Павел Юрьевич Гурушкин

Санкт-Петербургский государственный университет

p.gurushkin@spbu.ru

**Этические аспекты применения искусственного интеллекта в академической среде**

В статье рассмотрены этические последствия использования сервисов искусственного интеллекта и связанных с ним технологий в академических кругах, научных исследованиях и публикациях. В частности, вопросы авторского права, практики цитирования и потенциального влияния на «эффект Матфея» в научных трудах.

Ключевые слова: искусственный интеллект, этика, ChatGPT.

 Искусственный интеллект (ИИ), машинное обучение и обработка естественного языка – быстро развивающиеся области, которые оказывают значительное влияние на развитие современных социальных институтов [3]. ИИ включает в себя моделирование разумного поведения компьютерной программы с целью создания алгоритмов, способных имитировать человеческий интеллект во многих ключевых жизненных аспектах, таких как понимание языка, рассуждение и решение различных повседневных или рабочих проблем. Самообучение является ключевым компонентом ИИ, позволяющим системам получать новую информацию и совершенствовать свои суждения и выводы, основанные на знаниях, полученных с помощью накопленного опыта [3].

OpenAI – исследовательская лаборатория, которая внесла значительный вклад в область развития искусственного интеллекта, включая разработку высокоразвитой языковой модели GPT-4 и чат-бота (ChatGPT), который использует обработку естественного языка для генерации ответов на вводимые пользователем данные. Помимо прочего, ChatGPT может быть интегрирован в академическую отрасль и «способствовать распространению новых исследовательских идей посредством создания более качественных метаданных, индексирования и обобщения результатов научных изысканий» [3].

Несмотря на то, что огромный набор интернет-данных предлагает широкий спектр точек зрения и мнений, компьютерный лингвист Эмили Бендер, отметила, что объём не гарантирует разнообразия [1]. Кроме того, языковые модели, управляемые искусственным интеллектом, описываются как «стохастические попугаи», способные только собирать обучающие материалы, но не применять их. С этической точки зрения исследовательские работы, созданные с использованием ChatGPT, могут рассматриваться как неоригинальные и потенциально проблематичные. Несколько исследований показали, что данные обучения и процесс кодирования языковых моделей, таких как GPT-3 и GPT-4, которые обычно получают из больших наборов веб-данных, могут содержать предвзятость в отношении пола, расы, этнической принадлежности и статуса инвалидности [2].

Атрибуция авторства является важной проблемой, касающейся генерации новых знаний интеллектуальными агентами, таких как ИИ. Могут возникнуть вопросы о праве собственности на контент, созданный моделью. Если пользователь предоставляет входные данные программе, а она генерирует контент на их основе, можно утверждать, что пользователь владеет авторскими правами на сгенерированный материал. Однако если программа генерирует контент независимо от пользовательского ввода или предоставляемый ввод максимально ограничен (например, «написать эссе на такую-то тему»), то становится сложнее определить подлинное авторство.

Базы научных работ, собранные ChatGPT, могут повлиять на «эффект Матфея», который порождает тенденцию поощрения успешных исследователей с высоким количеством цитирований, в то время как менее известные исследователи остаются невостребованными. Такие платформы, как ChatGPT, использующие подсчет цитирования как фактор отбора материала, могут усугубить этот эффект.

Хотя ChatGPT и GPT-4 представляют собой важные достижения в области искусственного интеллекта, многие вопросы об этике использования ИИ в научных кругах и её влиянии на продуктивность исследований остаются без ответа. Необходимо регулярно проводить всесторонний обзор текущего состояния этих дискуссий и побудить академическое сообщество к дальнейшему изучению этических проблем, связанных с использованием GPT и аналогичных технологий в научных кругах.

Литература

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 2021, 610–623.

2. Lund, B. D. (2022). Is academic research and publishing still leaving developing countries behind? Accountability in Research, 29(4), 224–231.

3. Wamba, S. F., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. (2021). Are we preparing for a good AI society? A bibliometric review and research agenda, article 120482.