

УДК 373.3

ЦИФРОВАЯ ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБРАЗОВАНИИ

Мулюкин К.Э

*студент 2 курса, направление 44.04.01 Педагогическое образование
(Инновации в образовании), магистерская программа,*

*ФГБОУ ВО «Калужский государственный университет им. К. Э.
Циолковского»*

Калуга, Россия¹

Аннотация:

Статья посвящена анализу принципов цифровой этики искусственного интеллекта. Выявлены основные этические принципы ИИ. Искусственный интеллект стал глобальной тенденцией последних лет. Акцентируется внимание на возможных проблемах в сфере образования в связи с активным внедрением технологий искусственного интеллекта. Рассматриваются вопросы, связанные с отсутствием четких правовых норм и этических принципов, а также границами использования результатов, созданных с помощью искусственного интеллекта.

Ключевые слова: искусственный интеллект, цифровизация, образовательный процесс, нейронные сети.

DIGITAL ETHICS OF ARTIFICIAL INTELLIGENCE IN EDUCATION

Mulyukin K. E.

*2nd year student, direction 44.04.01 Pedagogical education (Innovations in
education), master's program,*

Kaluga State University named after K. E. Tsiolkovsky

Kaluga, Russia

¹ Научный руководитель: Коржуева Е. В. заведующий кафедрой Педагогики ФГБОУ ВО «Калужский государственный университет им. К. Э. Циолковского», Калуга, Россия
Scientific supervisor: Korzhueva E. V. Head of the Department of Pedagogy, Kaluga State University named after K. E. Tsiolkovsky, Kaluga, Russia

Annotation:

The article is devoted to the analysis of the principles of digital ethics of artificial intelligence. The main ethical principles of AI have been identified. Artificial intelligence has become a global trend in recent years. Attention is focused on possible problems in the field of education in connection with the active introduction of artificial intelligence technologies. The issues related to the lack of clear legal norms and ethical principles, as well as the boundaries of the use of results created with the help of artificial intelligence, are considered.

Keywords: artificial intelligence, digitalization, educational process, neural networks.

Этика ИИ - это развивающаяся и относительно новая область, возникшая в ответ на растущую обеспокоенность по поводу влияния искусственного интеллекта (ИИ) на людей и их социальные институты [1, 9, 10]. В свою очередь, этика искусственного интеллекта является частью более широкой области цифровой этики, которая решает аналогичные проблемы, возникающие в связи с разработкой и внедрением новых цифровых технологий, включая, конечно же, искусственный интеллект, аналитику больших данных и блокчейн-технологии [5].

Искусственный интеллект вызвал бурные дебаты о принципах и ценностях, которыми следует руководствоваться при его использовании. Опасения по поводу того, что искусственный интеллект может поставить под угрозу рабочие места людей [7], быть неправильно использован злоумышленниками [9] были в центре внимания современной научной литературы и средств массовой информации. Искусственный интеллект (ИИ) широко провозглашается как продолжающаяся «революция», полностью преобразующая науку и общество [8]. Подходы к ИИ, такие как машинное обучение, глубокое обучение и нейронные сети, меняют способы обработки и анализа данных, а автономные и полуавтономные системы всё активнее применяются в различных секторах, включая образование. Учитывая мощную трансформирующую силу ИИ и его глубокое влияние на различные сферы

Дневник науки | www.dnevniknauki.ru | СМИ Эл № ФС 77-68405 ISSN 2541-8327

общества, он стал предметом широких дебатов о принципах и ценностях, которые должны направлять его развитие и использование. Ряд исследований обсуждают тему этичного ИИ, особенно в связи с системными рисками и непреднамеренными негативными последствиями, такими как алгоритмическая предвзятость или дискриминация. Национальные и международные организации ответили на эти общественные опасения, создав специальные экспертные комитеты по ИИ. Среди них — группа высокоуровневых экспертов по искусственному интеллекту, назначенная Европейской комиссией, экспертная группа по ИИ в обществе Организации экономического сотрудничества и развития (ОЭСР), Консультативный совет по этическому использованию искусственного интеллекта и данных в Сингапуре и специальный комитет по искусственному интеллекту Палаты лордов Великобритании. В рамках своих институциональных назначений эти комитеты подготовили или, как сообщается, готовят отчёты и руководящие документы по ИИ. Аналогичные усилия предпринимаются в частном секторе, особенно среди компаний, использующих ИИ в своей деятельности. Только в 2018 году такие компании, как Google и SAP, публично опубликовали руководящие принципы по ИИ. Интенсивные усилия такого разнообразного круга заинтересованных сторон по изданию принципов и политик в области ИИ демонстрируют не только необходимость этического руководства, но и сильный интерес этих сторон формировать этику ИИ таким образом, чтобы она соответствовала их приоритетам. В ходе анализа документов по этике искусственного интеллекта (ИИ) было выявлено, что ни один этический принцип не является универсальным для всех источников [3, 4]. Тем не менее, наблюдается тенденция к консенсусу по пяти ключевым принципам: прозрачность, справедливость, равноправие, ответственность и конфиденциальность.

Принцип прозрачности является наиболее распространённым в современных этических руководствах по искусственному интеллекту (ИИ) [7, 10, 13].

Основные аспекты прозрачности в контексте ИИ:

- Объяснимость и интерпретируемость: Прозрачность часто связывается с необходимостью объяснения и интерпретации решений ИИ, а также с другими формами коммуникации и раскрытия информации;

Например: алгоритмы ИИ, оценивающие учеников, должны быть понятны учителям и родителям.

Принципы справедливости и равенства занимают центральное место в этических руководствах по искусственному интеллекту. Они охватывают широкий спектр тем, включая предотвращение предвзятости, обеспечение инклюзивности, равный доступ и возможность обжалования решений, принятых с использованием ИИ. Многие руководства подчеркивают необходимость выявления и устранения нежелательной предвзятости в данных и алгоритмах [12]. Это включает в себя мониторинг и смягчение предвзятости, а также предотвращение дискриминации, особенно в отношении уязвимых групп населения. Важно обеспечить равный доступ к технологиям ИИ, данным и их преимуществам для всех, независимо от социального или экономического статуса. Это особенно актуально в контексте здравоохранения, образования и трудоустройства [6]. Например, ИИ не должен предлагать разные задания мальчикам и девочкам из-за скрытых предубеждений в данных.

Принцип ответственности несмотря на широкое упоминание концепции «ответственного ИИ» [11] редко получает четкое определение. Тем не менее, конкретные рекомендации включают действия с «честностью» и разъяснение распределения ответственности и юридической ответственности. Возникают

дальнейшие разногласия относительно того, следует ли привлекать ИИ к ответственности по аналогии с человеком или же люди всегда должны быть единственными субъектами, которые в конечном итоге несут ответственность за технологические артефакты. Например, если алгоритм неправильно оценил работу ученика, ответственность лежит на разработчиках или школе? Вопрос ответственности в сфере ИИ остается недостаточно проработанным, с отсутствием четких определений по ключевым аспектам. Существуют различные рекомендации и подходы к обеспечению ответственности, включая разъяснение юридической ответственности и предотвращения вреда.

Принцип конфиденциальности в контексте этичного искусственного интеллекта (ИИ) рассматривается как ценность, которую следует поддерживать, и как право, которое необходимо защищать [1]. Хотя понятие конфиденциальности часто не имеет четкого определения, оно обычно связывается с защитой и безопасностью данных. В школе это касается информации об успеваемости, поведении или здоровье детей. Как и в случае с наиболее популярными технологиями ИИ, существуют опасения по поводу больших объемов персональных данных, собираемых для поддержки применения ИИ в образовании [2]. Кому принадлежат эти данные, кто может получить к ним доступ, каковы проблемы конфиденциальности и неразглашения, как следует анализировать, интерпретировать и передавать данные?

Таким образом, разработка и обсуждение принципов этичного искусственного интеллекта (ИИ) приобрело динамичный характер во многих странах и организациях. Несмотря на общее признание их значимости, интерпретация и применение этих принципов пока демонстрируют существенное разнообразие. Этические вопросы, связанные с ИИ, остаются актуальными и требуют более глубокого анализа, а также создания четких

стандартов и рекомендаций для гарантирования справедливого, прозрачного и ответственного развития и внедрения технологий ИИ.

Библиографический список:

1. Asilomar AI Principles. Future of Life Institute, 2017.
2. Brundage, M., et al. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Future of Humanity Institute; University of Oxford; Centre for the Study of Existential Risk; University of Cambridge; Center for a New American Security; Electronic Frontier Foundation; OpenAI, 2018.
3. Deutsche Telekom. Deutsche Telekom's Guidelines for Artificial Intelligence. 2018.
4. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center Research Publication, 2020.
5. Hagendorff, T. The Ethics of AI Ethics: An Evaluation of Guidelines. Mind & Machines, 30, 1–22, 2020.
6. IBM. Everyday Ethics for Artificial Intelligence: A Practical Guide for Designers & Developers. 2018.
7. IBM. Transparency and Trust in the Cognitive Era. 2017.
8. Internet Society. Artificial Intelligence & Machine Learning: Policy Paper. 2017.
9. Jobin, A., Ienca, M., Vayena, E. The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 2019.
10. Kazim, E., Koshiyama, A. A High-Level Overview of AI Ethics. 2020.
11. Special Interest Group on Artificial Intelligence. Dutch Artificial Intelligence Manifesto. 2018.
12. Telefónica. Our Artificial Intelligence Principles. 2018.
13. UNI Global. 10 Principles for Ethical AI. 2017.

Оригинальность 75%