



DOI: 10.18287/2782-2966-2024-4-1-24-30

НАУЧНАЯ СТАТЬЯ
УДК 165:004.8Дата поступления: 10.01.2024
рецензирования: 03.02.2024
принятия: 15.03.2024**Н.А. Ястреб**Вологодский государственный университет,
г. Вологда, Российская Федерация
E-mail: nayastreb@mail.ru
ORCID: <https://orcid.org/0000-0001-6635-6008>

Концепции этики искусственного интеллекта: от принципов к критическому подходу

Аннотация: в статье рассматривается проблема определения методологических и концептуальных оснований этики искусственного интеллекта. Показано, что подход, основанный на принципах, базируется на теории вложения ценностей, которая предполагает, что технические объекты могут либо сами быть носителями ценностей, либо, по меньшей мере, способствуют реализации определенных ценностей. При этом он находится в сильной зависимости от заинтересованных сторон и скорее декларирует этичность, чем обеспечивает ее. Личностно ориентированный подход опирается на идею персональной моральной ответственности, однако упирается в проблемы разрыва ответственности и непредсказуемости действий искусственного интеллекта. Предлагается критический подход, в рамках которого предметом этики искусственного интеллекта является воздействие технологий на представления и ценности людей, их поведение и принятие решений. Вводится и рассматривается понятие парадокса масштаба как эффекта от применения искусственного интеллекта, состоящего в том, что множество этически корректных случаев использования технологии может привести к этически неприемлемым последствиям. Показано, что одним из вариантов применения критического подхода может быть исследование установок и стереотипов, связанных в массовом сознании с искусственным интеллектом.

Ключевые слова: искусственный интеллект (ИИ); этика искусственного интеллекта; этические парадоксы; критический подход в этике; образы искусственного интеллекта.

Цитирование: Ястреб Н.А. Концепции этики искусственного интеллекта: от принципов к критическому подходу // Семиотические исследования. Semiotic studies. 2024. Т. 4. № 1. С. 24–30. DOI: <http://doi.org/10.18287/2782-2966-2024-4-1-24-30>.

Благодарности: автор выражает благодарность профессору В.А. Суровцеву за оказанную помощь в проведенном исследовании.

Информация о конфликте интересов: автор заявляет об отсутствии конфликта интересов.

© Ястреб Н.А., 2024 – доктор философских наук, доцент, директор института социальных и гуманитарных наук, Вологодский государственный университет, 160035, Российская Федерация, г. Вологда, ул. Ленина, д. 15.

SCIENTIFIC ARTICLE

N.A. YastrebVologda State University,
Vologda, Russian Federation

E-mail: nayastreb@mail.ru

ORCID: <https://orcid.org/0000-0001-6635-6008>

Artificial intelligence ethics: from principles to the critical approach

Abstract: the article discusses the problem of determining the methodological and conceptual foundations of the artificial intelligence ethics. It is shown that the principled approach is based on the theory of value embedding, which assumes that technical objects can either be the carriers of values themselves, or at least contribute to the realization of certain values. At the same time, it is highly dependent on stakeholders, and it rather declares ethics than ensures it. The person-centered approach is based on the idea of personal moral responsibility. The main problems of the personality-oriented approach are the gap in responsibility and the unpredictability of the actions of artificial intelligence. A critical approach is proposed, according to which the subject of artificial intelligence ethics is the impact of technology on people's ideas and values, their behavior and decision-making. The work introduces and discusses the concept of the scale paradox,

resulting from the artificial intelligence use. This concept states that many ethically correct cases of using technology can lead to ethically unacceptable consequences. It is shown that one of the options for applying a critical approach can be the study of attitudes and stereotypes associated with artificial intelligence in the mass consciousness.

Key words: artificial intelligence; ethics of artificial intelligence; ethical paradoxes; critical approach in ethics; images of artificial intelligence.

Citation: Yastreb, N.A. (2024), Artificial intelligence ethics: from principles to the critical approach, *Semioticheskie issledovanija. Semiotic studies*, vol. 4, no. 1, pp. 24–30, DOI: <http://doi.org/10.18287/2782-2966-2024-4-1-24-30>.

Acknowledgments: the author express their gratitude professor V. Surovtsev for the assistance provided in the study.

Information about conflict of interests: the author declares no conflict of interests.

© Yastreb N.A., 2024 – Doctor of Sciences in Philosophy, Associate Professor, Head of the Institute of Human and Social Sciences, Vologda State University, 15, Lenin st., Vologda, 160035, Russian Federation.

Введение

За последнее десятилетие практически всеми технологически развитыми странами и крупными IT-корпорациями были разработаны и приняты этические кодексы искусственного интеллекта либо другие документы, декларирующие этические принципы и установки в отношении таких систем. Этому способствовали рост производительности и расширение сферы применения искусственных нейронных сетей, создание моделей широкого назначения (AGI – общий искусственный интеллект), немаловажную роль сыграли и отдельные достижения, заставившие экспертов поверить в то, что искусственный интеллект уже существует. Одним из таких событий стала победа системы в чемпионате мира по игре в Go, победить человека в которой, как считалось ранее, машина никогда не сможет.

В то же время этическое и правовое регулирование систем искусственного интеллекта недостаточно проработано на концептуальном уровне. В литературе можно найти достаточно большое количество работ, посвященных отдельным этическим проблемам искусственного интеллекта, таким как конфиденциальность персональных данных, алгоритмическая предвзятость или охрана авторских прав на объекты, созданные при участии нейросетевых моделей. На этапе разработки первых кодексов мы видим активную экстраполяцию опыта биоэтики и правового регулирования разработки и эксплуатации машин и механизмов (в частности автомобилей) на сферу искусственного интеллекта. Такой подход позволил решить ряд проблем. Однако в силу того, что системы искусственного интеллекта онтологически и функционально радикально отличаются как от живых организмов, так и от механических устройств, построение этических подходов в данной сфере требует более тщательной концептуальной проработки.

На глобальном уровне влияние новых технологий на человека и общество «непрозрачно и

глобально» (Reijers 2023, p. 4) и без разработки методологических и концептуальных основ этики искусственного интеллекта значимый прогресс в данной области невозможен. В современных исследованиях этических проблем искусственного интеллекта можно выделить несколько подходов, каждый из которых имеет свои преимущества и ограничения.

Подход, основанный на принципах

Одной из первых этических систем искусственного интеллекта стал подход, основанный на формулировании и декларировании принципов. В метаисследовании, основанном на анализе 84 кодексов и рекомендаций по этическому искусственному интеллекту, авторы показали, что наиболее часто в документах выделяются принципы прозрачности, справедливости и честности, непричинения вреда, ответственности и конфиденциальности (Jobin 2019).

Подход, основанный на принципах, имеет ряд существенных недостатков. Прежде всего, набор принципов определяется конкретной компанией, государством или любой заинтересованной стороной на основе ее собственных представлений и предпочтений. Кроме того, само толкование принципов может сильно различаться в зависимости от контекста и целей интерпретатора. Также важно отметить, что большинство этих этических кодексов не являются законодательными политическими инструментами и служат руководящими принципами или «мягким правом», не имея обязательной юридической силы. Фактически такой подход позволяет разработчику или владельцу технологии заявить о наличии у него ценностей и этических установок, не требуя обязательного следования им.

Еще одним важным недостатком подхода, основанного на принципах, является то, что практическое применение подобных документов фактически переносит ответственность за последствия действий искусственного интеллекта на коллек-

тивные субъекты, прежде всего на компании-разработчики, делая невозможным установление персональной ответственности. Институциональная ответственность может позволить, например, взыскать с компании или запретить ту или иную систему, но моральная ответственность всегда возлагается на конкретного человека.

Практической реализацией подхода, основанного на принципах, стала концепция этического искусственного интеллекта, в рамках которой ставится задача наделения уже на этапе проектирования и разработки системы искусственного интеллекта ценностями, моральными принципами и установками. Этический ИИ базируется на теории вложения ценностей, которая предполагает, что технические объекты могут либо сами быть носителями ценностей, либо, по меньшей мере, способствуют реализации определенных ценностей (Klenk 2020). Приписывание ценностей самим техническим объектам, которое мы можем встретить в разнообразных декларациях принципов этического искусственного интеллекта, не является в полной мере обоснованным и может привести к серьезным негативным социальным и гуманитарным последствиям. В то же время в силу своей физической организации и функциональной направленности технические объекты конструируют многие виды деятельности человека, а некоторые делают возможными в принципе. Поэтому опосредованно они могут влиять на реализацию тех или иных ценностей, что обязательно должно учитываться в практике регулирования применения этих объектов.

Личностно ориентированный подход

Личностный поворот в этике искусственного интеллекта связан с осознанием того, что без определения персональной моральной ответственности конкретных субъектов за последствия действий искусственного интеллекта, в том числе непредсказуемые, невозможно построение работающих этических систем. Развитие данного подхода затрудняется рядом фундаментальных проблем, ключевыми из которых являются непропорциональные непредсказуемые последствия и разрыв ответственности.

Первая проблема связана с тем, что автономные системы искусственного интеллекта обучаются и действуют в условиях реальной окружающей среды, и внешние факторы могут радикально влиять на последствия их работы. Можно утверждать, что в целом в настоящее время достигнут консенсус относительно того, что системы искусственного интеллекта, вне зависимости от уровня сложности, не обладают субъектностью и не являются носителями моральной ответственности. Однако чем сложнее система и условия среды, в которых она действует, и чем выше уровень ее автономно-

сти, тем более непредсказуемыми могут быть последствия ее работы.

Разрыв в ответственности возникает при возложении ответственности за действия автономных искусственных агентов. С одной стороны, не имея субъектности и не осознавая принимаемые решения, машины не могут быть акторами, на которые возлагается моральная ответственность. С другой стороны, разработчики, инженеры, операторы, владельцы и другие субъекты, участвующие в создании и использовании системы, не могут в полной мере быть ответственны за конкретные последствия, так как степень участия каждого из них зачастую не может быть определена. Так, программист, создающий код для беспилотного оружия, не может прогнозировать все ситуации, в которых может оказаться система. Командир, принимающий решение об использовании беспилотника, не может приказом определить последовательность действий системы. Оператор, управляющий аппаратом, действует в режиме обратной связи в условиях пространственно-временных ограничений. Тот факт, что вина за те или иные негативные последствия часто возлагается на компанию или предприятия, еще сильнее затрудняет установление моральной ответственности за работу систем искусственного интеллекта.

Как отмечают Таддео и Бланшар, определение личной моральной ответственности каждого человека является необходимым условием возможности этики искусственного интеллекта, в частности, в случае применения автоматических систем вооружения (AWS). Важно, чтобы «дизайнеры, разработчики и специалисты по развертыванию взяли на себя значимую моральную ответственность за намеченные действия, осознавая при этом риск того, что могут возникнуть непредсказуемые результаты, и принимая на себя моральную ответственность также за непредсказуемые последствия, которые могут последовать за решением о развертывании автоматических систем вооружения» (Taddeo, Blanchard 2022, p. 16). Принимая на себя эту ответственность и продолжая разработку, люди-агенты совершают моральный гамбит. Они осознают, что в реальных условиях системы могут совершать непредсказуемые действия, но продолжают проектировать, разрабатывать и внедрять их, осознанно «принимая на себя моральную ответственность за непредвиденные результаты AWS и их последствия» (Taddeo, Blanchard 2022, p. 16).

Одним из условий решения данной проблемы является повышение уровня знаний и информированности относительно работы системы у всех участников процесса разработки и применения искусственного интеллекта. На данный момент это часто может быть затруднено из-за распределенного принципа организации разработки, ког-

да каждый участник занят частной задачей и не понимает особенности работы системы в целом, а также вследствие секретности данных, представляющих собой коммерческую или военную тайну. Только субъект, обладающий полнотой знаний о работе системы, может быть морально ответственным за нее, то есть осознавать последствия своих решений и быть способным влиять на них. В этом смысле сократовский принцип тождественности знания и добродетели в полной мере отражает фундаментальные положения этики искусственного интеллекта.

Установление персональной моральной ответственности сопряжено с рядом трудностей. Сам субъект должен не просто испытывать чувство вины, а знать и осознавать, какие именно его действия и решения привели к наступлению нежелательных последствий. Это означает, что как минимум такие причинно-следственные связи должны быть реконструируемы. Между тем в исследованиях искусственного интеллекта еще в начале разработки методов машинного обучения был выявлен феномен «непрозрачности» принятия решения нейронной сетью в большинстве случаев. Сложность взаимосвязей внутри многослойной нейронной сети делает практически невозможным пошаговое понимание последовательности проводимых в ней вычислений. Кроме того, для объективного обоснованного возложения ответственности на конкретного человека необходимо экспертное заключение, подтверждающее его роль в конечном результате. Для этого опять-таки необходима полнота знаний. То есть эксперт должен быть полностью осведомлен о принципах работы системы и обеспечен доступом ко всем данным о ее действиях. Учитывая инновационность разработок, а также распространенность ограничений из-за коммерческой и военной тайны, в реальности внешние эксперты часто не имеют возможности получения исчерпывающей информации. Институциональная система ограничения доступа к данным тем самым по умолчанию работает на защиту института, будь то коммерческая компания или армия.

Проблема личной ответственности приводит к консервативному парадоксу. Допустим, компания разработала алгоритм, позволяющий выявлять по цифровому следу пользователей, которые в ближайшее время могут совершить попытку суицида. Однако для ее применения необходима передача данных о пользователях в медицинские или психологические службы, которые могут помочь человеку преодолеть кризис. В результате компания, осознавая последствия самовольной передачи данных третьим лицам, не будет использовать систему, которая может спасти жизнь человека, потому что за нарушение законов о персональных данных она будет наказана, а за совершенный

пользователем суицид – нет. Как показывает опыт биотехнологических фармацевтических разработок, такой «потенциально смертельный консерватизм» (Chomanski 2023, p. 10) является сильным фактором, сдерживающим развитие и внедрение технологий.

Критический подход

Современные системы искусственного интеллекта используются для принятия решения о выдаче банковских кредитов, принятии человека на ту или иную должность, возможности условно-досрочного освобождения и многих других социально значимых действиях. Изначально привлечение искусственных систем было вызвано стремлением исключить «человеческий фактор», то есть избежать субъективности и предвзятости людей, а также возможности совершения ошибок из-за невнимательности. Пациенты заменяют психотерапевта-человека чат-ботами, покупатели доверяют рекомендациям, основанным на больших данных, больше, чем отзывам окружающих их людей, специалисты по подбору персонала все больше ориентируются на оценку кандидатов, сделанную на основе анализа их цифрового следа. Объективность, непредвзятость и системность в массовом сознании все больше начинают ассоциироваться с работой машин. Эти особенности искусственного интеллекта, как считалось, дадут возможность совершать рациональные действия, свободные от предрассудков, что повысит эффективность принятых решений. Однако такие системы принятия решений основываются на обучающей выборке и извлекают из нее закономерности, которые могут приводить к дискриминации. Например, машина может отказать в выдаче образовательного кредита на основе почтового индекса, если кандидат живет в неблагополучном районе, или счесть потенциального сотрудника неблагонадежным, если в его списке друзей в социальных сетях есть одноклассники, имеющие проблемы с законом.

Некритичность отношения к технологии проявляется в желании пользователей получать результат, не задумываясь, почему появилась эта технология, какие пробелы она закрывает и закреплению каких ценностей способствует. Как отмечают А. Ракиб и ее соавторы, «проблема состоит не в использовании искусственного интеллекта, а в неизбежной зависимости от технологий, когда становится естественным адаптироваться к технологическим ценностям, а не подвергать их сомнению» (Raquib 2022, p. 4).

Для формирования критического подхода в этике искусственного интеллекта мы должны задать вместо вопроса *как это работает?* спросить себя: *почему наши желания исполняются?* Критический подход к искусственному интеллекту направлен в первую очередь на субъекта, то есть

на человека (Нестеров 2021). Почему мне нравится использовать эту систему; какие мои слабости она использует, какие недостатки преодолевает; какие ценности поддерживает, какие нивелирует и, наконец, какие установки и стереотипы эксплуатирует? Искусственный интеллект не столько создает новые этические проблемы и парадоксы, сколько актуализирует, расширяет и делает заметным то, что уже свойственно человеку и присутствует в обществе. Когда идет речь об этичности беспилотных систем вооружения, уместно вспомнить Л. Мэмфорда с его концепцией мегамшины и задать вопрос о том, чем использование беспилотника принципиально отличается от отдачи командиром приказа солдату, если в значительной степени обучение последнего в армии было направлено на его превращение в орудие исполнения приказа? Тем, что беспилотник не может осудить негуманное решение командира или даже не согласиться с ним. То есть *предметом этики искусственного интеллекта должно быть воздействие технологий на представления и ценности людей, их поведение и принятие решений.*

Одной из методологических трудностей этического анализа искусственного интеллекта является проблема, которую мы назовем парадоксом масштаба (фактически он является вариантом парадокса кучи). Парадокс состоит в том, что множество этически корректных случаев использования технологии может привести к этически неприемлемым последствиям. Например, на рынке появляется образовательная технология, позволяющая реферировать тексты. Для ее использования в образовательном процессе педагог должен проконтролировать безопасность персональных данных обучающихся, соблюдение гигиенических и социальных норм. Допустим, в случаях каждого применения этого инструмента все эти требования выполняются. Через определенное время обнаруживается, что обучающиеся, регулярно применявшие систему, начали утрачивать навыки работы с большими текстами. Более того, они не видят смысла в чтении крупных текстов, ведь их за несколько секунд можно сжать до основных мыслей. Таким образом, применение технологии в образовательном процессе в целом навредило когнитивному развитию обучающихся.

Парадокс масштаба присутствует в истории практически всех значимых информационных технологий – от компьютерных игр до социальных сетей, чье воздействие на человека и общество становится заметным на глобальном уровне. В каждом конкретном случае использования системы искусственного интеллекта такие аспекты, как ценности, цели, идеи, являются слишком абстрактными, чтобы влиять на решение, но при изменении масштаба рассмотрения проблемы они конкретизируются, можно оценить динамику

и последствия технологии. Проблема в том, что процессы внедрения технологий искусственного интеллекта в большинстве случаев являются непрозрачными и необратимыми. Хотя доступ к ChatGPT обычные пользователи получили совсем недавно, уже сейчас мы видим глобальные последствия для системы образования и многих других, которые фактически принимаем как данность, так как отменить их не получится. К тому времени, когда вред технологии становится заметным, исправлять или отменять ее уже слишком поздно. Использование умных лент социальных сетей для манипулирования общественным мнением, распространение множества фейковых изображений, аудио- и видеофайлов, полученных с помощью нейросетей, и многие другие примеры демонстрируют, что, несмотря на кажущуюся безобидность конкретных случаев использования, «результаты применения технологий искусственного интеллекта зачастую непрозрачны, но в то же время радикальны, необратимы и широко распространены» (Raquib 2022, p. 5).

Одним из вариантов применения критического подхода может быть исследование установок и стереотипов, связанных в массовом сознании с искусственным интеллектом. Визуальные образы массовой культуры поддерживают ряд стереотипов, демонстрируя изображения искусственного интеллекта как сильного беспристрастного воина или героя картины Микеланджело, протягивающего руку помощи человеку. Более того, такие установки активно поддерживаются и транслируются в том числе и научным сообществом через визуальное сопровождение презентаций, сайтов и страниц, связанных с тематикой искусственного интеллекта. А. Ромеле называет само использование таких изображений неэтичным и обращает внимание на то, что «в то время как этика научной коммуникации обычно поощряет практику таких добродетелей, как скромность, смирение, искренность, прозрачность, открытость, честность и щедрость, стандартные изображения и другие популярные визуальные изображения ИИ являются высокомерными, напыщенными и самоуверенными» (Romele 2022, p. 16).

Представления людей об искусственном интеллекте формируют онтологическую основу их взаимоотношений с умными устройствами, и влияют через социальные ожидания на разработку самих интеллектуальных систем. Анализ представлений и ожиданий людей относительно искусственного интеллекта может дать точку отсчета для проработки концептуальных основ этики искусственного интеллекта. Вместо высмеивания страхов бунта машин и порабощения человечества можно выявить установки, лежащие в их основе. Осознание того, что «невозможно разработать какое-либо комплексное понимание

искусственного интеллекта, не принимая во внимание представления об искусственном интеллекте» (Romele 2022, p. 2), создает новое направление в этике, рассматривающее искусственный интеллект как «технологическое воображение и его последствия» (Romele, p. 3).

Заключение

Развитие систем искусственного интеллекта не просто актуализирует проблему этического регулирования в сфере высоких технологий. Внедрение интеллектуальных систем ставит задачу пересмотра этических норм в областях, выходящих за пределы сферы информационных технологий. Включаясь в практики повседневной коммуникации людей, заботы о здоровье, профессиональную деятельность и образование, искусственный интеллект «подсвечивает» и обостряет непроявленные аспекты человеческих взаимоотношений. Методологической основой этических исследований искусственного интеллекта должен стать критический анализ совместного действия и взаимовлияния человека и машин. Ответы на вопросы о том, почему люди стремятся передать машине ту или иную функцию или почему они доверяют боту больше чем врачу, позволят нам прогнозировать долгосрочные последствия внедрения технологии и формировать осознанное отношение людей к использованию искусственного интеллекта.

Библиографический список

- Chomanski, B. (2023), A Moral Bind? – Autonomous Weapons, Moral Responsibility, and Institutional Reality, *Philosophy & Technology*, vol. 36 (41), pp. 1–14, DOI: <https://doi.org/10.1007/s13347-023-00647-2>.
- Elder, A. (2020), Conversation from beyond the grave? A neo-confucian ethics of chatbots of the dead, *Journal of Applied Philosophy*, vol. 37 (1), pp. 73–88, DOI: <https://doi.org/10.1111/japp.12369>.
- Flick, C. and Worrall, K. (2022), The Ethics of Creative AI, *The Language of Creative AI. Springer Series on Cultural Computing*, Vear, C., Poltronieri, F. (eds), Springer, Cham: Springer International Publishing, pp. 73–91, DOI: https://doi.org/10.1007/978-3-031-10960-7_5.
- Foot, P. (1967), The Problem of Abortion and the Doctrine of the Double Effect, *Oxford: Review*, vol. 5, pp. 1–7.
- Howard, D. (2018), Technomoral Civic Virtues: a Critical Appreciation of Shannon Vallor’s Technology and the Virtues, *Philosophy & Technology*, vol. 31, pp. 293–304.
- Jobin, A., Ienca, M. and Vayena, E. (2019), The global landscape of AI ethics guidelines, *Nature Machine Intelligence*, 2019, vol. 1, pp. 389–399. DOI: <https://doi.org/10.1038/s42256-019-0088-2>.
- Klenk, M. (2020), How Do Technological Artefacts Embod Moral Values? *Philosophy & Technology*, vol. 2, pp. 524–544, DOI: <https://doi.org/10.1007/s13347-020-00401-y>.
- Nyholm, S. (2018), Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci, *Science and Engineering Ethics*, vol. 24, pp. 1201–1219, DOI: <https://doi.org/10.1007/s11948-017-9943-x>.
- Ochigame, R. (2019), The invention of “ethical AI”: how big tech manipulates academia to avoid regulation, *The Intercept*, [Online], available at: <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/> (Accessed 06 Jan 2024).
- Raquib, A., Channa, B., Zubair, T. et al. (2022), Islamic virtue-based ethics for artificial intelligence, *Discover Artificial Intelligence*, vol. 2 (11), pp. 1–16, DOI: <https://doi.org/10.1007/s44163-022-00028-2>.
- Reijers, W. (2023), Technology and Civic Virtue, *Philosophy & Technology*, vol. 36, pp. 1–22, DOI: <https://doi.org/10.1007/s13347-023-00669-w>.
- Romele, A. (2022), Images of Artificial Intelligence: a Blind Spot in AI Ethics, *Philosophy & Technology*, vol. 4 (35), DOI: <https://doi.org/10.1007/s13347-022-00498-3>.
- Taddeo, M. and Blanchard, A. (2022), Accepting moral responsibility for the actions of autonomous weapons systems – A moral gambit, *Philosophy & Technology*, vol. 35 (3), pp. 1–24.
- Vallor, S. (2016), *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford University Press, Oxford, 309 p.
- Нестеров А.Ю. Проблема субъекта в искусственной природе // Гуманитарный вектор. 2021. Т. 16. № 2. С. 22–28.

References

and the Virtues, *Philosophy & Technology*, vol. 31, pp. 293–304.

Jobin, A., Ienca, M. and Vayena, E. (2019), The global landscape of AI ethics guidelines, *Nature Machine Intelligence*, 2019, vol. 1, pp. 389–399. DOI: <https://doi.org/10.1038/s42256-019-0088-2>.

Klenk, M. (2020), How Do Technological Artefacts Embody Moral Values? *Philosophy & Technology*, vol. 2, pp. 524–544, DOI: <https://doi.org/10.1007/s13347-020-00401-y>.

Nyholm, S. (2018), Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci, *Science and Engineering Ethics*, vol. 24, pp. 1201–1219, DOI: <https://doi.org/10.1007/s11948-017-9943-x>.

Ochigame, R. (2019), The invention of “ethical AI”: how big tech manipulates academia to avoid regulation, *The Intercept*, Url: <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/> (дата обращения 06.01.2024).

Raquib, A., Channa, B., Zubair, T. et al. (2022), Islamic virtue-based ethics for artificial intelligence, *Discover Artificial Intelligence*, vol. 2 (11), pp. 1–16, DOI: <https://doi.org/10.1007/s44163-022-00028-2>.

Reijers, W. (2023), Technology and Civic Virtue, *Philosophy & Technology*, vol. 36, pp. 1–22, DOI: <https://doi.org/10.1007/s13347-023-00669-w>.

Romele, A. (2022), Images of Artificial Intelligence: a Blind Spot in AI Ethics, *Philosophy & Technology*, vol. 4 (35), DOI: <https://doi.org/10.1007/s13347-022-00498-3>.

Taddeo, M. and Blanchard, A. (2022), Accepting moral responsibility for the actions of autonomous weapons systems – A moral gambit, *Philosophy & Technology*, vol. 35 (3), pp. 1–24.

Vallor, S. (2016), *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, Oxford University Press, Oxford, 309 p.

Nesterov, A. (2021), The problem of the subject in artificial nature, *Gumanitarnyj Vektor*, vol. 16, no. 2, pp. 22–28.

Submitted: 10.01.2024

Revised: 03.02.2024

Accepted: 15.03.2024