**УДК 004**

**Ратнер Н.П.**

Ассоциация «It_этика»

(г. Владимир, Россия)

# ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ВЫЗОВЫ, РИСКИ И РЕШЕНИЯ

***Аннотация**: вся совокупность достижений в области искусственного интеллекта (ИИ) привлекли внимание общественности к цифровой этике и вызвали множество дискуссий по этическим проблемам, связанным с цифровыми технологиями. В данной статье анализируются аргументы сторонников и противников искусственного интеллекта, различные подходы к разработке систем искусственного интеллекта, этические проблемы, связанные с использованием технологий искусственного интеллекта, включая вопросы управления искусственным интеллектом и концепции активного и ответственного развития технологий искусственного интеллекта — общие принципы развития систем искусственного интеллекта, изложенные в учредительном документе. Методологической основой данной работы является диалектический метод, а в процессе исследования были приняты сравнительный метод и метод анализа литературы. Источником является то, что отечественные и зарубежные авторы занимаются исследованием различных этических проблем искусственного интеллекта, европейские «Руководящие принципы этики для надежного искусственного интеллекта» и российский «Кодекс этики в сфере искусственного интеллекта».*

*Выявляется важность морализаторства технологии искусственного интеллекта, то есть сознательной разработки технологий для формирования этического поведения и решений. Открытым является вопрос о том, как найти демократический способ морализировать технологии, поскольку технология отличается от закона тем, что она может ограничивать свободу человека, а не быть результатом демократических процессов. Утверждается, что существует необходимость создания независимой международной научной организации для выработки четкого научного взгляда на искусственный интеллект и создания независимого международного органа по регулированию искусственного интеллекта, который объединит методы стран, частных компаний и понимание этого.*

***Ключевые слова**: искусственный интеллект, цифровая этика, этика искусственного интеллекта.*

Искусственный интеллект (ИИ) – это совокупность множества различных технологий, связанных с моделированием интеллектуального поведения компьютерных систем. Достижения в области искусственного интеллекта привлекли внимание общественности к цифровой этике и являются движущей силой многих дискуссий по этическим вопросам, связанным с цифровыми технологиями. Что значит способность систем ИИ принимать решения? Каковы моральные последствия этих решений? Могут ли системы ИИ нести ответственность за свои решения? Как можно управлять этими системами? Эти и многие другие связанные с ними вопросы сейчас находятся в центре пристального внимания исследователей.

В данной статье анализируются аргументы сторонников и противников искусственного интеллекта, различные подходы к разработке систем искусственного интеллекта, этические проблемы, связанные с использованием технологий искусственного интеллекта, включая вопросы управления искусственным интеллектом и концепции активного и ответственного развития. технологий искусственного интеллекта — общие принципы развития систем искусственного интеллекта, изложенные в учредительном документе. Методологической основой данной работы является диалектический метод, а в процессе исследования были приняты сравнительный метод и метод анализа литературы. Диалектика рассматривает искусственный интеллект как сложное и противоречивое явление. Сравнительные методы используются для сравнения аргументов сторонников и противников ИИ, а также для анализа подходов к развитию систем ИИ.

Метод анализа документов применялся при изучении принятых Европейской комиссией «Руководящих принципов этики для надежного искусственного интеллекта» и разработанного ассоциацией «Альянс в сфере искусственного интеллекта» российского «Кодекса этики в сфере

искусственного интеллекта». Источниками являются отечественные и зарубежные ученые, занимающиеся исследованиями различных этических проблем в области искусственного интеллекта, европейских этических принципов и российских этических рекомендаций в области искусственного интеллекта. Этика искусственного интеллекта изучает моральную ответственность разработчиков интеллектуальных систем за последствия их функциональности. Выделяют следующие этические проблемы, возникающие при использовании технологий искусственного интеллекта [10, 20]:

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

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

3) проблема возможного будущего сверхразума искусственного интеллекта, ведущего к «технологической сингулярности», то есть моменту, когда развитие искусственного интеллекта станет неуправляемым и необратимым, что приведет к радикальному изменению характера человеческой цивилизации. Некоторые авторы акцентируют внимание на проблеме доказательства безопасности разрабатываемых интеллектуальных систем с учетом их способности к рекурсивному самосовершенствованию.

Это серьезная проблема, и ожидается, что даже если первоначальные версии интеллектуальных систем будут иметь существенные ограничения безопасности, будет сложно гарантировать, что последующие поколения систем сохранят эти ограничения. Другие утверждают, что исследования по разработке сильного ИИ предполагают, что правильно запрограммированные компьютеры могут думать, понимать и обладать другими когнитивными способностями, что по своей сути неэтично из-за боли, которую это может причинить ИИ.

Отмечается, что фиксированный набор моральных правил может привести к различным противоречиям и трудностям [15]. На примере беспилотных автомобилей, чтобы подчеркнуть некоторые из этих трудностей, утверждалось, что системы ИИ должны иметь встроенные моральные правила, соответствующие ценностям их владельцев, а не универсальный набор моральных ценностей. Например, универсальные ценности могут непреднамеренно дискриминировать определенные группы. Обсуждается необходимость гибких этических рамок в контексте автоматизированных военных систем и подчеркивается важность присутствия людей, которые могут решать, когда применять оружие, а когда не применять его.

Интересен спор между противниками и сторонниками искусственного интеллекта [5]. Аргументы против ИИ предполагают, что, объединив ИИ с большими данными, мы в конечном итоге окажемся в ситуации, когда ИИ, вероятно, будет представлять серьезную угрозу человечеству. Эти взгляды особенно разделяют такие дальновидные предприниматели, как Илон Маск и Билл Гейтс. Мы уже можем видеть, как нарушается конфиденциальность и контролируются люди в таких странах, как Сингапур, где компьютерные программы влияют на экономическую и иммиграционную политику, рынки недвижимости и школьные программы. Программные системы уже используют «убеждающие вычисления», чтобы программировать людей на определенное поведение, и если не будет принято законодательство, эта тенденция сохранится.

В связи с этим представляются важными дискуссии о морализации технологий [3, 23]. Морализация технологий относится к сознательному развитию технологий для формирования этического поведения и принятия решений. Люди должны и могут морализировать не только других, но и свое физическое окружение, включая разработанные и принятые технологии. Один из самых больших вопросов в сегодняшних дебатах заключается в том, можно ли морально морализировать технологию демократическим путем. Доводы в пользу

искусственного интеллекта демонстрируют чрезвычайно оптимистичный взгляд на искусственный интеллект [18].

Сторонники ИИ обеспокоены тем, что регулирование может помешать успеху ИИ, и советуют будущим исследователям рассматривать усилия по регулированию ИИ как «темный век» человеческого прогресса. Они обсудили отсутствие консенсуса по четкому определению искусственного интеллекта и высказали идею о том, что невозможно регулировать то, что невозможно определить. В результате они считают, что еще слишком рано рассматривать вопрос о регулировании искусственного интеллекта, особенно если такое регулирование будет препятствовать развитию, которое может быть важным для выживания человечества. Они также отмечают, что с точки зрения ответственности системы ИИ состоят из множества компьютерных программ, некоторые из которых могли быть написаны за много лет до появления элементов ИИ, поэтому было бы несправедливо привлекать к ответственности разработчиков этих программ. Нести ответственность за результаты, вызванные искусственным интеллектом системы. Также нет соглашения о создании систем искусственного интеллекта.

Например, одними исследователями рассматривается подход к разработке систем ИИ, который учитывает человеческие ценности и этику, и предлагается использовать принципы подотчетности, ответственности и прозрачности для усовершенствованного процесса разработки систем искусственного интеллекта [9, 15]. Другие исследователи придерживаются противоположного подхода и предполагают, что интересной альтернативой попыткам разработать безопасные системы ИИ с соблюдением этических норм является создание «злонамеренного» ИИ.

Они пришли к выводу, что если такой вредоносный ИИ возможен, то исследователи обязаны публиковать примеры любых проектов ИИ с отрицательными результатами и делиться данными, чтобы помочь понять, почему такие вещи происходят и как их предотвратить. Эту дискуссию можно

включить в дискуссию об искусственных моральных агентах, то есть машины могут в некотором смысле стать моральными агентами, ответственными за свои действия.

Эта идея связана с подходом, называемым машинной этикой, где машины рассматриваются как субъекты [7, 11]. Основные идеи машинной этики в настоящее время воплощены в современной робототехнике.

Учитывая потенциальный вред, который ИИ может нанести в различных сферах, предполагается, что междисциплинарный подход является ключом к успеху, и выделяются три области, на которых следует сосредоточиться:

1) этическое управление (рассмотрение наиболее важных вопросов этики ИИ, таких как справедливость, прозрачность и конфиденциальность);

2) объяснимость и интерпретируемость (эти две концепции можно рассматривать как возможные механизмы повышения алгоритмической справедливости, прозрачности и подотчетности);

3) этический аудит (для очень сложных алгоритмических систем механизмы подотчетности не могут полагаться исключительно на интерпретируемость, поэтому в качестве возможных решений предлагаются механизмы аудита). Существует также и проблема юридической ответственности ИИ [6].

Например, исследуется ответственность за ущерб, причиненный ИИ [14], и анализируются следующие варианты:

а) если системы ИИ рассматривать так же, как молоток или гаечный ключ (то есть «ИИ как инструмент» без собственного независимого волеизъявления), то в этом случае применяется субсидиарная ответственность за действия ИИ;

б) если ИИ рассматривается как полностью автономный («ИИ как человек»), то в этом случае системы ИИ должны знать о своих действиях и нести ответственность за них. Авторы приходят к выводу, что искусственный интеллект в настоящее время не признан юридическим лицом и поэтому применяется теория «искусственного интеллекта как инструмента», и поэтому

правила субсидиарной ответственности регулируют поведение искусственного интеллекта и что эта ответственность распространяется на разработчиков, пользователей и владельца. Система искусственного интеллекта. Искусственный интеллект все чаще используется практически во всех областях медицины: диагностика, принятие клинических решений, биомедицинские исследования и разработка лекарств, персонализированная медицина, телемедицина, медицинское образование и многое другое. Одним из основных вопросов в данном случае является вопрос ответственности, когда интеллектуальная программа ставит диагноз или выбирает лечение.

Кроме того, этические аспекты включают требования к прозрачности и объяснимости интеллектуальных алгоритмов, используемых при принятии медицинских решений. В последние годы возникла идея активно отвечать за развитие технологий в целом и технологий искусственного интеллекта в частности. Это означает не только предотвращение негативного воздействия технологий, но и реализацию некоторых положительных последствий. Один из способов проявить упреждающую ответственность — это проектирование, основанное на ценностях, что делает этические соображения обязательным требованием при проектировании технологий.

Когда к технологиям искусственного интеллекта применяются ценностно-ориентированные подходы, проблемы, связанные с выбором включения этических ценностей в эти сложные технологии, становятся еще более серьезными. Идея включения позитивных ценностей не лишена риска. В частности, существует множество негативных реакций на технологии искусственного интеллекта, созданные для управления поведением человека, в том числе перманентным поведением. Вероятное беспокойство заключается в том, что свобода человека находится под угрозой и демократия заменяется технократией. Идея о том, что власть контролируют технологии, а не люди, тесно связана с представлением о том, что уменьшение автономии представляет угрозу человеческому достоинству. Существует также риск того, что, когда принятие

этических решений делегируется машинам, люди могут стать ленивыми или даже неспособными принимать этические решения. Важно подчеркнуть, что технология отличается от закона тем, что технология ограничивает свободу человека, а не является результатом демократических процессов. Поэтому, как уже говорилось ранее, открытым остается вопрос о том, как найти демократический способ морализировать технологии.

В последнее время предпринимаются попытки сформулировать общие принципы разработки систем ИИ [8]. В 2019 году Европейской комиссией были опубликованы «Руководящие принципы этики для надежного искусственного интеллекта» (The Ethics Guidelines for Trustworthy AI), определяющие этические принципы и связанные с ними ценности, которые необходимо соблюдать при разработке, внедрении и использовании систем ИИ. Они четко заявили, что разработка, внедрение и использование систем искусственного интеллекта должны соответствовать этическим принципам уважения человеческой автономии, предотвращения вреда, справедливости и подотчетности. В России в октябре 2021 года появился «Кодекс этики в сфере искусственного интеллекта» [4].

Он разработан консорциумом в области искусственного интеллекта и подписан ведущими научно-исследовательскими институтами России и крупнейшими технологическими компаниями. Кодекс является рекомендательным документом для участников сферы искусственного интеллекта (российские и иностранные компании, государственные органы) и декларирует принципы ответственности, информационной безопасности, контроля рекурсивного самосовершенствования систем искусственного интеллекта. В результате искусственный интеллект стал одним из основных вопросов цифровой этики. Хотя многие исследователи и технологи воодушевлены потенциалом искусственного интеллекта, многие относятся к этому с осторожностью. Позитивное влияние искусственного интеллекта (беспилотные автомобили повышают безопасность, цифровые помощники,

роботы, выполняющие тяжелый ручной труд, мощные алгоритмы, которые делают полезные и важные выводы из больших объемов данных) и негативное влияние (автоматизация приводит к сокращению рабочих мест, увеличению неравенства) обсуждаются), угрозы конфиденциальности). Важным вопросом является регулирование взаимоотношений между сферами искусственного интеллекта и робототехники. Можно сказать, что развитие искусственного интеллекта достигло значительного прогресса благодаря четырем ключевым факторам: усовершенствованным статистическим моделям, огромным наборам данных, дешевой вычислительной мощности и широкому внедрению технологий в жизнь человека. Однако текущие исследования в области искусственного интеллекта в основном проводятся частными предприятиями, и необходимо решить проблему отсутствия социальной и политической ответственности и долгосрочного планирования.

Для этого необходимо создать независимый международный орган по регулированию искусственного интеллекта, который объединит подходы стран, частных компаний и научных кругов к пониманию искусственного интеллекта. Несмотря на то, что в настоящее время существует множество академических и государственно-частных платформ, поддерживающих правительства в продвижении исследований и разработок в области ИИ, все еще существует потребность в независимом органе, который поможет улучшить навыки политиков по вопросам, связанным с ИИ. Международная интеграция необходима, чтобы избежать конфликтов, связанных с различными национальными законодательными подходами. Кроме того, существует необходимость создания независимой международной научной организации для выработки четкого научного взгляда на ИКТ и искусственный интеллект.

Также необходимы расширение прав и возможностей каждого гражданина (например, посредством разработки новых инструментов, таких как цифровые помощники) и большая прозрачность (как на уровне частных

компаний, так и на уровне правительств), важна децентрализация услуг, данных и компьютерных систем.

## СПИСОК ЛИТЕРАТУРЫ:

1. Гуров О.Н. Этичное взаимодействие с интеллектуальными системами // Искусственные общества. – 2020. – Т. 15,
2. Гусев А.В., Добриднюк 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. Этикаи «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. Artifi cial 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 artifi cial intelligence: ethical, legal and technical opportunities and challenges // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. – 2018. – Vol. 376

**Ratner N.P.**

It_ethics Association

(Vladimir, Russia)

# ETHICS OF ARTIFICIAL INTELLIGENCE: CHALLENGES, RISKS AND SOLUTIONS

***Abstract**: the whole set of achievements in the field of artificial intelligence (AI) has attracted public attention to digital ethics and caused a lot of discussions on ethical issues related to digital technologies. This article analyzes the arguments of supporters and opponents of artificial intelligence, various approaches to the development of artificial intelligence systems, ethical issues related to the use of artificial intelligence technologies, including issues of artificial intelligence management and the concept of active and responsible development of artificial intelligence technologies — the general principles of the development of artificial intelligence systems set out in the founding document. The methodological basis of this work is the dialectical method, and in the process of research, the comparative method and the method of literature analysis were adopted. The source is that domestic and foreign authors are engaged in research on various ethical problems of artificial intelligence, the European "Ethics Guidelines for Reliable Artificial Intelligence" and the Russian "Code of Ethics in the field of Artificial Intelligence".*

*The importance of moralizing artificial intelligence technology, that is, the conscious development of technologies for the formation of ethical behavior and decisions, is revealed. The question of how to find a democratic way to moralize technology is open, since technology differs from the law in that it can restrict human freedom, and not be the result of democratic processes. It is argued that there is a need to create an independent international scientific organization to develop a clear scientific view of artificial intelligence and to create an independent international body for the regulation of artificial intelligence, which will combine the methods of countries, private companies and the understanding of this.*

***Keywords**: artificial intelligence, digital ethics, ethics of artificial intelligence.*