

Развитие самостоятельного искусственного интеллекта как угроза человеческой расе: философский аспект

Н. В. Даниелян

Национальный исследовательский университет «МИЭТ», Москва, Россия

vend22@yandex.ru

Аннотация. Рассматривается познавательный сдвиг в понимании субъект-объектного взаимодействия, происходящий в современной эпистемологии вследствие развития когнитивных технологий, являющихся неотъемлемой частью НБИК-конвергенции. Особое внимание уделено философской концепции искусственного интеллекта. На примерах автор рассматривает не только быстрый прогресс в сфере слабого, но и всё большее приближение к тому, что можно принимать за сильный искусственный интеллект. Обосновано, что в таких условиях целью познания всё больше становится не описание объективной реальности, так как субъективное начинает смещаться в сферу искусственного, а определенная организация ее субъективного восприятия, т. е. совершенство познавательных моделей, через которые «схватывается» реальность. Данное предположение полностью подтверждают концепция аутопоэза У. Матураны и Ф. Варелы и нейробиологический конструктивизм Г. Рота. Автором сделан вывод, что границы субъекта познания размываются вследствие стирания жестких границ между естественным и искусственным.

Ключевые слова: когнитивные технологии, искусственный интеллект, субъект, познание, объект, аутопоэз, замкнутая система, нейронная сеть

Для цитирования: Даниелян Н. В. «Развитие самостоятельного искусственного интеллекта как угроза человеческой расе: философский аспект». *Экономические и социально-гуманитарные исследования* 12.1 (2025): 89–94.

<https://doi.org/10.24151/2409-1073-2025-12-1-89-94> EDN: EWCCCM

Original article

The development of full artificial intelligence as threat to the human race: Philosophical aspect

N. V. Danielyan

National Research University of Electronic Technology, Moscow, Russia

vend22@yandex.ru

Abstract. The author examines a cognitive shift in understanding the subject-object interaction that takes place in modern epistemology due to the development of cognitive technologies being an integral part of NBIC convergence. Special attention is paid to the philosophical concept of artificial intelligence. The author gives examples of fast development of artificial narrow intelligence and also of forthcoming so-called artificial general intelligence. It has been reasoned that the purpose of cognition under such circumstances is not the description of the objective reality, as the subjective shifts more and more to the artificial under the influence of modern technologies, but a definite arrangement of its subjective perception. It means the perfection of cognitive models allowing a subject to capture the reality. Such a supposition is completely confirmed by the autopoiesis concept by H. Maturana and F. Varela and the neurobiological constructivism by G. Roth. The author has concludes that the borders of a cognition subject are blurred as a result of erasing strict borders between natural and artificial

Keywords: cognitive technologies, artificial intelligence, subject, cognition, object, autopoiesis, closed system, neural network

For citation: Danielyan N. V. “The Development of Full Artificial Intelligence as Threat to the Human Race: Philosophical Aspect”. *Ekonomicheskie i sotsial’no-gumanitarnye issledovaniya* = *Economic and Social Research* 12.1 (2025): 89–94. (In Russian).
<https://doi.org/10.24151/2409-1073-2025-12-1-89-94>

Введение

Очевидно, что ответить однозначно на вопрос, поставленный в интервью С. Хокинга корпорации BBC News в 2014 г.: «Угрожает ли искусственный интеллект концом человеческой расы?»¹ — невозможно. Сегодня человечество всё более погружается в мир создаваемых им технологий. Нас повсеместно окружают «умные» приложения: умный дом (Smart Home), Интернет вещей (IoT), приложения дополненной (AR) и виртуальной реальности (VR), всевозможные интеллектуальные приложения, которые способны за нас рисовать, сочинять, решать задачи, за-

полнять документы, переводить тексты и т. д. Технологии не только расширяют, но и в значительной степени замещают интеллектуальный труд человека в современном мире. Мы вынуждены конкурировать с тем, что сегодня создается в ходе НБИК-конвергенции (здесь речь идет о взаимопроникновении и взаимовлиянии нано-, био-, инфо- и когнитивных технологий). Остановимся более подробно на последних.

Как отмечает В. А. Лекторский, «когнитивные технологии, основанные на результатах нейробиологических исследований, используют информационные представления,

¹ Hawking S. “Stephen Hawking Warns Artificial Intelligence Could End Mankind”. Interview by R. Cellan-Jones. *BBC News*. BBC, 02 Oct. 2014. Web. 25 Jan. 2025. <<https://christusliberat.org/wp-content/uploads/2017/10/Stephen-Hawking-warns-artificial-intelligence-could-end-mankind-BBC-News.pdf>>.

а для экспериментального изучения процессов переработки информации на нейронном уровне используют биологические и нанотехнологии и т. д.» (Лекторский, 2014: 11). Технологии искусственного интеллекта (ИИ), базируясь на междисциплинарном подходе к самой идее мышления (в их основе лежит изучение сознания, познания, поведения с позиций естественно-научных дисциплин с учетом гуманитарной составляющей, так как любая искусственная интеллектуальная система имеет целью заместить мыслительный труд человека), уже могут имитировать мыслительную деятельность человека. Предполагается, что в будущем ИИ сможет воспроизводить основные принципы работы человеческого мозга. Если в 1960-х гг., с появлением компьютеров, укоренилось представление о том, что «можно понять все познавательные процессы, начиная от восприятия и кончая мышлением, как своеобразные процессы переработки информации с помощью определенных правил и алгоритмов по аналогии с тем, как это происходит в компьютере» (Лекторский, 2014: 4), то понимание познания как своего рода вычисления вполне легитимно для информатики и философии. Таким образом, философия и информатика ставят проблему создания и функционирования искусственных замкнутых систем (поскольку любое вычислительное устройство является замкнутой системой), конструирующих определенным образом поле возможных состояний для моделирования объективного.

Конструктивизм и возможности моделирования естественного в искусственном

Яркий пример подтверждения возможности имитировать объективное в искусственном можно найти в идеях радикального конструктивизма. Теоретики аутопоэза У. Матурана и Ф. Варела обосновывают, что в силу операциональной замкнутости человеческого мозга как системы «каждый акт познания рождает некий мир» (Матурана, Варела, 2019: 26). В работе «Древо познания» они

проводят глубокий и подробный анализ разнообразных паттернов окружающей среды, вызывающих в человеческом мозге определенные возмущения. Из концепции следует, что «рационально возможен только сконструированный наблюдателем мир» (Даниелян, 2014: 59). При рассмотрении процесса познания с позиции когнитивных операций, выполняемых системами наблюдения, можно предположить, что они не обязательно могут иметь естественную природу.

В нейробиологическом конструктивизме Г. Рот дополняет этот теоретический подход, предлагая корректировку аутопоэза (Roth, 1997). По его мнению, так как мозг является частью целостной аутопоэтической системы организма человека, он не может функционировать самостоятельно, поэтому его невозможно рассматривать изолированно. Однако именно мозг обладает нейронной сетью (попытки ее моделирования ученые и инженеры предпринимают при проектировании систем ИИ для имитации познания в неживой системе), которая способна принимать бесконечное число состояний, что делает мозг закрытой системой. Подобная информационная замкнутость обуславливает специфику функционирования человеческого мышления, выражающуюся в процессе конструирования действительности, что опять говорит в пользу моделирования его в искусственном.

Сильный и слабый искусственный интеллект: к сути вопроса

Актуальность вопроса о необходимости проведения исследования достижений в сфере создания ИИ с позиции перспективы его превращения в познающий субъект возрастает. Однако человечество пока не пришло к полному пониманию, что такое интеллект, из чего он состоит, каким образом формируется. А уж тем более нет ясности, как нам понимать ту форму интеллекта, которую принято называть «искусственной». Так, в предисловии к труду одного из классиков кибернетики Н. Нильсона область ИИ

определена следующим образом: «В самом общем смысле — это решение „интеллектуальных“ задач с помощью автоматических методов, в первую очередь вычислительных машин. Но какую деятельность следует считать интеллектуальной, а какую нет? Это не вполне ясно» (Фомин, 1973: 5).

Наиболее успешной попыткой ответа на поставленный вопрос, с точки зрения автора статьи, являются две модели ИИ: сильная и слабая. Сторонники концептуализации слабой модели ИИ считают, что соответствующим образом запрограммированный компьютер может только моделировать мыслительные акты человека, тогда как сторонники сильной допускают, что запрограммированные компьютерные устройства действительно мыслят и в силу этого могут находиться в соответствующих когнитивных состояниях. По мнению М. Ю. Гутенева, «сильная версия искусственного интеллекта предполагает, что компьютеры могут приобрести способность к рефлексивной мыслительной деятельности и к осознанию себя, пусть даже их мыслительный процесс будет происходить отлично от человеческого» (Гутенев, 2016: 10: 128).

На II конгрессе Российского общества истории и философии науки в 2020 г. тезис о разделении на сильный и слабый ИИ вызвал бурные споры со стороны присутствовавших специалистов в сфере создания ИИ. На взгляд автора, такое разграничение вполне обосновано, так как то, с чем мы имеем дело сегодня: голосовые помощники, системы генерации изображений, медицинские системы диагностики, чат-боты, системы обработки данных основано на слабом ИИ, — все эти системы ИИ пытаются программными методами и математическими моделями воспроизвести заложенные в них алгоритмы. Яркой иллюстрацией модели слабого ИИ является созданная в Китае «государственная инфраструктура надзора». Как известно, Китай — один из лидеров мировых держав в области разработки ИИ. Он выступает амбассадором использо-

вания больших массивов данных своих граждан, которые собираются, например, для системы социального рейтингования. Она обрабатывает информацию, включая данные социальных сетей, местных властей и данные об индивидуальной деятельности, чтобы подсчитать рейтинг благонадежности каждого гражданина. Можно полагать, что задача философской рефлексии этой модели — осмысление того, насколько этично в государственном масштабе собирать персональные данные с граждан страны и манипулировать ими по своему усмотрению. Очевидно, что возникает ряд угроз, таких как утечка информации, скрытая слежка, манипуляция, взлом и т. д. Ни о каком сильном ИИ речи не идет, в данном случае он необходим только как удобный и быстрый инструмент, исполняющий целенаправленный запрос.

Термин «сильный искусственный интеллект» впервые ввел в научный оборот американский философ Дж. Серл, выступив его противником в «аргументе китайской комнаты» (Searle, 1980). Человек, который не знает китайского языка, вполне способен после продолжительной тренировки, следуя составленным для него инструкциям, научиться правильно отвечать на вопросы на китайском. Однако это не означает, что он понимает их смысл. То же самое делает компьютер: несмотря ни на какие технологические прорывы, ИИ мыслить не способен, он следует заложенному в него алгоритму.

К настоящему времени совершён значительный прогресс в области технологий сильного искусственного интеллекта, поэтому можно полагать, что аргумент Дж. Серла не вполне справедлив. Так, 21 сентября 2024 г. газета *The Economist* опубликовала статью «Прорыв, который необходим ИИ». В ней, в частности, отмечается, что разработчики вносят изменения в программное обеспечение ИИ. В связи с этим на смену большим моделям, требующим больших вычислительных мощностей, приходят меньшие по размеру, более специализированные системы. В статье отмечается, что новейшая

модель компании OpenAI o1 более ориентирована на «размышление» (reasoning), чем на генерацию текста². Несмотря на широко обсуждаемые на разных сайтах недостатки этой модели ИИ, налицо огромный прогресс в области когнитивных технологий, которые создаются на основе нейросетей. Еще совсем недавно невозможно было подумать, что работу по генерации креативных идей, отладке программного кода, написанию сложных математических формул и прочие задачи можно поручить искусственной системе, основанной на компьютерной технологии, а не человеку. Это ведет к пониманию, что границы субъекта познания размываются, так как всё более стирается разграничение между естественным и искусственным.

Таким образом, в последнее время активно развиваются обе модели искусственного интеллекта: прогресс неумолим. То, что еще недавно казалось невозможным, сегодня становится обыденностью. В книге «Сознающий ум: в поисках фундаментальной теории», впервые опубликованной в 1996 г., Д. Чалмерс пишет: «Исследования в области искусственного интеллекта (или ИИ) в значительной своей части были нацелены на воспроизведение ментального в вычислительных машинах. О серьезном прогрессе говорить пока не приходится, но сторонники этого подхода доказывают, что у нас есть все основания считать, что в итоге компьютеры действительно обретут ум» (Чалмерс, 2015: 389).

Соответственно, возникает вполне закономерный вопрос: действительно ли «размышляет» компьютер? Не вникая в технологические подробности, ответим: вполне

очевидно, что разработчики системы применили аналитический подход к решению комплексных задач, состоящий из ряда шагов, направленных на имитирование мыслительной деятельности человека. В качестве примера можно привести первые аналоги искусственных нейронных сетей на основе программируемых логических интегральных микросхем (Лебедев, Белецкий, 2021). Они содержат в себе переход от программного продукта к физическому моделированию искусственной нейросети. Это означает, что субъективное всё более смещается в сферу искусственного, моделируемого и конструируемого самим человеком, но начинает функционировать и совершать познавательные и мыслительные операции согласно заложенным алгоритмам независимо от него.

Заключение

Современный мир движется в инновационной эпистемологической плоскости, где новая научная революция неизбежна. По своим кардинальным последствиям она способна превзойти все предыдущие этапы, так как базируется на коренном изменении традиционных представлений о субъекте — уже не только человеку, коллективу, но компьютеризированном устройстве, наделенном способностью мыслить. Поэтому ответ на вопрос С. Хокинга, взятый в качестве заглавного для размышления, как уже было сказано, содержится в разумности и обоснованности применения человечеством технологий ИИ. То, что кажется простым и полезным в ежедневной работе сегодня, может кардинальным образом изменить жизнь человечества как в лучшую, так и в худшую сторону уже завтра.

Список литературы и источников / References

- Гутенев М. Ю. «К вопросу об идентичности искусственного и естественного интеллекта». *Дискурсология: методология, теория, практика* 10 (2016): 123—134. EDN: WYXUDR.
 Gutenev M. Yu. “The Identity of Artificial and Natural Intelligence”. *Diskursologiya: metodologiya, teoriya, praktika* 10 (2016): 123—134. (In Russian).

Даниелян Н. В. *Научная рациональность и конструктивизм*. М.: МИЭТ, 2014. 100 с.

² “The Breakthrough AI Needs”. *The Economist*. The Economist Newspaper, 19 Sep. 2024. Web. 25 Jan. 2025. <<https://www.economist.com/leaders/2024/09/19/the-breakthrough-ai-needs>>.

- Danielyan N. V. *Scientific Rationality and Constructivism*. Moscow: MIET, 2014. 100 p. (In Russian).
- Лебедев М. С., Белецкий П. Н. «Реализация искусственных нейронных сетей на ПЛИС с помощью открытых инструментов». *Труды ИСП РАН* 33.6 (2021): 175—192. [https://doi.org/10.15514/ISPRAS-2021-33\(6\)-12](https://doi.org/10.15514/ISPRAS-2021-33(6)-12). EDN: IXEIBR.
- Lebedev M. S., Belecky P. N. “Artificial Neural Network Inference on FPGAs Using Open-Source Tools”. *Trudy ISP RAN = Proc. of the Institute for System Programming of the RAS* 33.6 (2021): 175—192. (In Russian). [https://doi.org/10.15514/ISPRAS-2021-33\(6\)-12](https://doi.org/10.15514/ISPRAS-2021-33(6)-12)
- Лекторский В. А. «Исследование интеллектуальных процессов в современной когнитивной науке: философские проблемы». *Естественный и искусственный интеллект: методологические и социальные проблемы*. Под ред. Д. И. Дубровского, В. А. Лекторского. М.: «Канон+»; РООИ «Реабилитация», 2014. 3—16.
- Lektorsky V. A. “Research of Intellectual Processes in Modern Cognitive Science: Philosophical Issues”. *Estestvennyy i iskusstvennyy intellekt: metodologicheskiye i sotsial’nyye problemy*. Ed. by D. I. Dubrovsky, V. A. Lektorsky. Moscow: Kanon+; ROOI “Reabilitatsiya”, 2014. 3—16. (In Russian).
- Матурана У., Варела Ф. *Древо познания: Биологические корни человеческого понимания*. 2-е изд., доп. Пер. Ю. А. Данилова. М.: URSS; ЛЕНАНД, 2019. 316 с. Синергетика: от прошлого к будущему 95.
- Maturana Humberto, Varela Francisco J. *Tree of Knowledge: The Biological Roots of Human Understanding*. Forew. by J. Z. Young. Boulder, CO: Shambhala, 1987. 263 p.
- Фомин С. В. Предисловие редактора перевода. *Искусственный интеллект: Методы поиска решений*. Н. Нильсон. Пер. с англ. В. Л. Стефанюка. М.: Мир, 1973. 5—6.
- Fomin S. V. Translation Editor’s Foreword. *Problem-Solving Methods in Artificial Intelligence*. By N. Nilsson. Transl. by V. L. Stefanyuk. Moscow: Mir, 1973. 5—6. (In Russian).
- Чалмерс Д. *Сознающий ум: в поисках фундаментальной теории*. Пер. с англ. В. В. Васильева. Изд. 2-е. М.: URSS; Либроком, 2015. 512 с. Философия сознания.
- Chalmers David J. *The Conscious Mind: In Search of a Fundamental Theory*. Rev. ed. New York: Oxford Up, 1997. 432 p. Philosophy of Mind.
- Roth G. *Das Gehirn und seine Wirklichkeit: Kognitive Neurobiologie und ihre philosophischen Konsequenzen*. Frankfurt a. M.: Suhrkamp, 1996. 384 S. (In German).
- Searle J. R. “Minds, Brains, and Programs”. *Behavioral and Brain Sciences* 3.3 (1980): 417—424. <https://doi.org/10.1017/S0140525X00005756>

Информация об авторе

Даниелян Наира Владимировна — доктор философских наук, доцент, профессор Института лингвистического и педагогического образования, профессор Института высокотехнологичного права, социальных и гуманитарных наук Национального исследовательского университета «МИЭТ» (Россия, 124498, Москва, Зеленоград, пл. Шокина, 1), vend22@yandex.ru, ORCID: 0000-0002-9613-2969.

Information about the author

Naira V. Danielyan — Dr. Sci. (Philos.), Assoc. Prof., Professor at the Institute of Linguistic and Pedagogical Education, Professor at the Institute of High-Tech Law, Social Sciences and Humanities, National Research University of Electronic Technology (Russia, 124498, Moscow, Zelenograd, Shokin sq., 1), vend22@yandex.ru, ORCID: 0000-0002-9613-2969.

Статья поступила в редакцию 06.02.2025, одобрена после рецензирования 10.03.2025.
The article was submitted 06.02.2025, approved after reviewing 10.03.2025.