Секция «Нейронные сети общения: как мозг формирует переговорную позицию»

Этические принципы и нормы применения технологий искусственного интеллекта

Ясковец Валерия Игоревна
Студент (магистр)
Московский государственный университет имени М.В.Ломоносова, Факультет глобальных процессов, Образовательная программа «Глобальные политические процессы и дипломатия», Москва, Россия
E-mail: lera.yaskovets@gmail.com

Развитие технологий искусственного интеллекта (ИИ) и их интеграция в различные сферы жизни вызвали необходимость создания нормативных документов, регулирующих этические вопросы, возникающие при использовании ИИ. Эти вопросы охватывают широкий спектр проблем – от конфиденциальности данных до угроз безопасности человека, что требует разработки универсальных стандартов. Несмотря на начальный этап формирования всеобъемлющего этического кодекса ИИ, различные организации, такие как ЮНЕСКО, IEEE, ISO и национальные регуляторы, уже предлагают нормативные подходы и рекомендации.

Этические обсуждения ИИ нередко сопровождаются путаницей, связанной с некорректными определениями и недостаточным пониманием специфики интеллектуальных и автономных систем (И/АС). В отличие от традиционных этических проблем, в И/АС критически важное значение имеет механизм принятия решений, поскольку такие системы способны самостоятельно определять последствия своих действий. Этические принципы должны служить дополнительным фильтром при выборе альтернативных решений, закладывая механизмы регулирования морально значимых ситуаций.

ЮНЕСКО предложило первый глобальный этический нормативный документ для ИИ, призванный обеспечить согласованный международный подход к разработке технологий. В рекомендациях содержится перечень ключевых ценностных установок, включая защиту прав человека, экологическую устойчивость, справедливость, инклюзивность и прозрачность. Особый акцент сделан на необходимости тотального контроля ИИ на всех этапах его жизненного цикла, а также законодательного закрепления этических норм на национальном уровне. Однако часть положений документа вызывает вопросы, поскольку некоторые пункты, например меры по гендерному равенству, выходят за рамки специфики ИИ и представляют собой элементы политической повестки.

Российский национальный кодекс этики ИИ, разработанный рядом ИТ-компаний, во многом повторяет принципы ЮНЕСКО, но носит более абстрактный и менее конъюнктурный характер. Его ключевые положения касаются недискриминации, соблюдения законодательства и защиты прав человека. Документ поощряет добровольную сертификацию ИИ-систем и требует информирования пользователей о взаимодействии с ИИ в критически важных сферах. Однако вопросы о реальном механизме контроля исполнения норм остаются открытыми.

IEEE разрабатывает стандарты, ориентированные на практическую реализацию этических принципов. В отличие от ЮНЕСКО и национальных кодексов, IEEE P7000 не дает строгого определения этичности ИИ, а предлагает методологию интеграции этических аспектов в процесс проектирования. Это делает стандарт более гибким, но при этом оставляет определение морали на усмотрение разработчиков. Серия стандартов IEEE P7000 также охватывает такие аспекты, как прозрачность алгоритмов, предвзятость данных и безопасность ИИ.

Правовое регулирование алгоритмической предвзятости становится важнейшим направлением правотворчества, требующим анализа российских и зарубежных концепций. Искусственный интеллект обладает многочисленными преимуществами, включая повышение творческого потенциала, усиление безопасности, улучшение качества жизни и повышение эффективности принятия решений. Однако его применение вызывает серьезные опасения в отношении автономии личности, конфиденциальности данных и соблюдения фундаментальных прав человека. Алгоритмическая предвзятость существует даже в отсутствие намерений разработчиков дискриминировать пользователей, поскольку системы машинного обучения выявляют закономерности, заложенные в исходных данных, что может приводить к усилению существующих социальных стереотипов.

Юридическое сообщество рассматривает различные способы минимизации рисков алгоритмической предвзятости. Среди ключевых подходов выделяется принятие международных деклараций, политик и стандартов, регулирующих разработку, тестирование и эксплуатацию систем искусственного интеллекта. Оставленные без внимания предвзятые алгоритмы могут приводить к решениям, оказывающим разрозненное, но значительное влияние на определенные группы людей, даже без намерения разработчиков внедрять дискриминацию. Государственная политика в области регулирования искусственного интеллекта часто оказывается недостаточной для эффективного выявления и устранения последствий такой предвзятости. Решение этой проблемы исключительно техническими средствами не приводит к необходимым результатам, что подчеркивает важность разработки комплексных подходов, включающих как правовые механизмы, так и этические нормы.

Мировое сообщество уже предпринимает шаги по внедрению стандартов и разработке этических принципов, направленных на обеспечение справедливого применения искусственного интеллекта. Создание специальных норм, ограничивающих алгоритмическую предвзятость, необходимо для предотвращения нарушений прав человека и недобросовестной конкуренции. Введение обязательных требований к структуре данных, исключающих явную или скрытую сегрегацию групп населения, может способствовать формированию более универсальных моделей искусственного интеллекта, учитывающих данные различных социально-правовых групп общества.

Этические вопросы, связанные с предвзятостью алгоритмов, приобретают особую актуальность в контексте функционирования рекомендательных систем и анализа пользовательских предпочтений. Эти системы, активно применяемые в бизнесе, государственных услугах и социальной политике, способны формировать однородные пользовательские группы, исключая определенные категории граждан из рекомендательных алгоритмов. Таким образом, искусственный интеллект, не обладая сознательной предвзятостью, закрепляет и усиливает уже существующее неравенство в обществе, влияя на доступ к товарам, услугам и возможностям.

Источники и литература

1) Вейценбаум Дж. Возможности вычислительных машин и человеческий разум: От суждений к вычислениям. – М.: Радио и связь, 1982. – 369 с.

2) Савельев А. И. Комментарий к Федеральному закону от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и защите информации» (постатейный). – М.: Статут, 2015. – 320 с.

3) Харитонова А. Р. Сохранность и анонимность персональных данных в социальных сетях // Предпринимательское право. – 2019. – № 4. – С. 48–55. [Приложение «Право и Бизнес»].

4) Харитонова Ю. С. Контекстная (поведенческая) реклама и право: точки пересечения // В кн.: Рожкова М. А. (ред.). Право в сфере Интернета: сб. ст. / М. З. Али, Д. В. Афанасьев, В. А. Белов [и др.]. – М.: Статут, 2018. – 528 с.

5) Харитонова Ю. С., Савина В. С. Технология искусственного интеллекта и право: вызовы современности // Вестник Пермского университета. Юридические науки. – 2020. – Вып. 3. – С. 524–549. DOI: 10.17072/1995-4190-2020-49-524-549.

6) Gauthier J., Grauwin S., Sobolevsky S., Hong Y., Ratti C. Quantifying the Effects of Social Influence on Facebook // Science. – 2015. – № 348(6239). – Pp. 1130–1132. DOI: 10.1126/science.aaa1160.

7) Barr A. Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms // The Wall Street Journal. – 2015. – № 1. URL: https://www.wsj.com/articles/BL-DGB-42522 (дата обращения: 03.02.2025).

8) Bartolini C., Lenzini G., Santos C. An Agile Approach to Validate a Formal Representation of the GDPR // JSAI: Annual Conference of the Japanese Society for Artificial Intelligence. – Springer, Cham, 2018. – November. – Pp. 160–176.

9) Bengio Y. Learning Deep Architectures for AI. – Now Publishers Inc., 2009.

10) Benkler Y. Don't Let Industry Write the Rules for AI // Nature. – 2019. – № 569(7754). – Pp. 161–162.

11) Buolamwini J., Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification // Proceedings of the Conference on Fairness, Accountability and Transparency. – PMLR, 2018. – January. – Vol. 81. – Pp. 77–91. URL: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf (дата обращения: 03.02.2025).

12) Burns N. Why Should We Expect Algorithms to Be Biased? // MIT Technology Review. – 2016. URL: https://www.technologyreview.com/s/601775/why-we-should-expect-algorithms-to-be-biased/ (дата обращения: 03.02.2025).

13) Гусев А. В., Шарова Д. Е. Этические проблемы развития технологий искусственного интеллекта в здравоохранении // Общественное здоровье. – 2023. – Т. 3. – № 1. – С. 42–50.

14) Леушина В. В., Карпов В. Э. Этика искусственного интеллекта в стандартах и рекомендациях // Философия и общество. – 2022. – № 3 (104). – С. 124–140.

15) Малышкин А. В. Интегрирование искусственного интеллекта в общественную жизнь: некоторые этические и правовые проблемы // Вестник Санкт-Петербургского университета. Право. – 2019. – Т. 10. – № 3. – С. 444–460.