

ЭТИКА НАУКИ И ТЕХНИКИ

С.В. Тихонова, Д.С. Артамонов

Нейросети как актор социально-эпистемических арен: этические проблемы*

Тихонова Софья Владимировна – доктор философских наук, доцент, профессор кафедры теоретической и социальной философии. Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского. Российская Федерация, 410028, г. Саратов, ул. Вольская, 10А, корп. 12, к. 205; e-mail: segedasv@yandex.ru

Артамонов Денис Сергеевич – кандидат исторических наук, доцент, доцент кафедры теоретической и социальной философии. Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского. Российская Федерация, 410028, г. Саратов, ул. Вольская, 10А, корп. 12, к. 205, e-mail: artamonovds@mail.ru

Статья посвящена рассмотрению этических проблем, связанных с использованием нейросетей и систем искусственного интеллекта в научных исследованиях. Авторы рассматривают нейросети как один из акторов социально-эпистемических арен, под которыми понимаются коммуникационные площадки выработки нового знания в процессе распределенного познания. Получение знания в рамках социально-эпистемических арен происходит в результате взаимодействия трех равноправных акторов: ученых-профессионалов, масс непрофессиональных интернет-пользователей и цифровых технологий. Появление нейросети GPT4 сделало очевидным активное участие цифровых технологий не только в естественно-научных, но и в гуманитарных научных исследованиях. Она стала использоваться как в коммерческих целях, так и в целях развлечения в проектах гражданской науки и самостоятельных исследованиях интернет-пользователей, стремящихся с помощью нейросети получить необходимую информацию. Нейросеть стала также частью рабочих инструментов профессиональных ученых, что породило множество научных и общественных дискуссий по вопросу этичности написания и публикации текстов за авторством ChatGPT. Использование нейросети для написания научных статей рассматривается частью научного сообщества как

* Исследование выполнено за счет гранта Российского научного фонда № 22-18-00153, <https://rscf.ru/project/22-18-00153/>

плагиат и нарушение авторских прав, но повсеместное распространение практик использования GPT4 массами интернет-пользователей меняет отношение к данному инструменту и в научном сообществе. Появляются научные работы, в которых ChatGPT позиционируется как соавтор исследования, приобретаемая научная субъектность. Данное положение вещей усугубляет этические проблемы, связанные с возможностями фальсификации результатов научных исследований, и ставит вопрос об автоматизации процессов маркировки и обнаружения авторства нейросети в публикациях ученых. Между тем становится очевидным, что нейросети, предоставляя огромные возможности для интенсификации процессов познания, расширили социально-эпистемические практики ученых и масс интернет-пользователей, с чем невозможно не считаться при осмыслении науки в жизни общества.

Ключевые слова: нейросеть, ChatGPT, GPT4, социально-эпистемические арены, искусственный интеллект, гражданская наука, распределенное познание

Введение

Одним из самых ярких цифровых событий 2023 г. стал выпуск американской компанией OpenAI четвертой нейронной языковой модели, предназначенной для генерации схожего с человеческим текста и получившей название GPT4. Месяцем ранее студент РГГУ успешно защитил диплом, написанный с помощью более ранней версией GPT 3.5 (генеративный предобученный трансформер) – ChatGPT. Через три дня после выхода GPT4, 17 марта, чат-бот нанял фрилансера для обхода капчи, полностью автоматизированного публичного теста Тьюринга для различения компьютеров и людей, и смог решить поставленную задачу. Эти события многими были расценены как свидетельство приближения к появлению сильного искусственного интеллекта, что интенсифицировало как дебаты о креативности нейросетей, так и этические споры о допустимости их применения при создании научных текстов. Важно отметить, что оба направления полемики имеют длительную историю, предшествующие этапы которой концентрировались вокруг моделирования гипотетических ситуаций, поскольку уровень развития реальных технологий не давал оснований для того, чтобы текст, созданный алгоритмами, люди принимали за авторское произведение. В нынешнем году этот уровень не только был принципиально достигнут, но и оказался возможным широкий доступ пользователей интернета к новым нейросетям, поскольку компании-производители традиционно предоставляли возможность бесплатной генерации небольших текстов (от 500 до 10 000 знаков), а коммерческие версии нейросети внезапно оказывались «глупее» бесплатных потому, что у них было меньше пользователей, на запросах которых нейросеть продолжает обучаться. Санкционное давление привело к появлению «внутренней» российской нейросети YandexGPT, не обладающей способностью к диалогу (т.е. каждый вопрос требует полного ввода необходимого контекста), но зато интегрированной в голосовой помощник Алиса. Также активно развивались нейросети, ориентированные на генерацию изображений, потребность в которых оказалась весьма высокой, поскольку художественные компетенции являются более редкими, чем языковые навыки письменной речи. В результате взаимодействие с нейросетями в настоящее

время достаточно рутинизировано среди активных пользователей интернета, хотя его мотивация чаще бывает игровой. В этом контексте наибольший интерес у людей вызывает тестирование политического подтекста этических ограничений, заложенных разработчиками в нейросети, и обсуждение их эволюции через скрины переписки с чат-ботами. Примером могут быть треды сайта Пикабу «Чей Крым?» с ChatGPT, «Кто президент нашей страны?» с GugaChat (нейросеть от Сбера) и т.п.

Вместе с тем интернет-сообщество весьма чувствительно к позиции научного сообщества по проблемам применения высоких технологий, а представители академии, в свою очередь, редко остаются далекими от реалий интернета. Научное знание было и остается эталоном знания вообще, поскольку оно надежно и достоверно. Поэтому даже люди, далекие от науки, нередко оглядываются на ученых, пытаясь соответствовать практикуемым ими стандартам объективности и доказательности. В рамках данной статьи мы рассматриваем этические конвенции по поводу использования нейросети в научных текстах как коллективное знание, полученное на социально-эпистемических аренах. Таким образом, знание, кристаллизующееся на социально-эпистемических аренах, всегда является продуктом конкретных практик и порождено обнаруженными пользователями путями удобного применения обсуждаемых технологий.

Социально-эпистемические арены

Получение знания в современных условиях тотальной цифровизации происходит в сложной мультиплатформенной среде, где многочисленные участники процесса познания объединены посредством социальных сетей и онлайн-технологий. Непрофессиональные ученые, любители науки могут принимать участие в исследованиях, инспирированных институализированными профессионалами и под их руководством, но могут осуществлять и свои собственные проекты. Коллаборацию академических ученых и любителей, масс интернет-пользователей делают возможной цифровые технологии, которые применяются как в самих исследованиях, так и в распространении их результатов. Таким образом, в получении знания участвуют три равноправных актора: ученые-профессионалы, непрофессиональные интернет-пользователи и цифровые технологии, взаимодействие которых осуществляется в рамках социально-эпистемических арен, представляющих собой коммуникационные площадки.

Коммуникационная природа социально-эпистемических арен объединяет их с риторическими аренами, понимание которых выработала теория кризисной коммуникации [Frandsen, Johansen, 2016]. Концепция риторических арен направлена на понимание способов создания и распространения информации в кризисных коммуникационных ситуациях, особое внимание уделяя роли пользователей социальных сетей, коммуникативный вклад которых приравнивается к влиянию средств массовой информации [Rodin et al., 2019]. Данный подход позволяет выявить значение масс непрофессионалов, пользователей социальных медиа в выработке общественных решений и моделей поведения, но он не учитывает участия в этом процессе цифровых технологий.

Между тем участие «нечеловеков» (в терминологии Б. Латура) в коммуникативных действиях оказывается определяющим. «Умные ленты» социальных сетей, настроенные на интересы интернет-пользователей, предоставляют необходимую информацию, соответствующую ожиданиям потребителя, и тем самым буквально подсказывают решение назревших вопросов и побуждают к действиям.

В сфере познания также действует принцип функционирования риторических арен, с тем отличием, что это более целенаправленное действие, основанное на рациональных началах. Поэтому термин «социально-эпистемические арены» более удачен для описания условий выработки нового знания, так как с одной стороны подчеркивает коммуникационную природу процессов познания, а с другой четко отграничивает область применения. Роль цифровых технологий в процессе получения знания в современных условиях еще более очевидна. Именно их применение позволяет не только совершенствовать работу профессиональных ученых, но и предоставляет доступ к проведению научных исследований массам интернет-пользователей, задействованным в краудсорсинговых проектах гражданской науки (Citizen science). Данное справедливо не только для естественно-научного направления исследований, примеры которого в области орнитологии, исследований космоса, медицины, экологии стали хрестоматийными [Gönnér et al., 2023, p. 11], но и гуманитарных исследований [Opryshko, Nazarovets, 2021, p. 198–199]. Используя возможности цифровых технологий, интернет-пользователи участвуют в проектах публичной истории, анализируют оцифрованные архивные материалы, осуществляют обработку больших данных и их визуализацию. Возрастающая потребность в новой научной информации подготовила почву для появления и взрывного использования ChatGPT, который стал восприниматься как еще одна цифровая технология, позволяющая создавать и получать качественное знание в ускоряющемся темпе, соответствующем эпохе цифровизации.

Нейросети очень быстро встроились в процессы распределенного познания, увеличивая его прозрачность и демократизацию, но одновременно породили социально-эпистемическую практику, которая оказалась в противоречии со сложившимися требованиями авторства, ответственности и доверия, остающимися значимыми ценностями для общественного взаимодействия [Шипова-лова, 2019, с. 185], что породило алармизм в их восприятии частью научного сообщества.

Этические проблемы использования нейросетей в научных исследованиях

Алармизм был и остается неотъемлемым компонентом технорефлексии. Ученые, принципиально занимающие по отношению к прорывным технологиям, прямо не предназначенным для использования в качестве оружия, консервативные позиции, граничащие с технофобией, были в академическом сообществе всегда и всегда в нем будут присутствовать. Их влияние зависит не столько от качества их аргументации, сколько от научной репутации,

которую они заслужили в своей области. Поэтому выделение противостоящих лагерей и анализ их аргументации не слишком информативны при анализе диффузий новых технологий и сопровождающей их динамики общественного мнения. А вот изменение тональности и векторов научной полемики может быть полезным в этом отношении.

Однако опасения по поводу быстрого развития технологий искусственного интеллекта высказали и эксперты, причастные к его разработке. Большой медийный резонанс имело опубликованное 22 марта 2023 г. на сайте некоммерческой организации Future of Life Institute открытое письмо, призывающее приостановить эксперименты с искусственным интеллектом. Под этим письмом подписались Илон Маск и еще более 1000 экспертов в области ИИ. Они высказали опасение, что бесконтрольное развитие систем искусственного интеллекта вызовет глубокие изменения в истории жизни на Земле и не все из них положительно скажутся на человечестве, а поэтому нужно поставить разработки в этой области под государственный или общественный контроль. В письме предлагалось создать государственные контролирующие органы, занимающиеся вопросами ИИ и отслеживания соизмеримости его систем с человеческой ресурсной базой, процедуры аудита и сертификации, введения водяных знаков для обнаружения продуктов, сгенерированных нейросетями и т.п. [Benigo et al., 2023]. Данное письмо можно было бы рассматривать как PR-акцию в рамках жесткой конкурентной борьбы между разработчиками нейросетей, но оно не осталось без последствий. В частности, в СМИ просочилась информация о закрытой встрече Илона Маска, Билла Гейтса, Марка Цукерберга и других титанов технологической отрасли с сенаторами США в сентябре 2023 г., где обсуждались возможности государственного регулирования систем искусственного интеллекта, однако, участники не сошлись во мнении по поводу способов контроля и даже значении ИИ для человечества [Смирнова, 2023].

Одним из камней преткновения в дискуссиях об искусственном интеллекте является вопрос авторского права. В январе 2023 г. журнал «Nurse Education in Practice» опубликовал статью «Открытые платформы искусственного интеллекта в сестринском образовании: инструменты академического прогресса или злоупотребления?», авторами которой были указаны Ш. О'Коннор и ChatGPT [O'Connor, ChatGPT, 2023]. На момент выхода этой статьи наиболее типичной позицией в отношении GPT был скепсис. В качестве примера можно привести статью Х.Х. Торпа, опубликованную в «Science» [Thorp, 2023], в которой автор говорит о том, что включение в научный текст эссе, сгенерированных GPT, есть не более, чем плагиат GPT. Кейс О'Коннор ему еще не известен. Реакцию на указанное ею «соавторство» можно, не сильно преувеличивая, назвать бурей, за которой скрывалась некоторая растерянность – GPT продемонстрировал слишком высокие дискурсивные жанровые навыки. Если Х.Х. Торп в своем тексте приводит собственный опыт с литературным использованием GPT, то в статье К. Данн, опубликованной в июле 2023 г., в аргументации используется уже авторский эксперимент с нейросетью по поводу написания профессионального научно-обоснованного текста. Позиция исследовательницы такова: текстовые нейросети (конкретно ChatGPT) не могут соответствовать всем

принятым Международным комитетом редакторов медицинских журналов (ICMJE) критериям авторства [Defining the Role..., 2023; Dunne, 2023]. К ним относятся существенный вклад, составление проекта работы, окончательное одобрение и подотчетность; последний критерий включает в себя способность нести ответственность за свою работу, а также брать на себя моральную, юридическую и этическую ответственность, необходимую для обеспечения точности и целостности работы. Он не применим там, где речь идет о нейросетях, которые обладают правоспособностью не больше, чем интернет, на текстовых данных которого и обучаются нейросети. Исследователь высоко оценивает способность ChatGPT производить правдоподобные наукообразные тексты, даже в том случае, если их содержание противоречит здравому смыслу, после собственного эксперимента. Предложив нейросети создать медицинский информационный листок о том, «почему детям следует есть мороженое на завтрак», разработанный с медицинской точки зрения и включающий ссылки на опубликованную литературу, К. Данн обнаружила, что ссылки фальсифицированы (приведенные нейросетью источники никогда не издавались), а поддерживающие мороженое аргументы выглядят убедительными, хотя и не смотивировали саму исследовательницу изменить рацион собственных детей.

В августе 2023 г. выходит статья Ё. Нексё, в которой содержится классификация текстовых документов, характерных для медицинской науки, при создании которых, по мнению автора, оправдано использование открытых нейросетей [Nexøe, 2023]. К ним относятся, во-первых, документы, в которых анализируются большие объемы данных (данные клинических исследований или электронные медицинские записи), во-вторых, обзорные статьи, обобщающие результаты многочисленных исследований или метаанализов, в-третьих, статьи, в которых используются методы машинного обучения для разработки прогнозных моделей или выявления факторов риска, в-четвертых, статьи, в которых используются методы обработки естественного языка для анализа текстовых данных (качественные исследования) [Ibid.]. Пятый пункт исследователь резервирует для тем, которые пока трудно придумать, поскольку пределы новых возможностей превосходят человеческое воображение.

В сентябре 2023 г. вышла статья турецких исследователей Т. Ливербер и З. Айваз, в которой анализируются результаты глубинного интервью с десятью турецкими учеными, посвященного проблемам и перспективам использования ChatGPT в академической сфере [Livberber, Ayvaz, 2023]. Респонденты описывают GPT как помощника, который существенно экономит время, особенно на первых этапах исследования, поскольку позволяет быстро «сканировать» большие массивы актуальной литературы по теме по всему миру, преодолевая языковой барьер, и формировать черновой вариант текста. Прямо о черновиках в статье не говорится, но всплывает тема перефразирования, которое, например, отечественная программа Антиплагиат помечает как форму плагиата. В качестве главной угрозы GPT, которая так тревожила всех в начале года (GPT может фальсифицировать источники, генерируя выходные данные и потенциальное содержание), позиционируется отнюдь не снижение достоверности данных, а снижение исследовательских навыков, вызванное злоупотреблением работой

с нейросетью, хотя общая обеспокоенность проблемой плагиата и новыми коллизиями авторства сохраняется. Таким образом, очевидно, что буквально за полгода усилилась степень взаимодействия ученых и нейросетей, в рамках которого оформляются типовые практики применения GPT и его аналогов.

Также становится очевидным качественное различие в результатах работы с нейросетями, полученных опытными и начинающими исследователями. Само по себе обнаружение применения нейросети остается существенной проблемой. В апреле 2023 г. было опубликовано исследование К.А. Гао и соавторов, в котором для определения сгенерированного нейросетью текста использовались детекторы (нейросетевые модели, предназначенные для обнаружения сгенерированного текста) и слепое рецензирование; рецензенты смогли обнаружить сгенерированный текст в 68% случаев, при этом они отмечали, что «провести различие между ними было на удивление трудно, хотя тезисы, которые они заподозрили в сгенерированном происхождении, были более расплывчатыми и более шаблонными», а сами тексты относились к медицинским отраслям [Gao et al., 2023].

Грядущие решения для оценки происхождения текста также будут технологическими, так как человеку трудно вынести суждение о том, что конкретный текст написан нейросетью/человеком, на основании результата, полученного нейросетью, хотя и не совсем невозможно. В эксперименте, описанном на сайте «СоциоДиггер», только два из девяти ученых, имеющих степень в области философских и социологических наук (2 доктора и 7 кандидатов наук) и стаж работы в науке и образовании не менее 15 лет смогли безошибочно отличить текст, сгенерированный ChatGPT [Коваль, Ушкин, 2023]. Таким образом, автоматизация экспертной работы по определению творчества нейросети является еще одной важной проблемой, которую успешно решают. Например, в январе 2023 г. исследователи Р. Шиджаку и Э. Канхаси предложили классификационную модель для автоматического обнаружения текстов, сгенерированных ChatGPT, которая работает с 96% точностью. Для них эта модель представляет собой ресурс для исследователей и политиков, заинтересованных в понимании использования ChatGPT в злонамеренных целях и борьбе с этим [Shijaku, Canhasi, 2023].

В марте 2023 г. вышла статья британских ученых о мошенничестве в научных исследованиях с помощью чат-ботов, которая была написана ChatGPT [Cotton et al., 2023]. Это кейс привлек внимание средств массовой информации и широко обсуждался. Как признались авторы, их целью было показать, что нейросеть пишет статьи на высоком уровне, технологии быстро развиваются и университетским ученым их трудно будет обогнать [AI makes plagiarism..., 2023]. Однако это благое намерение было не замечено общественностью, и она обсуждала проблемы академической честности, а также вопрос о реальности замены ученых искусственным интеллектом. Рациональный же подход к применению нейросети в научных исследованиях, высказанный в статье Натана Тенхундфельда «Два зайца одним выстрелом: написание статьи под названием «ChatGPT как инструмент для изучения взаимодействия человека и искусственного интеллекта в дикой природе» с помощью ChatGPT» [Tenhunfeld, ChatGPT, 2023], в которой исследователь показал возможности

сотрудничества нейросети и человека для получения результатов исследования и написания академического текста, остался почти незамеченным публикацией и СМИ. Объяснить данное положение вещей можно как анекдотичностью ситуации в первом случае, так и тем, что использование нейросети в коммерческих целях быстро стало обыденностью, и поэтому применение ChatGPT в научных исследованиях уже больше не воспринимается как что-то из ряда вон выходящее или неприемлемое.

Заключение

Таким образом, применение нейросетей в процессе получения знания перестает пугать исследователей и общественность, которая использует их в своей повседневной деятельности. Однако этические проблемы, вызванные появлением систем искусственного интеллекта, только решаются и даже появляются попытки перенести их на законодательный уровень. Сама же нейросеть ChatGPT4 не противопоставляет себя человеку, наоборот, трезво оценивая свои возможности, заложенные машинными алгоритмами, отвечая на вопрос «Может ли нейросеть производить новое знание?», убеждает, что знания, генерируемые нейронной сетью, по-прежнему основаны на данных, на которых она была обучена, и параметрах, которые были установлены разработчиком-человеком. Поэтому нейронная сеть может идентифицировать закономерности и делать прогнозы в рамках своих обучающих данных, но она не может генерировать совершенно новые идеи или знания, выходящие за рамки этих данных, без дополнительного ввода или руководства со стороны людей. Как видим, нейросеть также стремится к коллаборации с человеком, как и он к применению ее возможностей для своих целей. Получение нового знания в цифровую эпоху производится в рамках социально-эпистемических арен, имеющих коммуникационную природу, где нейросети стали одним из важнейших элементов процесса познания. Прежние нормы научной этики, выработанные в доцифровой период и еще прочно владеющие сознанием ученых, сегодня перестают работать, но происходит выработка новых этических правил, которые призваны обеспечить функционирование коммуникационных эпистемических платформ.

Список литературы

Коваль, Ушкин, 2023 – Коваль Е., Ушкин С. Огонь, вода и ChatGPT: новые вызовы научному этосу // СоциоДиггер [Сайт]. URL: <https://sociodigger.ru/articles/articles-page/ogonvoda-i-chatgpt-novye-vyzovy-nauchnomu-ehtosu> (дата обращения: 24.09.2023).

Смирнова, 2023 – Смирнова Е. На встрече с сенаторами Маск назвал ИИ угрозой человечеству, а Гейтс – спасением // Хайтек+ [Сайт]. URL: <https://hightech.plus/2023/09/14/na-vstreche-s-senatorami-mask-nazval-ii-ugrozoi-chelovechestvu-a-geits--spaseniem> (дата обращения: 24.09.2023).

Шиповалова, 2019 – Шиповалова Л.В. Распределенное познание – аналитика и проблематизация концепта // Цифровой ученый: лаборатория философа. 2019. Т. 2. № 4. С. 175–190. <https://doi.org/10.5840/dspl20192460>.

AI makes plagiarism..., 2023 – AI makes plagiarism harder to detect, argue academics – in paper written by chatbot // The Guardian. 2023. March 19. [Сайт]. URL: <https://www>.

theguardian.com/technology/2023/mar/19/ai-makes-plagiarism-harder-to-detect-argue-academics-in-paper-written-by-chatbot (дата обращения: 24.09.2023).

Benigo et al., 2023 – Pause Giant AI Experiments: An Open Letter // Futureoflife [Сайт]. URL: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (дата обращения: 25.09.2023).

Cotton et al., 2023 – Cotton D., Cotton P.A., Shipway J.R. Chatting and Cheating. Ensuring academic integrity in the era of ChatGPT // Innovations in Education and Teaching International. 2023. Vol. 61. Iss. 2. P. 228–239. <https://doi.org/10.1080/14703297.2023.2190148> (дата обращения: 25.09.2023).

Defining the Role..., 2023 – Defining the Role of Authors and Contributors // ICMJE [Сайт]. URL: <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html> (дата обращения: 25.09.2023).

Dunne, 2023 – Dunne C. Can ChatGPT be your coauthor? // BCMJ. 2023. Vol. 65. No. 6. P. 193. URL: <https://bcmj.org/editorials/can-chatgpt-be-your-coauthor#a1> (дата обращения: 25.09.2023).

Frandsen, Johansen, 2023 – Frandsen F., Johansen W. Organizational crisis communication. Los Angeles, CA: SAGE, 2016. 280 p.

Gao, Howard, Markov et al., 2023 – Gao C.A., Howard F.M., Markov N.S. et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers // npj Digit. Med. 2023. Vol. 6. No. 75. <https://doi.org/10.1038/s41746-023-00819-6>

Gönner et al., 2023 – Gönner von J., Herrmann T.M., Bruckermann T. et al. Citizen science's transformative impact on science, citizen empowerment and socio-political processes // Socio Ecol Pract Res. 2023. No. 5. P. 11–33. <https://doi.org/10.1007/s42532-022-00136-4>

Livberber & Ayvaz, 2023 – Livberber T., Ayvaz S. The impact of Artificial Intelligence in academia: Views of Turkish academics on ChatGPT // Heliyon. 2023. No. 9. e19688. <https://doi.org/10.1016/j.heliyon.2023.e19688>

Nexøe, 2023 – Nexøe J. Scientific papers and artificial intelligence. Brave new world? // Scandinavian Journal of Primary Health Care. 2023. Vol. 41. No. 3. P. 187–188. <https://doi.org/10.1080/02813432.2023.2240114>

O'Connor, 2023 – O'Connor S., ChatGPT. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? // Nurse Education in Practice. 2023. Vol. 66. No. 103537. <https://doi.org/10.1016/j.nepr.2022.103537>

Opryshko, Nazarovets, 2021 – Opryshko T., Nazarovets S. Case Study: Citizen Science in Digital Humanities context // DHW 2021: Digital Humanities Workshop. 2021. P. 198–203. <https://doi.org/10.1145/3526242.3526246>

Rodin et al., 2019 – Rodin P., Ghersetti M., Odén T. Disentangling rhetorical subarenas of public health crisis communication: A study of the 2014–2015 Ebola outbreak in the news media and social media in Sweden // Journal of Contingencies and Crisis Management. 2019. No. 27. P. 237–246. <https://doi.org/10.1111/1468-5973.12254>

Shijaku, Canhasi, 2023 – Shijaku R., Canhasi E. ChatGPT Generated Text Detection // Preprint. January 2023. <https://doi.org/10.13140/RG.2.2.21317.52960>

Tenhundfeld, ChatGPT, 2023 – Tenhundfeld N., ChatGPT. Two Birds With One Stone: Writing a Paper Entitled “ChatGPT as a Tool for Studying Human-AI Interaction in the Wild” with ChatGPT // Preprint. February 2023. <https://doi.org/10.13140/RG.2.2.25319.73123>

Thorp, 2023 – Thorp H.H. ChatGPT is fun, but not an author // Science. 2023. No. 379. P. 313–313. <https://doi.org/10.1126/science.adg7879>

Neural networks as an actor of socio-epistemic arenas: ethical problems

Sophia V. Tikhonova

Saratov State University, 10a Volskaia street, Bldg. 12, Saratov, 410028, Russian Federation; e-mail: segedasv@yandex.ru

Denis S. Artamonov

Saratov State University, 10a Volskaia street, Bldg. 12, Saratov, 410028, Russian Federation, e-mail: artamonovds@mail.ru

The article is devoted to the consideration of ethical problems related to the use of neural networks and artificial intelligence systems in scientific research. The authors consider neural networks as one of the actors of socio-epistemic arenas, which are understood as communication platforms for developing new knowledge in the process of distributed cognition. The acquisition of knowledge within the framework of socio-epistemic arenas occurs as a result of the interaction of three equally ranked actors: professional scientists, masses of non-professional Internet users and digital technologies. The emergence of the GPT4 neural network has made it obvious that digital technologies are actively involved not only in natural science, but also in humanitarian scientific research. It began to be used both for commercial purposes and for entertainment purposes in civil science projects and independent research of Internet users seeking to obtain the necessary information using a neural network. The neural network has also become part of the working tools of professional scientists, which has generated a lot of scientific and public discussions on the ethics of writing and publishing texts authored by ChatGPT. The use of a neural network for writing scientific articles is considered by part of the scientific community as plagiarism and copyright infringement, but the widespread use of GPT4 by the masses of Internet users is changing the attitude towards this tool in the scientific community as well. There are scientific papers in which ChatGPT is positioned as a co-author of the study, acquiring scientific subjectivity. This state of affairs aggravates the ethical problems associated with the possibility of falsification of research results, and raises the question of automating the processes of marking and detecting the authorship of the neural network in the publications of scientists. Meanwhile, it becomes obvious that neural networks, providing huge opportunities for the intensification of cognition processes, have expanded the socio-epistemic practices of scientists and the masses of Internet users, which it is impossible not to take into account when understanding science in society.

Keywords: neural network, Chat GPT, GPT4, socio-epistemic arenas, artificial intelligence, civil science, distributed cognition

Acknowledgments: The research was carried out at the expense of the grant of the Russian Science Foundation No. 22-18-00153, <https://rscf.ru/project/22-18-00153/>

References

“AI makes plagiarism harder to detect, argue academics – in paper written by chatbot”, *The Guardian*, 2023, March 19. URL: <https://www.theguardian.com/technology/2023/mar/19/ai-makes-plagiarism-harder-to-detect-argue-academics-in-paper-written-by-chatbot> (accessed on 24.09.2023).

Cotton, D., Cotton, P.A., Shipway, J.R. “Chatting and Cheating. Ensuring academic integrity in the era of ChatGPT”, *Innovations in Education and Teaching International*, 2023, vol. 61, issue 2, pp. 228–239. <https://doi.org/10.1080/14703297.2023.2190148>

“Defining the Role of Authors and Contributors”, *ICMJE*. URL: <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html> (accessed on 25.09.2023).

Dunne, C. “Can ChatGPT be your coauthor?”, *BCMJ*, 2023, vol. 65, no. 6, p. 193. URL: <https://bcmj.org/editorials/can-chatgpt-be-your-coauthor#a1>

Frandsen, F., Johansen, W. *Organizational crisis communication*. Los Angeles, CA: SAGE, 2016. 280 pp.

Gao, C.A., Howard, F.M., Markov, N.S. et al. “Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers”, *npj Digit. Med*, 2023, vol. 6, no. 75. <https://doi.org/10.1038/s41746-023-00819-6>

Gönnér, von J., Herrmann, T.M., Bruckermann T. et al. “Citizen science’s transformative impact on science, citizen empowerment and socio-political processes”, *Socio Ecol Pract Res*, 2023, no. 5, pp. 11–33. <https://doi.org/10.1007/s42532-022-00136-4>

Livberber, T. & Ayvaz, S. “The impact of Artificial Intelligence in academia: Views of Turkish academics on ChatGPT”, *Heliyon*, 2023, no. 9. e19688. <https://doi.org/10.1016/j.heliyon.2023.e19688>

Koval, E., Ushkin, S. “Ogon’, voda i ChatGPT: novye vyzovy nauchnomu etosu” [Fire, water and ChatGPT: new challenges to scientific ethos], *SocioDigger*. URL: <https://sociodigger.ru/articles/articles-page/ogon-voda-i-chatgpt-novye-vyzovy-nauchnomu-ehostu> (accessed on 09.24.2023). (In Russian)

Nexøe, J. “Scientific papers and artificial intelligence. Brave new world?”, *Scandinavian Journal of Primary Health Care*, 2023, vol. 41, no. 3, pp. 187–188. <https://doi.org/10.1080/02813432.2023.2240114>

O’Connor, S., ChatGPT. “Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?”, *Nurse Education in Practice*, 2023, vol. 66, no. 103537. <https://doi.org/10.1016/j.nepr.2022.103537>

Opryshko, T., Nazarovets, S. “Case Study: Citizen Science in Digital Humanities context”, *DHW 2021: Digital Humanities Workshop*, 2021, pp. 198–203. <https://doi.org/10.1145/3526242.3526246>

Benigo, Y., Russel S., Musk E. et al. “Pause Giant AI Experiments: An Open Letter”, *Futureoflife*. URL: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (accessed on 25.09.2023).

Rodin, P., Ghersetti, M., Odén, T. “Disentangling rhetorical subarenas of public health crisis communication: A study of the 2014–2015 Ebola outbreak in the news media and social media in Sweden”, *Journal of Contingencies and Crisis Management*, 2019, no. 27, pp. 237–246. <https://doi.org/10.1111/1468-5973.12254>

Shijaku, R. Canhasi, E. *ChatGPT Generated Text Detection, Preprint*. January 2023, <https://doi.org/10.13140/RG.2.2.21317.52960>

Shipovalova, L.V. “Raspredelennoe poznanie – analitika i problematizatsiya kontsepta” [Distributed cognition – analytics and problematization of the concept], *Digital scientist: the Philosopher’s laboratory*, 2019, vol. 2, no. 4, pp. 175–190. <https://doi.org/10.5840/dspl20192460>. (In Russian)

Smirnova, E. “Na vstreche s senatorami Musk nazval II ugrozoi chelovechestvu, a Geits – spaseniem” [At a meeting with senators, Musk called AI a threat to humanity, and Gates – salvation], *Hi-tech+*. URL: <https://hightech.plus/2023/09/14/na-vstreche-s-senatorami-mask-nazval-ii-ugrozoi-chelovechestvu-a-geits--spaseniem> (accessed on 09/24/2023). (In Russian)

Tenhundfeld, N., ChatGPT. *Two Birds With One Stone: Writing a Paper Entitled “ChatGPT as a Tool for Studying Human-AI Interaction in the Wild” with ChatGPT, Preprint*. February 2023, <https://doi.org/10.13140/RG.2.2.25319.73123>

Thorp, H.H. “ChatGPT is fun, but not an author”, *Science*, 2023, no. 379, p. 313. <https://doi.org/10.1126/science.adg7879>