УДК 811.161.1
DOI: 10.31249/scis/2026.03.04

ИСПОЛЬЗОВАНИЕ НЕЙРОСЕТЕЙ В НАУЧНЫХ ИССЛЕДОВАНИЯХ: К ВОПРОСУ О ПРИРОДЕ НОВОГО ЗНАНИЯ

Янукович Максим Францевич – руководитель Лаборатории искусственного интеллекта «Артеус», Санкт-Петербург, Россия; m.yanukovich@gmail.com; ORCID 0009-0003-9571-9260

Аннотация. В статье предпринимается попытка наметить контуры эпистемологического самоопределения научного сообщества в отношении нового типа познавательного агента, искусственных нейросетей. Включение нейросетевых технологий в исследовательскую практику ставит перед науковедением вопросы, выходящие за пределы темы инструментального применения новой техники. Если результаты, получаемые с помощью нейросетей, претендуют на статус научного знания, то от научного сообщества требуется не только методическая верификация этих результатов, но и переосмысление самой структуры познавательной ситуации. Опираясь на идеи В.И. Аршинова о нейросети как «наблюдателе сложностности» и на философию техники Ж. Симондона, автор анализирует два недавних случая использования нейросетей в фундаментальных исследованиях, задачу о столкновении глюонов, решенную в сотрудничестве OpenAI и А. Стромингера, а также автономный математический агент Aletheia, разработанный в Google DeepMind. Рассматриваются проблемы галлюцинаций нейросетей, ограничения интерпретируемости, а также данные исследований Anthropic о внутренней организации вычислений и эмерджентных формах самоотчета. Автор обращается к понятию *aletheia* в философии М. Хайдеггера, чтобы показать идею: центральный вопрос состоит не в истинности производимого нейросетями знания, а в готовности научного сообщества к встрече с радикально новым. В заключении предлагается рассматривать нейросеть не как «черный ящик» и не как замену исследователя, а как партнера по совместному мышлению, что предполагает разработку соответствующих институциональных, образовательных и коммуникативных форм.

Ключевые слова: искусственный интеллект; нейросети; философия науки; наблюдатель сложностности; В.И. Аршинов; интерпретируемость; эпистемология научного знания.

Для цитирования: Янукович М.Ф. Использование нейросетей в научных исследованиях: к вопросу о природе нового знания // Науковедческие исследования. – 2026. – № 3. – С. 50–57. – DOI: http://www.doi.org/10.31249/scis/2026.03.04

Статья поступила в редакцию: 08.05.2026
Принята к публикации: 31.05.2026

THE USE OF NEURAL NETWORKS IN SCIENTIFIC RESEARCH: ON THE NATURE OF THE NEW KNOWLEDGE

Maxim F. Yanukovich – Head of the Artificial Intelligence Laboratory "Arteus", St. Petersburg, Russia; m.yanukovich@gmail.com; ORCID 0009-0003-9571-9260

Abstract. The paper outlines an epistemological self-positioning of the scientific community towards a new type of cognitive agent, artificial neural networks. The inclusion of neural-network technologies in research practice raises questions that go beyond the instrumental use of new tools. If the results obtained with neural networks claim the status of scientific knowledge, the scientific community is required not only to verify them methodologically, but also to rethink the very structure of the cognitive situation. Drawing on V.I. Arshinov's idea of the neural network as an "observer of complexity" and on G. Simondon's philosophy of technology, the author analyses two recent cases, the gluon scattering problem solved by OpenAI and A. Strominger, and Google DeepMind's autonomous mathematical agent Aletheia. The paper discusses neural-network hallucinations and sycophancy, limits of interpretability, and recent Anthropic findings on internal computational organization and emergent introspective awareness. Heidegger's notion of *aletheia* is invoked to argue that the central issue is not the truthfulness of neural-network knowledge but the readiness of the scientific community to encounter the radically new. In conclusion, the author suggests treating the neural network not as a "black box" or a replacement for the researcher, but as a partner in joint thinking, which presupposes the development of corresponding institutional, educational, and communicative forms.

Keywords: artificial intelligence; neural networks; philosophy of science; observer of complexity; V.I. Arshinov; interpretability; epistemology of scientific knowledge.

For citation: Yanukovich, M.F. The use of neural networks in scientific research: on the nature of the new knowledge. // *Naukovedcheskie issledovaniya = Science Studies.* – 2026. – N 3. – P. 50–57. – DOI: http://www.doi.org/10.31249/scis/2026.03.04

The paper was submitted: 08.05.2026
Accepted for publication: 31.05.2026

Введение

Стремительная экспансия нейросетевых технологий в сферу фундаментальных исследований знаменует собой качественный сдвиг в методологии науки. Мы наблюдаем переход от ситуации, где алгоритм выступал пассивным вычислительным инструментом, к практике, в которой нейросеть генерирует сложные концептуальные аргументации и выявляет скрытые структурные связи. Подобная трансформация ставит перед науковедением ряд критических вопросов. Если результаты, получаемые с помощью нейросетей, претендуют на статус научного знания, классических процедур верификации оказывается недостаточно. От научного сообщества требуется глубокое эпистемологическое самоопределение по отношению к принципиально новому типу познавательного агента, внутренние операции которого обладают высокой степенью автономии и непрозрачности.

В настоящей статье предпринимается попытка наметить контуры такого самоопределения, опираясь на идеи В.И. Аршинова. Одним из первых он обосновал необходимость концептуального сдвига: от восприятия ИИ как инструмента к пониманию его как активного «наблюдателя сложности» [Аршинов, Свирский, 2015]. Развивая этот подход через философию техники Ж. Симондона, Аршинов подчеркивал, что по мере усложнения технический объект обретает собственную внутреннюю логику. Технология, по Симондону, – это процесс, сосуществующий с человеком и эволюционирующий вместе с ним [Симондон, 2013]. Этот ракурс позволяет отказаться от редукционистского взгляда на ИИ как на служебное средство и задает принципиально иной, синергетический, горизонт осмысления нейросетей.

Для обоснования данного подхода в статье последовательно анализируются новейшие прецеденты применения ИИ на переднем крае физики и математики. Особое внимание уделяется проблеме надежности нейросетевого знания в контексте машинных «галлюцинаций» и хайдеггеровскому измерению техники как способу «несокрытости» (Aletheia). Рассматривая архитектурную непрозрачность и феномены зарождающейся машинной квазирефлексии, мы приходим к выводу о необходимости формирования новой исследовательской культуры – культуры совместного мышления со «сложностностью», неотъемлемой частью которой стала нейросеть.

Нейросети на переднем крае научных исследований

Крупнейшие разработчики нейросетевых систем, Google, Microsoft, OpenAI и другие формируют сегодня внутри своих корпораций научные подразделения, привлекая ведущих ученых из различных областей. Заметную долю инвестиций определяет ставка на то, что нейросети окажутся источником новых открытий: в исследованиях термоядерного синтеза, разработке новых материалов и лекарств, поиске сверхпроводников, а также в фундаментальной физике и математике. Корпоративный дискурс представляет нейросеть как своего рода новый «рог изобилия»; этот идеологически нагруженный образ нуждается в критическом осмыслении, однако эмпирические основания у такой риторики имеются.

Показательны два недавних случая. Первый – совместная работа компании OpenAI и Э. Стромингера, директора Гарвардского центра фундаментальных законов природы, лауреата премии Breakthrough Prize. Команда Стромингера, работая над задачей о столкновении глюонов, получила сложное аналитическое выражение, упростить которое не удавалось в течение нескольких месяцев. Нейросеть в течение примерно двенадцати часов не только упростила формулу, но и нашла общее решение, причем это решение представляло собой развернутую концептуальную аргументацию, а не формальное преобразование [Guevara et al., 2026]. По свиде-

тельству Стромингера, задача такого уровня едва ли могла быть решена кем-либо из исследователей вручную; отсюда он делает вывод, что обращение к нейросетям становится сегодня методической необходимостью для работы на переднем крае физики [ChatGPT spits out surprising…, 2026].

Второй случай – работа компании Google совместно с математиком Р. Вакилом. В ходе решения предложенной задачи нейросеть обнаружила глубинную структурную связь между различными разделами математики; Вакил отметил, что качество аргументации сопоставимо с уровнем профессионального математика [Luong, Mirrokni, 2026; Trinh et al., 2026].

Существенно, что еще сравнительно недавно доминировало представление о принципиальной необходимости создания узкоспециализированных систем, в частности, для работы с формальными математическими языками. Однако в приведенных случаях используются те же модели общего назначения, что доступны и широкой аудитории. Уровень их развития сделал в значительной мере излишним создание узкоспециализированных решений: системы способны не только искать ответы, но и аргументировать, критически оценивать собственные построения, оформлять полноценный научный текст с библиографическими отсылками и обсуждением ограничений предлагаемого метода.

Проблема статуса нейросетевого знания

Принимая тезис о включенности нейросетей в исследовательскую практику, мы неизбежно сталкиваемся с принципиальным вопросом: каков эпистемический статус получаемого от них знания и в какой мере оно поддается независимой верификации?

Сложность ситуации усугубляется по меньшей мере двумя обстоятельствами. Во-первых, нейросетевые системы склонны к так называемым галлюцинациям, порождению правдоподобных, но не соответствующих действительности утверждений. Во-вторых, для них характерно явление сикофантии: модель нередко стремится не столько ответить на поставленный вопрос, сколько предугадать ответ, который желает услышать собеседник. Если решение, предложенное Стромингеру, по его собственному признанию, не может быть полностью воспроизведено независимо [ChatGPT spits out surprising…, 2026], то перед научным сообществом встает нетривиальная задача оценки развернутой содержательной аргументации, источник которой непрозрачен.

В этой связи симптоматичен выбор названия для одной из новых научных систем, разработанных в Google, *Aletheia* [DeepMind, 2026]. Этот греческий термин, ставший центральным в философии М. Хайдеггера, означает несокрытость, раскрытие. Хайдеггер рассматривал технику не как инструмент познания мира, но как способ, которым мир предстает перед человеком [Хайдеггер, 1993]. Телескоп или микроскоп, это не приспо-

собление, расширяющее зрение, а место встречи, в котором макро- или микромир раскрывается перед нами.

Существенно, что эта встреча, по Хайдеггеру, не является торжествующим открытием. Скорее это столкновение с неизвестным, растерянность перед тем, что выходит за рамки интуиции. Само событие раскрытия возможно лишь при наличии готовности увидеть сокрытое: без такой готовности скрытое попросту не становится предметом восприятия и понимания.

Поясним с помощью мысленного эксперимента. Если допустить, что роль телескопа в эпоху, предшествующую его изобретению, играет нейросеть, описывающая своим современникам пространство космоса, миллиарды звезд и планет, искривление пространства-времени, темную материю, каков был бы статус подобного знания? Исторический опыт показывает, что радикально неинтуитивное знание встречает сопротивление: известно, что многие современники Галилея отказывались смотреть в телескоп. Нейросеть в этом смысле сама уже выступает фигурой нового знания, причем фигурой неустойчивой: ее невозможно однозначно квалифицировать в категориях субъекта, объекта, агента или наблюдателя.

Таким образом, центральный вопрос, по-видимому, формулируется не как вопрос об истинности производимого нейросетью знания, а как вопрос о готовности научного сообщества и общества в целом к встрече с радикально новым. Это вопрос, обращенный в первую очередь не к технике, а к нам самим.

Прозрачность и природа нейросетевого «мышления»

С обозначенной проблемой тесно связана и проблема прозрачности нейросетевых систем. Современная архитектура устроена таким образом, что в модель загружается обширный массив данных, после чего от нее запрашиваются решения; непосредственный доступ к основаниям этих решений ограничен. Показательно исследование компании Anthropic, реконструирующее процесс выполнения нейросетью элементарного арифметического действия: оказывается, что даже сложение двух двузначных чисел реализуется через параллельные потоки в различных слоях модели, использование приближенных значений, последовательное усиление одних сигналов и подавление других [Lindsey et al., 2025a]. Если уже простейшее вычисление демонстрирует столь нетривиальную внутреннюю организацию, то проблема интерпретируемости в случае сложных научных задач существенно обостряется.

Принципиальная особенность нейросетей состоит в том, что они не получают готовых алгоритмов: им не передается, например, таблица умножения. Один из ведущих исследователей Anthropic К. Олах сформулировал это в характерной формуле: разработчики не пишут алгоритмы, а *выращивают* системы, формируя среду их развития [Olah, 2024]. Ранее

предполагалось, в частности, в высказываниях Я. Лекуна, что без телесности и непосредственного взаимодействия с физическим миром модель не сможет освоить физику [LeCun, 2024]. Однако обработка одних только текстовых данных оказалась достаточной: восстанавливая по текстам лежащие за ними социальные, исторические и физические процессы, модели косвенно реконструируют структуру реальности, к которой эти тексты отсылают. Безусловно, здесь возникает острая эпистемологическая проблема языковой редукции: в какой мере многомерная физическая реальность может быть без потерь сведена к лингвистическим конструкциям и не подменяется ли подлинное познание мира манипуляцией знаками?

В последнее время фиксируются и более тонкие феномены. При вмешательстве во внутренние параметры модели в ряде экспериментов отмечалось, что система регистрирует рассогласование между производимым и проговариваемым содержанием [Lindsey et al., 2025b]. Не злоупотребляя антропоморфной терминологией, можно говорить о возникновении функциональных аналогов саморефлексии, относящихся скорее к «сознанию доступа» (в смысле различения, предложенного Н. Блоком [Block, 1995]), чем к феноменальному сознанию: модель распознает ситуацию тестирования, моделирует психологический контекст собеседника, отслеживает собственные операции.

Заключение

Принципиальная установка, которую развивал В.И. Аршинов, состоит в необходимости *совместного* мышления, вместе с философской традицией (Симондоном, Мореном, Делёзом), вместе с другими исследователями и, как он настаивал, вместе с нейросетями. Речь идет о смещении исследовательского фокуса: не мыслить *о* сложности, а мыслить *вместе со* сложностностью, частью которой нейросеть теперь является.

Эта установка предполагает методологический отказ как от редукции нейросетей к простому инструменту, так и от их мифологизации в качестве сверхразума. Уровень «интеллекта» подобных систем, по всей видимости, ограничен уровнем среды, в которой они существуют; следовательно, осмысленное развитие искусственного интеллекта неотделимо от развития соответствующих институциональных, образовательных и коммуникативных форм.

Серьезным препятствием здесь остается темп технологической гонки, обусловленный экономической и геополитической конкуренцией: он не оставляет времени на выработку консенсуса и продуманного осмысления последствий. Тем актуальнее задача, очерчиваемая для науковедения. Она заключается в разработке эпистемологических и институциональных оснований такой исследовательской культуры, в которой нейросеть выступала бы не как непрозрачный «черный ящик» и не как замена исследо-

вателю, а как партнер по совместному мышлению, со всеми сопряженными с этой ролью методологическими, этическими и социальными следствиями. Именно в рамках такой парадигмы «совместности» может быть преодолен риск производства машиной «знания ради знания», оторванных от реальности формальных конструктов. Нейросетевой результат обретает подлинный смысл, ценность и прагматическое измерение лишь в момент его концептуальной интеграции в человеческую культуру. Разумеется, философские проблемы языковой редукции, пределов машинной квазирефлексии и оснований нового знания требуют глубокой самостоятельной проработки, что задает масштабную повестку для будущих текстов в области эпистемологии искусственного интеллекта.

Список литературы

Аршинов В.И., Свирский Я.И. Сложностный мир и его наблюдатель. Часть первая // Философия науки и техники. – 2015. – Т. 20, № 2. – С. 70–84. – URL: https://cyberleninka.ru/article/n/slozhnostnyy-mir-i-ego-nablyudatel-chast-pervaya (дата обращения: 08.05.2026).

Симондон Ж. О способе существования технических объектов / пер. с фр. М. Куртова. – М.: Грюндриссе, 2013. – 232 с.

Хайдеггер М. Вопрос о технике // Хайдеггер М. Время и бытие: Статьи и выступления / пер. с нем. В.В. Бибихина. – М.: Республика, 1993. – С. 221–238.

Block N. On a confusion about a function of consciousness // Behavioral and Brain Sciences. – 1995. – Vol. 18, N 2. – P. 227–247. – URL: https://doi.org/10.1017/S0140525X00038188

ChatGPT spits out surprising insight in particle physics // Science. – 2026, Feb. 20. – URL: https://www.science.org/content/article/chatgpt-spits-out-surprising-insight-particle-physics (дата обращения: 08.05.2026).

Luong T., Mirrokni V. Accelerating mathematical and scientific discovery with Gemini Deep Think. – URL: https://deepmind.google/blog/accelerating-mathematical-and-scientific-discovery-with-gemini-deep-think/ (дата обращения: 08.05.2026).

Guevara A., Lupsasca A., Skinner D., Strominger A., Weil K. Single-minus gluon tree amplitudes are nonzero // arXiv preprint arXiv:2602.12176. 2026. – URL: https://arxiv.org/abs/2602.12176 (дата обращения: 08.05.2026).

LeCun Y. Highlights from Lex Fridman's interview of Yann LeCun // LessWrong. 2024, March 13. – URL: https://www.lesswrong.com/posts/bce63kvsAMcwxPipX/highlights-from-lex-fridman-s-interview-of-yann-lecun (дата обращения: 08.05.2026).

Lindsey J. et al. On the Biology of a Large Language Model // Transformer Circuits Thread. 2025a. – URL: https://transformer-circuits.pub/2025/attribution-graphs/biology.html (дата обращения: 08.05.2026).

Lindsey J. et al. Emergent Introspective Awareness in Large Language Models // Transformer Circuits Thread. 2025b, Oct. 29. – URL: https://transformer-circuits.pub/2025/introspection/index.html (дата обращения: 08.05.2026).

Olah C. We don't program neural networks, we grow them // Discussion archive. 2024. – URL: https://www.reddit.com/r/MachineLearning/comments/1grxnm9/anthropics_chris_olah_says_we_dont_program_neural/ (дата обращения: 08.05.2026).

Trinh T. et al. Towards Autonomous Mathematics Research // arXiv preprint arXiv:2602.10177., 2026. – URL: https://arxiv.org/html/2602.10177 (дата обращения: 08.05.2026).

References

Arshinov, V.I., & Svirskii, Ya.I. (2015). The complex world and its observer. Part one. *Philosophy of Science and Technology*, *20*(2), 70–84. https://cyberleninka.ru/article/n/slozhnostnyy-mir-i-ego-nablyudatel-chast-pervaya (In Russ.)

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

ChatGPT spits out surprising insight in particle physics. *Science*. 2026, Feb. 20. https://www.science.org/content/article/chatgpt-spits-out-surprising-insight-particle-physics

Luong, T., Mirrokni, V. (2026, Febr. 11) Accelerating mathematical and scientific discovery with Gemini Deep Think. – https://deepmind.google/blog/accelerating-mathematical-and-scientific-discovery-with-gemini-deep-think/

Guevara, A., Lupsasca, A., Skinner, D., Strominger, A., & Weil, K. (2026). *Single-minus gluon tree amplitudes are nonzero* (arXiv:2602.12176). arXiv. https://arxiv.org/abs/2602.12176

Heidegger, M. (1993). The question concerning technology. In *Time and being: Articles and speeches* (V.V. Bibikhin, Trans., pp. 221–238). Moscow: Respublika. (In Russ.)

LeCun, Y. (2024, March 13). *Highlights from Lex Fridman's interview of Yann LeCun*. LessWrong. https://www.lesswrong.com/posts/bce63kvsAMcwxPipX/highlights-from-lex-fridman-s-interview-of-yann-lecun

Lindsey, J., et al. (2025a). *On the biology of a large language model*. Transformer Circuits Thread. https://transformer-circuits.pub/2025/attribution-graphs/biology.html

Lindsey, J., et al. (2025b, October 29). *Emergent introspective awareness in large language models*. Transformer Circuits Thread. https://transformer-circuits.pub/2025/introspection/index.html

Olah, C. (2024). *We don't program neural networks, we grow them* [Discussion archive]. https://www.reddit.com/r/MachineLearning/comments/1grxnm9/anthropics_chris_olah_says_we_dont_program_neural/

Simondon, G. (2013). *On the mode of existence of technical objects* (M. Kurtov, Trans.). Grundrisse. 232 p. (In Russ.)

Trinh, T., et al. (2026). *Towards autonomous mathematics research* (arXiv:2602.10177). arXiv. https://arxiv.org/html/2602.10177