences of Belarus (Republic of Belarus, 220072, Minsk, Surganova, 1/2), tutoriks@gmail.com, SPIN code: 8095-8273.

Статья поступила в редакцию 10.12.2024, одобрена после рецензирования 06.02.2025.

The article was submitted 10.12.2024, approved after reviewing 06.02.2025.