Секция «Онтология и теория познания»

Цифровая эпистемология: проблемы и перспективы в эпоху искусственного интеллекта

Научный руководитель – Даниелян Наира Владимировна

Гершунин Сергей Аркадьевич

Аспирант

Национальный исследовательский университет «МИЭТ», Москва, Россия

E-mail: finbaricus@yandex.ru

В современном мире искусственный интеллект (ИИ) все более широко применяется в производстве знаний. Он позволяет трансформировать подходы к генерации, обработке и использованию информации в различных профессиональных областях [7]. Однако работа систем ИИ в части генерации новых знаний определяется «непрозрачными» для пользователей алгоритмами в отличие от традиционных методов, основанных на человеческой интуиции в устоявшихся эпистемологических рамках, что естественно вызывает недоверие к получаемым знаниям и требует новых способов их проверки.

С приходом знаний, генерируемых ИИ, традиционная эпистемология нуждается в дополнении. Исторически эпистемология была сосредоточена на вопросах, касающихся сущности знания, того, как оно приобретается и при каких условиях может считаться достоверным [1]. В контексте ИИ эти вопросы приобретают новое измерение. К примеру, модели машинного обучения зачастую приходят к определенным выводам на основании исходных данных без явных констатаций причин и предпосылок, на основании которых данные выводы были получены [8]. Эта неявность получения результатов бросает вызов традиционным критериям оценки знаний, таким как ясность, логическая согласованность и эмпирическая проверяемость. Кроме того, системы ИИ также подвержены искажениям в силу ограниченности данных, предоставляемым им для обработки, что вызывает опасения в части справедливости и всеохватности получаемых ими результатов [6]. Для решения данных проблем представляется необходимым руководствоваться особой теоретико-познавательной установкой, ориентированной на учет уникальных характеристик знаний, генерируемых ИИ, в совокупности с базовыми эпистемологическими принципами надежности и достоверности знания. Сегодня данная установка находит отражение в концепции «цифровой эпистемологии» [5], которая представляется не просто очередной абстрактной философской разработкой, но отражает практическую потребность современного общества в пересмотре критериев проверки истинности получаемых знаний посредством систем ИИ. Например, в сфере здравоохранения уже используются для выявления заболеваний по медицинским изображениям и историям болезни диагностические инструменты, создаваемые на базе ИИ. Хотя такие системы и могут повысить точность диагностики, они также подвержены риску ошибочного диагноза в силу несовершенства алгоритмов и искаженности данных для машинного обучения [8]. В результате отсутствие ясных критериев оценки получаемых ИИ результатов в цифровых системах, участвующих в производстве знаний, подрывает доверие к таким системам и нивелирует их потенциальные преимущества.

Важнейшими для традиционной эпистемологии являются вопросы о достоверности знания, которая обычно заключается в степени логического обоснования и эмпирической проверяемости результатов. Исторически в эпистемологии подчеркивается связь между рационализмом, который опирается на логические суждения, и эмпиризмом, основанном на чувственных данных. Оба этих философских течения служат основой для оценки достоверности и надежности знаний в различных дисциплинах [3]. В традиционных системах знание оценивается посредством изучения его согласованности, соответствия реальности

и воспроизводимости. При этом предполагается, что оно создается в прозрачных и наблюдаемых процессах под контролем человека. Однако появление ИИ существенно меняет подходы к созданию и проверке знаний, изучением которых занимается цифровая эпистемология. Данное направление стремится адаптировать классические философские принципы познания к новым вызовам в эпоху ИИ. Его цель состоит в разработке методов и стандартов для оценки достоверности и надёжности знаний, получаемых с использованием ИИ в условиях цифровой среды. Основными задачами цифровой эпистемологии являются анализ влияния алгоритмов обработки больших данных на формирование «машинных знаний» [2], разработка подходов к их верификации, а также разрешение проблемы интерпретируемости получаемых системами ИИ результатов.

Таким образом, «цифровая эпистемология» сосредоточена на исследовании специфических характеристик систем ИИ. Прежде всего, к ним относятся: высокая степень зависимости от больших данных, играющих главенствующую роль в обучении и функционировании ИИ, вероятностный характер выводов систем ИИ и, наконец, сами алгоритмы машинного обучения, которые зачастую могут не соответствовать традиционным процедурам верификации знаний, принятым в классической эпистемологии [4]. Таким образом, ключевой здесь является проблема прояснения «непрозрачности» получения знаний, так как подавляющее число систем ИИ представляют собой «черные ящики», в которых внутренние механизмы их работы остаются не всегда понятными даже для их разработчиков.

Подводя итог, следует отметить, что принципиальное отсутствие прозрачности в работе систем ИИ значительно усложняет процесс оценки их результатов. При этом традиционные познавательные методы, такие как дедуктивное мышление и опытное подтверждение, оказываются совершенно непригодными для анализа и проверки выдаваемых ИИ знаний, что, в свою очередь, не позволяет проследить их логическую или эмпирическую основу. Поэтому традиционные эпистемологические подходы к оценке достоверности и надежности знания должны быть дополнены с учетом новых реалий, так чтобы стало возможным глубже понять природу познавательных процессов, совершаемых при использовании систем ИИ. На решении этой перспективной задачи и сосредоточена «цифровая эпистемология».

Источники и литература

1) Audi R. Epistemology. New York: Routledge, 2010. 432 p.
2) Bai H. The Epistemology of Machine Learning // Filosofija. Sociologija. 2022. Vol. 33, № 1. P. 40-48.
3) Fleisher W. Understanding, Idealization, and Explainable AI // Episteme. 2022. Vol. 19, № 4. P. 534-560.
4) Floridi L. The Logic of Information: A Theory of Philosophy as Conceptual Design. Oxford: Oxford University Press, 2019. 272 p.
5) Mahbubi M. Digital Epistemology: Evaluating the Credibility of Knowledge Generated by AI: 1 // YUDHISTIRA: Journal of Philoshopy. 2025. Vol. 1, № 1. P. 8-18.
6) O'Neil C. Weapons of math destruction: how big data increases inequality and threatens democracy. New York: Crown, 2016. 272 p.
7) Russo F., Schliesser E., Wagemans J. Connecting ethics and epistemology of AI // AI & Society: Knowledge, Culture and Communication. 2024. Vol. 39, № 4. P. 1585-1603.
8) Goodfellow I., Bengio Y., Courville A. Deep Learning. URL: https://www.deeplearningbook.org (дата обращения: 03.03.2025).