

ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ТЕНДЕНЦИИ, ВЫЗОВЫ, РЕШЕНИЯ

Тамерлан Гаджиев

докторант, Институт Философии и Социологии НАНА

tamerlan.haciyev.2000@mail.ru

<https://orcid.org/0009-0008-1268-1954>

Резюме. Этика связана с философией ИИ, так как философия ИИ занимается сознанием и мозгом, что определяет этику в любом ее измерении. В современном обществе внедрение искусственного интеллекта происходит на различных уровнях и в различных областях, что создает ряд этических вопросов и вызовов. Данная статья представляет обзор существующих тенденций в развитии искусственного интеллекта и анализирует связанные с этим проблемы, предлагаются практические решения для решения этических проблем, связанных с искусственным интеллектом. Основные аспекты, рассмотренные в статье, включают безопасность и конфиденциальность данных, а также роль международного сообщества и законодательства в регулировании применения искусственного интеллекта. Автор обращает внимание на необходимость разработки и внедрения этических стандартов в области искусственного интеллекта, чтобы обеспечить прозрачность и ответственность в использовании технологии.

Цель: исследовать современные этические проблемы искусственного интеллекта и дать практические рекомендации.

Методология: анализ, синтез, сравнение, прогнозирование.

Научная новизна заключается в кросс-дисциплинарном подходе к анализу проблемы, объединяя в себе знания из области информатики, этики, права, социологии и других дисциплин.

Ключевые слова: этика, искусственный интеллект, цифровая этика, этика искусственного интеллекта, "Кодекс этики в области искусственного интеллекта"

Введение

Искусственный интеллект (ИИ) представляет собой комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении задач результаты. Прогресс в области ИИ привлекают внимание общественности к цифровой этике и становятся ключевым фактором в дискуссиях об этических аспектах цифровых технологий. Каковы последствия способности ИИ систем принимать решения? Какие этические вопросы возникают из этих решений? Могут ли ИИ системы нести ответственность за свои действия? Каким образом можно контролировать эти системы? Эти и другие вопросы, связанные с этой темой, в настоящее время привлекают внимание исследователей.

Этика искусственного интеллекта изучает нравственную ответственность разработчиков интеллектуальных систем за последствия их работы, в итоге возникают следующие этические вопросы, связанные с применением технологий искусственного интеллекта [10, с. 20]:

1) этические аспекты взаимодействия с системами искусственного интеллекта как объектами, то есть инструментами, созданными и используемыми людьми (конфиденциальность, непрозрачность, предвзятость);

2) этические вопросы, связанные с системами искусственного интеллекта в качестве субъектов, то есть этикой самих ИИ (искусственная мораль, машинная этика);

3) задача будущего сильного ИИ приведет к "технологической сингулярности" — моменту, когда развитие искусственного интеллекта станет необратимым и неуправляемым, вызывая радикальные изменения в человеческой цивилизации.

Некоторые авторы сосредотачивают внимание на проблеме обеспечения безопасности разрабатываемых интеллектуальных систем, учитывая их способность к рекурсивному улучшению. Это весьма серьезная, прогнозируемая, проблема, и, если изначальные версии искусственного интеллекта будут иметь значительные ограничения в области безопасности, будет сложно гарантировать сохранение этих ограничений в последующих поколениях. Большинство ученых утверждает, что исследования в области разработки сильного искусственного интеллекта предполагают о правильно настроенных компьютерах, которые могут мыслить, понимать и обладать другими когнитивными способностями, что, по сути, является неэтичным из-за возможной боли, которая может причиняться искусственному интеллекту.

Противоречивые мнения относительно будущего искусственного интеллекта вызывают общественный диалог [5]. Критики ИИ считают, что сочетание его с большими массивами данных может привести к тому, что ИИ станет реальной угрозой для человечества. Такие бизнесмены, как Илон Маск и Билл Гейтс, разделяют эти опасения. Уже сегодня мы можем наблюдать за утратой конфиденциальности и усилением контроля над гражданами в странах, таких как Сингапур, где программные продукты оказывают влияние на экономическую стратегию, иммиграционные правила, рынок недвижимости и образовательные стандарты. Программные системы уже применяют «убеждающий вычислительный» метод, направленный на формирование определённого поведения у людей, и без введения соответствующего законодательства эта практика, судя по всему, останется без изменения.

Обсуждения о морализации технологий считаются важными в настоящее время [3, с. 23]. Под морализацией технологий понимается осознанное развитие технологий с целью формирования этического поведения и принятия решений. Люди имеют возможность и ответственность морализировать не только других людей, но и своё физическое окружение, включая разработанные и принятые технологии. Одним из ключевых вопросов в текущих дискуссиях является возможность морализации технологий по демократическим методам. Аргументы в пользу искусственного интеллекта представляют чрезвычайно оптимистичную перспективу его развития [13].

Приверженцы искусственного интеллекта выражают опасения относительно возможных негативных последствий регулирования для развития ИИ. Они предлагают будущим исследователям рассматривать усилия по контролю за ИИ как препятствие угроз на пути человеческого прогресса, подчеркивая отсутствие юридически закрепленного общего мнения относительно четкого определения искусственного интеллекта и высказывая мнение: «невозможно регулировать то, что не является четко определенным». Таким образом они считают, что пока еще рано говорить о необходимости регулирования ИИ, особенно если это может препятствовать развитию, которое имеет значение для будущего человечества, отмечая при этом, что системы искусственного интеллекта включают в себя множество компьютерных программ, написанных давно (до появления элементов ИИ), и поэтому несправедливо возлагать ответственность на разработчиков этих программ за последствия, вызванные действием системы ИИ, а также на отсутствие согласия относительно принципов создания систем искусственного интеллекта.

Некоторые ученые рассматривают новый подход к разработке систем искусственного интеллекта, который учитывает человеческие ценности и этику, предлагая использовать принципы открытости, ответственности и прозрачности для улучшения процесса создания таких систем [9, с. 15]. Другие исследователи, напротив, считают, что создание "злонамеренного" искусственного интеллекта представляет собой интересную альтернативу попыткам разработать безопасные системы ИИ с учетом этических норм. Вывод: если возможно создать вредоносный искусственный интеллект, то ученые должны публиковать примеры проектов ИИ с негативными результатами и делиться информацией, чтобы помочь разобраться в причинах таких ситуаций и способах их предотвращения. Эта дискуссия может быть включена в обсуждение о моральных агентах в области искусственного интеллекта, то есть о машинах, способных принимать моральные решения и нести ответственность за свои

действия. Эта концепция связана с подходом, известным как машинная этика, где машины рассматриваются как активные участники процесса [7, с. 11]. Основные принципы машинной этики в настоящее время реализованы в современной робототехнике.

Принимая во внимание возможные негативные последствия, которые искусственный интеллект может повлечь за собой в различных областях, считается, что междисциплинарный подход является ключом к успешной борьбе с этими проблемами. Выделяются три основных направления, на которые стоит обратить внимание:

1) этическое управление (рассмотрение основных этических аспектов использования искусственного интеллекта, таких как справедливость, прозрачность и конфиденциальность);

2) понятность и возможность интерпретации (эти два понятия могут рассматриваться как инструменты для повышения алгоритмической справедливости, прозрачности и ответственности);

3) этический аудит (для сложных алгоритмических систем необходимы механизмы ответственности, которые не могут полагаться только на интерпретируемость, поэтому предлагаются механизмы аудита в качестве возможных решений). Также существует проблема юридической ответственности за действия искусственного интеллекта [6].

Рассматривается вопрос о том, кто несет ответственность за ущерб, причиненный искусственным интеллектом, проанализируем следующие сценарии:

а) если рассматривать искусственный интеллект как обычный инструмент (подобно молотку или гаечному ключу, то есть как "ИИ как инструмент" без собственной воли), то в этом случае применяется дополнительная ответственность за действия искусственного интеллекта;

б) если рассматривать искусственный интеллект как полностью автономный (как "ИИ как разум"), то в этом случае системы искусственного интеллекта должны осознавать свои действия и нести за них ответственность. Авторы приходят к выводу, что в настоящее время искусственный интеллект не признается юридическим лицом, соответственно применяется концепция "искусственного интеллекта как инструмента", и поэтому правила дополнительной ответственности регулируют поведение искусственного интеллекта, а эта ответственность распространяется на разработчиков, пользователей и владельцев. Искусственный интеллект все шире используется в различных областях медицины: от диагностики и принятия клинических решений до биомедицинских исследований, разработки лекарств, персонализированной медицины, телемедицины, медицинского образования и многого другого. Одним из ключевых вопросов здесь является вопрос о том, кто несет ответственность, когда искусственный интеллект ставит диагноз или выбирает метод лечения. Кроме этого, этические вопросы также включают в себя необходимость прозрачности и понятности интеллектуальных алгоритмов, используемых в процессе принятия медицинских решений.

В последнее время стала актуальной идея активного участия в развитии технологий в общем и искусственного интеллекта, в частности. Это подразумевает не только предотвращение негативных последствий технологий, но и реализацию некоторых позитивных эффектов. Одним из способов продемонстрировать принципиальную ответственность является создание технологий на основе ценностей, что делает учет этических аспектов обязательным при разработке новых технологий. При использовании ценностно-ориентированных подходов к развитию технологий искусственного интеллекта возникают серьезные вопросы о включении этических принципов в эти сложные системы. Внедрение позитивных ценностей не лишено опасностей, так как возможны негативные реакции на технологии искусственного интеллекта, предназначенные для контроля над поведением человека (включая его постоянное поведение). Одной из основных опасностей является угроза свободы личности и замещение демократии технократией. Идея того, что технологии (а не люди) контролируют власть, напрямую связана с мыслью о том, что ограничение автономии человека угрожает его достоинству. Так же существует опасность того, что, когда принятие этических решений поручается машинам, люди могут потерять способность к самостоятельному принятию решений или стать ленивыми.

В настоящее время предпринимаются усилия для выработки общих принципов создания искусственного интеллекта [8]. В 2019 году Европейская комиссия опубликовала "Руководящие принципы этики для надежного искусственного интеллекта" (The Ethics Guidelines for Trustworthy AI), которые устанавливают этические принципы и связанные с ними ценности и которые необходимо соблюдать при разработке, внедрении и использовании систем ИИ. В них отмечается, что разработка, внедрение и использование систем искусственного интеллекта должны соответствовать этическим принципам уважения человеческой автономии, предотвращения вреда, справедливости и подотчетности.

Для решения различных проблем (включая этические, юридические, социальные и др.) я считаю важным учреждение независимого международного органа по регулированию искусственного интеллекта. Этот орган должен объединить подходы различных стран, частных компаний и научных сообществ в понимании и развитии искусственного интеллекта. Несмотря на наличие множества академических и государственно-частных инициатив, направленных на поддержку исследований и разработок в области ИИ, все еще необходимо создание независимого органа, который поможет улучшить понимание политиками вопросов, связанных с искусственным интеллектом. Международное сотрудничество необходимо для предотвращения конфликтов, возникающих из-за различий в национальных законодательных подходах. Также важно создание независимой международной научной организации для разработки четкого научного взгляда на информационно-коммуникационные технологии и искусственный интеллект.

Использованная литература

1. Гуров О.Н. Этичное взаимодействие с интеллектуальными системами // Искусственные общества. – 2020. – Т. 15
2. Гусев А.В., Добридюк С.Л. Искусственный интеллект в медицине и здравоохранении // Информационное общество. – 2021. – № 4–5. – с. 78–93.
3. Дедюлина М.А. «Морализация технологий»: от компьютерных артефактов к социальным практикам // Философские проблемы информационных технологий и киберпространства. – 2019. – № 2 (10). – с. 75–86.
4. Кодекс этики в сфере искусственного интеллекта / Альянс в сфере искусственного интеллекта. – URL: <https://a-ai.ru/ethics/index.html> (дата обращения: 24.05.2023).
5. Лапаев Д.Н., Морозова Г.А. Искусственный интеллект: за и против // Развитие и безопасность. – 2020. – № 3 (7). – с. 70–77.
6. Лаптев В.А. Понятие искусственного интеллекта и юридическая ответственность за его работу // Право. Журнал Высшей школы экономики. – 2019. – № 2. – С. 79–102.
7. Макулин А.В. Этический калькулятор: от философской «вычислительной морали» к машинной этике искусственных моральных агентов (ИМА) // Общество: философия, история, культура. – 2020. – № 11 (79). – С. 18–27.
8. Мамина Р.И., Ильина А.В. Искусственный интеллект: в поисках формализации этических оснований // Дискурс. – 2022. – Т. 8, № 6. – с. 17–30.
9. Шляпников В.В. Искусственный интеллект: эмпатия и подотчетность // Общество. Среда. Развитие. – 2022. – № 3 (64). – с. 100–103.
10. Этика и «цифра»: этические проблемы цифровых технологий – М: РАНХиГС, 2020 – 207 с.
11. Machine Ethics / ed. by M. Anderson, S. Anderson. – New York; Cambridge: Cambridge University Press, 2021. – 548 p.
12. Buch V.H., Ahmed I., Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities // British Journal of General Practice. – 2018. – Vol. 68, iss. 668. – P. 143–144.
13. Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. – 2018. – Vol. 376

SÜNI İNTELLEKTİN ETİKASI: MEYİLLƏR, ÇAĞIRIŞLAR, HƏLLƏR

Tamerlan Hacıyev

Xülasə: müasir cəmiyyətdə süni intellektin tətbiqi müxtəlif səviyyələrdə və müxtəlif sahələrdə baş verir ki, bu da bir sıra etik suallar və çağırışlar yaradır. Bu məqalə süni intellektin inkişafındakı mövcud tendensiyaların icmalını təqdim edir və bununla bağlı problemləri təhlil edir. Məqalədə müzakirə olunan əsas cəhətlər məlumatların təhlükəsizliyi və məxfiliyi və texnologiyaları öyrənmə alqoritmlərindəki ədalət məsələlərini əhatə edir.

Müəllif, texnologiyanın istifadəsində şəffaflıq və məsuliyyət təmin etmək üçün süni intellekt sahəsində etik standartların hazırlanması və tətbiq edilməsinin zəruriliyinə diqqət çəkir. Xüsusilə, Sİ alqoritmlərinin sosial və mədəni bərabərsizlikləri artırmamasını və ya yaymamasını necə təmin etmək barədə suallar müzakirə olunur.

Məqalədə müzakirə olunan digər vacib cəhət süni intellektin tətbiqinin tənzimlənməsində beynəlxalq ictimaiyyətin və qanunvericiliyin rolu, habelə onun istifadəsinə nəzarət və nəzarət mexanizmləridir.

Məqalədə süni intellektlə əlaqəli etik problemləri həll etmək üçün praktik həllər də təklif olunur. Ümumiyyətlə, məqalə süni intellekt etikası probleminin mövcud vəziyyətinin hərtərəfli icmalını təqdim edir və gələcəkdə bu problemlərin həlli üçün praktik tövsiyələr təqdim edir.

Məqsəd: süni intellektin müasir etik problemlərini araşdırmaq və praktik tövsiyələr vermək.

Metodologiya: analiz, sintez, müqayisə, proqnozlaşdırma.

Elmi yenilik kompüter elmləri, etika, hüquq, sosiologiya və digər fənlər sahəsindəki bilikləri birləşdirərək problemin təhlilinə çarpaz-intizam yanaşmasından ibarətdir.

Açar sözlər: etika, süni intellekt, rəqəmsal etika, süni intellekt etikası, "süni intellekt etikası məcəlləsi"

THE ETHICS OF ARTIFICIAL INTELLIGENCE: TRENDS, CHALLENGES AND SOLUTIONS

Tamerlan Hacıyev

Abstract: In today's society, artificial intelligence is being introduced at various levels and fields, creating a range of ethical concerns and challenges. This paper provides an overview of current developments in artificial intelligence and examines the issues associated with them. The key aspects discussed include data security and privacy, as well as fairness in machine learning algorithms. The author emphasizes the need for developing and implementing ethical standards in the field of artificial intelligence to ensure transparency and accountability in the use of this technology. They discuss how to prevent AI algorithms from exacerbating or perpetuating social and cultural inequalities.

Another important topic discussed in the article is the role of international cooperation and legislation in regulating AI use, as well as monitoring and oversight mechanisms. The article also proposes practical solutions to address ethical issues related to AI.

Overall, the article provides an overview of current ethical challenges in AI and offers recommendations for addressing them in the future.

Purpose is to investigate these issues and provide practical solutions.

Methodology: analysis, synthesis, comparison, forecasting

The scientific novelty of this study lies in its cross-disciplinary approach to the problem, combining knowledge from various fields such as computer science, ethics, law and sociology.

Keywords: ethics, artificial intelligence, digital ethics, the ethics of artificial intelligence, Code of Ethics for Artificial Intelligence.

Rəyçi: f.e.d., prof. İlham Məmmədzadə