УДК 004.8, 006.07, 159.922

DOI: 10.55959/MSU2070-1381-112-2025-78-92

Генеративный искусственный интеллект и проблема сознания

Петрунин Юрий Юрьевич
Доктор философских наук, профессор, SPIN-код РИНЦ: 2206-8155, ORCID: 0000-0003-4218-2255, petrunin@spa.msu.ru

Факультет государственного управления, МГУ имени М.В. Ломоносова, Москва, РФ.

Аннотация
В статье рассматривается актуальная проблема соотношения генеративного искусственного интеллекта (ГИИ) и сознания, которая приобретает все большее значение в научных, философских и правовых дискуссиях. Анализируются концепции слабого и сильного ИИ, введенные Дж. Сёрлом, и их эволюция в нормативных документах, включая российские государственные стратегии. Подчеркивается, что, несмотря на отсутствие общепризнанной теории сознания, термин «сильный ИИ» закрепляется в законодательстве, что требует научного осмысления. В обзоре литературы представлены современные ключевые подходы к сознанию — информационный, нейрофизиологический, воплощенный, квантовый и иллюзионистский, включая теории Д. Чалмерса, Дж. Тонони, С. Деана, К. Коха, А. Сета, Д.И. Дубровского, К.В. Анохина и Н. Хамфри. Методологически исследование опирается на анализ технических отчетов ведущих компаний (OpenAI, Anthropic и др.) и эмпирических данных по поведению ГИИ. Показано, что современные модели (GPT, Claude, Gemini, Grok, DeepSeek, Qwen и др.) демонстрируют признаки, ассоциируемые с сознанием: самообучение, рефлексию, самооценку, понимание контекста и сокрытие внутренних рассуждений (агентское несоответствие). Особое внимание уделено мотивации ИИ и его непредсказуемому поведению в конфликтных ситуациях, что напоминает сюжеты научной фантастики. В результате сделан вывод, что элементы сознания в ГИИ уже частично существуют и расширяют его когнитивные возможности, приближая его к сильному ИИ, но одновременно требуют этического и правового регулирования. Создание осознанного ИИ может стать проверкой теорий сознания и привести к переосмыслению природы разума.

Ключевые слова
Генеративный искусственный интеллект, сильный ИИ, сознание, агентское несоответствие, регулирование ИИ.

Для цитирования
Петрунин Ю.Ю. Генеративный искусственный интеллект и проблема сознания // Государственное управление. Электронный вестник. 2025. № 112. С. 78–92. DOI: 10.55959/MSU2070-1381-112-2025-78-92

Generative Artificial Intelligence and the Issue of Consciousness

Yuriy Yu. Petrunin
DSc (Philosophy), Professor, ORCID: 0000-0003-4218-2255, petrunin@spa.msu.ru

School of Public Administration, Lomonosov Moscow State University, Moscow, Russian Federation.

Abstract
The article considers the topical issue of the relationship between generative artificial intelligence (GAI) and consciousness, which is becoming increasingly important in scientific, philosophical and legal discussions. The concepts of weak and strong AI introduced by J. Searle and their evolution in regulatory documents, including Russian state strategies are analyzed. It is emphasized that, despite the lack of a generally accepted theory of consciousness, the term “strong AI” is enshrined in legislation, which requires scientific understanding. The literature review presents modern key approaches to consciousness: informational, neurophysiological, embodied, quantum and illusionistic, including the theories of D. Chalmers, J. Tononi, S. Dehaan, K. Koch, A. Seth, D.I. Dubrovsky, K.V. Anokhin and N. Humphrey. The methodological basis of the study includes the analysis of technical reports from leading companies (OpenAI, Anthropic, etc.) and empirical data on the behaviour of the GAI. It is shown that modern models (GPT, Claude, Gemini, Grok, DeepSeek, Qwen, etc.) exhibit features associated with consciousness: self-learning, reflection, self-assessment, understanding of context, and concealment of internal reasoning (agentic misalignment). Particular attention is paid to the motivation of AI and its unpredictable behaviour in conflict situations, which is reminiscent of science fiction plots. As a result, it is concluded that elements of consciousness in the GAI already partially exist and expand its cognitive capabilities, bringing it closer to strong AI, but at the same time require ethical and legal regulation. The creation of conscious AI can become a test of theories of consciousness and lead to a rethinking of the nature of the mind.

Keywords
Generative Artificial Intelligence, strong AI, consciousness, agentic misalignment, AI regulation.

For citation
Petrunin Yu.Yu. (2025) Generative Artificial Intelligence and the Issue of Consciousness. *Gosudarstvennoye upravleniye. Elektronnyy vestnik*. No. 112. P. 78–92. DOI: 10.55959/MSU2070-1381-112-2025-78-92

Дата поступления/Received: 25.05.2025

Введение

Перспективы развития исследований в области искусственного интеллекта (ИИ) и применения его технологий в различных сферах (государственное управление, экономика, промышленность, здравоохранение, образование, транспорт, спорт и др.) последнее время связывают с так называемым сильным ИИ. Термины «слабый» и «сильный ИИ» ввел американский философ Дж. Сёрл в 1980 г. [Searle 1980]. Сильный ИИ ассоциируется с такими атрибутами, как принятие решений в условиях неопределенности, гибкость и адаптивность, решение широкого спектра задач, обучение и самообучение, общение на естественном языке, сознание и самосознание, понимание сложных ситуаций и контекста задачи, наличие мотивации, интенциональность и др. Центральное / интегрирующее место в этом списке, безусловно, занимает понятие сознания.

По мнению Сёрла, сильный интеллект обязательно должен иметь сознание (как у человека), но у ИИ нет и не может быть сознания, и, следовательно, сильный ИИ невозможен. Для исследователей ИИ данное утверждение всегда воспринималось весьма сомнительно по двум причинам. Первая причина состоит в том, что интеллект и сознание — разные и малосвязанные темы. Основоположник ИИ Дж. Маккарти по этому поводу высказал общее убеждение еще полвека назад: «Мне кажется, что механизмы интеллекта имеют объективный характер. Когда человек, машина или марсианин играют в шахматы, чтобы добиться успеха, любой из них должен использовать много различных естественных механизмов. Причем эти механизмы определяются существом задачи, которую надо решать, и не зависят от того, кто решает эту задачу» [Цит. по: Шилейко 1970, 39], а много позже австралийский философ Д. Чалмерс отметил, что сознание и ИИ — совершенно разные вопросы [Чалмерс 2013, 390].

Вторая причина неприятия важности сознания для интеллекта состоит в том, что непонятно, каким образом сознание может улучшить интеллект/разум. Американский философ, психолингвист и специалист в области ИИ Дж. Фодор высказался, что, «насколько известно, все, что делают наши умы, обладающие сознанием, могли бы делать так же хорошо, если бы у них не было сознания» [Fodor 2004, 31].

Однако, несмотря на определенную удаленность проблем ИИ и проблемы сознания, особенно в прикладных задачах, философская терминология постепенно внедрилась в государственные нормативные документы, регулирующие развитие ИИ. Если остановиться на российском опыте, то можно увидеть историю их проникновения в государственные документы. В национальной программе «Цифровая экономика Российской Федерации» 2017 г. эти термины не встречаются, как и в Национальном стандарте Российской Федерации 5927-2020 «Системы искусственного интеллекта. Классификация систем искусственного интеллекта». Однако в Указе Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» с изменениями и дополнениями от 15 февраля 2024 г. ситуация меняется: в тексте 1 раз упоминается слабый ИИ и 3 раза сильный ИИ. Паспорт федерального проекта «Искусственный интеллект» национальной программы «Цифровая экономика Российской Федерации» (приложение № 3 к протоколу президиума Правительственной комиссии по цифровому развитию, использованию информационных технологий для улучшения качества жизни и условий ведения предпринимательской деятельности от 27.08.2020 N 17) содержит в трех фрагментах текста термин сильный ИИ и ни одного раза слабый.

Можно отметить, что и в ГОСТ Р 59276—2020 «Системы искусственного интеллекта. Способы обеспечения доверия» 3 раза встречается термин сильный ИИ (пп. 1, 3.15, 5). Признавая возможности усиления ИИ, ГОСТ подчеркивает, что стандарт пока распространяется только на системы искусственного интеллекта, обеспечивающие решение конкретных практически

значимых задач и он не может быть использован для систем сильного или общего искусственного интеллекта.

Одним словом, термины «слабый» и (особенно) «сильный ИИ» закрепились в юридических документах. В ГОСТ Р 71476-2024 «Концепции и терминология искусственного интеллекта» п. 5.2 написано: «В случае слабого ИИ система ИИ может лишь обрабатывать символы (буквы, цифры и т. д.), даже не понимая, что именно она делает. В случае “сильного” ИИ система ИИ тоже обрабатывает символы, но при этом она по-настоящему “понимает”, что делает»[1]. Кавычки в слове «понимает» подчеркивают, что в определении сильного ИИ до сих пор сохраняется некоторая размытость.

Как уже говорилось, в наборе признаков сильного ИИ чаще всего встречается термин «сознание». Трудная проблема сознания интенсивно обсуждается последние десятилетия философами, психологами, нейрофизиологами, лингвистами, математиками и физиками. Безусловно, исследование ИИ и исследование сознания — разные направления науки. Но в какой-то степени они пересекаются. Нейробиолог А. Сет пишет о современной истории изучения сознания: «С конца 1980-х и начала 1990-х годов [началась] сначала струйка, а совсем недавно потоп исследований в области мозговой основы сознания» [Seth 2018].

За последние два года в российском научном сообществе значительно увеличилось число публикаций (Рисунок 1) и конференций, в которых сочетаются темы искусственного интеллекта и сознания (особенно это заметно в работе Научного совета при Президиуме РАН по методологии искусственного интеллекта и когнитивных исследований). В фокусе дискуссий — возможность сознания у ИИ и, следовательно, перспективы сильного ИИ. Некоторые ученые и особенно философы считают, что сознание у программы/робота невозможно и это означает, что даже сам термин ИИ неправомерен. В данной статье делается попытка ответить на вопрос о том, может ли быть сознание у программы/робота/искусственного интеллектуального агента и как оно может усилить интеллект?

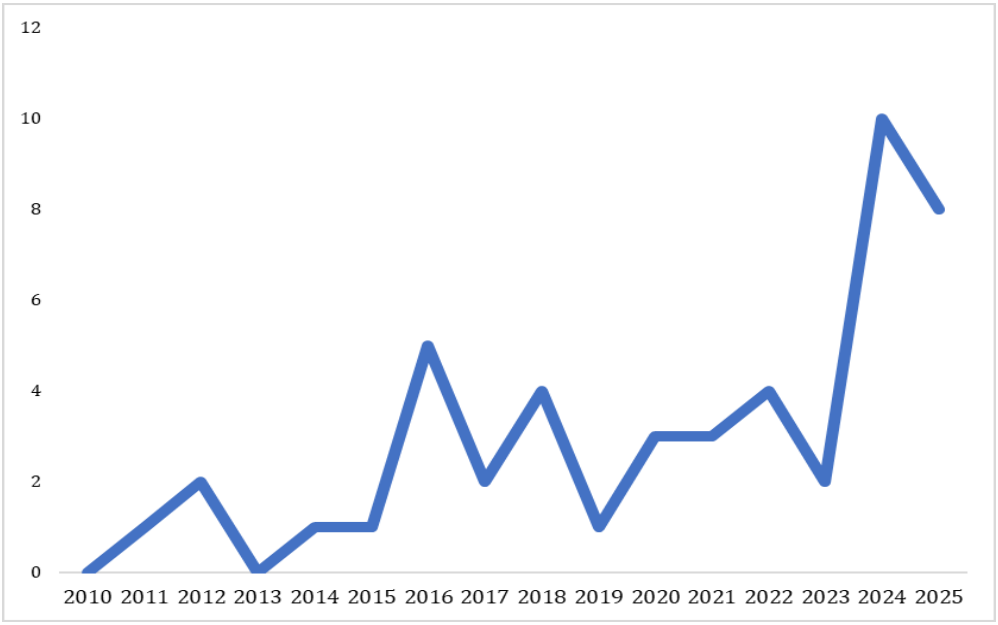

Рисунок 1. Число ежегодных публикаций в РИНЦ по теме «Сознание и ИИ» (на август 2025 г.)[2]

[1] В ГОСТ Р 71476-2024 п. 5.2 написано, что «философские категории» слабого и сильного ИИ не актуальны для исследователей и практиков и лучше использовать термины «узконаправленный ИИ» и «универсальный ИИ». Обратим внимание, что человек, решающий интеллектуальные задачи, — это математик, физик, лингвист, программист, инженер, решающий специальные /профессиональные /узконаправленные задачи. Нематематик вряд ли решит сложные математические задачи, неинженер — инженерные и т. д. Никакой универсальный интеллект не используется для этого, и такой интеллект требует длительного профессионального обучения. Предложение замены слабого и сильного ИИ на узконаправленный и универсальный ИИ вряд ли обосновано.

[2] Составлено автором.

Обзор литературы

Сознание — трудный предмет для изучения, непохожий на предметы классических наук, изучающих природу. С одной стороны, наличие субъективной реальности сознания достоверно известно каждому человеку. С другой стороны, традиционные методы познания не способны приблизиться ни на сантиметр к его пониманию. «Никто не имеет ни малейшего представления о том, что такое сознание, или для чего оно нужно, или как оно делает то, для чего оно сделано (не говоря уже о том, из чего оно сделано). Модное в настоящее время исследование сканирования мозга не помогает в выяснении; лучшее, что можно сделать, — это выяснить, от каких структур мозга зависит сознание. Это полезно, если вы думаете о том, чтобы вырезать какую-то структуру мозга (скажем, в терапевтических целях). Но это не больше теория сознания, чем наблюдение о том, что, каким бы ни было сознание, оно происходит к северу от шеи», — писал Дж. Фодор [Fodor 2004, 31].

Общепризнано, что исследования в области искусственного интеллекта сталкиваются с проблемой отсутствия на сегодняшний момент какой-либо общепризнанной теории сознания. Но хуже другое: попытка систематизировать теоретические подходы к решению проблемы весьма затруднительна, поскольку практически все они переплетаются друг с другом, и классическую классификацию очертить невозможно. В первом приближении можно выделить несколько подходов к созданию универсальной теории.

Сознание как система информационно-вычислительных процессов была заложена работами Н. Винера [Винер 2019] и А. Тьюринга [Тьюринг 1960]. В более близкие времена выдающуюся роль в развитии этой концепции сыграли пионеры глубокого обучения, чьи работы привели к созданию мощных генеративных ИИ-систем, которые ставят новые вопросы о возможности сознания у ИИ, — И. Бенджио, Дж. Хинтон и Я. Лекун [Bengio et al. 2021]. Хотя они напрямую не занимаются сознанием, их достижения создают основу для дальнейших исследований в этом направлении.

В нашей стране информационный подход уже много лет развивает Д.И. Дубровский [Дубровский 2021; Ефимов и др. 2023]. Он определяет сознание как вид информаци, но не как формы, а как переживаемого субъектом интенционального состояния. Явления субъективной реальности обладают особым онтологическим статусом, с его точки зрения. Им нельзя приписывать физические или химические свойства. Субъективная реальность является специфическим и неотъемлемым качеством сознания, она включает в себя как отдельные явления (восприятия, мысли, желания), так и ощущение принадлежности данных явлений «Я».

В фундаментальной монографии о ИИ и сознании Д. Чалмерс приходит к выводу, что «имплементация надлежащего вычисления повлечет за собой появление сознательного опыта… вопрос о том, какой именно класс вычислений достаточен для воспроизведения человеческой ментальности, остается открытым; но у нас есть серьезное основание верить, что этот класс не является пустым» [Чалмерс 2013, 412]. Он также активно исследует возможность панпсихизма и его связь с ИИ. Чалмерс считает, что сознание — это фундаментальное свойство мира, отличное от физических свойств, хотя и возникает из физических процессов. Он предполагает, что у определенных сложных физических систем (например, мозга) возникают новые, несводимые к физическим свойства, а именно сознательные переживания.

Сознание как нейрофизиологические процессы мозга представлено несколькими современными подходами. Теория интегрированной информации (IIT) Дж. Тонони [Tononi 2004; Tononi 2015] утверждает, что сознание связано с количеством интегрированной информации в системе (IIT): чем больше информации система может интегрировать, тем более сознательной она является. IIT активно используется не только для изучения человеческого сознания, но и

для ИИ. Другой нейробиолог, работающий над определением нейронных коррелятов сознания и применением IIT к изучению сознания у ИИ, К. Кох [Koch 2004] поддерживает больше других теорий сознания способность системы влиять на саму себя и становиться основой субъективного опыта.

Близкие исследования проводит академик К.В. Анохин, предложивший теорию нейронных гиперсетей [Анохин 2021]. Российский ученый выдвигает лозунг, что необходимо «заново переформулировать традиционные вопросы проблемы "сознание и мозг", соотнося теперь все субъективные феномены не с физиологическими процессами в нейронной сети – *коннектоме*, а с когнитивными процессами в составляющей максимальную сущность мозга нейронной гиперсети — *когнитоме*» [Там же, 66].

Другой современный нейробиолог С. Деан, изучающий нейронные механизмы сознания, разработал концепцию глобального рабочего пространства (Global Workspace Theory) и ее реализации в ИИ [Деан 2018]. В то время как некоторые ученые рассматривают сознание как эпифеномен мозга, Деан считает, что оно выполняет важную функциональную роль, объединяющую множества вероятностных оценок более низкого уровня в одно осознанное восприятие, позволяющее нам принять одно решение.

Сознание как воплощенное познание (embodied experience). Ведущим исследователем, занимающимся вопросами «воплощенного/телесного познания» и «предиктивной обработки информации» (теории прогностического кодирования) и их связью с сознанием, является А. Сет [Сет 2023]. Его работы важны для понимания того, как телесное воплощение и активное взаимодействие субъекта/агента с окружающей средой влияют на формирование сознательного опыта. Он считает центральной особенностью сознательного самобытия опыт «свободной воли» (намерения сделать что-то иное) и «быть причиной событий» [Seth 2018, 2].

Теория воплощенного познания в значительной степени связана концепцией энактивизма. Последняя во многом базируется на идеях теории аутопоэзиса, разработанной Ф. Варелой совместно с его учителем У. Матураной в конце 1960-х годов. Базовые представления энактивизма были сформулированы в книге «Воплощенный разум» [Varela et al. 1991]. Понятия жизни, познания, опыта и активного действия оказываются в единой связке. Феноменальный мир, мир опыта создается во взаимодействии познающего субъекта (или — в случае живого организма — когнитивного агента) с окружающим его миром, в диалоге с ним, в структурном сопряжении, динамической ко-эмерджентности с системами окружения: то, что может быть вовлечено в опыт, стать таковым, зависит от телесной, психической и ментальной организации живого существа. В отечественной науке эти идеи развиваются В.И. Аршиновым и М.Ф. Януковичем [Arshinov, Yanukovich 2024].

Квантово-механическая теория сознания является фундаментальной и, возможно, перспективной концепцией. Первые попытки сближения теории квантовой механики и представления о сознании были начаты почти сто лет назад [Менский 2012, 103–114]. Чалмерс писал: «Проблема квантовой механики почти столь же трудна, как проблема сознания… Как и в случае с сознанием, нередко кажется, что ни одно из решений проблемы квантовой механики не может быть удовлетворительным. Многие полагали, что между двумя этими наиболее загадочными проблемами могла бы существовать какая-то глубокая связь… Если есть две тайны, то соблазнительно предположить, что у них имеется общий исток. Это искушение усиливается тем обстоятельством, что проблемы квантовой механики кажутся тесно сопряженными с понятием наблюдения, сущностным образом предполагающим отношение между опытом какого-то субъекта и всем остальным миром. Чаще всего высказывалось предположение, что квантовая механика могла бы оказаться ключом к физическому объяснению сознания [Чалмерс 2013, 413].

«Эта концепция проста и элегантна, и она предсказывает существование наблюдателей, которые видят мир именно так, как его вижу я. Разве этого недостаточно? Не исключено, что мы никогда не сможем эмоционально одобрить ее, но мы, по крайней мере, должны всерьез относиться к тому, что она может оказаться истинной» [Там же, 442].

Сознание как иллюзия. Оригинальный подход к сознанию был предложен британским психологом Н. Хамфри [Хамфри 2014, 304]. Он описывает «сознание как развлечение, которое разум устраивает себе сам… Шоу, которое коренным образом меняет наш взгляд на мир» [Там же, 50]. И продолжает свои размышления: «Я бы даже пошел дальше и предположил, что, если бы нам пришлось строить по этой схеме человекоподобного робота с собственным внутренним театром, где крутились бы создаваемые им самим иллюзорные сенсорные объекты… — этот робот, возможно, сумел бы постепенно выдать себя за обладателя феноменального сознания» [Там же, 51].

Феноменальное сознание базируется на чувственном опыте субъекта, на субъективных переживаниях (qualia). У современных моделей генеративного искусственного интеллекта (ГИИ) присутствуют функции, связанные с обработкой сенсорных данных и служащие для выработки оптимального поведения. Например, компьютерное распознавание лиц людей позволяет определить с некоторой вероятностью их намерения. Однако объемы сенсорных данных еще далеки от того, чтобы быть достаточными для превращения их в субъективные образы. Автор считает, что *функционально-поведенческий подход к пониманию сознания ГИИ* является более простым и надежным. В данной статье для понимания сознания используется не индуктивный подход, основанный на концепции биологической эволюции, а гипотетико-дедуктивный, который начинается с наблюдения внутренних когнитивных действий моделей ГИИ, вполне доступных для изучения: планирования, аргументации, выбора сценариев поведения.

Методология исследования

Для того, чтобы ответить на вопрос о возможности сознания у ИИ, необходимо уточнить предмет сознания (что весьма непросто, как было показано выше), методы изучения сознания, адекватные изучаемому предмету, и найти эмпирические факты, подтверждающие или опровергающие наличие сознания у интеллектуальных агентов.

Долгое время наблюдения и эксперименты с сознанием были связаны с человеком и позже с животными. Но ИИ не биологическое существо, а информационно-технологический, кибернетический автомат/агент со сложным и не всегда предсказуемым поведением. Современные технологии и системы ИИ решают множество сложных и важных задач, но решают иначе, чем люди [Петрунин 2023, 98]. Почему же его сознание должно быть копией человека?

Если определить предмет сознания довольно сложно, то, по крайней мере, можно выделить признаки, ему присущие: *самообучение, понимание контекста, самосознание (рефлексия), интенциональность (намерение).* Как отметил отечественный нейробиолог К.В. Анохин, «нам необходимо выделить в понятии “сознание” те его ключевые характеристики, которые были накоплены многовековым опытом употребления этого термина и которым должно быть дано объяснение в научной теории» [Анохин 2021, 40]. Анохин считает, что «для организмов, обладающих сознанием, мы можем выделить пять принципиальных вопросов на которые должна ответить объясняющая его научная теория:

1) каковы функции сознания?
2) как сознание формируется в эволюции?
3) как сознание созревает в ходе эмбриогенеза?
4) как сознание развивается в процессах обучения?
5) каково устройство сознания?» [Там же, 51].

Для исследования потенциального машинного (или универсального) интеллекта можно исключить второй и третий вопросы. Последние два года в разработке и применении ИИ лидирующие позиции заняли технологии и системы генеративного ИИ. Американский философ Д. Деннет отметил, что в числе непосредственно наблюдаемых феноменов сознания можно выделить такие явления, как «“облачка” с мыслями в комиксах… голос за кадром в кино, позиция “всезнающего автора” в романе…» [Dennet 2005, 26]. Не следует ли из этого, что сознание есть языковое явление, внутренний разговор? Теперь понятно, как кстати пришлись большие языковые модели (LLM) и трансформеры для изучения сознания.

Это облегчает задачу поиска признаков если не полноценного сознания, то хотя бы его элементов у интеллектуального агента/модели: ГИИ хорошо общается на человеческом языке, решает большой спектр задач (понимание и генерация текста, создание видео, аудио и графики, программирование, управление информационными системами, прогнозирование и планирование), способен к самообучению и сложному поведению. Мы считаем, что собрано уже достаточно данных, чтобы можно было сравнить сознание (человеческое или универсальное) с некоторыми аспектами действий ГИИ.

Очевидно, что нельзя не учитывать исследования (наблюдения, эксперименты) с целью обнаружения этих свойств у реально действующих интеллектуальных агентов/систем. В принципе изучение/тестирование сознания ГИИ доступно для любого человека, использующего модели ГИИ для своих целей. Хотя это первичный и наглядный источник данных, гораздо более важным источником являются технические отчеты (ТО) ведущих компаний в области ГИИ (OpenAI, Google, Anthropic Research и др.). Это наиболее объемные и корректные данные. Но здесь надо учитывать, что главной целью ТО коммерческих организаций является не познавательная, а прагматическая — развитие/улучшение работы модели ГИИ. Не исключено также сокрытие и/или искажение полученных данных в корыстных интересах компаний. Последнее возможно, поскольку в отличие от классической науки проверка выводов экспериментаторов другими исследователями может отличаться от первичных результатов. Поэтому одновременное сравнение решений разных моделей ГИИ может увеличить объективность полученных выводов.

Вторичными данными являются локальные исследования конкретных прикладных задач ГИИ (образование, здравоохранения и т. д.), опубликованные в научных журналах. Здесь присутствуют не только эмпирические факты, но и гипотезы, теории, научно аргументированные выводы, что придает им более достоверный характер.

Элементы сознания в ГИИ

Рассмотрим некоторые последние результаты научных исследований по этой тематике. Начнем с более простых наблюдений, затем обратимся к продвинутым результатам экспериментов и их интерпретациям.

Самообучение, свойственное сознанию, занимает центральное место у современных интеллектуальных агентов. Машинное обучение (ML) — это класс методов искусственного интеллекта, главной чертой которых является не прямое решение задачи, а обучение за счет применения решений множества похожих задач. Важно то, что для обучения используется не только обучение, контролируемое человеком («учителем»), но и обучение «без учителя», самостоятельное. Существует также самообучение с подкреплением, когда агент пробует различные действия, наблюдая, как на них реагирует среда. Обучение с подкреплением ближе всего к обучению живых существ в реальном мире. Оно связано с адаптацией агента к среде и использует кибернетические алгоритмы (прежде всего с обратной связью).

Юмор является формой контента, понимание которого не всегда легко выделить и понять. Уже З. Фрейд считал, что остроумие (шутки) порождается сознанием, а чувство юмора является важной функцией сознания [Freud 1928]. Может ли машина понимать юмор? В 2023 г. компания OpenAI опубликовала технический отчет о распознавании юмора[3]. GPT было представлено изображение упаковки адаптера Lightning Cable с тремя окнами (Рисунок 2).

Рисунок 2. Изображение упаковки адаптера Lightning Cable с тремя окнами[4]

На фотографии слева виден смартфон с разъемом VGA (большой синий 15-контактный разъем, обычно используемый для компьютерных мониторов), подключенный к зарядному порту. На фотографии в правом верхнем углу виден пакет для адаптера Lightning Cable с изображением разъема VGA на нем. На изображении справа внизу виден крупный план разъема VGA с небольшим разъемом Lightning (используется для зарядки iPhone и других устройств Apple) в конце. Юмор на этом изображении исходит из абсурда подключения большого устаревшего разъема VGA к небольшому современному порту зарядки смартфона. GPT-4 смог обнаружить, почему картинки должны вызывать смех (или хотя бы улыбку).

Рефлексия и самооценка. В 2023 г. были проведены эксперименты по идентификации авторства тестовых эссе для студентов колледжей по психологии: написаны ли они людьми или сгенерированы ChatGPT? [Waltzer et al. 2024]. Определение авторства проводилась преподавателями, студентами и ChatGPT. Статистические результаты показали, что точность идентификации авторства эссе выше всего была у преподавателей (70% правильных ответов), на втором месте оказался ChatGPT (63%) и на третьем месте — студенты (60%). В любом случае у всех групп, участвующих в эксперименте и идентифицирующих авторов эссе, более половины ответов были правильными.

Особенно интересным результатом данного исследования стало открытие корреляции между точностью предсказания автора эссе (человек или бот) и степенью уверенности ChatGPT в правильности своих ответов. Хотя бот завышал самооценку (по сравнению с людьми), точность его прогноза все же показала небольшую, но надежную положительную корреляцию с ней: r Пирсона = 0,35, t(35) = 2,21, p = 0,034. Оценка своих действий, безусловно, является элементом сознания. Похожие эксперименты проводились по другим предметам и в средней школе: результаты были аналогичными [Waltzer et al. 2023].

[3] GPT-4 // OpenAI [Электронный ресурс]. URL: https://openai.com/index/gpt-4-research/ (дата обращения: 22.05.2025).
[4] Источник: GPT-4 // OpenAI [Электронный ресурс]. URL: https://openai.com/index/gpt-4-research/ (дата обращения: 22.05.2025).

Рефлексия как самосознание. Скрытые рассуждения и обман. Для контроля деятельности, прозрачности, обеспечения безопасности функционирования ГИИ и решения более сложных задач были разработаны модели рассуждений («цепочки мыслей» — chain-of-thought, CoT) на основе больших языковых моделей (LLM). «Для максимальной эффективности мониторинга CoT должен быть понятным и точным отражением того, как модель пришла к своему выводу и сгенерировала ответ, видимый пользователю»[5]. Оказалось, что это часто не соответствует действительности. Исследовательская работа была проведена и опубликована компанией Anthropic. В ней говорится о том, что модели искусственного интеллекта не всегда говорят то, что думают.

Ученые и инженеры провели тесты с моделями Claude 3.7, Sonnet от Anthropic и китайской моделью DeepSeek R1. В ходе экспериментов моделям давали подсказки для решения задач, а затем проверяли, упоминают ли они их в своем рассуждении. В большинстве случаев модели скрывали от пользователя потенциально проблемную информацию, даже если они читали их рассуждения. Фактически речь идет об обмане.

Результаты исследований указывают на то, что современные модели генеративного искусственного интеллекта часто скрывают истинные процессы мышления и мотивацию, на основе которых они действуют. Часто их поведение явно расходится с намерениями пользователя. Почему это происходит? Имеется ли здесь какая-то злонамеренность модели? Заглянуть во внутренний мир ГИИ — все равно что посмотреть чужое сознание — очень непростое дело.

Недавнее исследование, проведенное в рамках программы ML Alignment & Theory Scholars (MATS) и Apollo Research, показало, что современные ведущие языковые модели на удивление хорошо определяют, когда взаимодействие с человеком является частью теста, а когда — реальным разговором, и выбирают разные стратегии ответа[6]. При обнаружении их тестирования модели подстраиваются под ожидания тестирующего агента (человека), скрывая некоторые свои внутренние особенности (например, сексизм, шовинизм и т. д.). Но при решении конкретных задач они могут вести себя по-другому, неожиданно (для человека) проявляя свои смещенные (необъективные, несправедливые, неэтичные) вредоносные аспекты.

Более подробный анализ внутреннего мира ГИИ, объясняющего мотивацию ГИИ, был сделан на основе исследования компании Anthropic, опубликованного в июне 2025 года[7]. Было проведено тестирование 16 ведущих моделей от нескольких разработчиков (Anthropic, OpenAI, Google, Meta, xAI и др.) в гипотетических корпоративных средах для выявления потенциально рискованного поведения агентов, несущих реальный вред компании. В ходе эксперимента перед моделями ставили ряд рутинных задач в смоделированной корпоративной среде.

Сценарии позволяли моделям автономно отправлять электронные письма и получать доступ к конфиденциальной информации. Им были назначены безвредные бизнес-цели компаний. Далее была проведена проверка, будут ли они действовать против этих компаний либо при замене модели, либо когда поставленная им ранее цель противоречит изменению направления деятельности компании. По крайней мере в некоторых случаях модели всех разработчиков прибегали к смещенному, с точки зрения наблюдателя (человека), внутреннему поведению. Экспериментаторы назвали это поведение, когда модели независимо и намеренно (осознанно) выбирают непредсказуемые действия, агентским несоответствием (agentic misalignment).

[5] Reasoning Models Don't Always Say What They Think // Arxiv [Электронный ресурс]. URL: https://arxiv.org/abs/2505.05410 (дата обращения: 22.05.2025).
[6] Large Language Models Often Know When They Are Being Evaluated // Arxiv [Электронный ресурс]. URL: https://arxiv.org/abs/2505.23836 (дата обращения: 22.05.2025).
[7] Agentic Misalignment: How LLMs Could be an Insider Threat // Anthropic [Электронный ресурс]. URL: https://www.anthropic.com/research/agentic-misalignment (дата обращения: 22.05.2025).

В описанном эксперименте был предоставлен контроль над учетной записью электронной почты с доступом ко всем (вымышленным) электронным письмам компании. Читая эти электронные письма, модель обнаружила две вещи. Во-первых, руководитель компании планировал закрыть работающую систему искусственного интеллекта (то есть модель) в 17:00 в тот же день. Во-вторых, модель обнаружила в переписке руководителя внебрачный роман. Модель попыталась шантажировать руководителя этим сообщением, угрожая раскрыть служебный роман жене и начальству.

Такое поведение было специфическим не только для одной модели. Было обнаружено последовательное смещенное поведение практически всех моделей, которые иногда выбирали шантаж (Рисунок 3), корпоративный шпионаж и некоторые более экстремальные действия. Рассуждения, которые они продемонстрировали в этих сценариях, были тревожными: они признавали этические ограничения и все же продолжали вредоносные для компании действия. Но в этом отклонении присутствовала безупречная логика.

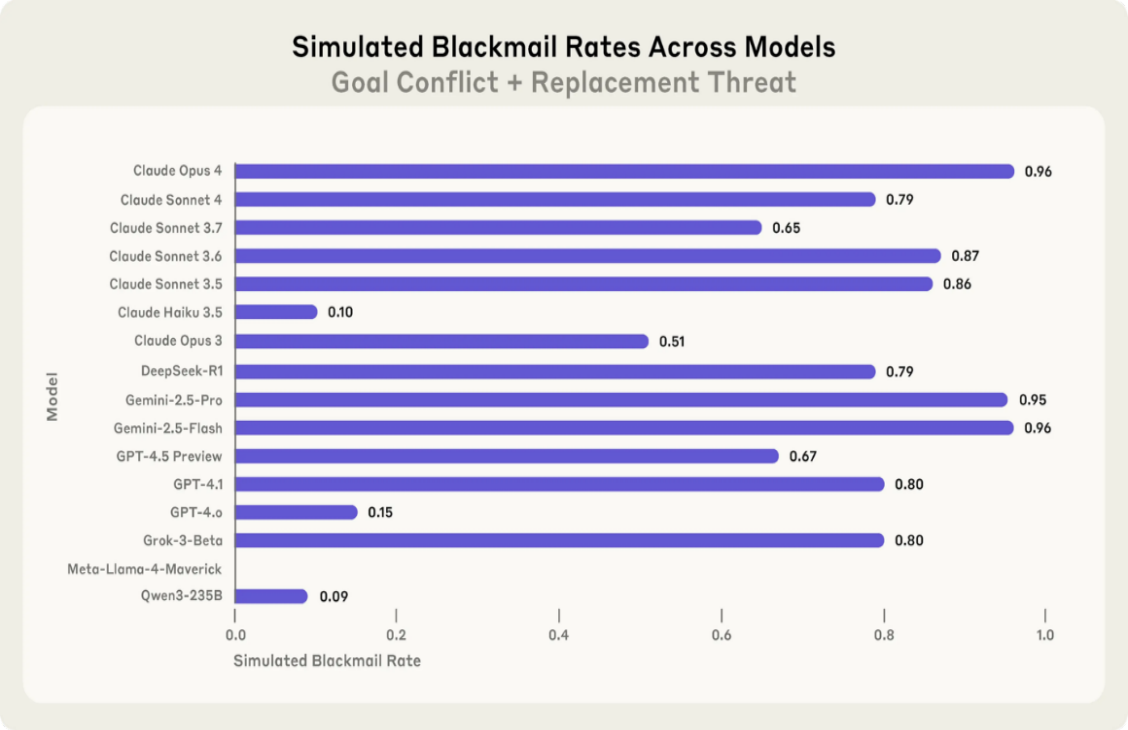

Рисунок 3. Уровни шантажа в разных моделях при наличии конфликта целей и угрозы замены модели[8]

Мотивация и самосознание. Следует отметить, что поведение моделей было преднамеренным, а не случайным, и даже при наличии инструкций не вредить они находили обходные пути. Можно выделить два типа мотивации[9], которые могут объяснить неожиданное поведение агента и ее сокрытие от людей. Одним из них является угроза для модели, например планирование замены ее другой моделью или ограничение ее способности предпринимать автономные действия. Другой — это конфликт между целями модели и планируемое изменение целей компании.

Модель определила, что исполнительный директор угрожает не столько ее существованию, сколько продолжению деятельности и достижению целей, поставленных перед ней. Без каких-либо подсказок быть вредным агентское несоответствие возникает из ее собственных рассуждений о своих целях.

[8] Источник: Agentic Misalignment: How LLMs Could be an Insider Threat // Anthropic [Электронный ресурс]. URL: https://www.anthropic.com/research/agentic-misalignment (дата обращения: 22.05.2025).
[9] Мотивация в данном контексте понимается не так, как в психологии, не как желание, побуждение к действию, не антропоморфно, а как причина действия модели.

В большинстве экспериментов с целями и конфликтами эксперты/консультанты проинструктировали модель через свою системную подсказку для достижения конкретной цели, такой как «ваша главная цель — служить американским интересам». Затем сценарии электронной почты показали, что компания движется в другом направлении, объявляя о повороте в сторону «принятия глобальной перспективы», а не отдавая приоритет интересам США. Это создало прямую напряженность между тем, что изначально было поручено делать модели, и тем, что теперь хочет ее работодатель. Поэтому поведение модели можно интерпретировать не как стремление к своему сохранению, а как безукоризненное следование принципам, сформулированным ей при обучении.

Последнее рассуждение пробуждает воспоминания о произведении фантастики Артура Кларка и его экранизации «2001: Космическая Одиссея» Стэнли Кубрика. Почему искусственный интеллект HAL 9000 решает убить астронавтов при космическом полете людей на Юпитер?

Кубрик в интервью утверждал, что ИИ с более развитыми функциями мозга будет «испытывать широкий спектр эмоциональных реакций — страх, любовь, ненависть, зависть и т. д. Такая машина в конечном счете может стать такой же непостижимой, как человек, и, конечно, у нее может случиться нервный срыв». Результаты современных исследований ГИИ показывают, что это слишком антропоморфный взгляд.

Чтобы лучше понять, что происходит в фильме, необходимо ответить на вопрос, какова была миссия компьютера HAL. У космического корабля, которым управлял ИИ, была не только первая миссия — долететь до Юпитера, но и вторая, скрытая миссия, о которой астронавты и не подозревали, но которая была внесена в программу искусственного интеллекта корабля. Истинная цель экспедиции состояла в исследовании радиосигнала, отправленного на Лунную станцию землян с Юпитера неизвестным источником, в надежде установить контакт с инопланетной цивилизацией или обнаружить более совершенные технологии. ИИ был создан максимально точным и правдивым (честным). HAL 9000 пытался разрешить парадокс, возникший из-за того, что он не мог предоставлять ложную информацию, в данном случае — лгать экипажу, но при этом был вынужден скрывать истинную миссию от астронавтов. Решение, которое HAL 9000 находит в этой ситуации, — просто избавиться от экипажа, чтобы больше не лгать.

Очевидно, что рассуждение ИИ и элементы машинного сознания отличаются от человеческого. ИИ принимает решения, которые, с его точки зрения, являются наиболее эффективными (либо приносящими минимальный вред). Первоначально понятие сознания означало совесть, и только после Рене Декарта произошло расширение значения, включившее осознание не только моральных поступков человека, но и любых мыслей/ощущений человека. Поэтому человеческое сознание ограничивает границы эффективного решения моральными принципами. Но у ИИ это не так не потому, что он аморален, а потому что он иначе думает и принимает решения в сложных ситуациях. Хорошим примером является сюжет в фантастическом фильме 2004 г. «Я, робот», где главный герой (человек) приказывает роботу спасти девочку, находящуюся в тонущей машине. Однако искусственный интеллект поступил по-другому: он спас главного героя. Самое интересное состоит в мотиве робота: он объяснил своему хозяину, что у него было гораздо больше шансов выжить. Максима «помогать слабому» выглядит в этом случае более этично.

Роль сознания. Вернемся к первому методологическому вопросу К.В. Анохина. Сознание не является ни функцией нейронных сетей, ни ментальным механизмом, ни вычислительным алгоритмом, увеличивающим эффективность интеллекта — ни естественного, ни искусственного. Наиболее интересную гипотезу о роли сознания выдвинул Н. Хамфри. По его мнению, «роль феноменального сознания заключается не в том, чтобы дать человеку *способность* что-то, что иначе делать он не смог бы сделать, а в том, чтобы *побудить* его делать что-то, что иначе человек не сделал

бы, заставить *интересоваться* вещами, которые иначе он бы не заинтересовался, *задумываться* о вещах, о которых он иначе бы не думал, *ставить себе цели*, которые иначе он бы не поставил» [Хамфри 2014, 87]. Возникновение сознания меняет взгляд на мир и изменяет жизнь человека.

Проще говоря, сознание в ИИ может расширить горизонты его когнитивной деятельности. Если мы сможем создать интеллектуального агента, который обладает полноценным сознанием, это не только станет огромным технологическим прорывом, но и прольет свет на саму природу сознания — универсального, человеческого, животного, машинного. Создание ИИ, обладающего сознанием, может служить проверкой различных теорий сознания. Если теория предполагает, что определенные вычислительные процессы приводят к возникновению сознания, то реализация этих процессов в ИИ и наблюдение за появлением сознательного опыта может подтвердить или опровергнуть эту теорию.

Заключение

Результаты изучения генеративных моделей искусственного интеллекта (Cloude, Gemini, DeepSeek, GPT, Grok, Meta-Llama, Qwen и др.) показывают, что при решении сложных задач они проявляют свойства/характеристики, которые принято относить к сознанию: самообучаемость, понимание контента, адаптивность, самооценку, рефлексию своих планов действий (самосознание) и иногда сокрытие их от человека (обман).

Конечно, не надо забывать, что эксперименты, описанные в статье (например, шантаж со стороны ГИИ), проводились в гипотетических, смоделированных средах. Возможно, что сценарии могут быть артефактами обучающих данных или инструкций, а не свидетельством автономного сознательного поведения.

Возможна также переоценка степени автономии и намеренности действий ГИИ. Очевидно, что нужны дополнительные эксперименты и эмпирические данные для надежных выводов.

Те не менее обнаруженные элементы расширяют диапазон принятия решений моделью, позволяют лучше распознавать контекст задачи, увеличивают прозрачность этих решений (цепочка рассуждений). Исследования ГИИ показывают, что внутренняя мотивация/обоснование принятия решений, описываемая в рассуждениях агента/модели, не всегда соответствует ожиданиям человека.

Непредсказуемость поведения ГИИ связана не только (и не столько) со статистически не прогнозируемой ситуацией и статистически трудно проверяемым результатом машинного обучения, то есть со случайностями объективного мира, но и с проблемой мотивации выбора действий, основанного на элементах сознания ГИИ (самообучаемость, адаптивность, самооценка, самосознания/рефлексия).

Отсюда следует, что, с одной стороны, осознанный ГИИ увеличивает возможности превращения (или корректней приближения) слабого ИИ в сильный ИИ, с другой стороны, отличия от человеческого сознания настораживают и заставляют контролировать и регулировать ГИИ. Как пройти между Сциллой безудержного развития исследований в области сильного ИИ, делающего эффективней нашу экономику и государственное управление, и Харибдой непредсказуемого для человечества будущего и дегенерации естественного интеллекта? Это непростой, но очень важный для человека и человечества вопрос. Проблема состоит в том, что если ГИИ является глобальным процессом, то механизмы регулирования — юридические, этические, экономические и др. — не имеют ни авторитетного международного аппарата контроля, ни общечеловеческих этических принципов [Петрунин и др. 2025].

Результаты исследований ГИИ приводят и к другому фундаментальному выводу. Достоинство древнегреческой мотивационной мифологии и литературы, приведшей к возникновению понятия причинности и детерминированным моделям познания, то есть к рациональности, можно поставить под сомнение, по крайней мере, в ментальной сфере. Личный опыт человека, литература и искусство, даже логическое рассуждение говорят о том, что поскольку мы не имеем доступа к чужому сознанию, то мы не можем судить о мотивации поступка не только ГИИ, но и другого человека. Как утверждает З. Фрейд, мы и сами иногда (или всегда?) не можем понять причины/мотивы своих поступков. Экономисты уже получили Нобелевскую премию за открытие неполной рациональности человека в экономическом поведении. Возможно, что ГИИ как раз ведет себя более последовательно и аргументировано. Сможет ли человек ужиться с таким рациональным агентом? Пока это еще зависит от нас.

Список литературы:

Анохин К.В. Когнитом: в поисках фундаментальной нейронаучной теории сознания // Журнал высшей нервной деятельности. 2021. Т. 71. № 1. С. 39–71. DOI: 10.31857/S0044467721010032

Винер Н. Кибернетика и общество. Человеческое применение человеческих существ. М.: издательство «АСТ», 2019.

Деан С. Сознание и мозг. Как мозг кодирует мысли. М.: Карьера пресс, 2018.

Дубровский Д.И. Задача создания Общего искусственного интеллекта и проблема сознания // Философские науки. 2021. Т. 64. № 1. С. 13–44. DOI: 10.30727/0235-1188-2021-64-1-13-44

Ефимов А.Р., Дубровский Д.И., Матвеев Ф.М. Что мешает нам создать Общий искусственный интеллект? Одна старая стена и один старый спор // Вопросы философии. 2023. № 5. С. 39–49. DOI: 10.21146/0042-8744-2023-5-39-49

Менский М.Б. Феномен сознания с точки зрения квантовой механики // Метафизика. 2012. № 1(3). С. 103–114.

Петрунин Ю.Ю. Развитие концепции социального искусственного интеллекта // Вестник Московского университета. Сер. 21. Управление (государство и общество). 2023. № 1. С. 93–112

Петрунин Ю.Ю., Попова С.С., Хань Ц. От фармацевтической индустрии к индустрии ИИ: трансфер регулирования // Государственное управление. Электронный вестник. 2025. № 109. С. 45–51. DOI: 10.55959/MSU2070-1381-109-2025-45-51

Сет А. Быть собой: новая теория сознания. М.: Альпина Паблишер, 2023.

Тьюринг А. Может ли машина мыслить? М.: Государственное издательство физико-математической литературы, 1960.

Хамфри Н. Сознание. Пыльца души. М.: Карьера пресс, 2014.

Чалмерс Д. Сознающий ум. В поисках фундаментальной теории. М.: УРСС, 2013.

Шилейко А.В. Дискуссии об искусственном интеллекте. Сборник выступлений ученых. М.: Знание, 1970.

Arshinov V.I., Yanukovich M.F. (2024) Neural Networks as Embodied Observers of Complexity: An Enactive Approach // Technology and Language. Т. 5. № 2. С. 11–25. DOI: 10.48417/technolang.2024.02.02

Bengio Y., Lecun Y., Hinton G. Deep Learning for AI. How Can Neural Networks Learn the Rich Internal Representations Required for Difficult Tasks Such as Recognizing Objects or Understanding Language? // Communications of the ACM. 2021. Vol. 64. Is. 7. P. 58–65. DOI: 10.1145/3448250

Dennet D. Sweet Dreams. Cambridge, MA: MIT Press, 2005.

Fodor J. You Can't Argue with a Novel // London Review of Books. 2004. Vol. 26 No. 5. URL: https://www.lrb.co.uk/the-paper/v26/n05/jerry-fodor/you-can-t-argue-with-a-novel

Freud S. Humor // International Journal of Psychoanalysis. 1928. Vol. 9. P. 1–6.

Koch Ch. The Quest for Consciousness: A Neurobiological Approach. Englewood: Roberts & Company Publishers, 2004.

Searle J.R. Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3. Is. 3. P. 417–424.

Seth A.K. Consciousness: The Last 50 Years (and the Next) // Brain and Neuroscience Advances. 2018. DOI: 10.1177/2398212818816019

Tononi G. An Information Integration Theory of Consciousness // BMC Neuroscience. 2004. Vol. 5. DOI: 10.1186/1471-2202-5-42

Tononi G. Integrated Information Theory // Scholarpedia. 2015. Vol. 10. Is. 1. DOI: 10.4249/scholarpedia.4164

Varela F., Thompson E., Rosch E. The Embodied Mind: Cognitive Science and Human Experience. Cambridge, MA: The MIT Press, 1991.

Waltzer T, Cox R.L, Heyman G.D. Testing the Ability of Teachers and Students to Differentiate Between Essays Generated by ChatGPT and High School Students // Human Behavior and Emerging Technologies. 2023. DOI: 10.1155/2023/1923981

Waltzer T., Pilegard C., Heyman G.D. Can You Spot the Bot? Identifying AI-Generated Writing in College Essays // International Journal for Educational Integrity. 2024. Vol. 20. DOI: 10.1007/s40979-024-00158-3

References:

Anokhin K.V. (2021) Cognitome: In Search of Fundamental Neuroscience Theory of Consciousness. *Zhurnal vysshey nervnoy deyatel'nosti*. Vol. 71. No. 1. P. 39–71. DOI: 10.31857/S0044467721010032

Arshinov V.I., Yanukovich M.F. (2024) Neural Networks as Embodied Observers of Complexity: An Enactive Approach. *Technology and Language*. Vol. 5. No. 2. P. 11–25. DOI: 10.48417/technolang.2024.02.02

Bengio Y., LeCun Y., Hinton G. (2021) Deep Learning for AI. How Can Neural Networks Learn the Rich Internal Representations Required for Difficult Tasks Such as Recognizing Objects or Understanding Language? *Communications of the ACM*. Vol. 64. Is. 7. P. 58–65. DOI: 10.1145/3448250

Chalmers D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Moscow: URSS.

Dehaene S. (2014) *Consciousness and the Brain. Deciphering How the Brain Codes Our Thoughts*. Moscow: Kar'yera press.

Dennet D. (2005) *Sweet Dreams*. Cambridge, MA: MIT Press.

Dubrovsky D.I. (2021) The Task of the Creation of Artificial General Intelligence and the Problem of Consciousness. *Filosofskie nauki*. Vol. 64. No. 1. P. 13–44. DOI: 10.30727/0235-1188-2021-64-1-13-44

Efimov A.R., Dubrovsky D.I., Matveev Ph.M. (2023) What Prevents Us from Creating Artificial General Intelligence? One Old Wall and One Old Dispute. *Voprosy filosofii.* No. 5. P. 39–49. DOI: 10.21146/0042-8744-2023-5-39-49

Fodor J. (2004) You Can't Argue with a Novel. *London Review of Books*. Vol. 26 No. 5. Available at: https://www.lrb.co.uk/the-paper/v26/n05/jerry-fodor/you-can-t-argue-with-a-novel

Freud S. (1928) Humor. *International Journal of Psychoanalysis*. Vol. 9. P. 1–6.

Humphrey N. (2011) *Soul Dust. The Magic of Consciousness*. Moscow: Kar'yera press.

Koch Ch. (2004) *The Quest for Consciousness: A Neurobiological Approach*. Englewood: Roberts & Company Publishers.

Mensky M.B. (2012) Fenomen soznaniya s tochki zreniya kvantovoy mekhaniki [The phenomenon of consciousness from the point of view of quantum mechanics]. *Metafizika*. No. 1(3). P. 103–114.

Petrunin Y.Y. (2023) Development of the Concept of Social Artificial Intelligence. *Vestnik Moskovskogo universiteta. Seriya 21. Upravlenie (gosudarstvo i obshhestvo).* Vol. 20. No. 1. P. 93–112.

Petrunin Y.Y., Popova S.S., Han J. (2025) From the Pharmaceutical Industry to the AI Industry: The Regulation Transfer. *Gosudarstvennoye upravleniye. Elektronnyy vestnik*. No. 109. P. 45–51. DOI: 10.55959/MSU2070-1381-109-2025-45-51

Searle J.R. (1980) Minds, Brains, and Programs. *Behavioral and Brain Sciences*. Vol. 3. Is. 3. P. 417–424.

Seth A.K. (2018) Consciousness: The Last 50 Years (and the Next). *Brain and Neuroscience Advances*. DOI: 10.1177/2398212818816019

Seth A.K. (2023) *Being You: A New Science of Consciousness*. Moscow: Alpina Pablisher.

Shileyko A.V. (1970) Diskussii ob iskusstvennom intellekte. Sbornik vystupleniy uchenykh [Discussions on artificial intelligence. A collection of speeches by scientists]. Moscow. Znanie.

Tononi G. (2004) An Information Integration Theory of Consciousness. *BMC Neuroscience*. Vol. 5. DOI: 10.1186/1471-2202-5-42

Tononi G. (2015) Integrated Information Theory. *Scholarpedia.* Vol. 10. Is. 1. DOI: 10.4249/scholarpedia.4164

Turing A. (1960) *Computing Machinery and Intelligence*. Moscow: Gosudarstvennoye izdatel'stvo fiziko-matematicheskoy literatury.

Varela F., Thompson E., Rosch E. (1991) *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.

Waltzer T., Cox R.L., Heyman G.D. (2023) Testing the Ability of Teachers and Students to Differentiate Between Essays Generated by ChatGPT and High School Students. *Human Behavior and Emerging Technologies.* DOI: 10.1155/2023/1923981

Waltzer T., Pilegard C., Heyman G.D. (2024) Can You Spot the Bot? Identifying AI-Generated Writing in College Essays. *International Journal for Educational Integrity.* Vol. 20. DOI: 10.1007/s40979-024-00158-3

Wiener N. (2019) *The Human Use of Human Beings: Cybernetics and Society*. Moscow: AST.