

ОНТОЛОГИЯ И ТЕОРИЯ ПОЗНАНИЯ

УДК 004.8

DOI: 10.25991/3033-604X.2026.37.83.001

Н. В. Крестиненко*

НАДЕЖНОСТЬ И НЕНАДЕЖНОСТЬ НЕЙРОСЕТЕЙ. ВЕРИФИКАЦИЯ ЛЖИ И ПРАВДЫ В ГУМАНИТАРНОЙ ОБЛАСТИ ЗНАНИЯ

***Аннотация:** В статье анализируется влияние технологий искусственного интеллекта (ИИ) на гуманитарные дисциплины. Особое внимание уделяется философским основаниям интереса к ИИ, историческим предпосылкам его появления и ключевым концепциям (включая позиции Н. Хомского и Б. Латура). Рассматриваются современные практики применения ИИ в переводе текстов, образовании, промышленности и гуманитарных исследованиях с конкретными примерами (например, системы машинного перевода DeepL и Google Translate, чат-боты и системы компьютерного зрения). Перечисляются преимущества ИИ (автоматизация рутинных задач, адаптивность, повышение эффективности), и иллюстрируются конкретными случаями (персонализация обучения, мгновенная поддержка студентов, контроль качества продукции). Затем обсуждаются методологические и этические проблемы внедрения ИИ в гуманитарной сфере: вопросы достоверности и правдивости генерируемого контента, идеологической однобокости, исторической фальсификации и потери оригинального авторского вклада. Анализируется заявление В. В. Путина, высказанное на съезде Русского географического общества (23.10.2025), о необходимости создания «суверенных моделей ИИ» на основе национальных культурно-исторических ценностей. На основе позиций современных исследователей делается вывод о том, что ИИ в гуманитарных науках должен использоваться осторожно, под контролем профессионалов, как инструмент, а не в качестве «замены» экспертов. Заключение*

* Крестиненко Нина Вячеславовна — ассистент кафедры теории и истории культуры РГПУ им. А. И. Герцена; krestinenko@bk.ru

Krestinenko Nina Vyacheslavovna — Assistant at the Department of Theory and History of Culture, Herzen State Pedagogical University of Russia; krestinenko@bk.ru

подчеркивает необходимость методологического осмысления ИИ и поиска синтеза между возможностями машинных систем и экспертным сообществом. Вопрос о создании «суверенного ИИ», ориентированного на национальные интересы, предлагается рассматривать как перспективную тему дальнейших исследований.

Abstract: *This article examines the impact of artificial intelligence (AI) technologies on the humanities. It foregrounds the philosophical foundations of interest in AI, outlines historical and conceptual premises, and contrasts critical positions (e. g., Noam Chomsky) with actor-network-inspired approaches (e. g., Bruno Latour). The paper surveys contemporary applications of AI in machine translation, education, industry and humanities research, highlights concrete benefits (automation of routine tasks, adaptive learning, increased efficiency), and documents methodological and ethical challenges: the veracity of AI-generated content, ideological bias, risks of historical falsification and threats to authorship and expert judgment. A recent presidential appeal for “sovereign AI” grounded in national culture is analyzed as a policy imperative. The article argues that AI should be employed in the humanities cautiously, under professional oversight, and as a tool subordinated to expert communities. It calls for dedicated methodological work and proposes the question of “sovereign AI” as a priority for future research.*

Ключевые слова: *искусственный интеллект, гуманитарные науки, философия науки, машинный перевод, образование, этика, историческая правда, суверенный ИИ.*

Keywords: *artificial intelligence; humanities; philosophy of science; machine translation; education; ethics; historical truth; sovereign AI; actor-network theory; AI governance.*

Введение

Интерес философов и гуманитариев к искусственному интеллекту объясняется фундаментальными вопросами о природе разума, культуры и исторического знания. Современный информационный мир порождает технологии, которые некогда были сюжетом научной фантастики. Так, появление генеративных нейросетей (например, ChatGPT) создало проблемную ситуацию для науки и образования: алгоритм способен мгновенно генерировать тексты и фабулы по любому запросу. Это породило этическую дилемму: можно ли и как использовать ИИ в научных исследованиях и учебном процессе, сохраняя научный подход и авторский замысел. Отмечается, что в 2022 г. уже возник скандал с выпускной работой, созданной ChatGPT, и в российском научном сообществе началась активная дискуссия о роли ИИ в образовании и гуманитаристике. В то же время появляются подходы к переосмыслению этой ситуации: одни критики указывают на ограниченность и «лингвистическую некачественность» подобных систем, в то время как сторонники интеграции предлагают рассматривать ИИ как элемент научной сети [4, с. 105–108]. Цель статьи — обобщить эти дискуссии в контексте гуманитарных наук. Сначала

исследуются истоки и философский замысел ИИ, оцениваются его конкретные применения в науках о человеке, а далее обсудим возникающие проблемы и ограничения. Вводная часть акцентирует философскую мотивацию интереса к ИИ как к новому феномену познания.

Появление ИИ и его философский замысел

Технологии ИИ возникают на основе математических моделей и машинного обучения, но их миссия выходит за рамки технического прогресса. Философы видят в ИИ продолжение спора о характере знания и разума. Ретроспективно можно отметить: крупнейшие мыслители XX в. (напр., Алан Тьюринг, Джон Серль) задавались вопросом, может ли машина мыслить. Сегодня нейросетевые модели не просто имитируют логику, но и генерируют тексты, изображения, даже музыку. В академическом дискурсе этот скачок осмысляется двояко. С одной стороны, консервативные критики (например, Н. Хомский в пересказе и анализе у Маляревича) трактуют ИИ как инструмент, лишенный подлинного понимания; по этому мнению, системы вроде нейросетей демонстрируют «лингвистическую некомпетентность» и не заменяют человеческое рассуждение [4, с. 102–103]. С его точки зрения, пока «неясно, кто действует»: искусственный интеллект лишь перерабатывает фрагменты данных и статистические связи. С другой стороны, в философии науки высказываются идеи о включении ИИ в научное сообщество. Так, Бруно Латур предлагал рассматривать науку не как монополию рационального субъекта, а как сеть акторов (людей, инструментов, идей), взаимодействующих друг с другом; с этой точки зрения нейросеть перестает быть «чужим телом» и становится участником дискурса. Латур замечал, что понятие «актор» снимает простую бинарность «субъект — объект» и требует учета сетевых взаимодействий; следовательно, нейросети можно включать в акторно-сетевую картину научной практики. М. Д. Маляревич развивает эту идею и предлагает трактовать ИИ как «встроенный элемент» научной деятельности, который при правильной методологии может служить инструментом, а не угрозой авторству и научной этике [4, с. 108]. Современный философский замысел ИИ состоит в попытке синтеза машинного интеллекта с человеческим знанием — поиска методологий, где одна из целей — не уничтожить человека, а расширить границы познания.

Преимущества применения ИИ

Технологии ИИ уже нашли широкое практическое применение в различных областях человеческой деятельности, в том числе в гуманитарных практиках. Ниже перечислены основные направления и конкретные примеры.

Машинный перевод. Современные переводчики на базе ИИ (DeepL, Google Translate и пр.) используют нейронные сети, обученные на больших корпусах параллельных текстов; это позволяет им ближе к живому языку передавать

смысл фраз и конструкций, хотя и не без ошибок. Автоматические системы существенно упрощают работу с большими объемами текстов и ускоряют междзыковую коммуникацию в научном обмене.

Образование и гуманитарные дисциплины. В образовательном процессе ИИ способствует персонализации обучения и автоматизации рутинных операций: адаптивные платформы подбирают материалы и задания по уровню студента, а чат-боты обеспечивают оперативную поддержку и консультации. Это расширяет возможности самообразования и снижает нагрузку на преподавателей, освобождая их время для глубокой методологической работы и дискуссий со студентами. При этом отмечается, что автоматизированное оценивание требует доработки и всегда должно сопровождаться экспертной проверкой [1, с. 60].

Промышленность и производство. ИИ-модули в виде систем компьютерного зрения и предиктивной аналитики уже применяются для контроля качества, мониторинга процессов и оптимизации ресурсов; на практике это приводит к снижению брака и увеличению эффективности производства. Эти достижения технической сферы косвенно влияют и на гуманитарные профили: например, специалисты по технике безопасности и организации труда получают дополнительные инструменты для анализа и предотвращения рисков [4, с. 102–103].

Гуманитарные и лингвистические исследования. Нейросети позволяют обрабатывать большие корпуса текстов, классифицировать темы, анализировать стилистику и эмоциональную окраску, распознавать рукописные документы и обрабатывать архивы [1, с. 59; 4, с. 104]. Это расширяет исследовательские возможности филологов и историков и открывает новые методы эмпирического исследования культурных массивов. Однако важно, чтобы такие инструменты использовались как вспомогательная платформа под контролем эксперта [6]. Применение ИИ в гуманитарном контексте обладает значительным потенциалом: персонализация обучения, ускорение коммуникаций и расширение исследовательских инструментов — все это реально достижимо при грамотной интеграции. Одновременно сохраняются ограничения: языковые модели не всегда учитывают контекст и культурные нюансы, а предвзятость обучающей выборки может воспроизводить искажения [2].

Постановка проблемы применения ИИ в гуманитаристике

Несмотря на преимущества, внедрение ИИ в гуманитарные науки сопряжено с серьезными проблемами. Главный вопрос — достоверность и правдивость генерируемого контента. Нейросети, обученные на тенденциозных данных, могут выдавать искаженную версию исторических событий: это представляет собой риск инструментальной фальсификации памяти и истории [3]. Подобные риски особенно опасны в контексте формирования мировоззренческих позиций и образовательного контента, где историческая корректность имеет ключевое значение.

Другая проблема — идеологическая зависимость и влияние на мировоззрение. Поскольку модели отражают те культурные и информационные массивы, на которых они обучены, существует риск навязывания готовых смыслов и ценностных установок. Без опоры на национальную культурно-историческую традицию создается риск «мировоззренческой зависимости» от чуждых алгоритмических решений [5; 4, с. 106–108]. Кроме того, возникают методологические и этические вопросы: кто отвечает за ошибки или фальсификации, как проверять источники, как сохранять авторство и оригинальность? Автоматически сгенерированные материалы могут содержать «галлюцинации» — убедительные, но ложные факты или несуществующие ссылки, что вводит в заблуждение исследователей и студентов [4, с. 103–104; 2, с. 69–71].

Краткий перечень ключевых рисков:

Достоверность информации: генеративные модели способны «выдумывать» факты и ссылки [4, с. 103–104].

Идеологическая однобокость: ограниченный или предвзятый тренировочный корпус приводит к систематическим искажениям исторического нарратива [3].

Фальсификация истории: автоматизация нарративов может использоваться в целевых информационных кампаниях [там же].

Риск потери экспертности: если аналитическая нагрузка полностью перенесется на ИИ, снизится качество критического мышления и научного анализа [6].

Анализ высказывания В. В. Путина на съезде РГО

Особое значение для формирования приоритетов в сфере ИИ имеет заявление президента РФ, сделанное на съезде Русского географического общества 23 октября 2025 г. В. В. Путин подчеркнул необходимость создания «суверенных моделей искусственного интеллекта», основанных на российских культурных и исторических ценностях. Дословно он заявил:

«Только опираясь на нашу культуру, историю, языковое богатство, традиции, традиционные ценности, мы сможем создать именно суверенные модели искусственного интеллекта, а не копировать чужие решения... Иначе это приводит к утрате суверенитета» [5].

Это высказывание содержит важные ориентиры. Во-первых, подчеркивается необходимость национальной специфики при разработке ИИ-систем: языковая, культурная и историческая специфика должны быть учтены в алгоритмических решениях. Во-вторых, акцент на «традиционных ценностях» означает, что ИИ-решения не должны противоречить основополагающим общественным идеалам, поскольку в противном случае искусственное «мировоззрение» окажется чуждым населению и способно подрывать культурную целостность [5; 4, с. 106–108].

Для гуманитарных наук это означает: при создании образовательных, аналитических и архивных систем ИИ необходимо обеспечить их согласование с национальной методологией и стандартами критической проверки. Такой подход представляет собой сочетание технологической независимости и ценностной ориентированности.

Необходимость осторожного применения ИИ в гуманитарной сфере

Современные исследователи сходятся на том, что ИИ в гуманитаристике должен интегрироваться с осторожностью и под контролем профессионалов. М. Д. Маляревич аргументирует, что нейросеть должна рассматриваться как один из акторов научного процесса, а ее использование — как инструмент, требующий человеческого надзора; только так сохраняется авторская и методологическая ответственность [4, с. 105–108]. Управляемая интеграция превращает ИИ из опасного «автора» в эффективный инструмент исследования.

Педагогическая аналитика показывает: при корректной настройке ИИ способствует персонализации и снижению нагрузки на преподавателя, однако требует присутствия эксперта в оценочной цепочке [1, с. 57–60]. Этические исследования подчеркивают риск предвзятости алгоритмов и необходимость прозрачности в механизмах принятия решений [2, с. 67–69]. Кроме того, практики по цифровой лингвистике указывают: основная задача специалиста — сохранить экспертность и научить ИИ работать по заданной методике, а не отдать анализ полностью машине [6]. Исходя из этого, формулируется практическое требование: любые ИИ-приложения в гуманитарных дисциплинах должны проходить экспертизу профессионального сообщества перед внедрением, а их результаты — подвергаться обязательной экспертной проверке. Это касается и образовательных продуктов, и научных публикаций, и цифровых архивов [1, с. 57–60; 2, с. 67–75].

Не менее важным элементом осторожного внедрения является введение комплексных процедур верификации и аудита моделей: ввод обязательных протоколов прозрачности (explainability), учета происхождения данных (data provenance) и ведения аудиторских следов для всех алгоритмических решений. Такие процедуры позволят установить, какие корпуса использовались для обучения, какие фильтры и корректировки применялись и кто несет ответственность за ошибочные или предвзятые выводы модели. Практика «человека в петле» (human-in-the-loop) при принятии окончательных решений должна стать обязательной для всех применений ИИ в учебных и исследовательских продуктах; это снизит риск автоматизированной фальсификации и сохранит за экспертом контроль над критически важными выводами [2, с. 67–69; 4, с. 105–108].

Следующим шагом должна стать институционализация профессионального надзора: создание междисциплинарных экспертных комиссий и сертификационных программ, которые будут оценивать готовые ИИ-решения для

гуманитарной сферы и давать рекомендации по их использованию. В состав таких комиссий логично включить историков, филологов, философов науки, специалистов по этике и IT-аудиту; их задача — не только проверять текущее состояние моделей, но и формировать методологические стандарты и учебные программы по цифровой грамотности для преподавателей и студентов. Практика пилотных внедрений с последующим экспертным ревью позволит развивать доверие к инструментам и одновременно повышать профессиональную компетентность пользователей [1, с. 57–60; 6].

Необходимо законодательное и организационное усиление мер по защите исторической правды и культурного кода: формирование «суверенных» корпусов данных, стандартизованных для обучения отечественных моделей, строгие требования к сохранению целостности архивов и санкции за манипуляции с историческим содержанием. Такая политика обеспечит технологическую независимость и культурную совместимость алгоритмов с национальными ценностями, сведя к минимуму риск внешнего информационного влияния. Защита памяти Отечества — это не только научная, но и гражданская обязанность; потому развитие ИИ следует вести так, чтобы он служил укреплению исторической истины и национального единства, а не ее подмене [3; 5].

Заключение

Искусственный интеллект уже стал неотъемлемой частью гуманитарного пространства, однако его роль требует осознанного методологического осмысления и институциональной регламентации. Необходимы специализированные методики работы с ИИ, междисциплинарные стандарты и этические нормы, позволяющие объединить опыт экспертов и вычислительные возможности машин [4, с. 108; 1, с. 60]. Идея «суверенного ИИ», о которой заявил президент, представляет собой отдельную перспективную тему: как создавать национальные модели ИИ, опираясь на отечественные научные традиции и культурный код; какие институты и кадровые ресурсы нужны для этого — все это требует отдельного междисциплинарного исследования [5]. Принципиальная позиция автора: ИИ в гуманитарных науках применим и полезен, но только как инструмент под контролем профессионального сообщества — сочетание технических возможностей и экспертной ответственности позволит сохранить историческую правду, культурную целостность и научную добросовестность [1, с. 60; 4, с. 108; 6].

Дальнейшее развитие ИИ в гуманитарной сфере должно опираться на три взаимосвязанные позиции: (1) строгий профессиональный контроль и экспертиза на всех этапах работы с данными и выводами моделей; (2) системная институциональная поддержка — сертификация, междисциплинарные комиссии и образовательные программы для повышения цифровой грамотности экспертов; (3) государственная стратегия по созданию суверенных, национально ориентированных корпусов данных и стандартов, гарантирующих защиту

исторической правды и культурного кода. Только сочетание ответственной науки, экспертного надзора и государственно-направленной политики даст возможность использовать мощь ИИ во благо общества, не допуская искажений памяти Отечества и утраты научной добросовестности.

ЛИТЕРАТУРА

1. *Везетиу Е. В.* Использование искусственного интеллекта в преподавании гуманитарных дисциплин // Проблемы современного педагогического образования. 2025. № 86. Ч. 1. С. 57–59.
2. *Докучаев И. И.* Феноменология знака. Избранные работы по семиотике и диалогике культуры. М.: Наука, 2010. 412 с.
3. Искусственный интеллект vs историческая правда: как новые технологии угрожают памяти о войне. Консорциум исследований безопасности ИИ. [Электронный ресурс] URL: <https://trust-ai.ru/news/otraslevye-novosti/iskusstvennyy-intellekt-vs-istoricheskaya-pravda-kak-novye-tekhnologii-ugrozhayut-pamyati-o-voyne/> (дата обращения: 30.10.2025).
4. *Маляревич Д. В.* ChatGPT и наука: взаимодействие научного сообщества и искусственного интеллекта // Векторы благополучия: экономика и социум. 2024. Т. 52. № 2. С. 99–109.
5. Путин призвал не копировать чужие решения в сфере ИИ // РИА Новости, 23.10.2025 [Электронный ресурс] URL: <https://ria.ru/20251023/putin-2050157613.html> (дата обращения: 28.10.2025).
6. *Смирнова А. Г.* Цифровые инструменты в лингвистике: как превратить ИИ в полезный инструмент, а не в конкурента // Новости Школы иностранных языков НИУ ВШЭ, 11.02.2025. [Электронный ресурс] URL: <https://lang.hse.ru/news/1015919565.html> (дата обращения: 29.10.2025).

REFERENCES

1. *Vezetiu, E. V.* *Ispol'zovanie iskusstvennogo intellekta v prepodavanii gumanitarnykh distsiplin* [The Use of Artificial Intelligence in Teaching Humanities]. In: *Problemy sovremennogo pedagogicheskogo obrazovaniia*, 2025, no. 86, part 1, pp. 57–59. (In Russian).
2. *Dokuchaev, I. I.* *Fenomenologiya znaka. Izbrannye raboty po semiotike i dialogike kul'tury* [Phenomenology of the Sign. Selected Works on Semiotics and Dialogics of Culture]. Moscow, Nauka Publ., 2010. 412 p. (In Russian).
3. *Iskusstvennyi intellekt vs istoricheskaya pravda: kak novye tekhnologii ugrozhayut pamiati o voine* [Artificial Intelligence vs Historical Truth: How New Technologies Threaten the Memory of War]. *Konsortsiy isledovaniy bezopasnosti II*. 2025. [Electronic resource]. URL: <https://trust-ai.ru/news/otraslevye-novosti/iskusstvennyy-intellekt-vs-istoricheskaya-pravda-kak-novye-tekhnologii-ugrozhayut-pamyati-o-voyne/> (accessed: 30.10.2025). (In Russian).
4. *Maliarevich, D. V.* *ChatGPT i nauka: vzaimodeistvie nauchnogo soobshchestva i iskusstvennogo intellekta* [ChatGPT and Science: Interaction between the Scientific Community and Artificial Intelligence]. In: *Vektory blagopoluchii: ekonomika i sotsium*, 2024, vol. 52, no. 2, pp. 99–109. (In Russian).
5. *Putin prizval ne kopirovat' chuzhie resheniia v sfere II* [Putin Urged Not to Copy Foreign Solutions in the Field of AI]. *RIA Novosti*, October 23, 2025. [Electronic resource]. URL: <https://ria.ru/20251023/putin-2050157613.html> (accessed: 28.10.2025). (In Russian).

6. Smirnova, A. G. *Tsifrovye instrumenty v lingvistike: kak prevratit' II v poleznyi instrument, a ne v konkurenta* [Digital Tools in Linguistics: How to Turn AI into a Useful Tool Rather Than a Competitor]. *Novosti Shkoly inostrannykh iazykov NIU VShE, February 11, 2025*. [Electronic resource]. URL: <https://lang.hse.ru/news/1015919565.html> (accessed: 29.10.2025). (In Russian).