

**Миндигулова Арина Александровна,**

старший преподаватель, кафедра философии, Новосибирский государственный технический университет  
E-mail: mindigulova@corp.nstu.ru

Данная статья нацелена на осмысление состояния этического регулирования применения технологий искусственного интеллекта (ИИ) в современном цифровом обществе. Обсуждается сама суть совместимости понятия этики и технологий ИИ с точки зрения отечественного и зарубежного опыта. Также в работе фиксируются ключевые факты эволюции понятия этики, причины, способствовавшие необходимому внедрению этического регулирования использования технологий ИИ, а также основные подходы к использованию этики в контексте ИИ. Кроме того, в статье приведён обзор основополагающих зарубежных и отечественных регламентирующих документов. Помимо этого, проанализированы конкретные критерии и принципы этичности применения технологий ИИ в отечественном и зарубежном опыте. Дан краткий обзор возможных причин недостаточной включенности этического регулирования в сфере ИИ.

**Ключевые слова:** искусственный интеллект, этика, цифровизация.

## Введение

Технологии искусственного интеллекта (ИИ) стали полноправными участниками отношений всех сфера жизнедеятельности современного общества. Несмотря на то, что в научном дискурсе наличествуют самые различные диаметрально противоположные суждения относительно такой повсеместности ИИ, налицо неуклонная и последовательная цифровизация общества. Развитие и рост технологий ИИ во всех сферах экономики ставит важные вопросы как перед представителями ИТ-сферы, так и человечеством в целом. Прежде всего – вопросы о том: какие/ чьи ценности должны согласовываться с внедрением систем ИИ [1]; должны ли для них в принципе существовать какие-либо нормы этики или сама их сущность с данным понятие несовместима? кто должен стать авторами этих норм? и стоит ли ставить знак равенства между моральными устоями авторов конкретных технологий ИИ и нормами самих «машин»?

Существует множество, порой полярных, точек зрения относительно совместимости понятий этики и искусственного интеллекта. Так, например, Ж. Брисон (J. Bryson) выдвигает – «создание роботов такими, чтобы они заслужили быть моральными индивидуумами, само по себе может быть истолковано как аморальное действие, особенно если учесть, что этого, очевидно, можно избежать, поскольку создание таких технологий – это всегда фактически выбор человека» [2]. В свою очередь, Дж. Генрихс (Jh. Heinrichs) считает, что «поскольку ответственность может взять на себя только тот, кого можно наказать, сама «машина» не подходит для принятия на себя ответственности. Поскольку она не испытывает страданий и ее поведение не может быть исправлено наказанием, не говоря уже о похвале или порицании, возлагать ответственность на машину просто невозможно или уместно» [3]. Интерес представляет точка зрения, зафиксированная в докладе Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации: «сегодня какого-либо технологического решения еще не было, завтра оно появится, а послезавтра оно будет в доме или на работе у каждого. Будущее наступает все быстрее, и разговор об этических аспектах цифровизации и использования технологий становится все актуальнее» [4].

## Исторический анализ понятия «этика»

Прежде чем приступить к осмыслению ряда возможных этических проблем цифровизации современного общества, в контексте интеграции в него

технологий ИИ, целесообразно прояснить основные понятия, концепции и подходы.

В общем контексте, понятие «этика» древнегреческого происхождения – этос (ethos, ἦθος), означающее «местопребывание», а именно человеческое жилище, звериное логово, птичье гнездо. Ранее оно имело другой смысл, ставший со временем превалирующим – устойчивая природа какого-либо явления, в том числе характер, внутренний нрав живых существ. В свое время, Гераклит писал об этосе человека, имея в виду то, что на русском языке созвучно с «образ жизни», «характер». Аристотель, исходя из значения «этос» в качестве характера, внутренней природы, нрава, образовал прилагательное «этический» или «этосный» (ethicos, ἠθικός) относящийся к этосу. Им он обозначил особый класс качеств, относящихся к характеру человека, описывающих его совершенное состояние – этические добродетели. Этические добродетели (мужество, умеренность, щедрость и другие) отличаются как от природных свойств человека, так и от качеств его ума. Далее, от прилагательного «этический» Аристотель пришел к существительному «этика» (ἠθικά), являющемуся, с одной стороны, обобщением соответствующего класса добродетелей, а с другой, обозначением той области знаний, которая изучает человеческие добродетели [5]. В настоящее время, под этикой принято понимать философскую дисциплину, изучающую поведение людей на основе определённых моральных мотивов. Этика раскрывает последствия принятия социумом каких-либо решений, приверженности каких-либо идей, а также предпосылки и перспективы подобных действий.

Этика – это своего рода рефлексия морали, которая выделяется тем, что регулируется не нормативными актами, а чувствами индивида. Основными категориями этики являются добро и зло, счастье и справедливость, смысл жизни и долг, честь, совесть и достоинство и т.д. Среди концепций этики выделяют гедонизм, эвдемонизм, абсолютизм, утилитаризм, релятивизм и формализм. Неразрывно с понятием этики связано понятие гуманизма, на идее которого основывалось создание основных документов, регулирующих в том числе и вопросы этики и морали, например, Всеобщей декларации прав человека. К основным принципам гуманизма следует отнести: а) гарантии основных прав человека как общее условие для подлинно частного существования; б) поддержка слабых, выходящая за рамки обычных представлений данного общества о справедливости; в) формирование социальных и нравственных качеств, позволяющих личности самореализоваться, используя общественные ценности; г) современная этика – это достаточно часто прикладная этика, которая находит применение в самых разных сферах; е) этика открытых моральных проблем призвана решать вопросы, на которые человечество пока не нашло ответов (например, проблему эвтаназии); ж) прикладная этика занимается анализом

общественных угроз, например, экологического кризиса или ядерной войны, и проводит социальную экспертизу проектов и их возможных последствий [4].

## Причины воззрений этических аспектов ИИ

Обратимся к прикладной этике, под которой понимается область знания и поведения, предметом которой являются практические моральные проблемы, имеющие пограничный и открытый характер [5]. Примечательно, что процесс взаимодействия человека и техники служит одной из ключевых проблематик длительного научного дискурса. Сложилось мнение о том, что моральный вызов человеку был брошен ещё в момент создания парового двигателя. Например, Гиффорд (Gifford) выдвигает идею о том, что ранняя автоматизация ткацких станков и железных дорог привела к «замене ранее существовавшего стандарта строгой ответственности за правонарушение режимом небрежности» и как широкое внедрение автомобилей сделало необходимым создание «законов о финансовой ответственности», которые требовали от владельцев автомобилей либо приобретения страховки, либо предоставления доказательств того, что они располагали достаточными финансовыми ресурсами для оплаты требований» [6]. В настоящее время, технологии искусственного интеллекта – это уже не просто локальная автоматизация производства или увеличение количества автомобилей на дороге, а реальность в которой мы живем, со всеми присутствующими ей проблемами, в том числе и этическими.

Первыми маркерами вышеозначенных проблем выступили автокатастрофы с самоуправляемыми автомобилями Tesla (2016) и Uber (2018), протест разработчиков компании Google против участия созданных ими программных продуктов в военных проектах Минобороны США, случаи манипулирования доступностью информации, сексизма и расизма в алгоритмах распознавания лиц и таргетированной рекламы с использованием ИИ. Также общеизвестными служат проблемы использования ИИ органами государственной власти для контроля за гражданами [7]. Всё вышеперечисленное, а также целый ряд других спорных, с точки зрения этики событий, связанных с ИИ, побудил мировое сообщество к нормативному регламентированию этических аспектов данных технологий. В конце 2021 г. Генеральная конференция ЮНЕСКО (UNESCO General Conference) приняла рекомендацию об этических аспектах искусственного интеллекта («recommendation on the ethics of artificial intelligence») [8] с целью уменьшить риски и трудности, связанные с искусственным интеллектом (ИИ), особенно с точки зрения усугубления существующего неравенства, а также последствий для прав человека. Ранее в 2019 г. в России была принята «Национальная стратегия развития искусственного интеллекта на период до 2030 года» [9], содержащая, среди прочего и этическую сторону вопроса, согласующаяся с «Рекомендаци-

ям» ЮНЕСКО в части этических принципов, а в октябре 2021 г. был разработан, опубликован и подписан первыми организациями-добровольцами кодекс этики ИИ [10]. Основываясь на вышеозначенных документах, а также на дополнительных сведениях и примерах, совершим попытку ответить на вопрос: какое «поведение» технологий ИИ можно назвать этическим? И каковы условия для этической работы ИИ? Мы не претендуем на способность ответить на все возможные вопросы в данной проблематике, но позволим себе задать некий вектор рассуждений в таком сложном вопросе об этическом применении технологий ИИ.

## Этика и ИИ

Под этическим применением ИИ понимается систематическое нормативное осмысление этических аспектов ИИ на основе эволюционирующей комплексной, всеобъемлющей и многокультурной системы взаимосвязанных ценностных установок, принципов и процедур, способное ориентировать общества в вопросах ответственного учёта известных и неизвестных последствий применения ИИ-технологий для людей, сообществ, окружающей природной среды и экосистем, а также служить основой для принятия решений, касающихся применения или отказа от применения технологий на основе ИИ [8].

Зафиксируем этические ценностные установки и принципы действия технологий ИИ. Несмотря на тот факт, что за рамками данной статьи остались аспекты, которые касаются исключительно технической реализации, все они, в конечном итоге, могут быть отнесены к какому-либо из нижеперечисленных пунктов. К этическим принципам технологий ИИ в общем случае могут быть отнесены следующие характеристики.

1. Уважение, защита и поощрение прав человека и основных свобод и человеческого достоинства. *Это означает, в процессе деятельности технологий не должна наблюдаться зависимость, какой-либо вред или ограничение свобод человека.*

2. Обеспечение разнообразия и возможности выбора. *Технологии ИИ должны предлагать человеку различные варианты своего использования, изменение параметров конкретных технологий, адаптацию их под собственные нужды.*

3. Жизнь в справедливых сообществах. *Любые процессы, относящиеся к технологиям ИИ не должны вести к дискриминации, ущемлению свободы или обезличиванию.*

4. Безопасность и защищённость, а также соразмерность и непричинение вреда. *При наличии какой-либо, пусть даже потенциальной, угрозы безопасности или жизни человека ИИ должен оценить возможные риски и предпринять максимально возможные действия для их предотвращения.*

5. Устойчивость. *Применение технологий ИИ должно проходить с постоянным соотношением их влияния на устойчивость развития. Под устойчивостью развития понимается набор целей, сфор-*

*мулированных Организацией Объединённых Наций (ООН) в качестве маркёров устойчивого развития общества.*

6. Право на неприкосновенность частной жизни и защиту персональных данных. *Эти права являются основополагающими для защиты человеческого достоинства и независимости личности. Крайне важно, чтобы на любом этапе работы технологий ИИ любая обработка данных осуществлялась в полном соответствии с нормативными документами.*

7. Подконтрольность, прозрачность и подотчётность. *Существует несколько теорий о том, что неподконтрольный ИИ способен, фактически, полностью поработить человечество [11]. Автору сложно судить о потенциале данного события, но тем не менее неоспоримым является тот факт, что в любой момент времени человек должен осознавать, что происходит в технологиях ИИ. Как они развиваются в каждом конкретном случае. Без такого понимания человек способен лишиться базового ощущения безопасности своего существования.*

В последние годы Мировое сообщество сделало достаточно серьёзный шаг вперёд в вопросах установления и контроля этики технологического применения технологий ИИ. Во многих государствах существует и даже устоялась документальная фиксация ценностей и норм, но вместе с тем постоянно в обществе возникают споры, касающиеся данной проблематики. *Во-первых*, несмотря на крайне высокую частоту употребления понятия ИИ, в обществе так и не устоялось его определение. Это демонстрирует явный разрыв между конкретным контекстом исследований, разработок и применения, с одной стороны, и этическим их осмыслением, с другой [12]. Специалисты, работающие с этическими проблемами ИИ, должны иметь общее понятие об технической стороне процессов, происходящих в процессе работы технологий ИИ. Очевидно, что понимая базовые принципы, например, генерации, записи, обработки, распространения и передачи данных [13] можно задать этические принципы данных процессов, а значит, в перспективе, и систем ИИ в целом. Другими словами, стоит перенести общую абстрактную этику на микроуровень.

В качестве *второй* причины можно указать отсутствие регламентированной ответственности за нарушение этических принципов ИИ. Этика ИИ содержит в себе добровольные рамки совместной работы специалистов по этике и технического персонала. Исследования и разработки ИИ осуществляются в «закрытых отраслевых условиях», где «согласие пользователей, конфиденциальность и прозрачность часто упускаются из виду в пользу беспрепятственной функциональности, которая поддерживает бизнес-модели, ориентированные на прибыль» [14].

В настоящее время массовой цифровизации и, следующего за ней повсеместного внедрения технологий ИИ, этическое регулирование крайне

необходимо. Несмотря на то, что нормативная база подобного регулирования неуклонно качественно и количественно растет, периодически имеют место прецеденты, демонстрирующие, что подобных мер недостаточно. В этике искусственного интеллекта всё ещё отсутствует подкрепление, отражающее реальную ответственность за несоблюдение её руководящих принципов. Попытка фиксации норм этики людьми, не имеющими непосредственного отношения к сфере, для которой последние создаются не приводит ни к чему, кроме как к созданию некоего шаблонного документа, не несущего реальной смысловой нагрузки. Соответственно, и его использование становится лишним, побочным. В результате мы получаем ситуацию, где с одной стороны прописанные нормы этики являются «оторванными» от реальной точки их приложения и с другой стороны, нарушение их с точки зрения закона ничего за собой не несёт. В итоге всё приводит к тому, что цели, для которых разрабатываются и применяются системы ИИ, не соответствуют общественным ценностям или основным правам, таким как благодеяние, непричинение вреда, справедливость и объяснимость [15], [16].

## Заключение

Резюмируем:

1. Этическое регулирование аспектов применения ИИ крайне необходимо в текущих условиях массовой цифровизации.
2. За последние годы существенным образом увеличилось количество документов, описывающих основные принципы этики ИИ. Но носят они, по большей части, рекомендательный характер.
3. Отсутствие предметно ориентированных специалистов и законодательного подкрепления зачастую переводит этику ИИ в своего рода маркетинговую стратегию для привлечения внимания.
4. Вышеперечисленные аспекты этики ИИ ограничивают её реальное применение в настоящий момент, но дают направление для развития исследований в будущем.

## Литература

1. I. Gabriel. Artificial Intelligence, Values, and Alignment. *Minds & Machines* 30, 411–437, 2020. <https://doi.org/10.1007/s11023-020-09539-2>
2. J.J. Bryson. Patience is not a virtue: the design of intelligent systems and systems of ethics. *Ethics Inf Technol* 20, 15–26, 2018. <https://doi.org/10.1007/s10676-018-9448-6>
3. JH. Heinrichs. Responsibility assignment won't solve the moral issues of artificial intelligence. *AI Ethics*, 2022. <https://doi.org/10.1007/s43681-022-00133-z>
4. Этика и «цифра»: Этические проблемы цифровых технологий. – М.: РАНХиГС, 207 с, 2020.

5. А.А. Гусейнов. Великие пророки и мыслители. Нравственные учения от Моисея до наших дней. / Вече, 2009.
6. D.G. Gifford. Technological triggers to tort revolutions: steam locomotives, autonomous vehicles, and accident compensation. *J. Tort Law* 11(1), 71–143. 2018. <https://doi.org/10.1515/jtl-2017-0029>
7. T. Simonite. AI experts want to end «black box» algorithms in government // *Wired Business*. URL: <https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government>
8. Рекомендация об этических аспектах искусственного интеллекта Режим доступа [https://d-russia.ru/wp-content/uploads/2021/11/unesco\\_ai\\_etica\\_16-11-2021.pdf](https://d-russia.ru/wp-content/uploads/2021/11/unesco_ai_etica_16-11-2021.pdf).
9. Национальная стратегия развития искусственного интеллекта на период до 2030 года. Режим доступа <https://d-russia.ru/wp-content/uploads/2019/10/ukaz-ob-AI.pdf>
10. Кодекс этики в сфере искусственного интеллекта. Режим доступа [https://d-russia.ru/wp-content/uploads/2021/10/kodeks\\_etiki\\_ii.pdf](https://d-russia.ru/wp-content/uploads/2021/10/kodeks_etiki_ii.pdf)
11. A.A. Smirnov Anthropology of nomadism (Anthropological trends and the Northern cultural code) // *Man. Culture. Education*, 2014. No. 3. p. 5.
12. T. Hagendorff. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines* 30, 99–120. 2020. <https://doi.org/10.1007/s11023-020-09517-8>
13. B. de Bruin, L. Floridi. The ethics of cloud computing. *Science and Engineering Ethics*, 23(1), 21–39. 2017.
14. A. Campolo, M. Sanfilippo, M. Whittaker, K. Crawford. AI now 2017 report. Retrieved October 02, 2018. [https://assets.ctfassets.net/8wprhvnpc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/\\_AI\\_Now\\_Institute\\_2017\\_Report\\_.pdf](https://assets.ctfassets.net/8wprhvnpc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf).
15. M. Taddeo, L. Floridi. How AI can be a force for good. *Science*, 361(6404), 751–752, 2018.
16. A. Tegmark. *Life 3.0: Being human in the age of artificial intelligence*. New York: Alfred A. Knopf. 2017.

## ETHICS AND ARTIFICIAL INTELLIGENCE: PROBLEMS AND CONTRADICTIONS

Mindigulova A.A.

Novosibirsk State Technical University Novosibirsk

This article is aimed at understanding the state of ethical regulation of the use of artificial intelligence (AI) technologies in modern digital society. The very essence of the compatibility of the concept of ethics and AI technologies from the point of view of domestic and foreign experience is discussed. The paper also records the key facts of the evolution of the concept of ethics, the reasons that contributed to the necessary introduction of ethical regulation of the use of AI technologies, as well as the main approaches to the use of ethics in the context of AI. In addition, the article provides an overview of the fundamental foreign and domestic regulatory documents. In addition, the specific criteria and principles of ethics of the use of AI technologies in domestic and foreign experience are analyzed. A brief overview of the possible reasons for the lack of inclusion of ethical regulation in the field of AI is given.

**Keywords:** Artificial intelligence, ethics, digitalization.

## References

1. I. Gabriel. Artificial Intelligence, Values, and Alignment. *Minds & Machines* 30, 411–437, 2020. <https://doi.org/10.1007/s11023-020-09539-2>
2. J.J. Bryson. Patience is not a virtue: the design of intelligent systems and systems of ethics. *Ethics Inf Technol* 20, 15–26, 2018. <https://doi.org/10.1007/s10676-018-9448-6>
3. JH. Heinrichs. Responsibility assignment won't solve the moral issues of artificial intelligence. *AI Ethics*, 2022. <https://doi.org/10.1007/s43681-022-00133-z>
4. Ethics and "digit": Ethical problems of digital technologies. – M.: RANEPa, 207 p, 2020.
5. A.A. Huseynov. Great prophets and thinkers. Moral teachings from Moses to the present day. / Veche, 2009.
6. D.G. Gifford. Technological triggers to tort revolutions: steam locomotives, autonomous vehicles, and accident compensation. *J. Tort Law* 11(1), 71–143. 2018. <https://doi.org/10.1515/jtl-2017-0029>
7. T. Simonite. AI experts want to end «black box» algorithms in government // *Wired Business*. URL: <https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government>
8. Recommendation on ethical aspects of artificial intelligence Access mode [https://d-russia.ru/wp-content/uploads/2021/11/unesco\\_ai\\_etiic\\_16-11-2021.pdf](https://d-russia.ru/wp-content/uploads/2021/11/unesco_ai_etiic_16-11-2021.pdf).
9. National strategy for the development of artificial intelligence for the period up to 2030. Access mode <https://d-russia.ru/wp-content/uploads/2019/10/ukaz-ob-AI.pdf>
10. Code of Ethics in the field of artificial Intelligence. Access mode [https://d-russia.ru/wp-content/uploads/2021/10/kodeks\\_etiki\\_ii.pdf](https://d-russia.ru/wp-content/uploads/2021/10/kodeks_etiki_ii.pdf)
11. A.A. Smirnov Anthropology of nomadism (Anthropological trends and the Northern cultural code) // *Man. Culture. Education*, 2014. No. 3. p. 5.
12. T. Hagendorff. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines* 30, 99–120. 2020. <https://doi.org/10.1007/s11023-020-09517-8>
13. B. de Bruin, L. Floridi. The ethics of cloud computing. *Science and Engineering Ethics*, 23(1), 21–39. 2017.
14. A. Campolo, M. Sanfilippo, M. Whittaker, K. Crawford. AI now 2017 report. Retrieved October 02, 2018. [https://assets.ctfassets.net/8wprhvnnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/\\_AI\\_Now\\_Institute\\_2017\\_Report\\_.pdf](https://assets.ctfassets.net/8wprhvnnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf).
15. M. Taddeo, L. Floridi. How AI can be a force for good. *Science*, 361(6404), 751–752, 2018.
16. A. Tegmark. *Life 3.0: Being human in the age of artificial intelligence*. New York: Alfred A. Knopf. 2017.