Эпистемология и философия науки
2026. Т. 63. № 3. С. 220–233
УДК 167.7

Epistemology & Philosophy of Science
2026, vol. 63, no. 3, pp. 220–233
DOI: https://doi.org/10.5840/eps202663350

Наука в цифровую эпоху: как ИИ меняет представление о научном знании и его создании

Лисенкова Анастасия Алексеевна – доктор культурологии,
директор.
Высшая школа общественных наук
Гуманитарного института
Санкт-Петербургского политехнического университета
Петра Великого.
Российская Федерация, 195251, г. Санкт-Петербург, ул. Политехническая, д. 29;
e-mail: oskar46@mail.ru

Быльева Дарья Сергеевна – кандидат политических наук, доцент.
Санкт-Петербургский политехнический университет
Петра Великого.
Российская Федерация, 195251, г. Санкт-Петербург, ул. Политехническая, д. 29;
e-mail: bylieva_ds@spbstu.ru

В современной философии науки и техники стремительное развитие технологий искусственного интеллекта (ИИ) выдвигает на первый план фундаментальные вопросы о природе научного познания и его трансформации под влиянием алгоритмических систем. Обзор философских дискуссий об интеллектуальных системах в науке демонстрирует, что за три десятилетия произошел фундаментальный сдвиг точки консенсуса: если на начальном этапе они рассматривались исключительно как инструмент усиления человеческих когнитивных способностей, то в 2022–2023 гг. философское сообщество оказалось перед необходимостью легитимизировать ИИ в качестве нетривиального эпистемического участника. Традиционный вопрос о том, можно ли считать ИИ когнитивным агентом, дополнился проблемами, связанными с практиками его использования: под влиянием больших данных и нейросетей происходит трансформация критериев научности, кризис субъект-объектного различения перед лицом алгоритмической непрозрачности и поиск новых моделей эпистемической ответственности. ИИ утвердился в качестве функционально нередуцируемого участника познавательных практик, однако ни одна из предложенных концепций – от расширенного инструментализма до распределенной или коллаборативной агентности – не устраняет полностью концептуальное напряжение, заданное разрывом между эффективным выполнением ИИ когнитивных операций и отсутствием у него интенциональных состояний. Это делает невозможным прямое применение традиционных критериев обоснованности. В то же время основные дискуссии остаются в рамках эпистемологии первого порядка, сосредоточенной на вопросе о (квази)субъектности машин. Однако данная исследовательская программа, при всей ее продуктивности, упускает из виду фундаментальную трансформацию: приобретение ИИ метанаучной роли, в которой он не только активно участвует в познании, но и во многом перестраивает конфигурацию исследовательского поля.

Ключевые слова: искусственный интеллект, научное познание, нейросети, алгоритмическая прозрачность, когнитивные акторы, этические вызовы, автоматизация науки, искусственный интеллект в науке





© Лисенкова А.А., 2026
© Быльева Д.С., 2026



SCIENCE IN THE DIGITAL AGE: HOW AI IS CHANGING THE UNDERSTANDING OF SCIENTIFIC KNOWLEDGE AND ITS CREATION

Anastasia A. Lisenkova –
DSc in Cultural Studies.
Head of the Graduate School of Social Sciences at the Institute of Humanities, Peter the Great St. Petersburg Polytechnic University.
29 Politekhnicheskaya St., Saint Petersburg 195251 Russian Federation;
e-mail: oskar46@mail.ru

Darya S. Bilyeva –
PhD in Political Science, Associate Professor.
Graduate School of Social Sciences at the Institute of Humanities, Peter the Great St. Petersburg Polytechnic University.
29 Politekhnicheskaya St., Saint Petersburg 195251 Russian Federation;
e-mail: bylieva_ds@spbstu.ru

In contemporary philosophy of science and technology, the rapid development of artificial intelligence (AI) technologies brings to the forefront fundamental questions about the nature of scientific cognition and its transformation under the influence of algorithmic systems. A review of philosophical discussions on intelligent systems in science demonstrates that over three decades there has been a fundamental shift in the point of consensus: whereas at the initial stage they were considered exclusively as a tool for enhancing human cognitive abilities, in 2022–2023 the philosophical community found itself facing the need to legitimize AI as a nontrivial epistemic participant. The traditional question of whether AI can be regarded as a cognitive agent has been supplemented by problems related to the practices of its use: under the influence of big data and neural networks, the criteria of scientificity are being transformed, the subject – object distinction is facing a crisis in the face of algorithmic opacity, and new models of epistemic responsibility are being sought. AI has established itself as a functionally irreducible participant in cognitive practices; however, none of the proposed conceptions – from extended instrumentalism to distributed or collaborative agency – completely eliminates the conceptual tension created by the gap between AI's effective performance of cognitive operations and its lack of intentional states. This makes the direct application of traditional criteria of justification impossible. At the same time, the main discussions remain within the framework of first-order epistemology, focused on the question of the (quasi-)subjectivity of machines. However, this research program, for all its productivity, overlooks a fundamental transformation: AI's acquisition of a metascientific role, in which it not only actively participates in cognition but also largely restructures the configuration of the research field.

Keywords: artificial intelligence, scientific cognition, neural networks, algorithmic transparency, cognitive actors, ethical challenges, automation of science, artificial intelligence in science

Введение

Цифровая эпоха поставила перед философией науки вопрос, который еще полвека назад казался сугубо футурологическим: каким образом системы, не обладающие сознанием, участвуют в производстве того, что мы привыкли считать знанием? Искусственный интеллект (ИИ), первоначально рассматривавшийся как вычислительный инструмент, постепенно занял позицию, которую трудно описать в терминах классической эпистемологии.

Современное развитие технологий искусственного интеллекта (ИИ) представляет собой не просто новый этап в научно-техническом прогрессе, а глубинную трансформацию самих основ научной деятельности, затрагивающую ее эпистемологические, методологические и социальные аспекты. ИИ перестал быть вспомогательным инструментом для обработки больших массивов данных или автоматизации рутинных операций – он становится активным участником познавательных процессов. Этот качественный сдвиг требует переосмысления традиционных представлений о природе научного знания, его методах и критериях валидности, поскольку алгоритмические системы вносят в познавательный процесс новые инструменты, основанные на вероятностных моделях, машинном обучении и обработке данных в масштабах, недоступных человеческому разуму. Целью статьи является выявление основных линий дискуссий о цифровых технологиях и ИИ в науке и демонстрация трансформации представлений, связанных с расширением роли последнего.

Роли ИИ в научном исследовании

Практическое применение ИИ в науке уже сегодня демонстрирует как его потенциал, так и сопутствующие риски: возникают вызовы, связанные с надежностью и интерпретируемостью результатов, полученных с помощью ИИ. Для анализа ИИ необходимо более четко представить, какие функциональные роли он может играть в науке.

Наиболее ранней ролью ИИ в науке можно считать обработку данных и выявление паттернов (классификация, кластеризация, регрессия, извлечение признаков из больших массивов данных). Системы ИИ такого типа выявляют корреляции, структуры и аномалии, которые затем интерпретируются человеком и способствуют развитию новых исследовательских парадигм, таких как data-driven science, где данные и вычислительные модели становятся центральными элементами научного метода. С помощью нейросетевой обработки данных Аргоннской национальной лаборатории удалось сделать ускоренное, масштабируемое и воспроизводимое обнаружение гравитационных волн, которые были предсказаны еще Эйнштейном, но не были обнаружены [Huerta et al., 2021]. Благодаря ИИ значительно продвинулось исследование белков. Обученная на сотнях тысяч экспериментально определенных последовательностей белка нейросеть AlphaFold смогла предсказывать трехмерные структуры с точностью, сопоставимой с экспериментальными методами и значительно превосходящей все иные известные до этого методы [Jumper et al., 2021]. Темп последующих открытий в области молекулярной биологии был также беспрецедентным. Открытие тысячи

потенциально новых семейств белков стало стимулом прогресса в генетических, эволюционных, медицинских исследованиях [Callaway, 2022, p. 236]. В гуманитарных науках нейросети используются для анализа исторических документов, литературных произведений, философских трактатов, выявляя скрытые лексические паттерны, структуру текста, его стилистику и даже подтексты. Так, например, ИИ позволил осуществить анализ литературных произведений на основе применения теории графов и NER-алгоритмов (в т.ч. выявив, в каких отношениях находились свыше 880 наименованных сущностей в книгах Д.Р. Толкина) [Yashchenko, Makarov, 2022]. В исторической науке, кроме оцифровки архивных документов и фондов, технологии ИИ успешно используются для расшифровки древних текстов. Например, Transkribus или языковые модели Google AI позволяют восстанавливать шумерские и ассирийские тексты. Проект Сары Паркак (“National Geographic”) использует машинное обучение для анализа спутниковых изображений и поиска археологических объектов, что помогло обнаружить сотни скрытых поселений в Египте. Проект “Pelagios” связывает древние карты с современными географическими данными. В культурологии и искусствоведении предобученные модели выявляют стилистические особенности, характерные для отдельных художников или художественных школ, что меняет процесс атрибуции и аутентификации произведений, сводя к минимуму субъективность.

Среди новых продуктов, создаваемых с помощью ИИ, можно в первую очередь отметить лекарства и новые материалы. Так, в результате анализа библиотеки химических соединений нейросеть обнаружила новые антибиотики, наиболее интересным из которых является халицин, воздействующий на бактерии, вызывающие псевдомембранозный колит и туберкулез [Stokes et al., 2020, p. 694]. Понятно, что для создания лекарственных средств недостаточно открытия активного вещества, требуется найти структуру лекарства, провести доклинические и клинические испытания. Благодаря работе нейросетей, одна из которых ищет, на что должно быть направлено воздействие для лечения заболевания, а вторая подбирает для этого молекулы, была найдена молекула INS018_055 для борьбы с легочным фиброзом. Искусственный интеллект активно способствует созданию материалов с уникальными свойствами, которые находят применение в широком спектре отраслей – от биомедицины до космической промышленности, эффективно прогнозируя их свойства (биосовместимость, прочность, электро- и теплопроводность и т.д.). Например, нейросеть PauliNet способна определить основные свойства вещества для произвольных молекул [Hermannet al., 2020, p. 893], сочетая подходы квантовой механики с алгоритмами глубокого обучения, что позволяет предсказывать энергетические и структурные характеристики молекул с высокой точностью.

Все более активно ИИ используется для генерации гипотез и новых теорий – в том числе в математических исследованиях [Davies et al., 2021, p. 73] для создания теорий принятия решений [Peterson et al., 2021, p. 1211] и др. Одно из направлений – это автоматизированное доказательство теорем с использованием модели трансформеров. Используя технологию трансформеров, система перебирает все возможные математические операции и построения дерева решений, сокращая время на проверку сложных доказательств и позволяя обнаруживать новые математические теории и решения. Сложные системы способны проектировать эксперименты, выполнять их в автоматизированных лабораториях и итеративно изменять результаты – так называемые «самоуправляемые лаборатории».

В целом для ИИ не существует разницы между прошлым и будущим, между тем, что известно человечеству, и новым. Индифферентность ИИ по отношению к той точке развития науки, на которой находится человечество, означает возможность «знать все». Для чего достаточно, как в фантастических произведениях, задать правильный вопрос или, в контексте научных исследований, вектор приложения сил и нужные базы данных [Быльева, 2025]. Поскольку сегодня ИИ генерирует гипотезы, разрабатывает эксперименты, собирает и интерпретирует большие наборы данных и открывает то, что, «вероятно, было бы невозможно получить с использованием только традиционных научных методов» [Wang et al., 2023, p. 47], можно предположить, что искусственные системы способны к выявлению закономерностей в областях, недоступных для человека, в частности обнаружению причинно-следственных связей между далеко отстоящими друг от друга событиями в исторической перспективе [Анохин и др., 2022, с. 97], а также выявлению прогнозных моделей и закономерностей без проведения реальных опытов.

При этом на более ранних этапах ИИ мог использоваться только институциональными акторами, крупными исследовательскими инфраструктурами. Эпистемическая агентность оставалась институционально опосредованной: доступ к инструментам, данным и вычислительным мощностям определял то, кто может выступать в роли исследователя. Распространение больших языковых моделей (БЯМ) привело к появлению новой массовой роли ИИ – автора, когда любой пользователь может использовать ИИ для написания статей, рецензий и т.п. При этом не существует возможности с абсолютной уверенностью определить, является ли текст сгенерированным, существуют только технические оценки вероятности. Статистика использования слов позволяет предположить, что не менее 10% научных статей созданы с помощью БЯМ. PubMed ежегодно публикует около 1,5 млн статей, из которых не менее 150 000 статей произведены с помощью искусственного интеллекта [Liang et al., 2024]. В марте 2024 г. в издательстве Elsevier была издана статья, начинавшаяся



со слов, хорошо известных всем пользователям ChatGPT: “Certainly, here is a possible introduction for your topic…” («Конечно, вот возможное введение для вашей темы…»), а в статье от июня 2024 г. можно было найти заявление: «Мне очень жаль, но у меня нет доступа к информации в реальном времени или данным о пациенте, поскольку я являюсь языковой моделью искусственного интеллекта». БЯМ могут быть полезны при выполнении рутинных интеллектуальных функций, однако расширяющееся использование ИИ для написания научных текстов значительно повышает риск наличия в статьях так называемых «галлюцинаций» и сфальсифицированных данных. В результате, с одной стороны, повышая скорость обработки данных и помогая составлять структуры, аннотации и подбирая списки литературы, ИИ сокращает время на работу со статьей на 40% [Alkaissi, McFarlane, 2023], с другой, это приводит к стандартизации и стилистической однотипности текстов на основе шаблонности построения фраз ИИ, а в целом к стремительному увеличению доли генерированных текстов в научном дискурсе.

Форматы использования ИИ демонстрируют функциональный плюрализм, не позволяющий зафиксировать его в какой-либо одной эпистемологической категории. Массовизация доступа к генеративным моделям дополнительно размывает институциональные границы научной деятельности. Генерация научных текстов с помощью ИИ приводит к «эпистемической инфляции», когда возможности проверки (экспертная оценка, институциональный контроль и т.п.) не соответствуют скорости создания.

Философская дискуссия до и в «эпоху генеративных моделей»

Философские размышления о науке, основанной на больших данных и интеллектуальных машинах, сформировались до «бума» БЯМ. По мере развития вычислительных систем все чаще возникает вопрос об их статусе и о том, можно ли рассматривать их как когнитивных агентов. Дискуссии о «разрушении субъекта» в науке встраиваются в более широкий контекст философии коллективного познания [List, Pettit, 2011] и споры об «искусственном субъекте». В рамках science and technology studies (STS) агентность понимается как эффект сетевых конфигураций, где нечеловеческие акторы участвуют в производстве знания наравне с людьми, перераспределяя эпистемические функции. Согласно функционализму в философии информации [Bynum, Moor, 1998; Floridi, 2013] любая система, демонстрирующая когнитивные функции (распознавание образов, вывод, обобщение и т.п.), может считаться когнитивным агентом.

Критики функционализма (например, Т. Мецингер) утверждают, что функциональное поведение не подразумевает интенциональности. ИИ не обладает самосознанием, не понимает смыслов и не способен к подлинной рефлексии [Metzinger, 2022]. В конкретно научном контексте вопрос об интеллектуальной деятельности и агентности машин разбивается на подмножество вопросов, связанных с возможностью полноценно участвовать в тех или иных научных процессах, традиционно закрепленных за человеком. В частности, уже с 1990-х гг. появилось понятие «компьютерное открытие» (например, [Langley, 1987; Thagard, 1988]). Однако существовал философский консенсус относительно цифровых технологий как инструментов, которые лишь усиливают когнитивные способности человека, не изменяя при этом структуру эпистемической активности.

Философы науки предпринимали попытки встроить использование «больших данных» в традиционные представления о научном методе. Вероятностная природа машинного обучения, в которой выводы основаны на статистических корреляциях, а не детерминированной логике, внесла дополнительную неопределенность в научные исследования. Одним из эпизодов в философии ИИ в науке стала дискуссия о машинном обучении как форме индукции (в русле старых дебатов между Карлом Поппером и Рудольфом Карнапом). Д. Гиллис утверждал, что существование успешных программ машинного обучения демонстрирует реальность индукции [Gillies, 1996]. Критика тезиса Гиллиса строилась на том, что системы машинного обучения работают на принципах, отличающихся от классических представлений об индукции, опираясь на оптимизацию, статистический вывод и, в случае глубокого обучения, автоматическое извлечение признаков исследуемого объекта. Далее дискуссия о вычислительном подходе к индуктивному выводу перешла от исследования, основанного на гипотезах, к исследованию, основанному на данных [Gillies, Gillies, 2022]. Сторонники «четвертой парадигмы» в науке выдвинули предположение, что корреляция может заменить причинно-следственное объяснение [Hey, 2009] – утверждение, вызвавшее резкую философскую критику [Leonelli, 2012; Pietsch, 2016]. Позже традиционные критерии, такие как воспроизводимость и фальсифицируемость, стали дополняться ИИ-специфическими метриками, такими как вариативность модели при повторном обучении или чувствительность к возмущениям входных данных [Jumper et al., 2021].

С появлением глубоких нейронных сетей, принципы функционирования которых недоступны даже их разработчикам, начались активные дискуссии об алгоритмической непрозрачности и проблеме «черного ящика». «Черный ящик» породил этические противоречия, связанные с ответственностью за решения, принимаемые ИИ. В философии науки в этой связи стали говорить о разделении понимания

и объяснения [Sullivan, 2022; Zednik, 2021], о «коллективной эпистемической агентности» с использованием ИИ [Tollefsen, 2015], «эпистемической непрозрачности» [Humphreys, 2009], стремясь предложить новый ракурс субъектности ученого в производстве научного знания с участием ИИ.

Бум нейросетей в 2022–2025 гг. стал не просто технологическим прорывом, но и катализатором изменений в представлениях об основах человеческого познания. Как пишут К. Митчем и Т. Лю, ИИ может заменить некоторые функции ученых, тогда как ученые не могут заменить некоторые функции ИИ (например, обработать петабайт данных) [Liu, Mitcham, 2024, p. 92]. Реальность в науке изменила точку консенсуса по поводу роли ИИ, «утвердив» его способность создавать новое [Stoliarova, 2025]. Генеративные модели не только классифицируют или предсказывают, но и продуцируют новые данные. Д. Хосрови и Ф. Финн утверждают, что при определенных условиях они могут выполнять функцию de novo synthetic evidence – свидетельства, сопоставимого по эпистемическому статусу с эмпирическими данными [Khosrowi, Finn, 2025].

Тем не менее наиболее дискуссионными остаются вопросы о том, может ли ИИ-система быть субъектом или квазиагентом научного знания, а также каковы эпистемологические последствия и этические аспекты ее участия в научной деятельности. Если первоначально этические принципы требовали отказа от использования ИИ-генерации текстов, то современные исследования предлагают разнообразные модели «эпистемической ответственности» [Lloyd, 2025]. Исследования границ применения ИИ в научном исследовании [Введенская, 2025; Зубковская, 2024; Шалак, 2024] сопровождаются концепциями ИИ как эпистемического посредника или артефактного агента, участвующего в производстве знаний и при этом не обладающего преднамеренными состояниями [Cruz-Aguilar, 2025]. Вместе с тем, не будучи сознательными, действия ИИ имеют эпистемические последствия и значение [Barassi, 2024]. В связи с этим современные исследователи развивают концепцию «масштабно-разделенной агентности» [Palermos, 2025], рассматривают ИИ-системы как эпистемические коллаборации [Greco, 2026], предлагая концепцию эмерджентной когнитивной структуры, в которой ИИ не является ни инструментом, ни автономным агентом, а границы субъекта временно размываются. Значимой проблемой в дискуссиях является феномен «галлюцинаций» ИИ, который, с одной стороны, может служить фактическим подтверждением безразличия генеративных моделей к истине [Herrmann, Levinstein, 2024], с другой, указывать на то, что ИИ создают ощущение ложной активности [Шевченко, 2025] и псевдокреативности нового порядка [Álvarez et al., 2025].

Дискуссии отражают противоречия между эффективным применением ИИ в научных практиках и невозможностью концептуализи-

ровать это участие в привычных эпистемологических категориях. В условиях, когда производство знания оказывается распределенным, гетерогенным и частично непрозрачным для познающих субъектов, трансформация научной рациональности требует особого внимания исследователей.

Метанаучная роль ИИ

Технологии ИИ анализируют современную науку как сложный объект с разветвленной структурой исследований, собственными уникальными социальными практиками и дискурсивными сетями. На основе выявленных трендов они генерируют новые гипотезы, отслеживают перспективные направления исследований, формируют исследовательские программы. ИИ-методы используются для анализа научной литературы в целях выявления тематических кластеров, динамики перехода между различными научными парадигмами, «горячих» и «заброшенных» областей, а также разрывов и лакун в имеющемся знании. Технологии NLP, глубокого обучения и графовые нейросети позволяют строить динамические графы знаний и цитат [Baghel et al., 2025], повышая эффективность прогностических научных моделей. Однако их применение создает замкнутую систему, ограничивающую интеллектуальный поиск, а наукометрические базы «алгоритмически» сужают многообразие подходов и результатов [Бажанов, 2023, с. 200].

Хасок Чанг представляет метанаучный уровень как пространство для критического осмысления науки, где история и философия науки работают в связке, позволяя видеть науку как динамичный процесс решения проблем [Chang, 2004, p. 235–250]. В роли метанаучного наблюдателя ИИ отслеживает изменения в программах исследований, сдвиги в нормативных рамках и накопленных аномалиях, фиксирует структуру современного научного дискурса и практики. Научный ландшафт представляется как гиперграф, учитывающий людей, методы, объекты, цели исследований в качестве основания прогнозной цепи для последующих открытий [Sourati, Evans, 2023]. По мнению Х. Китано, такие возможности анализа и прогнозирования должны существенно расширить возможности и ограничения, сдерживаемые пределами когнитивных возможностей человека [Kitano, 2021]. ИИ-ученый сможет генерировать и проверять огромное множество гипотез, независимо от их значимости и востребованности в научном дискурсе, так как может анализировать структуру цитирований и терминологическую близость для поиска скрытых связей в междисциплинарных исследованиях. С. Фуллер считает, что ИИ способен выявлять «неоткрытое публичное знание» (по Д. Свонсону)



[Delikatnyi, 2025], выступая не как (квази)агент, участвующий в познании наряду с человеком, а как конструктор исследовательского поля.

Заключение

Эволюция роли ИИ в науке от инструмента обработки данных и выявления паттернов до генератора гипотез, соавтора текстов и, в пределе, конструктора исследовательских полей представляет собой функциональную экспансию, которая не укладывается в классическую дихотомию «инструмент vs субъект» и делает несостоятельными попытки зафиксировать ИИ в рамках одной эпистемологической категории.

Философская рефлексия сталкивается с необходимостью пересмотреть базовые эпистемологические различения в ситуации, когда знание производится системами, не обладающими сознательными состояниями, оперирующими «нечеловеческими» способами получения знания и не подпадающими под действие привычных критериев обоснованности. Если классические критерии научности формировались в расчете на интенционально контролируемые процедуры, то в условиях, когда научный результат генерируется частично непрозрачными алгоритмами, ИИ становится не только участником познавательных практик, но и средством анализа науки на метауровне. Однако эта метанаучная функция сопряжена с фундаментальным ограничением: опора на исторические данные и сложившиеся паттерны цитирования создает риск алгоритмического сужения эпистемического многообразия, подмены плюрализма оптимизацией. Представляется, что дальнейшее движение философии науки лежит не в плоскости уточнения статуса ИИ в существующих категориях, а в пересмотре самих этих категорий. Необходим понятийный аппарат, способный описывать распределенные, гетерогенные и частично непрозрачные конфигурации акторов, участвующих в производстве знания.

Список литературы

Анохин и др., 2022 – *Анохин К.В., Новоселов К.С., Смирнов С.К., Ефимов А.Р., Матвеев Ф.М.* Искусственный интеллект для науки и наука для искусственного интеллекта // Вопросы философии. 2022. № 3. С. 93–105. DOI: 10.21146/0042-8744-2022-3-93-105

Бажанов, 2023 – *Бажанов В.А.* Искусственный интеллект, технологии Big Data (больших данных) и особенности современного политического процесса // Философия. Журнал Высшей школы экономики. 2023. Т. 7. № 3. С. 193–210.

Быльева, 2025 – *Быльева Д.С.* Философский вызов и онтолитический эффект искусственного интеллекта. М.: Инфра-М, 2025. 240 с.

Введенская, 2025 – *Введенская Е.В.* Критика искусственного интеллекта: Философская оптика // Технологос. 2025. № 1. С. 87–97. DOI: 10.15593/perm.kipf/2025.1.07.

Зубковская, 2024 – *Зубковская А.А.* Искусственный интеллект и научная объективность // Науковедческие исследования. 2024. № 2. С. 42–49.

Шалак, 2024 – *Шалак В.И.* Избавление от иллюзий ИИ на примере ChatGPT // Technology and Language. 2024. Т. 5. № 2. С. 26–39. DOI: 10.48417/technolang.2024.02.

Шевченко, 2025 – *Шевченко А.А.* «Галлюцинации» ИИ как новая форма эпистемической ошибки // Respublica Literaria. 2025. Т. 6. № 4. С. 93–98.

References

Alkaissi, McFarlane, 2023 – Alkaissi, H., McFarlane, S.I. “Artificial Hallucinations in ChatGPT: Implications in Scientific Writing”, *Cureus*, 2023. https://doi.org/10.7759/cureus.35179

Álvarez et al., 2025 – Álvarez, A., Villalba, M., Dumit, J. “Speculative Trainers: Large Language Models and Techniques of Affirmative Speculation”, *Technology and Language*, 2025, no. 18 (1), pp. 223–251. https://doi.org/10.48417/technolang.2025.01.15

Anokhin, K.V., Novoselov, K.S., Smirnov, S.K., Efimov, A.R., Matveev, F.M. “Iskusstvennyj intellekt dlya nauki i nauka dlya iskusstvennogo intellekta” [Artificial Intelligence for Science and Science for Artificial Intelligence], *Voprosy filosofii*, 2022, no. 3, pp. 93–105. https://doi.org/10.21146/0042-8744-2022-3-93-105. (In Russian)

Baghel et al., 2025 – Baghel, R.R.R.S., Rajput, P.S., Patel, M.K. “The Rise of AI–Driven Scientometrics: A Decade of Transformative Growth (2012–2023)”, *Annals of Library and Information Studies*, 2025, no. 72 (3). https://doi.org/10.56042/alis.v72i3.18487

Barassi, 2024 – Barassi, V. “Toward a Theory of AI Errors: Making Sense of Hallucinations, Catastrophic Failures, and the Fallacy of Generative AI”, *Harvard Data Science Review*, 2024, no. Special Issue 5. https://doi.org/10.1162/99608f92.ad8ebbd4

Bazhanov, 2023 – Bazhanov, V.A. “Iskusstvennyj intellekt, tehnologii Big Data (bol’shih dannyh) i osobennosti sovremennogo politicheskogo processa” [Artificial Intelligence, Big Data Technologies and Peculiarities of the Modern Political Process], *Filosofija. Zhurnal Vysshej shkoly jekonomiki*, 2023, no. 7 (3), pp. 193–210. (In Russian)

Byl’eva, D.S. *Filosofskij vyzov i ontoliticheskij effekt iskusstvennogo intellekta* [Philosophical Challenge and Ontolythic Effect of Artificial Intelligence]. Moscow: Infra-M Publ., 2025. (In Russian)

Bynum, Moor, 1998 – Bynum, T.W., Moor, J.H. (eds.) *The Digital Phoenix: How Computers Are Changing Philosophy*. Oxford: Blackwell Publ., 1998.

Callaway, 2022 – Callaway, E. “What’s Next for AlphaFold and the AI Protein-Folding Revolution”, *Nature*, 2022, no. 604 (7905), pp. 234–238. https://doi.org/10.1038/d41586-022-00997-5

Chang, 2004 – Chang, H. *Inventing Temperature: Measurement and Scientific Progress*. New York: Oxford University Press, 2004, 304 pp.

Cruz-Aguilar, 2025 – Cruz-Aguilar, M.A. “The Epistemic Revolution of AI: Reconfiguring the Foundations of Scientific Knowledge”, *AI & SOCIETY*, 2025. https://doi.org/10.1007/s00146-025-02658-3

Davies et al., 2021 – Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., … Kohli, P. “Advancing Mathematics by Guiding Human Intuition with AI”, *Nature*, 2021, no. 600 (7887), pp. 70–74. https://doi.org/10.1038/s41586-021-04086-x

Delikatnyi, 2025 – Delikatnyi, K. “Interview with Steve Fuller on Social Epistemology and the Future of Education in an Age of Artificial Intelligence”, *Social Epistemology Review and Reply Collective*, 2025, no. 14 (5), pp. 48–51.

Floridi, 2013 – Floridi, L. *The Philosophy of Information*. Oxford: Oxford University Press, 2013.

Gillies, 1996 – Gillies, D. *Artificial Intelligence and Scientific Method*. Oxford: Oxford University Press, 1996, 432 pp.

Gillies, Gillies, 2022 – Gillies, D., Gillies, M. “Artificial Intelligence and Philosophy of Science from the 1990s to 2020”, in: W.J. Gonzalez (ed.). *Current Trends in Philosophy of Science*. Cham: Springer Publ., 2022, pp. 65–79. https://doi.org/10.1007/978-3-031-01315-7_4

Greco, 2026 – Greco, J. “The Transmission of Knowledge via Large-Scale Technology: A Shared Agency Account”, *Social Epistemology*, 2026, no. 40 (1), pp. 87–100. https://doi.org/10.1080/02691728.2025.2463054

Hermann et al., 2020 – Hermann, J., Schätzle, Z., Noé, F. “Deep-Neural-Network Solution of the Electronic Schrödinger Equation”, *Nature Chemistry*, 2020, 12, pp. 891–897. https://doi.org/10.1038/s41557-020-0544-y

Herrmann, Levinstein, 2024 – Herrmann, D.A., Levinstein, B.A. “Standards for Belief Representations in LLMs”, *Minds and Machines*, 2024, no. 35 (1), art. 5. https://doi.org/10.1007/s11023-024-09709-6

Hey et al., 2009 – Hey, T., Tansley, S., Tolle, K. (eds.) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research, 2009.

Huerta et al., 2021 – Huerta, E.A., Khan, A., Huang, X., Tian, M., Levental, M., Chard, R., … Foster, I. “Accelerated, Scalable and Reproducible AI–Driven Gravitational Wave Detection”, *Nature Astronomy*, 2021, no. 5 (10), pp. 1062–1068. https://doi.org/10.1038/s41550-021-01405-0

Humphreys, 2009 – Humphreys, P. “The Philosophical Novelty of Computer Simulation Methods”, *Synthese*, 2009, no. 169 (3), pp. 615–626. https://doi.org/10.1007/s11229-008-9435-2

Jumper et al., 2021 – Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … Hassabis, D. “Highly Accurate Protein Structure Prediction with AlphaFold”, *Nature*, 2021, no. 596 (7873), pp. 583–589. https://doi.org/10.1038/s41586-021-03819-2

Khosrowi, Finn, 2025 – Khosrowi, D., Finn, F. “Can Generative AI Produce Novel Evidence?”, *Philosophy of Science*, 2025, pp. 1–12. https://doi.org/10.1017/psa.2025.10123

Kitano, 2021 – Kitano, H. “Nobel Turing Challenge: Creating the Engine for Scientific Discovery”, *npj Systems Biology and Applications*, 2021, no. 7 (1), art. 29. https://doi.org/10.1038/s41540-021-00189-3

Langley et al., 1987 – Langley, P., Simon, H.A., Bradshaw, G.L., Zytkow, J.M. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, MA: MIT Press, 1987.

Leonelli, 2012 – Leonelli, S. “Classificatory Theory in Data-Intensive Science: The Case of Open Biomedical Ontologies”, *International Studies in the Philosophy of Science*, 2012, no. 26 (1), pp. 47–65. https://doi.org/10.1080/02698595.2012.653119

Liang et al., 2024 – Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., … Zou, J.Y. “Mapping the Increasing Use of LLMs in Scientific Papers”, *arXiv*, 2024, no. 2404.01268.

List, Pettit, 2011 – List, C., Pettit, P. *Group Agency: The Possibility, Design and Status of Corporate Agents*. Oxford: Oxford University Press, 2011.

Liu, Mitcham, 2024 – Liu, T., Mitcham, C. “Toward Practical Hermeneutics of Fourth Paradigm AI for Science”, *Technology and Language*, 2024, no. 14 (1), pp. 89–105. https://doi.org/10.48417/technolang.2024.01.07

Lloyd, 2025 – Lloyd, D. “Epistemic Responsibility: Toward a Community Standard for Human-AI Collaborations”, *Frontiers in Artificial Intelligence*, 2025, no. 8. https://doi.org/10.3389/frai.2025.1635691

Metzinger, 2022 – Metzinger, T. “Towards a Global Artificial Intelligence Charter”, in: S. Voeneky, P. Kellmeyer, O. Mueller, W. Burgard (eds.). *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*. Cambridge: Cambridge University Press, 2022, pp. 167–175. https://doi.org/10.1017/9781009207898.013

Palermos, 2025 – Palermos, S.O. “Knowledge from AI”, *Social Epistemology Review and Reply Collective*, 2025, no. 14 (11).

Peterson et al., 2021 – Peterson, J.C., Bourgin, D.D., Agrawal, M., Reichman, D., Griffiths, T.L. “Using Large-Scale Experiments and Machine Learning to Discover Theories of Human Decision-Making”, *Science*, 2021, no. 372 (6547), pp. 1209–1214. https://doi.org/10.1126/science.abe2629

Pietsch, 2016 – Pietsch, W. “The Causal Nature of Modeling with Big Data”, *Philosophy & Technology*, 2016, no. 29 (2), pp. 137–171. https://doi.org/10.1007/s13347-015-0202-2

Shalak, V.I. “Izbavlenie ot illyuzij II na primere ChatGPT” [Delusion of AI on the Example of ChatGPT], *Technology and Language*, 2024, no. 5 (2), pp. 26–39. https://doi.org/10.48417/technolang.2024.02. (In Russian)

Shevchenko, A.A. “Gallyutsinatsii II kak novaya forma epistemicheskoj oshibki” [AI ‘Hallucinations’ as a New Form of Epistemic Error], *Respublica Literaria*, 2025, no. 6 (4), pp. 93–98. (In Russian)

Sourati, Evans, 2023 – Sourati, J., Evans, J.A. “Accelerating Science with Human-Aware Artificial Intelligence”, *Nature Human Behaviour*, 2023, no. 7 (10), pp. 1682–1696. https://doi.org/10.1038/s41562-023-01648-z



Stokes et al., 2020 – Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., … Collins, J.J. "A Deep Learning Approach to Antibiotic Discovery", *Cell*, 2020, no. 180 (4), pp. 688-702, e13. https://doi.org/10.1016/j.cell.2020.01.021

Stoliarova, 2025 – Stoliarova, O. "Techno-Contexts and the Birth of Novelty: Questioning the AI on Hermeneutics", *Technology and Language*, 2025, no. 6 (2), pp. 151–160. https://doi.org/10.48417/TECHNOLANG.2025.02.13

Sullivan, 2022 – Sullivan, E. "Understanding from Machine Learning Models", *British Journal for the Philosophy of Science*, 2022, no. 73 (1), pp. 109–133. https://doi.org/10.1093/bjps/axz035

Thagard, 1988 – Thagard, P. *Computational Philosophy of Science*. Cambridge, MA: MIT Press, 1988, 240 pp.

Tollefsen, 2015 – Tollefsen, D. *Groups as Agents*. Cambridge: Polity Press, 2015, 167 pp.

Vvedenskaya, E.V. "Kritika iskusstvennogo intellekta: Filosofskaya optika" [Critique of Artificial Intelligence: A Philosophical Optics], *Tekhnologos*, 2025, no. 1, pp. 87–97. https://doi.org/10.15593/perm.kipf/2025.1.07 (In Russian)

Wang et al., 2023 – Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., … Zitnik, M. "Scientific Discovery in the Age of Artificial Intelligence", *Nature*, 2023, no. 620 (7972), pp. 47–60. https://doi.org/10.1038/s41586-023-06221-2

Yashchenko, Makarov, 2022 – Yashchenko, A., Makarov, I. "Knowledge Retrieval and Relation Mining from Tolkien's History of Middle Earth", in: *2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI–MACRo)*. New York: IEEE, 2022, pp. 257–264. https://doi.org/10.1109/CINTI-MACRo57952.2022.10029506

Zednik, 2021 – Zednik, C. "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence", *Philosophy & Technology*, 2021, no. 34 (2), pp. 265–288. https://doi.org/10.1007/s13347-019-00382-7

Zubkovskaya, A.A. "Iskusstvennyj intellekt i nauchnaya ob"ektivnost'" [Artificial Intelligence and Scientific Objectivity], *Naukovedcheskie issledovaniya*, 2024, no. 2, pp. 42–49. (In Russian)