

УДК 004.8

ЭТИЧЕСКИЕ ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Ададуров В.С.¹*студент,**Поволжский государственный университет телекоммуникаций и**информатики,**Самара, Россия*

Аннотация

Статья посвящена рассмотрению этических проблем, возникающих при разработке и использовании систем искусственного интеллекта (ИИ). Анализируются вопросы ответственности разработчиков и пользователей, прозрачности алгоритмических решений, защиты персональных данных, дискриминации и алгоритмической предвзятости, а также риски нарушения авторских прав. Особое внимание уделяется применению ИИ в образовании, профессиональной деятельности и информационной среде. Рассматривается проблема соотношения эффективности автоматизированных систем и необходимости человеческого контроля. Обосновывается, что ответственное использование ИИ требует прозрачности, оценки потенциального вреда, защиты прав человека и возможности обжалования решений, принятых с участием алгоритмов.

Ключевые слова: искусственный интеллект, этика ИИ, алгоритмическая предвзятость, конфиденциальность, персональные данные, авторское право, ответственность, прозрачность.

¹ Научный руководитель: **Куваева Е.Н.**, ст. преподаватель, Поволжский государственный университет телекоммуникаций и информатики, Самара, Россия.

Kuvaeva E.N., Senior Lecturer, Povolzhskiy State University of Telecommunications and Informatics, Samara, Russia.

ETHICAL PROBLEMS OF ARTIFICIAL INTELLIGENCE USE

Adadurov V.S.

Student,

Povolzhskiy State University of Telecommunication and Informatics,

Russia, Samara

Abstract

The article examines ethical problems arising from the development and use of artificial intelligence (AI) systems. It considers the responsibility of developers and users, algorithmic transparency, personal data protection, discrimination and bias, and copyright risks. Particular attention is paid to the use of AI in education, professional activities and the information environment. The article discusses the balance between the efficiency of automated systems and the need for meaningful human oversight. Responsible AI use requires transparency, assessment of potential harm, protection of human rights and effective mechanisms for challenging decisions made with the participation of algorithms.

Keywords: artificial intelligence, AI ethics, algorithmic bias, privacy, personal data, copyright, responsibility, transparency.

Развитие систем искусственного интеллекта (ИИ) стало одним из наиболее заметных технологических процессов последних лет. Алгоритмы машинного обучения используются для обработки больших массивов данных, генерации текста и изображений, поддержки принятия решений, автоматизации производства и предоставления цифровых услуг. Одновременно с расширением возможностей ИИ возрастает число ситуаций, в которых технологическое решение непосредственно затрагивает интересы человека. Это делает этические

вопросы не дополнительным аспектом разработки, а важной частью оценки самой технологии. [7]

Этическая проблема использования ИИ заключается прежде всего в том, что алгоритмические системы способны оказывать влияние на человека, не обладая человеческим пониманием возможных последствий своих решений. Решение, принятое моделью, может восприниматься как объективное благодаря своей математической основе, однако его результат зависит от используемых данных, целей обучения, выбранных критериев качества и условий применения системы. Поэтому автоматизация не снимает вопрос ответственности, а, наоборот, требует определить, кто несёт ответственность за полученный результат: разработчик, владелец системы, оператор или пользователь. Одним из принципов ответственного использования ИИ является прозрачность. Пользователь должен понимать, когда он взаимодействует с автоматизированной системой и в каких ситуациях её результат используется при принятии значимых решений. Особенно большое значение объяснимость алгоритмов имеет в тех сферах, где ошибка может повлечь материальный или социальный ущерб. Если система оценивает кандидата, формирует рекомендации, определяет приоритет обращения или предлагает определённый контент, человек должен иметь возможность получить доступное и понятное объяснение причин такого результата. [1]

С прозрачностью тесно связана проблема человеческого контроля. Передача решения алгоритму может создавать иллюзию его безусловной правильности, поскольку вычислительная система воспринимается как более нейтральная, чем отдельный человек. Однако ИИ не является независимым от человеческих установок: они отражаются в постановке задачи, выборе данных и способе интерпретации результата. Поэтому автоматизированная рекомендация не должна автоматически превращаться в окончательное решение там, где затрагиваются права и существенные интересы человека. [8]

Отдельную группу этических рисков представляет алгоритмическая предвзятость. Модели обучаются на данных, содержащих статистические особенности реального общества, в том числе уже существующие формы неравенства. Если такие данные используются без критической проверки, система может воспроизводить и усиливать нежелательные различия между группами пользователей. Проблема проявляется, например, при автоматизированном отборе кандидатов, оценке кредитоспособности, распознавании изображений и персонализации информационных потоков. Наличие предвзятости трудно свести только к технической ошибке. Даже при высокой средней точности система может работать значительно хуже для отдельных групп, а математически оптимальная метрика не всегда совпадает с представлением о справедливом результате. Следовательно, разработка ИИ требует проверки качества не только в целом, но и в отношении разных категорий пользователей. Необходимы тестирование на неоднородных выборках, документирование ограничений модели и регулярный контроль результатов после внедрения. [6]

Серьёзные этические вопросы возникают при обработке персональных данных. Для обучения и функционирования интеллектуальных систем могут использоваться большие объёмы информации о поведении, интересах и действиях людей. Даже если отдельные сведения кажутся безобидными, их объединение способно сформировать подробный цифровой профиль. Пользователь должен понимать, какие данные собираются, для какой цели они применяются, кто получает к ним доступ и как долго они сохраняются. Минимизация собираемой информации и ограничение целей её использования снижают потенциальный вред. [3]

Особое значение проблема конфиденциальности приобретает при использовании генеративного ИИ. Пользователь может передать в сервис рабочие документы, персональные сведения или внутреннюю информацию организации, не оценивая последствия такой передачи. В результате конфиденциальный материал может оказаться в инфраструктуре внешнего

поставщика. Поэтому применение генеративных систем должно сопровождаться правилами работы с данными, разграничением доступа и понятными требованиями к тому, какую информацию разрешено отправлять во внешние сервисы. [4]

Другой важный вопрос связан с авторским правом и происхождением результатов, создаваемых ИИ. Генеративные модели способны создавать тексты, изображения и иные материалы на основе закономерностей, изученных на больших массивах данных. Возникает необходимость разграничивать вклад человека и автоматизированной системы, определять допустимость использования защищённых произведений и устанавливать правила указания источников. Этический подход предполагает не только стремление получить результат быстрее, но и уважение к труду авторов, чьи произведения участвуют в цифровой экосистеме. [7]

Проблема ответственности особенно заметна в ситуациях, когда система выдаёт ошибочную или недостоверную информацию. Генеративные модели могут формировать убедительные ответы, содержащие фактические ошибки, вымышленные сведения или некорректные ссылки. Опасность заключается в несоответствии между убедительной формой ответа и реальной надёжностью содержания. Пользователь, который воспринимает результат без проверки, переносит ошибку ИИ в свою деятельность. Поэтому необходима практика проверки критически важной информации человеком. В образовательной среде ИИ создаёт одновременно возможности и новые этические дилеммы. Интеллектуальные инструменты могут помогать объяснять сложные темы, структурировать материал, исправлять языковые ошибки и предоставлять персонализированную обратную связь. Вместе с тем без ясных правил их использование может затруднить оценку самостоятельности обучающегося. Проблема состоит не в самом факте применения ИИ, а в отсутствии прозрачных границ между допустимой помощью, совместной работой человека и полной передачей интеллектуальной задачи алгоритму. [1]

Для образования особенно важен принцип равенства доступа. Если часть обучающихся имеет постоянный доступ к более мощным или платным системам ИИ, а другая часть располагает только базовыми инструментами, технологическое преимущество может влиять на результаты обучения. Кроме того, различия в цифровой грамотности способны усиливать неравенство. Поэтому образовательные организации должны не только устанавливать ограничения, но и формировать навыки осознанного использования ИИ, включая проверку результатов, указание факта применения инструмента и критическое отношение к его рекомендациям. [3]

Этические риски присутствуют и в профессиональной сфере. Автоматизация отдельных операций способна повысить производительность, однако чрезмерная зависимость от алгоритмов может привести к потере важных навыков и ослаблению профессионального контроля. Если специалист принимает решение исключительно на основании рекомендации системы, ответственность становится размытой. Рациональным подходом является распределение функций: ИИ выполняет обработку и предлагает варианты, а человек оценивает контекст, учитывает исключения и принимает окончательное решение в ситуациях, требующих профессионального суждения. [8]

В информационной среде использование ИИ затрагивает вопросы достоверности и доверия. Генерация большого количества синтетических текстов, изображений и видео снижает стоимость производства контента, но одновременно облегчает распространение недостоверных материалов. Пользователь может столкнуться с содержанием, происхождение которого трудно установить. Это повышает значение цифровой грамотности, маркировки синтетического контента, механизмов проверки источников и ответственности платформ за правила распространения материалов. [2]

Важным этическим принципом является возможность обжалования решений, принятых с участием ИИ. Если алгоритмическая система оказывает существенное влияние на человека, недостаточно сообщить ему итоговый

результат. Должен существовать понятный механизм проверки, исправления ошибки и обращения к ответственному специалисту. Такой подход особенно важен там, где автоматизированная оценка может ограничить доступ к услуге, образованию, работе или иной значимой возможности. [4]

Ответственное использование искусственного интеллекта требует также документирования ограничений системы. Описание модели должно включать не только показатели точности, но и сведения о предполагаемом назначении, характерных ошибках, качестве данных и условиях, при которых результаты могут становиться менее надёжными. Такая информация позволяет пользователю соотнести возможности технологии с реальной задачей и не использовать инструмент за пределами его компетенции. На основании рассмотренных проблем можно выделить несколько принципов этически приемлемого применения ИИ. Принцип 1 - ответственность: должна быть определена конкретная сторона, отвечающая за последствия применения системы. Принцип 2 - прозрачность: пользователю необходимо сообщать о применении ИИ и существенных ограничениях результата. Принцип 3 - справедливость: система должна проверяться на наличие систематических различий в качестве работы для разных групп. Принцип 4 - конфиденциальность: сбор и использование персональных данных должны быть ограничены необходимой целью. Принцип 5 - человеческий контроль: в значимых ситуациях окончательное решение должно оставаться за человеком, способным оценить контекст и исправить ошибку алгоритма. [5]

Таким образом, этические проблемы использования искусственного интеллекта не сводятся к отдельным техническим сбоям. Они затрагивают ответственность, права человека, равенство, конфиденциальность, авторство и качество принимаемых решений. Чем шире ИИ включается в повседневные процессы, тем важнее рассматривать его не только как инструмент повышения эффективности, но и как технологию, требующую общественных и профессиональных ограничений. Ответственный подход предполагает

сочетание технических мер, понятных правил использования, человеческого контроля и постоянной оценки возможного вреда. [1]

Перспективным направлением является формирование культуры осознанного взаимодействия с ИИ. Пользователям необходимо понимать сильные и слабые стороны алгоритмов, разработчикам - учитывать этические последствия уже на этапе проектирования, а организациям - создавать процедуры контроля и реагирования на ошибки. Только при таком распределении ответственности искусственный интеллект сможет использоваться как средство поддержки человека, не подменяя его право на самостоятельное решение и не снижая уровень защиты его основных интересов. [3]

Библиографический список

1. Флориди Л., Каулс Дж., Бельтраметти М. и др. AI4People — этическая основа для благополучного общества с искусственным интеллектом: возможности, риски, принципы и рекомендации // *Minds and Machines*. — 2018. — Т. 28, № 4. — С. 689–707.
2. Жобин А., Иенка М., Вайена Э. Глобальный ландшафт руководящих принципов этики искусственного интеллекта // *Nature Machine Intelligence*. — 2019. — Т. 1. — С. 389–399.
3. ЮНЕСКО. Рекомендация по этике искусственного интеллекта. — Париж: ЮНЕСКО, 2022. — 42 с.
4. Европейская комиссия. Этические рекомендации для заслуживающего доверия ИИ. — Брюссель: Европейская комиссия, 2019. — 39 с.
5. OECD. Recommendation of the Council on Artificial Intelligence. — Paris: OECD Publishing, 2019.
6. Капур С., Брайан К., Каплан Р. Этические проблемы и риски алгоритмического принятия решений // *Artificial Intelligence and Society*. — 2021. — Т. 36. — С. 1–15.

7. Бостром Н., Юдковский Э. Этика искусственного интеллекта // The Cambridge Handbook of Artificial Intelligence / под ред. К. Франка, У. Рамсея. — Cambridge: Cambridge University Press, 2014. — С. 316–334.
8. Рассел С. Совместимое с человеком: искусственный интеллект и проблема контроля. — М.: Альпина Пабlishер, 2021. — 480 с.