# ЭТИКА

DOI: 10.17212/2075-0862-2023-15.2.2-365-376 УДК 17:008

# НЕКОТОРЫЕ ПРОБЛЕМЫ ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

**Шляпников Виктор Валерьевич**,
*кандидат философских наук, доцент,*
*доцент кафедры философии и социальных наук Санкт-Петербургского университета ГПС МЧС России,*
*Россия, 196105, г. Санкт-Петербург, Московский проспект, 149*
ORCID: 0000-0002-6502-5810
SPIN-код (РИНЦ): 4210-5441
AuthorID (РИНЦ): 797851
shlyapnikovv@mail.ru

**Аннотация**

Достижения в области искусственного интеллекта (ИИ) привлекли внимание общественности к цифровой этике и являются движущей силой многих дискуссий по этическим вопросам, связанным с цифровыми технологиями. В настоящей статье анализируются аргументы сторонников и противников искусственного интеллекта, различные подходы к разработке систем ИИ, этические проблемы, связанные с использованием технологий ИИ, в том числе проблема управления искусственным интеллектом и идея активной ответственности за развитие технологий ИИ, общие принципы разработки систем ИИ, сформулированные в основополагающих документах. Методологической основой работы послужил диалектический метод, в процессе исследования использовались сравнительный метод и метод анализа документов. Источниками послужили исследования отечественных и зарубежных авторов, посвященные различным этическим проблемам искусственного интеллекта, европейские «Руководящие принципы этики для надежного искусственного интеллекта» и российский «Кодекс этики в сфере искусственного интеллекта». Показана важность морализации технологий искусственного интеллекта, то есть сознательного развития технологий для формирования нравственных действий и принятия решений. Одна из открытых проблем – найти демократичный способ морализировать технологии, поскольку технологии отличаются от законов тем, что могут ограничивать свободу человека, не являясь при этом результатом демократического процесса. Утверждается необходимость создания независимой международной научной организации для выработки четкого научного взгляда на искусственный интеллект, а также независимого международного органа по регулированию искусственного интеллекта, который бы объединил подходы к пониманию данного феномена со стороны государств, частных компаний и научных кругов.

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

**Библиографическое описание для цитирования:**
*Шляпников В.В.* Некоторые проблемы этики искусственного интеллекта // Идеи и идеалы. – 2023. – Т. 15, № 2, ч. 2. – С. 365–376. – DOI: 10.17212/2075-0862-2023-15.2.2-365-376.

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 настоящей статье анализируются аргументы сторонников и противников искусственного интеллекта, различные подходы к разработке систем ИИ, этические проблемы, связанные с использованием технологий ИИ, в том числе проблема управления искусственным интеллектом и идея активной ответственности за развитие технологий ИИ, общие принципы разработки систем ИИ, сформулированные в основополагающих документах.

Методологической основой настоящей работы послужил диалектический метод, в процессе исследования использовались сравнительный метод и метод анализа документов. Диалектический метод предполагал рассмотрение искусственного интеллекта как сложного и противоречивого явления. Сравнительный метод использовался при сопоставлении аргументов сторонников и противников искусственного интеллекта, а также при анализе подходов к разработке систем искусственного интеллекта. 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 послужили исследования отечественных и зарубежных авторов, посвященные различным этическим проблемам искусственного интеллекта, европейские принципы этики и российский кодекс этики в сфере искусственного интеллекта.

Этика искусственного интеллекта изучает моральную ответственность разработчиков интеллектуальных систем за последствия их функционирования. 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 фокусируются на проблеме доказательства безопасности разрабатываемых интеллектуальных систем с учетом способности их рекурсивного самосовершенствования [26]. Это серьезная проблема, и предполагается, что даже если в первоначальной версии интеллектуальной системы будут предусмотрены значительные ограничения, связанные с безопасностью, чрезвычайно сложно гарантировать, что последующие поколения системы сохранят эти ограничения. Утверждается также, что исследования в области разработки сильного ИИ, предполагающие, что соответствующим образом запрограммированные компьютеры могут мыслить, понимать и иметь другие когнитивные способности, по своей сути неэтичны, поскольку это может привести к страданиям ИИ.

Обращается внимание на то, что фиксированный набор этических правил может привести к различным противоречиям и затруднениям [15]. Используя пример самоуправляемых автомобилей, чтобы осветить некоторые из этих затруднений, утверждается, что системы искусственного интеллекта должны иметь встроенные этические правила, которые соответствуют ценностям их владельцев, а не универсальному набору этических ценностей. Например, универсальные ценности могут невольно дискриминировать определенные группы. Обсуждается необходимость наличия гибкой этической системы в контексте автоматизированных систем военного назначения и подчеркивается важность присутствия человека, который может решать, когда применять оружие, а когда нет.

Интересна аргументация противников и сторонников ИИ [5]. Аргументы против ИИ предполагают, что в результате соединения искусственного интеллекта с большими данными мы окажемся в ситуации, в которой велика вероятность того, что ИИ может представлять серьезную угрозу для человечества [19]. Эти взгляды, в частности, разделяют такие визионеры-предприниматели, как Илон Маск и Билл Гейтс. Мы уже можем наблюдать, как нарушается неприкосновенность частной жизни, и в результате люди находятся под контролем в таких странах, как Сингапур, где ком-

пьютерные программы влияют на экономическую и иммиграционную политику, рынок недвижимости и школьные программы. Программные системы уже используют «убедительные вычисления» для программирования людей на определенное поведение, и эта тенденция сохранится, если не будут приняты законодательные меры [27]. В связи с этим представляются важными дискуссии о морализации технологий [3, 23]. Морализация технологий – это сознательное развитие технологии для формирования нравственных действий и принятия решений. Люди должны и могут морализировать не только других людей, но и свою материальную среду, включая разработанные и принятые технологии. Одна из самых больших проблем сегодняшних дебатов заключается в том, можно ли демократическим путем морализировать технологии.

Аргументы в пользу ИИ представляют исключительно оптимистичный взгляд на искусственный интеллект [18]. Сторонники создания ИИ обеспокоены тем, что нормативные акты могут задушить успех ИИ, и предполагают, что будущие исследователи увидят в усилиях по регулированию ИИ «темную эпоху» человеческого прогресса. Они обсуждают проблему отсутствия консенсуса в отношении четкого определения ИИ и выдвигают идею о том, что невозможно регулировать то, что не может быть определено. Поэтому они утверждают, что сегодня еще слишком рано думать о регулировании ИИ, особенно если такое регулирование будет препятствовать развитию, которое может оказаться важным для человеческого существования. Они также отмечают, что с точки зрения ответственности системы ИИ состоят из ряда компьютерных программ, некоторые из них могли быть написаны за много лет до того, как были введены элементы ИИ, поэтому было бы несправедливо возлагать на разработчиков этих программ ответственность за результаты, вызванные системами ИИ.

Нет согласия и по вопросу создания систем искусственного интеллекта [1]. 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 попытке разработать безопасную систему ИИ с хорошей встроенной этикой является создание «злонамеренного» ИИ [21]. Они приходят к выводу, что если такой злонамеренный ИИ возможен, то исследователи обязаны публиковать любые примеры проектов ИИ с негативными результатами и делиться данными, чтобы помочь понять, почему такие вещи происходят и как их предотвратить.

Это обсуждение может быть включено в дискуссию об искусственных моральных агентах, то есть о том, что машины могут в некотором смысле быть этическими агентами, ответственными за свои действия [24]. Эта идея связана с подходом, называемым машинной этикой, где машины рассматриваются как субъекты [7, 11]. Основная идея машинной этики в настоящее время находит отражение в современной робототехнике [24].

Особое внимание вопросам соблюдения этических норм уделяется при рассмотрении проблемы управления искусственным интеллектом [13]. 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 на данный момент ИИ не признан юридическим лицом, поэтому применяется теория «ИИ как инструмента», и, следовательно, правила субсидиарной ответственности регулируют поведение ИИ, и эта ответственность распространяется на разработчиков, пользователей и владельцев систем ИИ.

Искусственный интеллект всё чаще используется практически во всех областях медицины [2, 12, 22]: диагностике, принятии клинических решений, биомедицинских исследованиях и разработке лекарств, персонализированной медицине, телемедицине, медицинском образовании и многом другом. Одной из центральных проблем в данном случае является проблема ответственности, когда интеллектуальные программы ставят диагноз или выбирают лечение. Кроме того, к этическим аспектам относятся тре-

бования прозрачности и объяснимости интеллектуальных алгоритмов, используемых при принятии решений в медицине.

В последние годы возникла идея активной ответственности за развитие технологий в целом и технологий искусственного интеллекта в частности. Это означает не только предотвращение отрицательных эффектов технологии, но и реализацию некоторых положительных эффектов. Одним из способов реализации активной ответственности является проектирование с учетом ценностей, при котором моральные соображения используются в качестве требований для проектирования технологий [17]. Когда ценностно-чувствительный подход применяется к технологиям искусственного интеллекта, проблемы, связанные с выбором включения моральных ценностей в эти сложные технологии, становятся более серьезными. Идея включения положительных ценностей не лишена риска. В частности, существует множество негативных реакций на технологии искусственного интеллекта, созданные для управления человеческим поведением (в том числе и во благо) [23]. Возможные опасения заключаются в том, что свобода человека находится под угрозой и что демократия заменяется технократией. Идея о том, что не люди, а технологии контролируют ситуацию, тесно связана с восприятием сокращения автономии как угрозы человеческому достоинству. Существует также риск того, что когда моральные решения делегируются машинам, люди могут стать ленивыми в принятии моральных решений или даже неспособными к ним. Подчеркивается, что технологии отличаются от законов тем, что ограничивают свободу человека, не являясь при этом результатом демократического процесса. Итак, как уже отмечалось, одна из открытых проблем – найти демократичный 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) [16], определяющие этические принципы и связанные с ними ценности, которые необходимо соблюдать при разработке, внедрении и использовании систем ИИ. Они разъясняют, что системы искусственного интеллекта должны разрабатываться, внедряться и использоваться в соответствии с этическими принципами уважения автономии человека, предотвращения ущерба, справедливости и подотчетности.

В России в октябре 2021 года появился «Кодекс этики в сфере искусственного интеллекта» [4]. Он разработан Альянсом в сфере искусственного интеллекта и подписан ведущими научно-исследовательскими организациями и крупнейшими технологическими компаниями России. Кодекс является рекомендательным документом для участников отношений в сфе-

ре ИИ (российских и иностранных компаний, государственных структур) и провозглашает принципы ответственности, информационной безопасности, контроля рекурсивного самосовершенствования систем искусственного интеллекта.

Таким образом, искусственный интеллект стал одной из центральных проблем цифровой этики. В то время как немало исследователей и экспертов в области технологий воодушевлены потенциалом ИИ, многих других он настораживает. Обсуждаются как положительные эффекты ИИ (беспилотные автомобили, повышающие безопасность, цифровые помощники, роботы для тяжелой физической работы, мощные алгоритмы для получения полезных и важных выводов из больших объемов данных), так и отрицательные (автоматизация, ведущая к потере рабочих мест, растущее неравенство, угрозы конфиденциальности). Важной проблемой является регулирование отношений в сфере технологий искусственного интеллекта и робототехники.

Можно констатировать, что развитие ИИ достигло значительного прогресса с появлением четырех ключевых факторов: улучшенных статистических моделей, больших массивов данных, дешевой вычислительной мощности, повсеместного внедрения технологий в жизнь человека. Однако в настоящее время исследования в области искусственного интеллекта проводятся в основном частными компаниями, и, следовательно, существует дефицит социальной и политической ответственности и долгосрочного планирования, который необходимо устранить. Для этого необходимо создать независимый международный орган по регулированию ИИ, который бы объединил подходы к пониманию искусственного интеллекта со стороны государств, частных компаний и научных кругов. Хотя сейчас и существует ряд как академических, так и совместных площадок государственного и частного секторов, которые поддерживают правительства в продвижении исследований и разработок в области искусственного интеллекта, тем не менее независимый орган необходим, чтобы помочь повысить квалификацию политиков в вопросах, связанных с ИИ. Международная интеграция необходима во избежание конфликтов, связанных с различными национальными подходами к законодательству. Кроме того, требуется создать независимую международную научную организацию для выработки четкого научного взгляда на информационно-коммуникационные технологии 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.

**Литература**

1. *Гуров О.Н.* Этичное взаимодействие с интеллектуальными системами // Искусственные общества. – 2020. – Т. 15, вып. 3. – DOI: 10.18254/S207751800010905-4.

2. *Гусев А.В., Добриднюк С.А.* Искусственный интеллект в медицине и здравоохранении // Информационное общество. – 2017. – № 4–5. – С. 78–93.

3. *Дедюлина М.А.* «Морализация технологий»: от компьютерных артефактов к социальным практикам // Философские проблемы информационных технологий и киберпространства. – 2015. – № 2 (10). – С. 75–86. – DOI: 10.17726/philIT.2015.10.2.130.2.

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 А.Н., Морозова Г.А.* Искусственный интеллект: за и против // Развитие и безопасность. – 2020. – № 3 (7). – С. 70–77. – DOI: 10.46960/2713-2633_2020_3_70.

6. *Лаптев В.А.* Понятие искусственного интеллекта и юридическая ответственность за его работу // Право. Журнал Высшей школы экономики. – 2019. – № 2. – С. 79–102. – DOI: 10.17323/2072-8166.2019.2.79.102.

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. – DOI: 10.24158/fik.2020.11.2.

8. *Мамина Р.П., Ильина А.В.* Искусственный интеллект: в поисках формализации этических оснований // Дискурс. – 2022. – Т. 8, № 6. – С. 17–30. – DOI: 10.32603/2412-8562-2022-8-6-17-30.

9. *Шляпников В.В.* Искусственный интеллект: эмпатия и подотчетность // Общество. Среда. Развитие. – 2022. – № 3 (64). – С. 100–103. – DOI: 10.53115/19975996_2022_03_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, 2011. – 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. – DOI: 10.3399/bjgp18X695213.

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, iss. 2133. – DOI: 10.1098/rsta.2018.0080.

14. *Čerka P., Grigiene J., Sirbikyte G.* Liability for damages caused by artificial intelligence // Computer Law & Security Review. – 2015. – Vol. 31, iss. 3. – P. 376–389. – DOI: 10.1016/j.clsr.2015.03.008.

15. *Dignum V.* Ethics in artificial intelligence: introduction to the special issue // Ethics and Information Technology. – 2018. – Vol. 20. – P. 1–3. – DOI: 10.1007/s10676-018-9450-z.

16. Ethics Guidelines for Trustworthy AI // Shaping Europe's digital future: website. – 2019, 08 April. – URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed: 24.05.2023).

17. *Friedman B., Hendry D.* Value Sensitive Design: Shaping Technology with Moral Imagination. – Cambridge, MA: The MIT Press, 2019. – 229 p. – DOI: 10.1080/17547075.2019.1684698.

18. *Gurkaynak G., Yilmaz I., Haksever G.* Stifling artificial intelligence: Human perils // Computer Law & Security Review. – 2016. – Vol. 32, iss. 5. – P. 749–758. – DOI: 10.1016/j.clsr.2016.05.003.

19. Will Democracy Survive Big Data and Artificial Intelligence? / D. Helbing et al. // Towards Digital Enlightenment / ed. by D. Helbing. – Cham: Springer, 2019. – P. 73–98. – DOI: 10.1007/978-3-319-90869-4_7.

20. *Müller V.* Ethics of Artificial Intelligence and Robotics // Stanford Encyclopedia of Philosophy / ed. by E. Zalta. – Palo Alto, California: CSLI, Stanford University, 2020. – P. 1–70.

21. *Pistono F., Yampolskiy R.* Unethical Research: How to Create a Malevolent Artificial Intelligence // The Age of Artificial Intelligence: An Exploration / ed. by S. Gouveia. – Wilmington: Vernon Press, 2020. – P. 303–318.

22. *Rigby M.J.* Ethical Dimensions of Using Artificial Intelligence in Health Care // AMA Journal of Ethics. – 2019. – Vol. 21. – P. 121–124. – DOI: 10.1001/amajethics.2019.121.

23. *Verbeek P.-P.* Moralizing Technology: Understanding and Designing the Morality of Things. – Chicago: University of Chicago Press, 2011. – 196 p. – DOI: 10.7208/chicago/9780226852904.001.0001.

24. Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems / A. Winfield, K. Michael, J. Pitt, V. Evers // Proceedings of the IEEE. – 2019. – Vol. 107, iss. 3. – P. 509–517. – DOI: 10.1109/JPROC.2019.2900622.

25. *Wynsberghe A. van, Robbins S.* Critiquing the Reasons for Making Artificial Moral Agents // Science and Engineering Ethics. – 2019. – Vol. 25. – P. 719–735. – DOI: 10.1007/s11948-018-0030-8.

26. *Yampolskiy R.* Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach // Philosophy and Theory of Artificial Intelligence / ed. by V. Müller. – Berlin; Heidelberg: Springer, 2013. – P. 389–396. – (Studies in Applied Philosophy, Epistemology and Rational Ethics; vol. 5). – DOI: 10.1007/978-3-642-31674-6_29.

27. *Zuboff S.* The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. – New York: PublicAffairs, 2019. – 704 p.

Статья поступила в редакцию 09.02.2023.
Статья прошла рецензирование 28.03.2023.

DOI: 10.17212/2075-0862-2023-15.2.2-365-376

# SOME PROBLEMS WITH ARTIFICIAL INTELLIGENCE ETHICS

**Shlyapnikov, Viktor**,
*Cand. of Sc. (Philosophy), Associate Professor,*
*Associate Professor at the Department of Philosophy and Social Sciences,*
*Saint-Petersburg University of State Fire Service of EMERCOM of Russia,*
*149 Moskovskiy prospect, St. Petersburg, 196105, Russian Federation*
ORCID: 0000-0002-6502-5810
SPIN-код (RISC): 4210-5441
AuthorID (RISC): 797851
shlyapnikovv@mail.ru

**Abstract**

Advances in artificial intelligence (AI) have brought digital ethics to the public's attention and are the driving force behind many 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 AI systems, ethical issues associated with the use of AI technologies, including the problem of managing artificial intelligence and the idea of active responsibility for the development of AI technologies, general principles for the development of AI systems formulated in the founding documents. The methodological basis of this work was the dialectical method; in the process of research the author used the comparative method and the method of document analysis. The sources were the studies by domestic and foreign authors on various ethical issues of artificial intelligence, the European "Ethics Guidelines for Trustworthy AI" and the Russian "AI Ethics Code". The author demonstrates the importance of the moralization of artificial intelligence technologies, that is, the conscious development of technologies for the formation of moral actions and decision-making. One of the clear problems is finding a democratic way to moralize technology, since technology differs from laws in that it can restrict human freedom without being the result of a democratic process. It is argued that it is necessary to create an independent international scientific organization to develop a clear scientific view of artificial intelligence, as well as an independent international body for the regulation of artificial intelligence, which would unite approaches to understanding this phenomenon from different points of view (states, private companies and academia).

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

**Bibliographic description for citation:**
Shlyapnikov V. Some Problems with Artificial Intelligence Ethics. *Idei i idealy = Ideas and Ideals*, 2023, vol. 15, iss. 2, pt. 2, pp. 365–376. DOI: 10.17212/2075-0862-2023-15.2.2-365-376.

## References

1. Gurov O.N. Etichnoe vzaimodeistvie s intellektual'nymi sistemami [Ethical interaction with intellectual systems]. *Iskusstvennye obshchestva = Artificial Societies*, 2020, vol. 15, iss. 3. DOI: 10.18254/S207751800010905-4.

2. Gusev A.V., Dobridnyuk S.L. Iskusstvennyi intellekt v meditsine i zdravookhranenii [Artificial intelligence in medicine and healthcare]. *Informatsionnoe obshchestvo = Information Society*, 2017, no. 4–5, pp. 78–93.

3. Dedyulina M.A. «Moralizatsiya tekhnologii»: ot komp'yuternykh artefaktov k sotsial'nym praktikam ["Moralization of technologies": from computer artefacts to social practices]. *Filosofskie problemy informatsionnykh tekhnologii i kiberprostranstva = Philosophical Problems of Information Technology and Cyberspace*, 2015, no. 2 (10), pp. 75–86. DOI: 10.17726/philIT.2015.10.2.130.2.

4. AI Alliance Russia. *Kodeks ehtiki v sfere iskusstvennogo intellekta* [AI Ethics Code]. (In Russian). Available at: https://a-ai.ru/ethics/index.html (accessed 24.05.2023).

5. Lapaev D.N., Morozova G.A. Iskusstvennyi intellekt: za i protiv [Artificial intelligence: pros and cons]. *Razvitie i bezopasnost' = Development and Security*, 2020, no. 3 (7), pp. 70–77. DOI: 10.46960/2713-2633_2020_3_70.

6. Laptev V.A. Ponyatie iskusstvennogo intellekta i yuridicheskaya otvetstvennost' za ego rabotu [Artificial intelligence and liability for its work]. *Pravo. Zhurnal Vysshei shkoly ekonomiki = Law. Journal of the Higher School of Economics*, 2019, no. 2, pp. 79–102. DOI: 10.17323/2072-8166.2019.2.79.102.

7. Makulin A.V. Eticheskii kal'kulyator: ot filosofskoi «vychislitel'noi morali» k mashinnoi etike iskusstvennykh moral'nykh agentov (IMA) [Ethical calculator: from philosophical «computational morality» to machine ethics of artificial moral agents (AMA)]. *Obshchestvo: filosofiya, istoriya, kul'tura = Society: Philosophy, History, Culture*, 2020, no. 11 (79), pp. 18–27. DOI: 10.24158/fik.2020.11.2.

8. Mamina R.I., Il'ina A.V. Iskusstvennyi intellekt: v poiskakh formalizatsii eticheskikh osnovanii [Artificial Intelligence: in Search for Formalization of Ethical Foundations]. *Diskurs = Discourse*, 2022, vol. 8, no. 6, pp. 17–30. DOI: 10.32603/2412-8562-2022-8-6-17-30.

9. Shlyapnikov V.V. Iskusstvennyi intellekt: empatiya i podotchetnost' [Artificial Intelligence: Empathy and Accountability]. *Obshchestvo. Sreda. Razvitie = Society. Environment. Development*, 2022, no. 3 (64), pp. 100–103. DOI: 10.53115/19975996_2022_03_100-103.

10 *Etika i «tsifra»: eticheskie problemy tsifrovykh tekhnologii* [Ethics and digital: ethical issues of digital technologies]. Moscow, Russian Presidential Academy of National Economy and Public Administration Publ., 2020. 207 p.

11. Anderson M., Anderson S., eds. *Machine Ethics.* New York, Cambridge, Cambridge University Press, 2011. 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, pp. 143–144. DOI: 10.3399/bjgp18X695213.

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, iss. 2133. DOI: 10.1098/rsta.2018.0080.

14. Čerka P., Grigiene J., Sirbikyte G. Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, 2015, vol. 31, iss. 3, pp. 376–389. DOI: 10.1016/j.clsr.2015.03.008.

15. Dignum V. Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, 2018, vol. 20, pp. 1–3. DOI: 10.1007/s10676-018-9450-z.

16. Ethics Guidelines for Trustworthy AI. *Shaping Europe's digital future*: website, 2019, 08 April. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 24.05.2023).

17. Friedman B., Hendry D. *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA, The MIT Press, 2019. 229 p. DOI: 10.1080/17547075.2019.1684698.

18. Gurkaynak G., Yilmaz I., Haksever G. Stifling artificial intelligence: Human perils. *Computer Law & Security Review*, 2016, vol. 32, iss. 5, pp. 749–758. DOI: 10.1016/j.clsr.2016.05.003.

19. Helbing D., et al. Will Democracy Survive Big Data and Artificial Intelligence? *Towards Digital Enlightenment*. Ed. by D. Helbing. Cham, Springer, 2019, pp. 73–98. DOI: 10.1007/978-3-319-90869-4_7.

20. Müller V. Ethics of Artificial Intelligence and Robotics. *Stanford Encyclopedia of Philosophy*. Ed. by E. Zalta. Palo Alto, California, CSLI, Stanford University, 2020, pp. 1–70.

21. Pistono F., Yampolskiy R. Unethical Research: How to Create a Malevolent Artificial Intelligence. *The Age of Artificial Intelligence: An Exploration*. Ed. by S. Gouveia. Wilmington, Vernon Press, 2020, pp. 303–318.

22. Rigby M.J. Ethical Dimensions of Using Artificial Intelligence in Health Care. *AMA Journal of Ethics*, 2019, vol. 21, pp. 121–124. DOI: 10.1001/amajethics.2019.121.

23. Verbeek P.-P. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago, University of Chicago Press, 2011. 196 p. DOI: 10.7208/chicago/9780226852904.001.0001.

24. Winfield A., Michael K., Pitt J., Evers V. Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems. *Proceedings of the IEEE*, 2019, vol. 107, iss. 3, pp. 509–517. DOI: 10.1109/JPROC.2019.2900622.

25. Wynsberghe A. van, Robbins S. Critiquing the Reasons for Making Artificial Moral Agents. *Science and Engineering Ethics*, 2019, vol. 25, pp. 719–735. DOI: 10.1007/s11948-018-0030-8.

26. Yampolskiy R. Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach. *Philosophy and Theory of Artificial Intelligence*. Ed. by V. Müller. Berlin, Heidelberg, Springer, 2013, pp. 389–396. DOI: 10.1007/978-3-642-31674-6_29.

27. Zuboff S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, PublicAffairs, 2019. 704 p.

The article was received on 09.02.2023.
The article was reviewed on 28.03.2023.