

Мария Александровна Колесник¹

Юлия Сергеевна Замараева¹

¹ Сибирский федеральный университет, Красноярск, Россия

К вопросу о культурной (африканской) специфике «ответственного искусственного интеллекта»

Аннотация. В статье рассмотрены проблемы, связанные с социально-культурной спецификой разработки практических принципов «ответственного искусственного интеллекта (ИИ)». В настоящее время большое внимание исследователей привлекает африканский континент, стереотипы о котором были связаны с отсталостью и неразвитостью. В настоящее время страны Африки, эффективно выстраивая свои экономические стратегии, применяют при этом и технологии ИИ. Более подробно рассмотрена концепция профессора университета Умео Вирджинии Дигнум, изложенная ею в одном из разделов книги «Ответственный ИИ в Африке. Проблемы и возможности» (Шпрингер, 2023). Основная логика теории профессора Дигнум заключается в том, что все последствия применения технологии ИИ зависят не от ИИ, а от социокультурных особенностей социотехноантропологической среды, в которых эти технологии применяются. На примере философии Убунту Вирджиния Дигнум показывает, как можно применить социокультурные философские аспекты «незападной» философии к созданию новых социоэтических принципов, подчеркивающих культурное разнообразие современной этики ИИ. Представляется важным положение теории Вирджинии Дигнум о социокультурной специфике этики «ответственного ИИ», которое, на наш взгляд, может быть адаптировано для различных социокультурных пространств, в той или иной степени не совпадающих с англо-американской парадигмой экономической цивилизации.

Ключевые слова: искусственный интеллект, этика ИИ, Африка, «ответственный ИИ», специфика этики ИИ, философия Убунту, культурное многообразие этики ИИ.

Исследование выполнено за счет гранта Российского научного фонда № 23-28-00255,

<https://rscf.ru/project/23-28-00255/>

Maria Aleksandrovna Kolesnik¹

Yulia Sergeevna Zamaraeva¹

¹ Siberian Federal University, Krasnoyarsk, Russia

On the issue of cultural (African) specifics of "responsible artificial intelligence"

Abstract. The article deals with the problems associated with the socio-cultural specifics of the development of practical principles of "responsible artificial intelligence (AI)". At present, much attention of researchers is attracted by the African continent, stereotypes about which were associated with backwardness and underdevelopment. Currently, African countries, while effectively building their economic strategies, are also using AI technologies. The concept of Umeå University professor Virginia Dignum, which she outlined in one of the sections of the book "Responsible AI in Africa. Problems and Opportunities" (Springer, 2023). The main logic of Professor Dignum's theory is that all the consequences of using AI technology do not depend on AI, but on the socio-cultural characteristics of the socio-techno-anthropological environment in which these technologies are applied. Using Ubuntu philosophy as an example, Virginia Dignum shows how the sociocultural philosophical aspects of "non-Western" philosophy can be applied to the creation of new socioethical principles that emphasize the cultural diversity of modern AI ethics. The position of Virginia Dignum's theory on the sociocultural specifics of the ethics of "responsible AI" seems to be important, which, in our opinion, can be adapted for various sociocultural spaces that, to one degree or another, do not coincide with the Anglo-American paradigm of economic civilization.

Keywords: artificial intelligence, AI ethics, Africa, "responsible AI", specificity of AI ethics, Ubuntu philosophy, cultural diversity of AI ethics.

The study was supported by the Russian Science Foundation Grant No. 23-28-00255, <https://rscf.ru/project/23-28-00255/>

С 2018 г. издательство Springer публикует серию книг по теме «Социальные и культурные исследования роботов и ИИ». Это по-настоящему инновационная серия, проблематика которой показывает наиболее актуальные и востребованные тренды в социологии искусственного интеллекта. Сегодня издано 10 книг, каждая из которых репрезентирует отдельное направление в гуманитаристике ИИ. Как правило, это коллективные монографии, чьи авторы аффилированы с различными научными организациями мира. Некоторые издания представляют собой труд одного исследователя, чаще всего связанный с его исследованиями прошлого времени, где теоретическая или

методологическая проблематика перенесена на новую социальную и гуманитарную предметность – роботов и искусственный интеллект.

Так, Кэтлин Ричардсон в 2018 г. открывает эту серию монографией «Сложная социальность. Антропология роботов, аутизма и привязанности». Здесь она исследует, как роботы представлены как занимающее уникальное пограничное пространство между человеческим и нечеловеческим, как роботы используются для помощи детям с аутизмом в развитии их социального поведения и навыков. Монография помогает создать новый нарратив о стигме, расстройствах аутистического спектра и общественном отношении к этим проблемам (рис.1).

Автор монографии Кэтлин Ричардсон, профессор этики культуры и ИИ в Университете Де Монфор – автор термина Representational Technologies of the Human (RTH), что может быть переведено как русский язык как «Технологии репрезентации человека» (рис.2). Под ее редакторством в этой же серии вышла книги «Искусственные женщины. Сексуальная политика секс-кукол и секс-роботов» (2022).

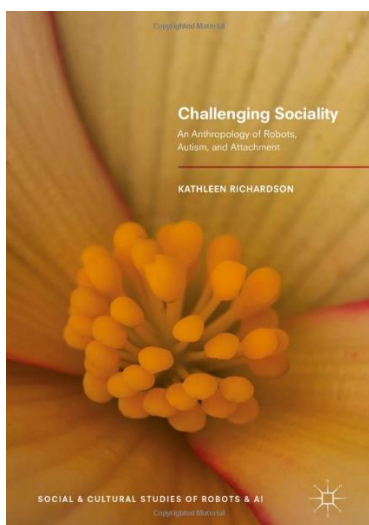


Рис. 1. Обложка книги Кэтлин Ричардсон «Сложная социальность. Антропология роботов, аутизма и привязанности». Источник изображения: <https://www.amazon.com>



Рис. 2. Кэтлин Ричардсон, профессор этики культуры и ИИ в Университете Де Монфор, автор книги «Сложная социальность. Антропология роботов, аутизма и привязанности». Источник изображения: <https://kathleenrichardson.org/>

В серии «Социальные и культурные исследования роботов и ИИ» в 2023 году вышла книга под редакторством Дамиана Окаибеди Эке, Кутомы Вакунумы, Симисолы Акинтойе «Responsible AI in Africa. Challenges and Opportunities», что может быть переведено на русский язык как «Ответственный ИИ в Африке. Проблемы и возможности» (рис.3).

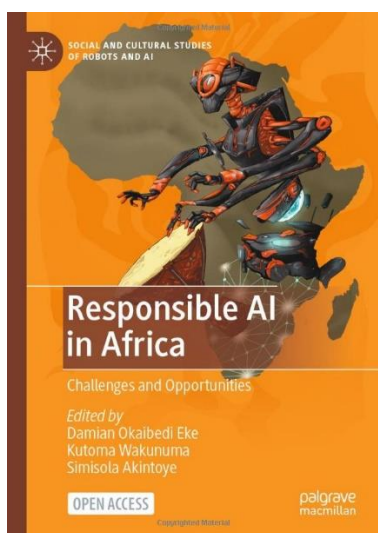


Рис.3. Обложка книги «Ответственный ИИ в Африке. Проблемы и возможности». источник изображения: <https://www.amazon.com>



Рис.4. Дамиан Окаибеди Эке, редактор книги «Ответственный интеллект в Африке. Проблемы и возможности». Источник изображения: <https://www.dmu.ac.uk>

Редакторы книги: Дамиан Окаибеди Эке (рис.4), Кутума Вакунома (рис.5), Симисола Акинтойе (рис.6) – профессора Университета Де Монфор (Великобритания).



Рис.5. Кутума Вакунома, редактор книги «Ответственный ИИ в Африке. Проблемы и возможности». Источник изображения: <https://www.dmu.ac.uk>



Рис.6. Симисола Акинтойе, редактор книги «Ответственный ИИ в Африке. Проблемы и возможности». Источник изображения: <https://www.dmu.ac.uk>

Книга полностью открыта и доступна на сайте издательства, что представляет собой значительную редкость и объясняется, скорее всего, социальной направленностью и актуальностью, связанной с современными африканскими исследованиями. В общем виде книгу можно охарактеризовать как монографию, где обсуждается этика ИИ с теоретической и практической точки зрения континента Африки. Ее авторы сосредоточены на истории ИИ, этике, конфиденциальности данных, слежке, вторжениях крупных технологий и глобализации / колонизации. Они исследуют социокультурный контекст ответственного ИИ, чувствительного к африканским культурам и обществам. С нашей точки зрения книга важна в концептуальном и методологическом отношении, поскольку здесь обсуждаются не только глобальные, но и социально-культурные контексты ИИ, которые «привязаны» к конкретным территориям, нациям, различным сообществам.

Вступительную статью, представляющее новое научное издание, написал Марк Кокельберг — профессор философии медиа и технологий и заместитель декана факультета философии и образования Венского университета. Ранее он был президентом Международного общества философии и технологий и членом Экспертной группы высокого уровня по ИИ Европейской комиссии. Он является автором многочисленных книг, в том числе «Этика ИИ» (издательство MIT Press), «Введение в философию технологий» (издательство Оксфордского университета) и «Политическая философия ИИ» (издательство «Polity Press»). Он пишет:

«Этика искусственного интеллекта (ИИ) — популярная тема. Но Африка обычно не попадает в поле зрения, когда речь идет об академических дискуссиях об этике и политике ИИ, даже когда речь идет о «глобальных» и межкультурных подходах. Это «забвение», вероятно, связано с предубеждениями и стереотипами об Африке со стороны западных собеседников. Для тех, кто считает, что Африка имеет мало общего с высокими технологиями и инновациями, такое название, как «Ответственный ИИ в Африке», звучит почти как оксюморон и в лучшем случае кажется маргинальной темой.

В течение многих десятилетий, если не столетий, Африка была синонимом проблем в западных СМИ и воображении. Я помню, как в 1980-х телевизионные кадры голодающих детей в Африке заполнили нашу гостиную. Это было первое, что я услышал об Африке в детстве. Африка стала синонимом нищеты, засух, войн и (других) нарушений прав человека. В разработке. Проблемы. Катастрофа. Сегодня ведущие американские и европейские СМИ почти не освещают Африку. Если специально не искать, то вообще об этом не услышишь. Следовательно, для большинства людей на Западе, включая политиков, Африка является своего рода *terra incognita*. Мы не знаем об этом, и, как счастливые, но и несколько трагические наследники Просвещения, мы рискуем увидеть его как место, где

процветает невежество. Многие жители Запада не знают ни одного африканского философа или ученого. Африка часто отсутствует на западных картах глобального ландшафта знаний и инноваций. Предполагается, что разработка и развертывание высоких технологий, таких как ИИ, будет происходить где-то еще: в Силиконовой долине, в Азии, иногда в Европе. А с точки зрения глобальной этики все гуманитарные науки, о которых говорят люди, — это колониализм и неоколониализм. Еще раз: Африка связана с проблемами. В худшем — это грустный сеттинг ужасных нарративов о рабстве, колониализме и эксплуатации.

В лучшем случае есть надежда на то, что Африка, как спящая красавица, может проснуться и развиваться в далеком будущем — возможно, поцелованная принцем иностранных инвестиций.

Хотя в этой картине Африки есть доля правды, она также очень односторонняя и искаженная. Африка также является местом, где ИИ используется и развивается: в социальных сетях, на предприятиях, в здравоохранении, в сельском хозяйстве, в образовании и в академических кругах. Конечно, как показано в нескольких главах этого тома, проблемы есть. Иногда наблюдается недостаточная цифровая грамотность, отсутствие инфраструктуры данных, неадекватная политика и нехватка средств. И, как и в западных странах, существуют риски, связанные с предвзятостью и дискриминацией, ответственностью, будущим работы и изменением климата (Coeckelbergh 2020). Но ИИ становится все более популярным, и, как и везде, ИИ уже меняет мир — в том числе и в Африке. Есть стартапы, есть исследования, есть инновации. Африка — это также место, где создается будущее» (<https://link.springer.com/content/pdf/bfm:978-3-031-08215-3/1?pdf=chapter%20toc>).

В этом прекрасном вступительном слове Марк Кокельберг обозначил основные аспекты, связанные с социокультурными стереотипами по отношению к Африке, где, кажется, и речь не может быть о передовых технологиях. Но эти стереотипы стремительно устаревают. Сегодня Африка — это пространство, где ИИ меняется людей, технологии, территории.

Структура книги представлена 10 разделами, как формальными, так и содержательными:

1. Вступительное слово Марка Кокельберга.
2. Представляем ответственный ИИ в Африке. Дамиан Окаибеди Эке, Кутума Вакунума, Симисола Акинтойе.
3. Эпистемическая справедливость и динамическая этика ИИ в Африке. Эмма Рутткамп-Блум.

4. Ответственный ИИ в Африке: проблемы и возможности. Чинаса Т. Около, Кехинде Арулеба, Джордж Обайдо.

5. Работа с роботами в качестве коллег: кенийские взгляды на этические проблемы возможной интеграции коботов на рабочих местах. Том Кванья.

6. ИИ в Африке: новые вызовы. Абехиде Аде-Ибихола, Чинеду Оконкво.

7. Использование гендерных чат-ботов в Нигерии: критические точки зрения. Фавор Борокини, Кутома Вакунума, Симисола Акинтойе.

8. Политика ИИ как ответ на этику ИИ? Решение этических проблем при разработке политики ИИ в Северной Африке. Бернд Карстен Шталь, Тони Лич, Олуйинка Ойениджи, Джордж Огох.

9. На пути к формированию будущего ответственного ИИ в Африке. Дамиан Окаибеде Эке, Шмидт Шилукобо Чинту, Кутома Вакунума.

10. Ответственный ИИ: рекомендации и извлеченные уроки. Вирджиния Дигнум.

Каждый из этих разделов заслуживает внимательного рассмотрения. Однако из-за ограничения объема научной статьи остановимся на понятии «ответственного ИИ», а также на разделе, который написала Вирджиния Дигнум «Ответственный ИИ: рекомендации и извлеченные уроки» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9).

«Ответственный ИИ» — это перевод на русский язык англоязычного выражения «Responsible AI». На сайте <https://www.accenture.com/> в разделе «Этика и управление ИИ» дается следующее определение ответственному ИИ:

«ИИ приносит бизнесу беспрецедентные возможности, но также и невероятную ответственность. Его прямое влияние на жизнь людей вызвало серьезные вопросы, связанные с этикой ИИ, управлением данными, доверием и законностью. Фактически, исследование Accenture Tech Vision 2022 года показало, что только 35% потребителей во всем мире доверяют тому, как организации внедряют ИИ. А 77% считают, что организации должны нести ответственность за неправильное использование ИИ.

Давление нарастает. По мере того, как организации начинают расширять использование ИИ для получения преимуществ для бизнеса, им необходимо помнить о новых и ожидаемых нормативных актах, а также о шагах, которые они должны предпринять, чтобы убедиться, что их организации соответствуют требованиям. Вот тут-то и появляется ответственный ИИ.

Итак, что такое ответственный ИИ?

Ответственный ИИ — это практика проектирования, разработки и развертывания ИИ с добрыми намерениями для расширения возможностей сотрудников и предприятий, а также справедливого воздействия на клиентов и общество, что позволяет

компаниям вызывать доверие и уверенно масштабировать ИИ» (<https://www.accenture.com/us-en/services/applied-intelligence/ai-ethics-governance#:~:text=Responsible%20AI%20is%20the%20practice,and%20scale%20AI%20with%20confidence>).

Вирджиния Дигнум – профессор департамента вычислительной техники университета Умео (Швеция) (рис. 7). Профессор Дигнум сосредоточена на анализе сложных взаимосвязей и взаимозависимостей между людьми, организациями и технологиями. Ее работа варьируется от разработки практических приложений и моделирования до разработки формальных теорий, объединяющих деятельность и организацию, и включает в себя сильный методологический компонент проектирования.



Рис. 7. Вирджиния Дигнум – профессор департамента вычислительной техники университета Умео (Швеция), автор раздела «Ответственный ИИ: рекомендации и извлеченные уроки». Источник изображения: <https://www.umu.se/en/staff/virginia-dignum/>

Основная проблематика исследований Вирджинии Дигнум описана ею самой следующим образом:

«Ответственный искусственный интеллект. Какие морально-этические проблемы возникают при работе автономных интеллектуальных агентов в командах? Как можно регулировать активность? Как разумные агенты и другие синтетические сущности могут рассуждать и принимать решения по моральным вопросам.

Анализ и формализация социального взаимодействия. Цель состоит в том, чтобы изучить влияние социальной и организационной структуры с учетом автономии и неоднородности участников и социальных и правовых ценностей, удерживаемых в контексте. С этой целью мы разрабатываем формальные теории и вычислительную архитектуру для рассуждений агентов на основе социальных практик.

Проектирование и оценка совместной работы человека и агента. Центральным исследовательским вопросом здесь является то, как люди взаимодействуют

(договариваются, доверяют, сотрудничают) с автономными когнитивными сущностями в социальной среде и разработке агентных симуляций сложных социотехнических областей.

Учитывая растущую важность понимания влияния ИИ на социальном, этическом и правовом уровне, я активно участвую в нескольких международных инициативах по политике и стратегиям исследований и приложений ИИ. В связи с этим я являюсь членом GPAI, Глобального партнерства по ИИ, Голландского альянса ИИ (ALLAI-NL), Совета Всемирного экономического форума по ИИ и Экспертной группы высокого уровня по реализации Рекомендации ЮНЕСКО по этике ИИ. Ранее я был членом Экспертной группы высокого уровня Европейской комиссии по искусственному интеллекту, Глобальной инициативы IEEE по этически согласованному проектированию автономных и интеллектуальных систем, Делфтского института дизайна для ценностей, Европейского глобального форума по ИИ (AI4People), Фонда ответственной робототехники. , и фонда ADA-AI.

Я являюсь программным председателем AAMAS'24, 23-й Международной конференции по автономным агентам и мультиагентным системам, и председателем по этике AAAI'24, 38-й конференции AAAI по искусственному интеллекту.

В знак признания моих исследований в 2020 году меня избрали членом Шведской королевской академии инженерных наук (IVA), а в 2018 году я стал членом Европейской ассоциации искусственного интеллекта (EURAI). В 2006 году я получил престижный грант Veni от NWO (Голландская организация научных исследований) за свою работу над организационными структурами на основе вычислительных агентов. Я часто говорю о социальном и этическом воздействии ИИ на конференциях, собраниях и других мероприятиях, много публиковался по этой теме и являюсь членом программных комитетов большинства крупных журналов и конференций по ИИ.

Я также связана с факультетом технологической политики и управления Делфтского технологического университета» (<https://www.umu.se/en/staff/virginia-dignum/>).

Уже за первые два месяца 2023 года количество скачиваний раздела, который написала Вирджиния Дигнум, составляет свыше 4000. Она пишет, что концепция и этика ответственного ИИ может и должна развиваться на «незападной» философии и указывает на соответствующий концептуальный и методологический потенциал философии Убунту. Эта концепция в настоящее время в русскоязычной исследовательской литературе развивается в основном в контексте философии образования (Михалина, 2010, Тимчук, 2017, Иванова, 2019, Ситаров и Миталов, 2021 и др.).

Убунту — это термин на языке нгуни-банту, означающий «человечество». Иногда его переводят как «Я есть, потому что мы есть» (также «Я есть, потому что вы есть») или

«гуманность по отношению к другим». Существуют различные определения слова «Ubuntu». Самое последнее определение было предоставлено Африканским журналом социальной работы. Журнал определил Убунту как:

«Совокупность ценностей и обычаев, которые люди африканского или африканского происхождения рассматривают как делающие людей подлинными людьми. Хотя нюансы этих ценностей и практик различаются у разных этнических групп, все они указывают на одно: подлинное индивидуальное человеческое существо является частью более крупного и значимого родственного, коллективного, социального, экологического и духовного мира» (Mugumbate and Chereni, 2020).

По словам Майкла Оньебучи Эзе, суть Убунту лучше всего можно описать следующим образом:

«Человек – это человек, через других людей совершающий утверждение своей человечности через признание «другого» в его уникальности и отличии. Это требование творческого интерсубъективного образования, в котором «другой» становится зеркалом (но только зеркалом) моей субъективности. Этот идеализм подсказывает нам, что человечность не заложена в моей личности исключительно как в индивидууме; моя человечность сосущностно дана другому и мне. Человечность — это качество, которым мы обязаны друг другу. Мы создаем друг друга и должны поддерживать это инаковое творение. И если мы принадлежим друг другу, мы участвуем в наших творениях: мы есть, потому что ты есть, а поскольку ты есть, то я определенно есть» (Eze, 2010).

Вирджиния Дигнум полагает, что новая этика ИИ может быть построена на принципах этой африканской философии, антропологии взаимного сосуществования. Во введении к разделу «Ответственный ИИ: рекомендации и извлеченные уроки» она пишет:

«В настоящее время искусственный интеллект (ИИ) используется практически повсеместно. Мы вряд ли сможем открыть газету или включить программу новостей, не услышав какую-нибудь историю об ИИ. Но ИИ означает разные вещи для разных людей.

Поскольку ИИ все больше влияет на многие аспекты жизни, осознание того, что он может повлиять на нашу жизнь и наш мир так, как это не удавалось ни одной другой технологии, по праву вызывает много вопросов, касающихся его этических, правовых, социальных и экономических последствий. Однако, несмотря на то что опасности и риски применения ИИ без должного учета его социальных, этических или правовых последствий получают все большее признание, нельзя сбрасывать со счетов потенциал ИИ, способствующего благополучию человека и общества. Всесторонний анализ роли ИИ в достижении «Целей устойчивого развития» (Vinuesa et al. 2020), в котором я участвовала, пришел к выводу, что он может влиять на достижение всех 17 целей, положительно влияя

на 134 задачи по всем целям, но он также может препятствовать достижению 59 целей (Vinueza et al., 2020).

Обеспечение ответственной разработки и использования ИИ становится основным направлением исследований и практики ИИ. Правительства, корпорации и международные организации в равной степени выступают с предложениями и декларациями о своей приверженности подотчетному, ответственному и прозрачному подходу к ИИ, в котором главенствуют человеческие ценности и этические принципы. Это очень нужная разработка, которой я посвятила свои усилия и исследования в последние несколько лет.

В настоящее время существует более 600 политических рекомендаций, руководств или стратегических отчетов, связанных с ИИ, которые были выпущены известными межправительственными организациями, профессиональными органами, комитетами национального уровня и другими общественными организациями, неправительственными и частными коммерческими компаниями. Недавнее исследование глобального ландшафта руководств по этике ИИ показывает, что наблюдается глобальное сближение вокруг пяти этических принципов: прозрачность, справедливость и справедливость, непричинение вреда, ответственность и конфиденциальность (Jobin et al., 2019). Тем не менее, даже несмотря на то, что организации соглашаются с необходимостью учитывать эти принципы, их интерпретация и применение на практике значительно различаются в разных рекомендательных документах.

В то же время растущий ажиотаж вокруг «ИИ» размывает его определение и впихивает в ту же кучу концептов и приложений самых разных мастей. Необходимый первый шаг в ответственной разработке и использовании ИИ — обеспечить надлежащее повествование об ИИ, которое демистифицирует возможности и процессы технологий ИИ и позволяет всем участвовать в обсуждении роли ИИ в обществе. Понимание возможностей и устранение рисков, связанных с искусственным интеллектом, требует четкого понимания того, что это такое, как он применяется, какие возможности и риски связаны с ним.

Работа организована следующим образом. После краткого обсуждения различных взглядов на то, что представляет собой система ИИ, я представляю в разделе «Обеспечение ответственной разработки и использования ИИ» текущие усилия по ответственной и надежной разработке и использованию ИИ. В разделе «От индивидуалистической к социальной концепции ИИ» я описываю необходимость расширения текущих исследований ИИ от индивидуалистической к социальной концепции ИИ» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Вирджиния Дигнум обозначает проблему размывания понятий об ИИ и предлагает собственное понимание как этого феномена, так и практических принципов, которые

должны быть воплощены для максимального снижения рисков внедрения технологии ИИ. В следующей главе своего раздела «Что такое ИИ и почему нас это должно волновать?» она конкретизирует свои тезисы следующим образом:

«Технологические разработки принесли множество потенциальных преимуществ, но в то же время все чаще сообщается о рисках и проблемах, связанных с приложениями на основе ИИ. Все эти многочисленные рекомендации по управлению и регулированию ИИ рискуют стать недействительными без понимания того, что такое ИИ и что он может и чего не может делать. Нынешние нарративы об ИИ рассказывают о преимуществах и рисках и описывают ИИ по-разному: от очевидного следующего шага в оцифровке до своего рода волшебства. Магия в том смысле, что она может знать о нас все и использовать это знание, чтобы принимать решения о нас или за нас, возможно, неожиданным образом, либо решая все наши проблемы, либо разрушая мир в процессе. Реальность, как обычно, где-то посередине.

В настоящее время ИИ в основном ассоциируется с машинным обучением (ML). Машинное обучение и, в частности, нейронные сети или глубокое обучение — это подмножество методов искусственного интеллекта, в которых используются статистические методы, позволяющие компьютерам воспринимать некоторые характеристики своей среды. Современные методы особенно эффективны при восприятии изображений, письменного или устного текста, а также во многих приложениях структурированных данных. Анализируя многие тысячи примеров (обычно несколько миллионов), система может выявить общие черты в этих примерах, что затем позволяет ей интерпретировать данные, которые она никогда раньше не видела, что часто называют прогнозированием» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Таким образом конкретизируется понятие ИИ через его технологии: машинное обучение, в том числе, нейронные сети или глубокое обучение. Обращается внимание на то, что современные технологии ИИ связаны в основном с различными текстами: вербальными, изобразительными. При этом система ИИ может дать некоторую интерпретацию и выполнить генерирование изображения или текста, которого ранее не существовало.

Далее профессор Диггум разоблачает некоторые мифы об ИИ, в том числе те, которые связаны с данным термином:

«ИИ не интеллектуален. Джон Маккарти, который первоначально ввел термин «искусственный интеллект», определил его как «изучение и проектирование интеллектуальных агентов». В этом определении, которое до сих пор является одним из

наиболее распространенных определений ИИ, понятие интеллекта относится к способности компьютеров выполнять задачи, обычно связанные с разумными существами, т. е. людьми или другими нечеловеческими животными. Остается вопрос, что такое человеческий (или животный) интеллект. Обычно ассоциируется со способностью разума делать правильные выводы о том, что верно, а что ложно, и о том, как решать проблемы (Colman 2015), не существует единого общепринятого определения интеллекта. Более того, интеллект — многогранное понятие. Психологи обсуждают такие вопросы, как типы интеллекта, роль природы и воспитания в интеллекте, как интеллект представлен в мозгу и значение групповых различий в интеллекте. Основные теории включают триархическую теорию Штернберга (Sternberg, 1984), теорию множественного интеллекта Гарднера (Gardner, 2011) и теорию развития Пиаже (Piaget, 1964). Многие характеризуют человеческий интеллект как нечто большее, чем просто аналитический процесс, включающий в себя творческие, практические и другие способности. Эти способности, в значительной степени связанные с социокультурным фоном и контекстом, далеки от возможности воспроизвести системами ИИ, даже если они могут приблизиться к аналитическому интеллекту для некоторых (простых) задач» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

В следующей небольшой главе Вирджиния Дигнум дезавуирует миф об «искусственности» ИИ:

«ИИ — это не магия. Он не решит всех наших проблем и не может существовать без использования природных ресурсов и труда легионов людей. В недавней книге «Атлас ИИ» (Crawford, 2021) Кейт Кроуфорд описывает поле как набор карт, которые позволяют читателю перемещаться по местам, их отношениям и их влиянию на ИИ как на инфраструктуру. От шахт, где производятся основные компоненты аппаратного обеспечения, до складов, где люди являются простыми слугами, до автоматизированной структуры, в беспокойном напоминании о «Новых временах» Чаплина, до трудностей классификации данных низкооплачиваемыми рабочими на фермах по маркировке данных, Кроуфорд раскрывает суровую реальность скрытой стороны успеха ИИ. Завершая важным напоминанием о том, что ИИ не является объективной, нейтральной и универсальной вычислительной техникой, а глубоко укоренился в социальной, политической, культурной и экономической реальности тех, кто его создает, использует и в основном контролирует» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

«ИИ — это не алгоритм. ИИ основан на алгоритмах. Понятие «алгоритм» достигает волшебных пропорций, используется справа и слева для обозначения многих вещей, де-

факто рассматривается как синоним ИИ. Но, хотя ИИ использует алгоритмы, как и любая другая компьютерная программа или инженерный процесс, ИИ не является алгоритмом.

Самый простой способ понять алгоритм — это рецепт, набор точных правил для достижения определенного результата. Каждый раз, когда вы складываете два числа, вы используете алгоритм, как и когда печете яблочный пирог. Однако сам по себе рецепт никогда не превращался в яблочный пирог; и конечный результат вашего пирога во многом зависит от ваших навыков выпечки и выбора ингредиентов, а также от выбора конкретного рецепта. То же самое относится и к алгоритмам ИИ: в значительной степени поведение и результаты системы зависят от ее входных данных и от выбора, сделанного теми, кто разработал, обучил и выбрал алгоритм. Точно так же, как у нас есть выбор использовать органические яблоки для приготовления нашего пирога, в ИИ у нас также есть выбор использовать данные, которые уважают и обеспечивают справедливость, конфиденциальность, прозрачность и все другие ценности, которыми мы дорожим. Это то, что касается ответственного ИИ, и включает в себя требование тех же требований от тех, кто разрабатывает системы, которые влияют на нас» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Следующее важное определение Вирджиния Дигнум делает по поводу ответственного ИИ:

«**Ответственный ИИ.** ИИ — это прежде всего технология, которая может автоматизировать простые, меньшие задачи. В настоящее время системы ИИ в значительной степени неспособны понимать смысл. Система искусственного интеллекта может правильно идентифицировать кошек на фотографиях или раковые клетки на сканированных изображениях, но она понятия не имеет, что такое кошка или раковая клетка. Более того, система ИИ может делать это только в том случае, если есть достаточное количество людей, выполняющих задачи (классификация, сбор, обслуживание...), которые необходимы для того, чтобы система функционировала, что может ввести в заблуждение, автономным образом.

Но это гораздо больше, как с точки зрения используемых методов, так и с точки зрения воздействия на общество и участия человека. Таким образом, ИИ лучше всего понимать как социально-техническую экосистему. Чтобы понять ИИ, необходимо осознать взаимодействие между людьми и технологиями, а также то, как сложные инфраструктуры влияют и зависят от общества и человеческого поведения.

Таким образом, ИИ — это не только автоматизация решений и действий, способность учиться на изменениях, затрагиваемых в окружающей среде, и интерактивность, необходимая для того, чтобы быть чувствительным к действиям и целям

других агентов в этой среде и решать, когда следует действовать. сотрудничать или конкурировать. В основном речь идет о структурах власти, участии и доступе к технологиям, которые определяют, кто может влиять на то, какие решения или действия автоматизируются, какие данные, знания и ресурсы используются для обучения и как взаимодействуют между теми, кто принимает решения, и теми, кто принимает решения. затронутые определяются и поддерживаются.

Ответственный, этический подход к ИИ обеспечит прозрачность того, как осуществляется адаптация, ответственность за уровень автоматизации, на котором система способна рассуждать, а также ответственность за результаты и принципы, которые определяют ее взаимодействие с другими, что наиболее важно, с людьми. Кроме того, и прежде всего, *ответственный подход к ИИ ясно показывает, что системы ИИ — это артефакты, созданные людьми для какой-то цели, и что те, кто их создает, имеют право принимать решения об использовании ИИ. Пришло время обсудить, как властные структуры определяют ИИ и как ИИ устанавливает и поддерживает властные структуры, а также баланс между теми, кто получает выгоду от использования ИИ, и теми, кто страдает от его использования* (Crawford, 2021) (курсив – Ю. Замараева и М. Колесник)» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Данное понимание разворачивается в определенную концептуальную последовательность.

«Обеспечение ответственной разработки и использования ИИ. Этический ИИ — это не способ возложить на машины какую-то «ответственность» за их действия и решения, как могут утверждать некоторые, и в процессе снять с людей и организаций их ответственность. Напротив, этический ИИ дает вовлеченным людям и организациям большую ответственность и большую ответственность: за решения и действия приложений ИИ, а также за их собственное решение об использовании ИИ в данном контексте приложения. При рассмотрении эффектов и управления ими технология или артефакт, в котором эта технология реализована, не может быть отделена от социотехнической экосистемы, компонентом которой она является. Руководящие принципы и стратегии, обеспечивающие доверие и ответственность в отношении ИИ, должны быть направлены на социально-техническую экосистему, в которой разрабатывается и используется ИИ. Это не артефакт или приложение ИИ, которые должны быть этическими, заслуживающими доверия или ответственными. Скорее, это *социальный компонент этой экосистемы, который может и должен взять на себя ответственность и действовать с учетом этических рамок, чтобы общество могло доверять всей системе* (курсив – Ю.Замараева и М.

Колесник). Утверждая это, управление может быть достигнуто несколькими способами, более мягкими или жесткими. В настоящее время исследуются несколько направлений, основные из которых выделены в оставшейся части этого раздела. Будущие исследования и опыт позволят определить, какие подходы являются наиболее подходящими, но, учитывая сложность проблемы, весьма вероятно, что потребуется сочетание подходов» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Первый принцип, который должен иметь практическое значение – это регулирование:

«Регулирование. Регулирование ИИ — горячая тема со многими сторонниками и противниками. Недавнее предложение Европейской комиссии предусматривает подход к регулированию, основанный на оценке риска, который гарантирует, что люди могут доверять тому, что технология ИИ используется безопасным образом и в соответствии с законом, включая уважение основных прав человека.

Предложение реализует большинство из 7 требований Руководства по этике для надежного ИИ в конкретные требования к ИИ с «высоким риском». Однако в нем прямо не рассматриваются вопросы инклюзивности, недискриминации и справедливости. Сведение к минимуму или устранение дискриминационной предвзятости или несправедливых результатов — это больше, чем исключение использования данных низкого качества. Дизайн любого артефакта, такого как система ИИ, сам по себе представляет собой совокупность вариантов выбора, а выбор по своей природе предвзят, поскольку предполагает выбор одного варианта над другим. Технические решения на уровне набора данных должны дополняться социотехническими процессами, которые помогают избежать любых дискриминационных или несправедливых результатов ИИ.

Более того, успешное регулирование требует четкого выбора того, что подлежит регулированию: сама технология, воздействие или результаты ее применения? Сосредоточив внимание на технологиях или методах, т. е. регулируя системы, основанные на «машинном обучении, логике или статистических подходах», таких как описано в определении ИИ, используемом в предложении Европейской комиссии, мы рискуем увидеть организации, уклоняющиеся от регулирования, просто классифицируя их приложения по-разному. И наоборот, существует множество приложений, основанных, например, на статистике, которая не является ИИ.

Ориентированное на будущее регулирование должно быть сосредоточено на результатах систем, независимо от того, соответствуют ли эти системы текущему пониманию того, что такое «ИИ». Если кого-то неправильно идентифицируют, ему отказывают в правах человека или доступе к ресурсам, или его принуждают верить или

действовать определенным образом, не имеет значения, является ли система «ИИ» или нет. Это просто неправильно. Более того, регулирование должно также учитывать входные данные, процессы и условия, при которых разрабатывается и используется ИИ, которые не менее важны. Много было сказано об опасностях необъективных данных и дискриминирующих приложений. Внимание к социальным, экологическим и климатическим издержкам систем ИИ возрастает. Все это должно быть включено в любые усилия по обеспечению ответственной разработки и использования ИИ.

В то же время системы ИИ являются компьютерными приложениями, т.е. являются артефактами, и как таковые подчиняются существующим ограничениям, законодательству, в отношении которых применяются обязательства и обязательства должной осмотрительности. То есть уже сейчас ИИ не действует в незаконном пространстве. Прежде чем определять дополнительные правила, нам нужно начать с понимания того, что уже охвачено существующим законодательством.

Подход к регулированию, основанный на оценке рисков, предложенный Европейской комиссией, является правильным направлением, но он должен основываться на четком понимании того, что является источником этих рисков. Более того, необходимо сосредоточиться не только на технических решениях на уровне алгоритмов или наборов данных, но и на разработке социально-технических процессов и корпоративной ответственности, чтобы избежать и смягчить любые дискриминационные или несправедливые результаты. Независимо от того, называем мы систему «ИИ» или нет» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Следующий практикоориентированный принцип – это стандартизация.

«Стандартизация. Стандарты — это основанные на консенсусе согласованные способы ведения дел, предоставляющие то, что они считают минимальными общепризнанными спецификациями. Доказано, что отраслевые стандарты полезны для организаций и частных лиц. Стандарты могут помочь снизить затраты и повысить эффективность организаций за счет обеспечения согласованности и показателей качества, создания общего словаря, хороших методологий проектирования и архитектурных рамок. В то же время стандарты обеспечивают потребителям уверенность в качестве и безопасности продукции и услуг.

Большинство стандартов считаются мягким управлением; то есть необязательное для выполнения. Тем не менее, часто в интересах компаний следовать им, чтобы продемонстрировать должную осмотрительность и, следовательно, ограничить свою юридическую ответственность в случае инцидента. Более того, стандарты могут обеспечить удобную интеграцию продуктов (Theodorou and Dignum, 2020).

Работа над стандартами ИИ для поддержки управления разработкой и использованием ИИ продолжается в ISO и IEEE, двух ведущих органах по стандартизации. Такие стандарты могут поддерживать цели политики ИИ, в частности, когда речь идет о безопасности, защищенности и надежности ИИ, гарантиях объяснимости и средствах уменьшения предвзятости в алгоритмических решениях (Cihon, 2019).

Совместно с МЭК ИСО учредила Комитет по стандартам искусственного интеллекта (SC 42). Текущие усилия SC 42 пока носят ограниченный и предварительный характер (Cihon 2019). С другой стороны, глобальная инициатива Ассоциации стандартов IEEE по этически согласованному проектированию активно работает над концепцией и рекомендациями по учету ценностей и намерений, а также над юридическими и техническими реализациями автономных и интеллектуальных систем для обеспечения благополучия человека (IEEE 2016). Это совместная работа более 700 международных исследователей и практиков. В частности, серия P7000Footnote2 направлена на разработку стандартов, которые в конечном итоге послужат основой и опорой для будущих норм и стандартов в рамках новой структуры этического управления проектированием ИИ/АС. В настоящее время рабочие группы P7000 работают над стандартными рекомендациями-кандидатами для решения таких разнообразных проблем, как проектирование системы, прозрачность в автономных системах, алгоритмическая предвзятость, управление персональными данными, данными о детях, студентах и работодателях, подталкивание или идентификация и оценка достоверности данных. источники новостей. Примечательно, что усилия по оценке влияния автономных и интеллектуальных систем на благополучие человека теперь доступны в виде стандарта IEEE» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

«Оценка. Ответственный ИИ — это больше, чем установка некоторых этических «галочек» или разработка некоторых дополнительных функций в системах ИИ. Тем не менее, разработчики и пользователи могут воспользоваться поддержкой и конкретными шагами для понимания соответствующих правовых и этических стандартов и соображений при принятии решений об использовании приложений ИИ. Инструменты оценки воздействия обеспечивают пошаговую оценку воздействия систем, методов или инструментов на такие аспекты, как конфиденциальность, прозрачность, объяснение, предвзятость или ответственность (Taddeo and Floridi, 2018 г.).

Важно понимать, как описано у Taddeo and Floridi (2018), что, хотя эти подходы «никогда не могут отобразить весь спектр возможностей, рисков и непредвиденных последствий систем ИИ, они могут определить предпочтительные альтернативы, ценные направления действий, вероятные риски и стратегии их снижения. Это имеет двойное

преимущество. В качестве стратегии возможностей методологии предвидения могут помочь использовать этические решения. Как форма управления рисками они могут помочь предотвратить или смягчить дорогостоящие ошибки, избегая решений или действий, которые являются этически неприемлемыми».

В настоящее время прилагаются большие усилия для разработки инструментов оценки. Руководство ЕС по достоверному ИИ сопровождается комплексной системой оценки, которая была разработана на основе процесса консультаций с общественностью.

Наконец, важно понимать, что любые требования к надежному ИИ необходимы, но недостаточны для разработки ИИ, ориентированного на человека. То есть такие требования необходимо понимать и реализовывать с контекстуальной точки зрения, т. е. должна быть возможность корректировать реализацию требования, такого как прозрачность, в зависимости от контекста, в котором используется система. То есть такие требования, как прозрачность, не должны иметь одного фиксированного определения для всех систем ИИ, а должны определяться на основе того, как используется система ИИ. В то же время любой метод искусственного интеллекта, используемый при разработке и реализации, должен учитывать все этические требования. Например, должна быть возможность объяснить (или показать), как система пришла к определенному решению или поведению.

Инструменты оценки должны учитывать эту контекстуализацию, а также обеспечивать соответствие существующим структурам и требованиям с точки зрения других типов оценки, чтобы оценка доверия и ответственности систем ИИ приносила дополнительную пользу тем, кто их разрабатывает и использует, а не добавлять еще одну бюрократическую нагрузку.

Кодексы поведения и консультативные советы. Профессиональный кодекс поведения — это публичное заявление, разработанное для профессиональной группы и отражающее общие принципы практики, поведения и этики тех, кто занимается профессией; описать качество поведения, отражающее ожидания профессии и общества; предоставить обществу четкое заявление об этих ожиданиях и дать профессионалам возможность обдумать свои собственные этические решения.

Кодекс поведения помогает профессионалам оценивать и решать сложные профессиональные и этические дилеммы. В то время как в случае этических дилемм нет правильного решения, профессионалы могут дать отчет о своих действиях, ссылаясь на кодекс. В соответствии с другими социально чувствительными профессиями, такими как врачи или юристы, то есть с сопутствующей сертификацией «этического ИИ», может поддерживать доверие. Несколько организаций работают над разработкой кодексов поведения для профессий, связанных с данными и ИИ, с особыми этическими

обязанностями. Совсем недавно АСМ, Ассоциация вычислительной техники, крупнейшая международная ассоциация профессионалов в области вычислительной техники, обновила свой кодекс поведения. Этот добровольный кодекс представляет собой «сборник принципов и руководств, призванных помочь профессионалам в области вычислительной техники принимать этически ответственные решения в профессиональной деятельности. Он переводит широкие этические принципы в конкретные положения о профессиональном поведении». В этом кодексе прямо рассматриваются вопросы, связанные с разработкой систем ИИ, а именно вопросы возникающих свойств, дискриминации и конфиденциальности. В частности, он призывает технических специалистов обеспечить инклюзивность и доступность систем для всех и требует, чтобы они были осведомлены о вопросах конфиденциальности.

В то же время роль специалиста по этике ИИ становится горячей темой, поскольку крупные предприятия все больше зависят от ИИ, а влияние этих систем на людей и общество становится все более очевидным, и не всегда в лучшую сторону. Недавние скандалы, связанные как с влиянием ИИ на предвзятость и дискриминацию, так и с тем, как предприятия справляются со своей собственной ответственностью, особенно в отношении роли и обращения с осведомителями, повысили спрос на четкие и четкие организационные структуры для борьбы с воздействием ИИ.

С тех пор многие организации учредили роль главного специалиста по этике ИИ или аналогичную должность. Другие, признавая, что социальные и этические проблемы, возникающие в связи с ИИ, сложны и многогранны и, следовательно, требуют понимания и опыта из многих различных дисциплин и открытого участия различных заинтересованных сторон, создали советы по этике ИИ или консультативные группы.

Осведомленность и участие. Инклюзивность и разнообразие являются более широкой социальной проблемой и имеют центральное значение для развития ИИ. Поэтому важно, чтобы как можно более широкая группа людей имела базовые знания об ИИ, о том, что можно (и нельзя) делать с помощью ИИ и как ИИ влияет на индивидуальные решения и формирует общество. Хорошо известной инициативой в этой области является Elements of AI, Footnote, инициированная в Финляндии с целью обучить один процент граждан ЕС основам искусственного интеллекта, тем самым укрепив цифровое лидерство в ЕС.

Параллельно с этим исследования и разработки систем искусственного интеллекта должны основываться на разнообразии во всех смыслах разнообразия и, очевидно, включая пол, культурное происхождение и этническую принадлежность. Более того, ИИ больше не является инженерной дисциплиной, и в то же время появляется все больше свидетельств того, что когнитивное разнообразие способствует лучшему принятию решений. Поэтому в

состав разрабатывающих команд должны входить социологи, философы и другие, а также обеспечивать гендерные, этнические и культурные различия. Не менее важно разнообразить дисциплину и опыт тех, кто работает над ИИ, чтобы включить в него специалистов по ИИ, обладающих знаниями, среди прочего, в философии, социальных науках, праве и экономике. Регулирование и кодексы поведения могут определять цели и задачи, а также стимулы, чтобы способствовать разнообразию в командах ИИ» (Dignum, 2020).

От индивидуалистической к социальной концепции ИИ. До сих пор доминирующим подходом к ИИ был индивидуалистический, рациональный подход. Классический учебник Рассела и Норвига по ИИ определяет ИИ в двух измерениях (Russell and Norvig, 2010): как он рассуждает (по-человечески или рационально) и что он «делает» (думает или действует). Человекоподобные подходы направлены на то, чтобы понять и смоделировать работу человеческого разума, а рациональные подходы направлены на разработку систем, обеспечивающих оптимальный уровень выгоды или полезности для человека. Оба подхода хорошо согласуются с утверждением западной философии «Я мыслю, следовательно, существую», фундаментально концептуализируя систему ИИ как индивидуальную сущность.

Интеллектуальные агенты обычно характеризуются как ограниченно-рациональные, действующие в своих собственных предполагаемых интересах. Например, идентифицируя и применяя шаблоны в (создаваемых человеком) данных, системы машинного обучения имитируют и расширяют человеческие рассуждения и действия, заложенные в этих данных, в то время как подходы символической логики (так называемый «старый добрый ИИ», или GOFAI) стремятся уловить законы рационального мышления и действия, что приводит к идеализированной модели человеческого мышления.

Человекоподобный или рациональный? То есть рациональность часто является центральным допущением для обсуждения агентом (Dignum, 2017). Более того, ожидается, что интеллектуальные системы будут придерживаться согласованных взглядов на мир (верований) и оптимизировать действия и решения на основе набора заданных предпочтений (часто точность имеет наивысший приоритет). Этот взгляд на рациональность подразумевает, что от агентов ожидается и они должны действовать рационально в том смысле, что они выбирают наилучшие доступные средства для достижения данной цели и поддерживают согласованность между желаемым и выбранным (Lindenberg, 2001).

Основными преимуществами предположения о рациональности являются их экономичность и применимость к очень широкому кругу ситуаций и сред, а также их

способность генерировать фальсифицируемые, а иногда и эмпирически подтверждаемые гипотезы о действиях в этих средах. Это дает традиционным подходам рационального выбора сочетание общности и предсказательной силы, которых нет в других подходах.

К сожалению, этот тип рационального поведения в основном соответствует стратегическому выбору, когда вся информация доступна или может быть собрана по желанию. На самом деле это не подходит большинству людей, которые основаны на решениях за доли секунды, на привычках, на социальных условностях и структурах власти. Когда целью систем ИИ является разработка моделей социального поведения или разработка систем, способных взаимодействовать с людьми в социальных условиях, рациональности недостаточно для моделирования человеческого поведения. Это было продемонстрировано ранее во всех прикладных областях, где рациональное поведение должно сочетаться с различными типами поведения, чтобы быть эффективным. В действительности человеческое поведение не является ни простым, ни рациональным, а проистекает из сложного сочетания умственных, физических, эмоциональных и социальных аспектов. Более того, реалистичные приложения должны учитывать ситуации, в которых нельзя предвидеть все альтернативы, последствия и вероятности событий. Таким образом, «рационально» оптимизировать полезность невозможно, так как функция полезности не известна полностью, как не известны и критерии оптимизации. Это делает подходы к рациональному выбору неспособными точно моделировать и предсказывать широкий спектр человеческого поведения. Уже в 2010 году Dignum и Dignum (2010) показали, как различные типы вариаций и моделей подходят для различных приложений, в то время как не существует универсальной модели, которая служила бы основой для всех моделей.

И человекоподобный, и рациональный взгляды на ИИ подходят для ориентированного на задачи взгляда на назначение систем ИИ. То есть ожидается, что система оптимизирует результат своих действий для конкретной цели. Несмотря на то, что он способен воспринимать окружающую среду и соответствующим образом адаптироваться, он в основном не осознает своей роли в этой среде и того факта, что его действия способствуют изменениям. Учитывая большое влияние ИИ на общество, *необходима новая парадигма моделирования, способная учитывать эту петлю обратной связи решение — действие — контекст* (курсив – Ю. Замараева и М. Колесник). **То есть моделирование ИИ должно следовать социальной парадигме.**

Социальный ИИ. Западные философии, и, в частности, Убунту, занимают общественную, а не индивидуальную позицию, что вызывает вопрос о том, как можно определить ИИ с точки зрения мысли Убунту. Не пытаюсь описать или полностью понять

философию Убунту, ниже я применю некоторые из ее основных принципов, чтобы показать, как их можно применить к концепциям ИИ и подходам к разработке.

Убунту выражает глубоко укоренившиеся африканские идеалы личности, уходящие корнями во взаимосвязь с другими, и подчеркивает нормы межличностных отношений, которые способствуют социальной справедливости, такие как взаимность, самоотверженность и симбиоз. Сообщество лежит в основе Убунту, уделяя особое внимание взаимосвязи и заботе о совместной жизни, подкрепленной ценностями сотрудничества и сотрудничества (Mugumbate and Nyanguru, 2013). Солидарность, которая требует, чтобы люди осознавали потребности окружающих и обращали на них внимание, а не сосредотачивались только на своих собственных потребностях, поэтому занимает центральное место в Ubuntu с упором на заботу, заботу и контекст (van Breda, 2019).

Таким образом, философия Убунту по существу является реляционной и определяет морально правильные действия как те, которые соединяют, а не разделяют (т. е. уважают общественные отношения, уменьшают разногласия или способствуют дружеским отношениям). должны направлять то, чего хочет большинство, или какие моральные нормы становятся центральными (Ewuoso and Hall, 2019). Это не означает, что индивидуальные права подчинены, но что люди преследуют свое собственное благо, стремясь к общему благу (Lutz, 2009).

Права человека устанавливают основополагающую ценность человеческого достоинства с точки зрения автономии. Эта точка зрения, в значительной степени восходящая к кантовской философии, устанавливает права человека как высший способ уважительного отношения к нашей внутренней ценности способности к самоуправлению. Поэтому утверждалось, что коллективистские основания мышления Убунту противоречат этому взгляду на индивидуальную автономию. Согласно Мецу, «в то время как кантовская теория состоит в том, что люди обладают превосходной ценностью, потому что они обладают способностью к автономии, нынешняя, вдохновленная Убунту версия состоит в том, что они имеют такую ценность, потому что они обладают способностью общаться с другими на общинной основе. (Мец 2011, 544). Или, как также описывает Мец, «нарушения прав человека — это способы серьезного неуважения к способности людей к общественным отношениям, понимаемым как идентичность и солидарность [...]» (там же, 545). В Ubuntu человеческая природа особенная и неприкосновенная благодаря своей способности к гармоничным отношениям. В то же время ни один человек не обладает большими правами, чем другой, поэтому каждый человек в сообществе, включая детей и взрослых, важен и должен быть услышан и уважаем (Osei-Hwedie, 2007).

Что касается этики разработки и использования ИИ, приведенная выше формулировка человеческого достоинства как способности человека общаться с другими на основе общности может объяснить или оправдать решение моральных дилемм, когда автономия вступает в противоречие с благодеянием или любым другим, других принципов, также предложенных в отношении биоэтики и медицины, или для обоснования целей ООН в области устойчивого развития (Ewuoso and Hall, 2019).

В частности, формулировка Убунту, описанная выше, может быть использована для оправдания решений перед лицом этической дилеммы, например, когда такое решение благоприятствует действию, которое укрепляет общественные отношения, или способности к тому же (там же). Таким образом, эта структура может с пользой дополнить утилитарный, индивидуалистический и деонтологический подходы, которые часто используются при принятии решений по этике ИИ. Как было предложено для клинического контекста, принятие этических решений в контексте систем ИИ также может быть дополнено правилами, которые гласят: «Нарушение этического принципа оправдано, если с точки зрения вероятностей такое нарушение более вероятно, улучшит общественные отношения (...)» (Ewuoso and Hall 2019).

На пути к социальной парадигме ИИ. Как обсуждалось в разделе «Что такое ИИ и почему нас это должно волновать?», ИИ не является ни «искусственным», ни «интеллектуальным», а продуктом выбора, включающего теорию и ценности. Текущая парадигма ИИ, как показано в разделе «Человекоподобный или рациональный?», опирается на связанные, но индивидуалистические теории интеллекта, мышления, рациональности и человеческой природы. Как таковая, эта парадигма поддерживает реализацию различных подходов к рассуждениям и действиям, соответствующих индивидуальному пониманию контекстов и различным реакциям агентов на контексты. Однако ИИ, как и все технологии, влияет на наш мир и изменяет его, что, в свою очередь, меняет нас. Необходимы новые парадигмы, которые учитывают коллективное понимание и влияние изменений в контексте, а также петлю обратной связи от изменений обратно к индивидуальным и коллективным рассуждениям и поведению. Моделирование этой петли обратной связи признает, что речь идет не только о том, какое действие выполняется, но и о том, какие рассуждения приводят к этому действию, и какие ценности и восприятия ведут к наблюдению за контекстом.

Учитывая транснациональный характер ИИ, также необходимо рассмотреть способы, которыми ИИ может повлиять или быть принятым обществом в различных регионах мира. В частности, необходимо позиционировать африканский континент в глобальных дебатах и выработке политики в рамках ответственного ИИ. Например, такие

инициативы, как Responsible AI Network—AfricaFootnote и Африканская обсерватория ответственного AIFootnote, направлены на то, чтобы понять, как ИИ может повлиять или быть принятым обществом в различных регионах мира, углубить понимание ИИ и его воздействия на макрорегион Сахару в Африке, а также содействовать разработке и внедрению соответствующих местным условиям политик в области искусственного интеллекта, основанных на фактических данных, и соответствующего законодательства.

Однако в то же время не менее важно расширить существующие концепции ИИ реляционным мировоззрением, которое характеризует африканскую мысль, например, встроенную в философию Убунту. Современные парадигмы ИИ укрепляют существующие структуры власти и препятствуют по-настоящему общественному пониманию воздействия и проблем ИИ для человечества и общества. Учитывая влияние, которое системы ИИ могут оказывать на людей, межличностные взаимодействия и общество в целом, кажется уместным рассмотреть относительную позицию при подходе к спецификации, разработке и анализу систем ИИ. Глубоко укоренившееся африканское видение собственной личности, уходящей корнями во взаимосвязь, выраженное в философии Убунту, может способствовать интеграции такой социальной перспективы в ИИ — с точки зрения того, как он рассуждает (по-человечески или рационально) и что он «делает» (подумайте или действие), что приведет к новой парадигме, которая рассматривает социальное/коллективное мышление и включает изменение или реакцию в качестве третьего возможного результата ИИ, наряду с мышлением и действием. На рис. 8 показан этот взгляд на ИИ, расширяющий хорошо известные измерения, определенные в (Рассел и Норвиг, 2010), здесь они показаны серым цветом.

	Human-like	Rational	Social
Think	Think humanly	Think rationally	Think socially
Act	Act humanly	Act rationally	Act jointly
Change/react	Enhance human performance	Rational institutions	Social engagement

Рис. 8. Социальные перспективы ИИ. Таблица, иллюстрирующая раздел Вирджинии Дигнум «Ответственный ИИ: рекомендации и извлеченные уроки». Источник изображения: https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9/figures/1

Коллективное или социальное рассуждение касается моделирования социальных ценностей и норм, а также того, как они обосновывают и влияют на человеческое и рациональное мышление. В социальном действии результаты соотносятся с действиями других, где другие рассматриваются не только как противники или препятствия на пути принятия решений, но и как позитивная сила для достижения совместных усилий.

Рассмотрение изменений с человеческой точки зрения приводит к подходам к ИИ, направленным на повышение, а не замену человеческой деятельности, а с рациональной точки зрения это касается разработки и оптимизации институциональных инфраструктур, которые максимизируют эффекты рационального поведения.

Все эти точки зрения необходимо объединить, чтобы учесть влияние ИИ на социально обоснованную и заинтересованную перспективу. Это нелегкий подвиг, для которого нет ни единых моделей, ни простых подходов. Это потребует междисциплинарного и многостороннего участия и признания того, что любое решение всегда зависит от обстоятельств и контекста» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Далее Вирджинии Дигнум делает ряд выводов, связанных с принципами ответственного ИИ, учитывающими и общечеловеческие моменты, и социокультурные, связанные в данном случае с африканской философией Убунту:

«Системы искусственного интеллекта все чаще будут принимать решения, которые влияют на нашу жизнь в большей или меньшей степени. *Во всех областях применения ИИ должен уметь учитывать общественные ценности, моральные и этические соображения, взвешивать соответствующие приоритеты ценностей, которых придерживаются различные заинтересованные стороны и в мультикультурном контексте, объяснять свои рассуждения и гарантировать прозрачность* (курсив – Ю.Замараева и М. Колесник). По мере роста возможностей автономного принятия решений, возможно, самым важным вопросом для рассмотрения является необходимость переосмысления ответственности. Будучи по своей сути инструментами, системы ИИ полностью находятся под контролем и ответственностью своих владельцев или пользователей. Однако их потенциальная автономность и способность к обучению требуют, чтобы при разработке учитывались принципы подотчетности, ответственности и прозрачности явным и систематическим образом. До сих пор разработка алгоритмов искусственного интеллекта велась с целью повышения производительности, что привело к созданию непрозрачных черных ящиков. Помещение человеческих ценностей в основу систем искусственного интеллекта требует изменения взглядов исследователей и разработчиков на повышение прозрачности, а не производительности, что приведет к новым и интересным методам и приложениям. В

частности, это требует дополнения преобладающего в настоящее время индивидуалистического взгляда на системы ИИ таким, который признает и включает коллективные, социальные и этические ценности, лежащие в основе проектирования, разработки и использования систем ИИ.

Биологическая эволюция уже давно пересматривается с точки зрения «лестницы»: однолинейная прогрессия от «примитивного» к «продвинутому». Тот же пересмотр наблюдается и в антропологии: была показана идея о том, что культурная эволюция следует модели лестницы с небольшими децентрализованными обществами внизу и иерархическими, государственными, обществами наверху, где вершина была бы технологически более развитой, было показано, что это не только уничижительным, но и неточно (Eglish 1999). Эти поля уже давно перешли к более динамичной модели с ветвящимся типом. Настало время, чтобы модели ИИ приняли такое ветвящееся представление. Только тогда ИИ сможет соответствовать разнообразию, которое действительно отражает мировые различия в культурном и философском мышлении. Точно так же, как биология или культура, интеллект не является линейным, он ветвящийся» (https://link.springer.com/chapter/10.1007/978-3-031-08215-3_9#author-information).

Выводы

Подводя итоги рассмотрению концепции Вирджинии Дигнум, изложенной ею в разделе «Ответственный ИИ: рекомендации и извлеченные уроки» коллективной монографии «Ответственный ИИ в Африке: проблемы и возможности» (2023), можно увидеть определенный методологический принцип, опираясь на который профессор Дигнум приходит к социокультурному рассмотрению проблем ИИ. Этот принцип связан 1) с указанием на то, что ИИ вписан в определенную социальную и культурную экосистему и все его возможности и ограничения проистекают не столько из него самого, сколько из особенностей социальной и культурной экосистемы, в которой он разворачивается; 2) ИИ действует сегодня в различных регионах глобального мира, нет такой территории, где бы не применялись эти технологии, поэтому везде есть собственный социокультурный опыт применения технологии ИИ, который может быть научно проанализирован и учтен; 3) на примере философии Убунту показано, как международная и национальная этика ИИ, делающая его «ответственным ИИ» может быть обогащена за счет т.н. «незападных», «деонтологизированных» философских подходов.

Применяя эти принципы для анализа различных стран, регионов и сообществ, возможно не только выполнить актуальное научное моделирование, но и реализовать прогностическую функцию, в том числе, предложить какие-то уже существующие в сообществе философские принципы, которые могли бы содействовать улучшению

социальных коммуникаций во всем их культурном многообразии, которое не поддается «простой» унификации и стандартизации.

Список литературы

Anderson J., Rainie L., Luchsinger A. Artificial intelligence and the future of humans //Pew Research Center. – 2018. – V. 10. – №. 12. URL: <http://tony-silva.com/eslefl/miscstudent/downloadpagearticles/AIhumanfuture-pew.pdf>

Cihon P. Standards for AI governance: international standards to enable global coordination in AI research & development //Future of Humanity Institute. University of Oxford. – 2019. 41 p. URL: https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf

Coeckelbergh, M. AI Ethics. – Cambridge, MA: MIT Press, 2020. 248 p.

Colman A. M. A dictionary of psychology. – Oxford quick reference, 2015. 878 p.

Crawford K. The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. – Yale University Press, 2021. 315 p.

Dignum F., Dignum V. How to center AI on humans //NeHuAI 2020, 1st International Workshop on New Foundations for Human-Centered AI, Santiago de Compostella, Spain, September 4, 2020. – 2020. – P. 59-62.

Dignum V. Social agents: Bridging simulation and engineering // Commun. ACM. – 2017. – V. 60 – P. 32–34.

Dignum V. The myth of complete AI-fairness //Artificial Intelligence in Medicine: 19th International Conference on Artificial Intelligence in Medicine, AIME 2021, Virtual Event, June 15–18, 2021, Proceedings. – Springer International Publishing, 2021. – P. 3-8. DOI https://doi.org/10.1007/978-3-030-77211-6_1

Eke D. O., Wakunuma K., Akintoye S. Responsible AI in Africa: Challenges and Opportunities. – Springer Nature, 2023. 216 p. DOI <https://doi.org/10.1007/978-3-031-08215-3> URL: <https://link.springer.com/book/10.1007/978-3-031-08215-3>

Ewuoso C., Hall S. Core aspects of Ubuntu: A systematic review // South Afr. J. Bioeth. Law – 2019. – № 12. – P. 93–103.

Eze M. O. Toward an African Renaissance //The Politics of History in Contemporary Africa. – 2010. – P. 181-196.

Gardner H. E. Frames of mind: The theory of multiple intelligences. – Basic books, 2011. 528 p.

Hanemaayer A. (ed.). Artificial intelligence and its discontents: Critiques from the social sciences and humanities. – Springer Nature, 2022. 275 p. DOI <https://doi.org/10.1007/978-3-030-88615-8>

Jobin A., Ienca M., Vayena E., 2019. The global landscape of AI ethics guidelines // Nat. Mach. Intell. – V.1. – P. 389–399. <https://doi.org/10.1038/s42256-019-0088-2>

Lindenberg S. Social rationality versus rational egoism // Handbook of Sociological Theory. Springer, 2001. – P. 635–668.

Lutz D.W. African Ubuntu philosophy and global management // J. Bus. Ethics – 2009. – V. 84. – P. 313–328.

Metz T. Ubuntu as a moral theory and human rights in South Africa // Afr. Hum. Rights Law J. – 2011. – V. 11. – P.532–559.

Mugumbate J. R., Chereni A. Now, the theory of Ubuntu has its space in social work //African Journal of Social Work. – 2020. – V. 10. – №. 1.

Mugumbate J., Nyanguru A.Exploring African philosophy: The value of Ubuntu in social work // Afr. J. Soc. Work – 2013. - № 3 – P. 82–100.

Odlind C., Richardson K. The Sexual Politics of Sex Dolls and Sex Robots. – Springer, 2022. 199 p. DOI <https://doi.org/10.1007/978-3-031-19381-1>

Osei-Hwedie K.2007. Afro-centrism: The challenge of social development // Soc. Work. Werk. – 2007. – V. 43 – № 2. DOI: <https://doi.org/10.15270/43-2-279> URL: <https://socialwork.journals.ac.za/pub/article/view/279>

Piaget J. Part I: Cognitive development in children: Piaget development and learning // J. Res. Sci. Teach. – 1964. – № 2. – P. 176–186.

Richardson K. Challenging sociality: An anthropology of robots, autism, and attachment. – Springer, 2018. 156 p. DOI <https://doi.org/10.1007/978-3-319-74754-5>

Russell S., Norvig P., 2010. Artificial intelligence: A modern approach. Third Edit. Prentice Hall 10, B978-012161964. 1154 p. URL: <https://scholar.alaqsa.edu.ps/9195/1/Artificial%20Intelligence%20A%20Modern%20Approach%20%283rd%20Edition%29.pdf%20%28%20PDFDrive%20%29.pdf>

Sternberg R.J., 1984. Toward a triarchic theory of human intelligence // Behav. Brain Sci. – № 7. – P. 269–287.

Taddeo M., Floridi L. How AI can be a force for good // Science. – 2018. – V. 361 – P.751–752.

Theodorou A., Nieves J. C., Dignum V. Good AI for Good: How AI Strategies of the Nordic Countries Address the Sustainable Development Goals //arXiv preprint arXiv:2210.09010. – 2022. <https://doi.org/10.48550/arXiv.2210.09010>

van Breda, A.D., 2019. Developing the notion of Ubuntu as African theory for social work practice // Soc. Work – 2019. – № 55. – P. 439–450.

Vinuesa R., Azizpour H., Leite I., Balaam M., Dignum V., Domisch S., Felländer A., Langhans S.D., Tegmark M., Fuso Nerini F., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat. Commun. 11, 233. <https://doi.org/10.1038/s41467-019-14108-y>

Ермаков, Т. К. Эстетические возможности нейросетей в современном искусстве / Т. К. Ермаков, О. А. Темникова, Е. Е. Шишкова // Социология искусственного интеллекта. – 2021. – Т. 2, № 4. – С. 14-29. – DOI 10.31804/2712-939X-2021-2-4-14-29. – EDN YCAWNO.

Иванова М. Г. Сходства и различия архетипов культур африканских стран и России (на примере философии Убунту) // Африка в контексте формирования новой системы международных отношений. – 2019. – С. 104-119.

Кирко, В. И. Рецензия на книгу “Artificial Intelligence for Business: Innovation, Tools and Practices” (редактор Ана Ландета Эчеберрия, издательство Springer, 2022) / В. И. Кирко // Социология искусственного интеллекта. – 2022. – Т. 3, № 3. – С. 28-40. – DOI 10.31804/2712-939X-2022-3-3-28-40. – EDN ICBKGO.

Кирко, В. И. Рецензия на книгу К. Н. Галсона, С. Селлара, П. Т. Уэбба "Алгоритмы образования: как датафикация и искусственный интеллект формируют политику" (Миннесота пресс, 2022) / В. И. Кирко // Социология искусственного интеллекта. – 2022. – Т. 3, № 2. – С. 34-43. – DOI 10.31804/2712-939X-2022-3-2-34-43. – EDN CDDGQX.

Кирко, В. И. Рецензия на книгу Стюарта Дж. Рассела и Питера Норвига "Искусственный интеллект: современный подход" (4 издание, 2021 г.) / В. И. Кирко // Социология искусственного интеллекта. – 2021. – Т. 2, № 4. – С. 30-35. – DOI 10.31804/2712-939X-2021-2-4-30-35. – EDN OGVVFP.

Колесник, М. А. Искусственный интеллект как инструмент и соавтор в творчестве современных художников: примеры художественных практик и анализ произведений визуального искусства / М. А. Колесник, А. А. Ситникова, Я. Д. Андрюшина // Социология искусственного интеллекта. – 2023. – Т. 4, № 1. – С. 37-51. – DOI 10.31804/2712-939X-2023-4-1-37-51. – EDN UOQRVI.

Копцева, Н. П. Рецензия на книгу "Artificial intelligence and the future of warfare: The USA, China, and strategic stability", автор Джеймс Джонсон (Manchester University Press, 2021) / Н. П. Копцева, А. А. Шпак // Цифровизация. – 2023. – Т. 4, № 1. – С. 50-58. – EDN LMBEAX.

Копцева, Н. П. Рецензия на книгу Джеффри Мур "Зона победы. Управление в эпоху цифровой трансформации" (М.: Манн, Иванов и Фербер, 2020) / Н. П. Копцева //

Цифровизация. – 2022. – Т. 3, № 3. – С. 61-67. – DOI 10.37993/2712-8733-2022-3-3-61-67. – EDN SXAGOP.

Копцева, Н. П. Современные исследования в области социологии искусственного интеллекта: базовые подходы. Часть 6.1 / Н. П. Копцева, Ю. С. Замараева // Социология искусственного интеллекта. – 2023. – Т. 4, № 1. – С. 8-25. – DOI 10.31804/2712-939X-2023-4-1-8-25. – EDN QSHRUJ.

Лещинская, Н. М. Дайджест новостей в сфере развития искусственного интеллекта / Н. М. Лещинская, Н. А. Сергеева // Социология искусственного интеллекта. – 2022. – Т. 3, № 3. – С. 41-51. – DOI 10.31804/2712-939X-2022-3-3-41-51. – EDN INAALZ.

Лузан, В. С. Роль цифрового образовательного контента в деятельности художественных музеев России / В. С. Лузан, О. А. Темникова // Цифровизация. – 2021. – Т. 2, № 3. – С. 19-28. – DOI 10.37993/2712-8733-2021-2-3-19-28. – EDN AOLMST.

Михалина О. А. Африканская философия образования в поисках смысла // Вестник Российского университета дружбы народов. Серия: Философия. – 2010. – №. 3. – С. 39-46.

Омелик, А. А. Возможности использования невзаимозаменяемых токенов (NFT) в художественной и институциональной практике в Красноярске / А. А. Омелик // Цифровизация. – 2022. – Т. 3, № 2. – С. 96-108. – DOI 10.37993/2712-8733-2022-3-2-96-108. – EDN RMXVLW.

Омелик, А. А. Развернутая рецензия на коллективную монографию "Философские проблемы развития искусственного интеллекта" / А. А. Омелик, Ю. В. Квашнина // Социология искусственного интеллекта. – 2022. – Т. 3, № 4. – С. 129-141. – DOI 10.31804/2712-939X-2022-3-4-129-141. – EDN ТПАМІ.

Омелик, А. А. Факторы влияния искусственного интеллекта на творческий процесс / А. А. Омелик // Социология искусственного интеллекта. – 2023. – Т. 4, № 1. – С. 52-63. – DOI 10.31804/2712-939X-2023-4-1-52-63. – EDN YTVWUK.

Резникова, К. В. Искусственный интеллект в Американском кинематографе конца XX - начала XXI веков / К. В. Резникова, Е. А. Сертакова, А. А. Ситникова // Социология искусственного интеллекта. – 2022. – Т. 3, № 1. – С. 42-49. – DOI 10.31804/2712-939X-2022-3-1-42-49. – EDN HZKFIC.

Резникова, К. В. Особенности использования цифровых технологий коренными малочисленными народами / К. В. Резникова // Цифровизация. – 2021. – Т. 2, № 4. – С. 14-23. – DOI 10.37993/2712-8733-2021-2-4-14-23. – EDN BRTGEF.

Сергеева, Н. А. Дайджест новостей в сфере развития искусственного интеллекта / Н. А. Сергеева, А. А. Омелик, Ю. С. Замараева // Социология искусственного интеллекта. – 2023. – Т. 4, № 1. – С. 26-36. – DOI 10.31804/2712-939X-2023-4-1-26-36. – EDN HWZDKB.

Сертакова, Е. А. Компьютерное искусство 1960-1980-х годов / Е. А. Сертакова, А. А. Ситникова, М. А. Колесник // Социология искусственного интеллекта. – 2022. – Т. 3, № 3. – С. 69-90. – DOI 10.31804/2712-939X-2022-3-3-69-90. – EDN JUGALR.

Ситаров В. А., Маралов В. Г. Южноафриканская философия жизни Ubuntu и ее значение для понимания ненасилия как общечеловеческой ценности // Alma mater (Вестник высшей школы). – 2021. – №. 1. – С. 101-106.

Ситникова, А. А. Художественный образ искусственного интеллекта в анимации XXI века / А. А. Ситникова, Е. А. Сертакова // Социология искусственного интеллекта. – 2022. – Т. 3, № 2. – С. 57-70. – DOI 10.31804/2712-939X-2022-3-2-57-70. – EDN FNLDPE.

Тимкук Д. А. Диалогический аспект Африканского коммунитаризма в философии убунту // Социодинамика. – 2017. – №. 9. – С. 113-122.

Шпак, А. А. Проблемы применения технологий искусственного интеллекта в социологическом аспекте / А. А. Шпак // Социология искусственного интеллекта. – 2021. – Т. 2, № 4. – С. 43-46. – DOI 10.31804/2712-939X-2021-2-4-43-46. – EDN QVBASW.

Шпак, А. А. Роботы наступают: развитие технологий и будущее без работы. Рецензия на книгу автора Мартина Форда / А. А. Шпак // Социология искусственного интеллекта. – 2023. – Т. 4, № 1. – С. 64-70. – DOI 10.31804/2712-939X-2023-4-1-64-70. – EDN DRQHWW.