

Фундаментальные исследования в сфере информационного общества**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РАКУРСЕ КОГНИТИВНОЙ АГЕНТНОСТИ****Хмелевская Светлана Анатольевна**

Доктор философских наук, профессор

Московский государственный университет им. М.В.Ломоносова, философский факультет, кафедра философии естественных факультетов, профессор

Москва, Российская Федерация

xmelevsk@mail.ru

Ермаков Дмитрий Николаевич

Доктор политических наук, доктор экономических наук, профессор

Московский государственный университет им. М.В. Ломоносова, исторический факультет, кафедра истории общественных движений и политических партий, профессор, Заслуженный работник Высшей школы РФ

Москва, Российская Федерация

dermakow@mail.ru

Аннотация

Развитие искусственного интеллекта (ИИ) поднимает проблему надления его когнитивной агентностью, которая в настоящее время понимается по-разному. В связи с этим в статье ставятся задачи раскрыть не только, что такое когнитивная агентность, но и показать специфику его проявления у ИИ. В статье обосновывается идея, что когнитивная агентность еще не дает основания для рассмотрения ИИ как субъекта познания. Вместе с тем допускается возможность такого рассмотрения в дальнейшем по мере формирования AGI (Artificial general intelligence).

Ключевые слова

искусственный интеллект; когнитивная агентность; познание; информация; субъект познания

Введение

Тема когнитивной агентности стала одной из тех, которая в последнее время активно обсуждается в научной литературе. Основные аспекты когнитивной агентности включают: восприятие окружающей среды, обработку данных и использование полученной информации, принятие решений и действия на их основе, обучение на основе опыта.

Исследования познания в биологии, когнитивистике в целом, эволюционной эпистемологии расширили представления о когнитивном агенте, заставили по-новому взглянуть на его субстратную основу, подтвердив, тем самым, что познавательные процессы могут реализовываться как на нейронной, так и ненеуронной основах. В связи этим некоторые авторы стали утверждать о нетрадиционных воплощениях когнитивной агентности на базе искусственно созданных образований (например, ксеноботов, киборгов и пр.) [1], что, в свою очередь, потребовало переосмысления такой агентности и привело к идее построения общей теории познания, которая включала бы многочисленных агентов, функционирующих на общих основах когнитивной активности.

Задачи, которые ставятся при этом исследователями, заключаются в том, чтобы изучить процессы когнитивной сложности, начиная от простейших проявлений когнитивной активности, затем применить полученные результаты к более сложным организмам и далее - к искусственному интеллекту (ИИ) [2]. Переход к изучению ИИ в данном случае вполне логичен, поскольку речь идет

© Хмелевская С. А., Ермаков Д. Н., 2025.

Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «С указанием авторства - С сохранением условий версии 4.0 Международная» (Creative Commons Attribution – ShareAlike 4.0 International; CC BY-SA 4.0). См. <https://creativecommons.org/licenses/by-sa/4.0/legalcode.ru>

https://doi.org/10.52605/16059921_2025_04_12

о когнитивных механизмах, работающих на ненейронной основе. Но является ли ИИ самостоятельным когнитивным агентом? Что обозначает такая самостоятельность? К каким последствиям она может привести? Именно поиски ответов на эти вопросы и послужили предметом изучения в данной статье¹. Ее методологической основой выступил философский подход, в центре которого рефлексия над полученными результатами изучения ИИ как когнитивного агента в частнонаучном познании и их интерпретация с позиций философского анализа.

1 Теоретико-методологические проблемы изучения когнитивной агентности

Сложность изучения анализируемой темы во многом связана с неразработанностью категориально-понятийного аппарата. Развитие когнитивистики и иных сопряженных с ней наук идет столь стремительными темпами, что изменение указанного аппарата не успевает отразить полученные результаты исследований. Так, исследователи сталкиваются с нечеткостью, во многом порожденной многозначностью термина «познание», трактовка которого во многом зависит от примененных для его изучения подходов. Например, при использовании антропогенного подхода познание отождествляется с социальным познанием, тем самым, делая некорректным применение этого термина к процессам, происходящим вне деятельности и общения людей («будучи самостоятельной реальностью, познание вместе с тем пронизывает все аспекты человеческого мира и лишь в абстракции может быть выделено из него» [4, с.707]). Хотя при этом не исключается, что современное социальное познание является результатом длительного биологического и общественного развития.

При применении биологического подхода к познанию оно не ограничивается продвинутыми способностями, такими как планирование, самосознание, метапознание, язык и т. д., а рассматривается в широком смысле в соответствии с его непрерывным развитием и эволюционным происхождением [5]. Среди такого рода формулировок познания можно выделить следующее определение: «Познание включает в себя сенсорные и другие механизмы обработки информации, которые организм имеет для ознакомления, оценки и продуктивного взаимодействия с особенностями своей окружающей среды [исследования, эксплуатации, уклонения] с целью удовлетворения экзистенциальных потребностей, основными из которых являются выживание/устойчивость, рост/процветание и воспроизводство» [5]. В этом определении понятия «знание» и «информация» совпадают, а познание трактуется как процесс, связанный с переработкой информации.

В более широком варианте, выходящем за пределы непосредственно биологических систем, познание рассматривается как процесс, суть которого состоит в функциональных вычислениях, происходящих «между восприятием и действием, которые позволяют актору охватывать более широкий диапазон времени (за счет памяти и способности к прогнозированию, какими бы большими он ни обладал), чем его непосредственное сейчас» [5]. Такое определение познание может быть применено и по отношению к ИИ. Тем самым, формируется «континуум развития и разнообразия возможностей обработки информации» [5]. В связи с этим интерес представляет концепция М.Левина, названная им «ТАМЕ» («Technological Approach to Mind Everywhere»), цель которой в раскрытии когнитивных процессов на вариативной субстратной основе. Однако в ракурсе философского анализа определения познания как процесса получения, переработки информации возникает проблема, обозначенная Дж.Серлом в его знаменитом мысленном эксперименте с «китайской комнатой», – понимание значения полученной информации (манипулирование символами не то же самое, что понимание их семантического значения) [6]. Это поднимает вопрос о понимании значения полученной информации когнитивным агентом.

2 Понятие когнитивной агентности и особенности ИИ как когнитивного агента

Требуется прояснения и понятие когнитивной агентности. Для этого весьма эвристичным представляется совокупное применение биологического, психологического и информационного

¹ Искусственный интеллект (ИИ) в статье понимается как область исследований и разработок, которая фокусируется на создании машин, способных проявлять интеллект, то есть понимать и обучаться на основе сложной информации, принимать решения и решать проблемы. Дальнейшее развитие ИИ приводит к появлению синтетического интеллекта, что делает его способным думать и вести себя как люди. «Применяя принципы когнитивных вычислений к синтетическому интеллекту, исследователи могут создавать интеллектуальные машины, которые могут демонстрировать когнитивные способности, подобные человеческим» [3].

подходов. В общем плане такую агентность можно определить, как способность систем, функционирующих как на нейронной, так и ненеуронной основах, воспринимать, интерпретировать и обрабатывать информацию, а также принимать решения на базе этой информации. Обозначенный подход предполагает наличие множества когнитивных агентов, проявляющих разный уровень когнитивной сложности, что отличает его от иных подходов, связывающих когнитивную агентность только с биологическими агентами, имеющими нервную систему, или с человеком, обладающим сознанием. Иными словами, когнитивная агентность обозначает способность осуществлять познавательные процессы различными структурами, имеющими разную субстратную основу. Такой подход позволяет рассматривать ИИ также в качестве когнитивного агента, способного выполнять задачи, требующие интеллектуальных усилий, которые обычно ассоциируются с человеческим мышлением. К таким задачам относятся: сбор, обработка, анализ данных из различных источников; использование машинного обучения и других методов для извлечения значений и паттернов из больших объемов информации; принятие решений, основываясь на алгоритмах и данных; обучение и адаптация, взаимодействие с окружающей средой с помощью использования сенсоров и интерфейсов в целях получения информации о внешнем мире и реагирования на нее в реальном времени.

Однако специфика ИИ как когнитивного агента состоит в том, что он все больше и больше становится участником общественной жизни и предлагает решения, которые могут оказать на нее существенное влияние. В условиях дальнейшего развития ИИ, когда он в полной мере приобретет способность к самопрограммированию, возникнет вопрос о его сопоставимости со сложными когнитивными агентами. Так, в настоящее время на базе междисциплинарного подхода, известного как когнитивный ИИ, создаются синтетические системы (синтетический интеллект), которые могут не только понимать и интерпретировать данные, но также думать, рассуждать и адаптироваться к новым ситуациям, решать сложные проблемы [3], а применение в таких системах машинного обучения даст им возможность самообучаться и принимать решения аналогично людям и без явного программирования.

Вместе с тем, на наш взгляд, следует различать понятия «агент познания» и «субъект познания». Агент познания наделен способностью к познанию, в то время как субъект познания обладает еще и собственными побудительными мотивами к такому познанию, социальному по своему типу, может ставить и реализовывать собственные цели. Субъект познания инициирует познавательную активность в целях удовлетворения собственных потребностей, интересов и целей [7, с.391]. Вероятно, развитие ИИ приведет к такому рубежу, когда его будет допустимо отнести к когнитивным субъектам. Хотя уже сейчас можно обнаружить сходство человеческого сознания и ИИ, как то: способность кодировать и обрабатывать символы, быть перепрограммируемыми, воспринимать многомасштабную причинность и др.

В случаях с ИИ представляется некорректным утверждать о существовании гибридного субъекта, состоящего «из людей и не-людей, каждый из которых в сети отвечает за одну или несколько областей деятельности» [8, с.44]. Речь идет не о гибридном когнитивном субъекте как едином целостном образовании, а о его многосоставной основе, включающей самостоятельного когнитивного субъекта (человека) (или субъектов) и ИИ как когнитивного агента. Причем под субъектом познания в данном случае могут подразумеваться, с одной стороны, субъект-разработчик программного обеспечения; с другой стороны, субъект-пользователь информацией, предоставленной ИИ.

ИИ как когнитивный агент обладает определенным набором качеств, которые свойственны субъектам познания («подобие целеустремленности, своего рода рефлексивность, коммуникативность и упрощенные элементы социальности» [9, с.58]), но одновременно определять его в полном объеме как субъекта познания нельзя (некоторые авторы называют его псевдосубъектом, имитирующим когнитивные функции человека [8, с.43]).

3 Дилеммы, порожденные пониманием искусственного интеллекта как когнитивного агента

ИИ как когнитивный агент порождает ряд дилемм, требующих нахождения их решений социумом. Одна из таких дилемм состоит в том, что, с одной стороны, ИИ можно рассматривать как результат программирования людьми, которые заложили его функциональное предназначение и базовые возможности. Но, с другой стороны, по мере развития ИИ становится самопрограммируемым, что

«позволяет ему делать то, что его первоначальные программисты-люди не представляли или не намеревались [10, р.64]. Отсюда, возникает проблема субъектности и, в частности правосубъектности когнитивных агентов. Причем уже в настоящее время по этому вопросу имеется разная юридическая практика (например, Япония предоставила чат-боту «Shibuya Mirai» резидентство, а в ЮАР существует прецедент выдачи ИИ патента [11]; в России ИИ правосубъектностью не обладает). С правосубъектностью связана и дилемма правовой ответственности: должен ли ее нести ИИ или нет? Возникают и проблемы наказания искусственного агента, его отрицательной мотивации.

В связи с когнитивным ИИ возникает и проблема истинности продуцированных им знаний. Так, из реальной практики общения с ChatGPT следует, что можно получить информацию, которая не является корректной, а иногда и просто ложной. В конечном итоге, все опять же упирается в дилемму: верить спродуцированной ИИ информации или нет?

Еще один ракурс этой дилеммы отметил в своей работе Ст.Фуллер. Если исходить из того, что успешное прохождение теста Тьюринга можно считать критерием отделения человеческого интеллекта от иных форм воплощения разума, то как быть в случае такого прохождения ИИ? Следует ли считать, что «машины, прошедшие тест, действительно в каком-то смысле "люди"?» [10, р.65]. Получается, что иметь интеллект, равный человеческому интеллекту, еще не означает иметь человеческий облик.

Еще одна дилемма видится в том, что ИИ как когнитивный агент начинает выполнять социальные роли в разных сферах общественной жизни, создавая конкуренцию человеку (например, в сфере занятости, что ведет к потере рабочих мест, пересмотру требований к знаниям и навыкам сотрудников). Как совместить все большее проникновение ИИ в деятельность социума и сохранение в нем главенствующей позиции человека?

Заключение

Тема ИИ как когнитивного агента поднимает вопросы, связанные не только непосредственно с его осмыслением в этой роли, но и фундаментальные философские проблемы, такие как, что такое познание, как отличить структуры, обладающие когнитивной способностью и пр. Информационный подход познанию достаточно эвристичен, но и он требует корректировки с учетом понимания того, как формируется осмысленная, интерпретированная информация, которая извлекается и обобщается из сигналов получателем/наблюдателем и интегрируется в его будущие действия [12]. Остаются непроясненными и вопросы, связанные с формированием когнитивной сложности, как и понимание эмерджентной сложности. Так, следует отметить эпистемологическую непрозрачность ИИ, в частности, для исследователя (как и для разработчика программы) остается проблемой узнать критерии, по которым классифицируются отдельные точки данных в наборе данных.

Что касается ИИ как когнитивного агента, то он способен весьма эффективно обрабатывать информацию, предлагать на ее основе решения, прогнозировать результаты и пр. Но когнитивистика до сих пор не сформулировала единого мнения о том, что такое «по-настоящему понимать», и в этой связи можно ли применить такое понимание к ИИ.

Несмотря на успехи ИИ в ракурсе его когнитивной агентности вопрос о наделении его субъектностью остается открытым [13]. Возможно, вернуться к этому вопросу можно будет по мере формирования AGI (Artificial general intelligence) - искусственного интеллекта, способного выполнить любую интеллектуальную и творческую человеческую задачу на том же уровне мастерства, что и человек [14].

Литература

1. Levin M. Why We Fear Diverse Intelligence Like AI. 2024. September 3. URL:<https://www.noemamag.com/why-we-fear-diverse-intelligence-like-ai/?ref=feedle.world> (дата обращения: 10.05.2025).
2. Lyon P., Keijzer F., Arendt D., Levin M. Reframing cognition: Getting down to biological basics // Philosophical Transactions of the Royal Society B: Biological Sciences. 2021. Vol. 376. No 1820. P. 1-11.

3. AI for social good. The future of cognitive science and artificial intelligence – unlocking the potential of intelligent machines. 11.01.2024. URL:<https://aiforsocialgood.ca/blog/the-future-of-cognitive-science-and-artificial-intelligence-unlocking-the-potential-of-intelligent-machines> (дата обращения: 10.05.2025).
4. Касавин И.Т. Познание /Энциклопедия эпистемологии и философии науки. М.: «Канон+» РООИ «Реабилитация», 2009. С. 707.
5. Levin M. Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds // *Frontiers in Systems Neuroscience*. 2022. Vol. 16.
URL:https://www.researchgate.net/publication/359441984_Technological_Approach_to_Mind_Everywhere_An_Experimentally_Grounded_Framework_for_Understanding_Diverse_Bodies_and_Mind doi.org/10.3389/fnsys.2022.768201 (дата обращения: 10.05.2025).
DOI:10.3389/fnsys.2022.768201
6. Searle J. Minds, brains, and programs // *Behavioral and brain sciences*. 1980. Vol. 3, No 3 (September). P. 417-424. DOI:10.1017/S0140525X00005756
7. Момджян К.Х. Социальная философия. Деятельностный подход к анализу человека, общества, истории. Часть 1. М.: Издательство Московского университета, 2013. 400 с.
8. Николаева Е.М., Камалева А.М., Николаев М.С. Новая субъектность в сетевом пространстве человеческих и не-человеческих объектов // *Информационное общество*. 2025. № 2. С. 41-48. https://doi.org/10.52605/16059921_2025_02_41
9. Дубровский Д.И., Ефимов А.Р., Лепский В.Е., Славин Б.Б. Фетиш искусственного интеллекта. *Философские науки*. 2022. Т. 65. № 1. С. 44-71. <https://doi.org/10.30727/0235-1188-2022-65-1-44-71>.
10. Fuller St. Six Challenges for Science, Technology, and Society in the Future // *Social Epistemology Review and Reply Collective*. Vol. 14. No 3. P. 64-66. <https://wp.me/p1Bfg0-9H1>
11. Закон.ru. Ралько В. Правосубъектность искусственного интеллекта: возможно ли? Правовой и морально-этический аспект. 17.10.2023.
URL:https://zakon.ru/blog/2023/10/17/pravosubektnost_iskusstvennogo_intellekta_vozmozhno_li_pravovoj_i_moralno-eticheskij_aspekt (дата обращения: 10.05.2025).
12. Levin M. The computational boundary of a "self": Developmental bioelectricity drives multicellularity and scale-free cognition // *Frontiers in Psychology*. 2019. Vol. 10.
https://www.researchgate.net/publication/337930182_The_Computational_Boundary_of_a_Self_Developmental_Bioelectricity_Drives_Multicellularity_and_Scale-Free_Cognition (дата обращения: 10.05.2025). <https://doi.org/10.3389/fpsyg.2019.02688>
13. Хмелевская С.А. Человек как био-социо-техническое существо: влияние новых технологий на природу человека /В сб. Цифровое общество как культурно-исторический контекст развития человека. Под общ. ред. Р.В.Ершовой. Коломна: Государственный социально-гуманитарный университет. 2018. С. 407-412.
14. Ellis D., Bhalla J. The spiritual existential threats posed by AI // *Social Epistemology Review and Reply Collective*. 2024. Vol. 13. No 9. P. 9-14. <https://social-epistemology.com/2024/09/04/the-spiritual-existential-threats-posed-by-ai-david-ellis-and-jag-bhalla/>

ARTIFICIAL INTELLIGENCE FROM THE PERSPECTIVE OF COGNITIVE AGENCY

Khmelevskaya, Svetlana Anatolyevna

Doctor of philosophy, professor

Lomonosov Moscow State University, Faculty of Philosophy, Department of philosophy of natural sciences

Moscow, Russian Federation

xmelevsk@mail.ru

Ermakov, Dmitry Nikolaevich

Doctor of political science, doctor of economics, professor, Honored Worker of the Higher School of the Russian Federation

Lomonosov Moscow State University, Faculty of history, Department of history of social movements and political parties, professor

Moscow, Russian Federation

dermakow@mail.ru

Abstract

The development of artificial intelligence (AI) raises the problem of endowing it with cognitive agency, which is currently understood in different ways. In this regard, the article aims to reveal not only what cognitive agency is, but also to show the specifics of its manifestation in AI. The article substantiates the idea that cognitive agency does not yet provide a basis for considering AI as a subject of cognition. At the same time, the possibility of such consideration is allowed in the future as the AGI (Artificial general intelligence) is formed.

Keywords

artificial intelligence; cognitive agency; knowledge; information; subject of cognition

References

1. Levin M. Why We Fear Diverse Intelligence Like AI. 2024. September 3. URL:<https://www.noemamag.com/why-we-fear-diverse-intelligence-like-ai/?ref=feedle.world> (дата обращения: 10.05.2025).
2. Lyon P., Keijzer F., Arendt D., Levin M. Reframing cognition: Getting down to biological basics // *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2021. Vol. 376. No 1820. P. 1-11.
3. AI for social good. The future of cognitive science and artificial intelligence – unlocking the potential of intelligent machines. 11.01.2024. URL:<https://aiforsocialgood.ca/blog/the-future-of-cognitive-science-and-artificial-intelligence-unlocking-the-potential-of-intelligent-machines> (дата обращения: 10.05.2025).
4. Kasavin I.T. Poznanie /Enciklopediya epistemologii i filosofii nauki. M.: «Kanon+» ROOI «Reabilitaciya», 2009. S. 707.
5. Levin M. Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds // *Frontiers in Systems Neuroscience*. 2022. Vol. 16. URL:https://www.researchgate.net/publication/359441984_Technological_Approach_to_Mind_Everywhere_An_Experimentally_Grounded_Framework_for_Understanding_Diverse_Bodies_and_Mind doi.org/10.3389/fnsys.2022.768201 (дата обращения: 10.05.2025). DOI:10.3389/fnsys.2022.768201
6. Searle J. Minds, brains, and programs // *Behavioral and brain sciences*. 1980. Vol. 3, No 3 (September). P. 417-424. DOI:10.1017/S0140525X00005756
7. Momdzhyan K.H. Social'naya filosofiya. Deyatel'nostnyj podhod k analizu cheloveka, obshchestva, istorii. Chast' 1. M.: Izdatel'stvo Moskovskogo universiteta, 2013. 400 с.
8. Nikolaeva E.M., Kamaleeva A.M., Nikolaev M.S. Novaya sub"ektnost' v setevom prostranstve chelovecheskih i ne-chelovecheskih ob"ektov // *Informacionnoe obshchestvo*. 2025. № 2. S. 41-48. https://doi.org/10.52605/16059921_2025_02_41

9. Dubrovskij D.I., Efimov A.R., Lepskij V.E., Slavin B.B. Fetish iskusstvennogo intellekta. *Filosofskie nauki*. 2022. T. 65. № 1. S. 44-71. <https://doi.org/10.30727/0235-1188-2022-65-1-44-71>.
10. Fuller St. Six Challenges for Science, Technology, and Society in the Future // *Social Epistemology Review and Reply Collective*. Vol. 14. No 3. P. 64-66. <https://wp.me/p1Bfg0-9H1>
11. Zakon.ru. Ral'ko V. Pravosub'ektnost' iskusstvennogo intellekta: vozmozhno li? Pravovoj i moral'no-eticheskij aspekt. 17.10.2023.
URL:https://zakon.ru/blog/2023/10/17/pravosubektnost_iskusstvennogo_intellekta_vozmozhno_li_pravovoj_i_moralno-eticheskij_aspekt (data obrashcheniya: 10.05.2025).
12. Levin M. The computational boundary of a "self": Developmental bioelectricity drives multicellularity and scale-free cognition // *Frontiers in Psychology*. 2019. Vol. 10. https://www.researchgate.net/publication/337930182_The_Computational_Boundary_of_a_Self_Developmental_Bioelectricity_Drives_Multicellularity_and_Scale-Free_Cognition (дата обращения: 10.05.2025). <https://doi.org/10.3389/fpsyg.2019.02688>
13. Hmelevskaya S.A. Chelovek kak bio-socio-tehnicheskoe sushchestvo: vliyanie novyh tekhnologij na prirodu cheloveka /V sb. Cifrovoe obshchestvo kak kul'turno-istoricheskij kontekst razvitiya cheloveka. Pod obshch. red. R.V.Ershovoj. Kolomna: Gosudarstvennyj social'no-gumanitarnyj universitet. 2018. S. 407-412.
14. Ellis D., Bhalla J. The spiritual existential threats posed by AI // *Social Epistemology Review and Reply Collective*. 2024. Vol. 13. No 9. P. 9-14. <https://social-epistemology.com/2024/09/04/the-spiritual-existential-threats-posed-by-ai-david-ellis-and-jag-bhalla/>