УДК 004.8+172.4 DOI: 10.17506/18179568_2022_19_3_157

ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РОССИИ И ЕВРОПЕЙСКОМ СОЮЗЕ: ОБЩЕЕ В РИСК-ОРИЕНТИРОВАННЫХ ПОДХОДАХ*

Екатерина Дмитриевна Сорокова,

Московский государственный институт международных отношений (университет) МИД России, Москва, Россия,
sorokova.e.d@my.mgimo.ru

Статья поступила в редакцию 04.04.2022, принята к публикации 29.08.2022

Для цитирования: Сорокова Е.Д. Этика искусственного интеллекта в России и Европейском союзе: общее в риск-ориентированных подходах // Дискурс-Пи. 2022. Т. 19. № 3. С. 157–169. https://doi.org/10.17506/18179568_2022_19_3_157

Аннотация

Повышенное внимание к этическим аспектам создания и применения искусственного интеллекта во многом продиктовано интенсивностью социально-экономической трансформации под влиянием цифровых технологий. Приоритеты и направления развития технологий искусственного интеллекта зафиксированы в ряде стратегических документов Российской Федерации и Европейского союза. В статье выделяются некоторые общие черты риск-ориентированных подходов России и ЕС к разработке, внедрению и использованию искусственного интеллекта на примере двух документов, выпущенных в 2021 г.: Кодекса этики в сфере искусственного интеллекта и проекта Регламента о гармонизации правил, применяемых

* Доклад был сделан автором на Всероссийской научной конференции с международным участием «Философские контексты современности: искусственный интеллект и интеллектуальная интуиция» (ФИКОС-2022), 25–26 февраля 2022 г., Ижевск, Россия.

к системам искусственного интеллекта. Автор анализирует эти документы по ряду критериев: человеко-центрированность и гуманистичность; степени риска; система взаимодействия акторов и место саморегулирования в ней; опора на экспертные знания; тенденция к эволюции «мягких» форм регулирования в более обязывающие. По результатам исследования выявляется ряд общих черт российского и европейского подходов к этике искусственного интеллекта, отраженных в данных документах (при этом формальные и содержательные различия не являются объектом анализа в данной статье). Делается вывод, что этическое регулирование выступает способом реагирования государственных и негосударственных акторов на глобальные технологические риски, связанные с искусственным интеллектом. Оно гармонично дополняет развитие нормативно-правового регулирования и позволяет снизить неопределенность для всех акторов искусственного интеллекта, определяя на высоком уровне обобщения рамки допустимых антропологических и иных гуманитарных последствий их действий.

Ключевые слова:

этика искусственного интеллекта, Европейский союз, Россия, международное регулирование искусственного интеллекта, глобальные технологические риски.

UDC 004.8+172.4 DOI: 10.17506/18179568_2022_19_3_157

AI ETHICS IN RUSSIA AND THE EUROPEAN UNION: COMMONALITIES IN RISK-ORIENTED APPROACHES

Ekaterina D. Sorokova,

Moscow State Institute of International Relations (MGIMO University),
Moscow, Russia,
sorokova.e.d@my.mgimo.ru

Article received on April 4, 2022, accepted on August 29, 2022

For citation: Sorokova, E.D. (2022). AI Ethics in Russia and the European Union: Commonalities in Risk-Oriented Approaches. *Discourse-P*, *19* (3), 157–169. (In Russ.). https://doi.org/10.17506/18179568_2022_19_3_157

Abstract

Socio-economic transformation propelled by digital technologies has drawn increased attention to ethical aspects of the creation and application of artificial intelligence.

Priorities and directions of AI technologies development are fixed in a range of strategic documents both in the Russian Federation and in the European Union. The article highlights several common features of risk-oriented approaches of Russia and the EU to the development, implementation, and use of AI on the example of two documents published in 2021 – the Code of Ethics in the Field of Artificial Intelligence and the Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence. The author analyzes the documents following a set of criteria: human-centricity and humanism; degrees of risk; the system of interaction of AI actors, and the role of self-regulation in it; reliance on expert knowledge; the tendency for the evolution of "soft law" into more binding norms. The results of the study reveal a few commonalities between the Russian and European risk-oriented approaches to AI ethics (formal and substantive differences are not the object of analysis in this article). It is concluded that ethical regulation is a way state and non-state actors respond to global technological risks associated with artificial intelligence. It harmoniously complements the development of legal regulation and reduces uncertainty for all AI actors, as it defines the scope of permissible anthropological and other humanitarian consequences of their actions at a high level of generalization.

Keywords:

AI ethics, the European Union, Russia, international regulation of artificial intelligence, global technological risks.

Введение

2021 г. стал знаменательным для развития этики искусственного интеллекта (ИИ). В апреле Европейская комиссия представила проект Регламента о гармонизации правил, применяемых к системам ИИ (далее – Регламент), в основу которого был положен риск-ориентированный подход; в октябре в России крупнейшие технологические компании, университеты, исследовательские центры и фонды подписали Кодекс этики в сфере ИИ (далее – Кодекс); в ноябре был принят важный международный документ – Рекомендации ЮНЕСКО в сфере этики ИИ.

Актуальность этических исследований в области развития и использования ИИ во многом продиктована динамичной цифровой трансформацией всей социально-экономической сферы жизни общества. С середины 2010-х гг. развитие и внедрение технологий ИИ стали неотъемлемой частью государственных стратегий цифровизации. И Россия, и Евросоюз ставят перед собой достаточно амбициозные цели. В частности, в ЕС к 2030 г. планируется повысить долю компаний, использующих облачные вычисления, большие данные и ИИ, до 75% (актуальный уровень составляет чуть выше 20%)[1]. Россия к этому же сроку намерена довести до «цифровой зрелости» ключевые отрасли экономики

[1] *2030 digital compass: The European way for the digital decade* (2021). Retrieved March 8, 2022, from https://ec.europa.eu/info/strategy/priorities-2019–2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en

и социальной сферы и уже к 2024 г. обеспечить использование продуктов и услуг, основанных преимущественно на отечественном ИИ[2].

И в России, и в Евросоюзе существует достаточно разветвленная система стратегических и нормативных документов, регулирующих развитие ИИ-технологий.

В нашей стране основные цели, задачи, направления и целевые показатели развития ИИ зафиксированы в ряде документов стратегического планирования, основные из которых – национальная программа «Цифровая экономика» (2019 г.), Национальная стратегия развития искусственного интеллекта на период до 2030 г. (2019 г.), Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 г. (2020 г.), а также федеральные проекты «Нормативное регулирование цифровой среды» (2019 г.) и «Искусственный интеллект» (2020 г.). Примечательно, что адаптация нормативного регулирования, исследование этических проблем и разработка специальных принципов в сфере взаимодействия человека с системами ИИ обозначены в качестве отдельного направления как в Национальной стратегии (ст. 48)[3], так и в федеральном проекте (пп. 1.1 и 2.1 паспорта)[4]. В 2020 г. с предложением о разработке свода правил, содержащих этические принципы и ценностные ориентиры работы в сфере ИИ, на конференции *AI Journey* выступил Президент Российской Федерации В.В. Путин[5]. Непосредственно разработкой Кодекса занимались представители российского бизнеса, экспертного и профессионального сообщества.

Развитие цифровизации является одним из приоритетов Комиссии У. фон дер Ляйен на 2019–2024 гг.[6]. Видение, цели и приоритетные направления цифровой трансформации стран ЕС до 2030 г. сформулированы в сообщении Еврокомиссии «Цифровой компас – 2030: европейский путь в цифровое десятилетие» (2021 г.). Документ определил четыре направления, которые странам ЕС рекомендовано учитывать при подготовке национальных планов цифровизации: цифровые навыки, цифровая трансформация бизнеса, безопасная и устойчивая цифровая инфраструктура, цифровизация государственных услуг. Особое внимание уделено сфере облачных технологий, этичному использованию ИИ, обеспечению безопасной цифровой идентификации пользователей, укреплению инфраструктуры связи и высоких технологий[7].

---

[2] *Национальные проекты России. Цифровая экономика* (2020). Взято 8 марта 2022, с https://digital.gov.ru/uploaded/presentations/prezentatsiya-tse_mjl6o1Q.pdf

[3] *Указ Президента Российской Федерации № 490 «О развитии искусственного интеллекта в Российской Федерации»* (2019, 10 октября). Взято 9 июля 2022, с http://www.kremlin.ru/acts/bank/44731

[4] *Паспорт Федерального проекта «Искусственный интеллект» Национальной программы «Цифровая экономика Российской Федерации»* (2020). Взято 9 июля 2022, с https://sudact.ru/law/pasport-federalnogo-proekta-iskusstvennyi-intellekt-natsionalnoi-programmy/

[5] *Конференция по искусственному интеллекту* (2020, 4 декабря). Взято 4 мая 2022, с http://www.kremlin.ru/events/president/news/64545

[6] *6 Commission priorities for 2019–24* (2019, July 16). Retrieved March 8, 2022, from https://ec.europa.eu/info/strategy/priorities-2019-2024_en

[7] *Europe's digital decade* (n. d.). Retrieved March 8, 2022, from https://digital-strategy.

Евросоюз ставит перед собой задачу достигнуть лидерства в сфере этичного ИИ. Стратегия развития ИИ изложена в сообщении Еврокомиссии «Об искусственном интеллекте для Европы» (2018 г.). Позже специально созданная Экспертная группа высокого уровня совместно с Европейским альянсом по ИИ, Группой высокого уровня в рамках инициативы по цифровизации европейской промышленности и Европейской группой по этике науки и новых технологий разработала Руководящие принципы по этике для надежного ИИ (2019 г.). С привлечением экспертов и общественности были составлены методика оценки надежности ИИ (ALTAI) и «Белая книга» (2020 г.), где впервые представлен риск-ориентированный подход. В 2021 г. Еврокомиссия предложила обновленный концептуальный документ «Поддержка европейского подхода к искусственному интеллекту»[8] и опубликовала проект Регламента.

Степень исследованности проблемы

При рассмотрении данной темы можно выделить несколько пластов научной литературы. Первый связан с *исследованиями рисков в целом и технологических рисков в частности*. Они проводятся с 70-х гг. XX в. преимущественно в рамках социологического подхода. Основными методологическими направлениями социологической теории риска называют модернистское, поведенческое, перцептивистское и социально-управленческое (Зубков, 1999), причем технологические риски подробно рассматривают представители первых двух направлений.

«Модернисты» (Бек, 2000; Гидденс, 2011) связывали риски с негативными последствиями модернизации. Основоположник поведенческого подхода Н. Луман поставил под сомнение существующие методы риск-менеджмента и прогнозирования, основанные на рациональном просчитывании, выявлении и предупреждении рисков. По его мнению, ряд факторов, особенно при рассмотрении технологических и экологических рисков, все равно остается неучтенным, когда построенная модель переносится на мир, «который в целом не вписывается в рациональный расчет и реагирует неожиданным образом» (Луман, 2007).

Попытку систематизации рискологических исследований, классификации рисков по степени неопределенности и серьезности последствий, а также разработки «дерева решений» для руководителей и управленцев предприняли немецкие специалисты А. Клинке и О. Ренн. Они выделили пять основных тем дискуссий экспертов-рискологов: реалистское и конструктивистское понимание рисков; учет общественного восприятия рисков в процессе управления ими; решение проблемы неопределенности; сочетание риск-ориентированного и превентивного подходов к управлению рисками; оптимальное совмещение аналитических процессов и общественных/экспертных обсуждений (Klinke, Renn, 2002). Специалист по теории принятия решений М. Меркхофер сделала акцент на исследовании процесса генерации рисков для выявления оптимальной стратегии управления ими (Гришаев, 2002).

---

ec.europa.eu/en/policies/europes-digital-decade

[8] *A European approach to artificial intelligence* (n. d.). Retrieved March 8, 2022, from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Второй пласт исследований посвящен глобальным вопросам, которые связаны с *экзистенциальными рисками для человечества*, создаваемыми новыми технологиями и непосредственно развитием и применением ИИ. Например, онтологический подход философа М. Хайдеггера предполагает связь основного риска современности – «ухода Онтологического Человека» – с феноменом техники. При этом техника считается явлением амбивалентным – объединяющим «опасность полной утраты человеком связи с истиной и бытием» и «ростки спасительного», которые требуют концептуального выявления (Хоружий, 2016). Российский философ С.С. Хоружий, основатель синергийно-антропологического подхода, занимался вопросами глубинных трансформаций человека при бесконтрольном погружении его в технологии и фактическом слиянии с ними (Хоружий, 2016). В частности, он изучал риски виртуализации общения, перехода к Виртуальному человеку и последующего качественного изменения социума и общественного сознания, а также фундаментальные моральные, этические и философские вопросы, связанные с переходом к Постчеловеку (Хоружий, 2008). Исследованиями экзистенциальных рисков (Cotton-Barratt et al., 2020), а также глобальных угроз и этических аспектов, связанных с использованием ИИ (Bostrom, 2019), занимается оксфордский Институт будущего человечества под руководством Н. Бострома.

Интенсивное развитие новых технологий и соответствующие общественные изменения закономерно приводят к появлению новых озабоченностей в сфере безопасности, связанных с недружественным применением ИИ и его непредсказуемыми последствиями. Отдельные аспекты военного применения ИИ (Сорокова, 2021) рассматриваются в контексте международной информационной безопасности, обеспечения кибербезопасности предприятий и объектов инфраструктуры.

Однако в последнее время как в научной литературе, так и в политике можно наблюдать «поворот» от изучения «традиционных» военно-технологических и инфраструктурных рисков *к анализу социально-гуманитарных последствий широкого внедрения и использования новейших технологий*. Оценка гуманитарного воздействия ИИ призвана сделать последствия интенсивной трансформации более управляемыми.

В последние годы количество и разнообразие междисциплинарных исследований по данной тематике увеличивается. Одно из таких направлений – *этика ИИ*. В рамках статьи это понятие трактуется в прикладном ключе: совокупность моральных и ценностных ориентиров и норм поведения, применяемых общественной или профессиональной группой (Ибрагимов и др., 2021). Они отражаются в руководящих принципах, рекомендациях и правилах поведения, которые добровольно принимают на себя акторы ИИ. Среди российских специалистов, занимающихся этим вопросом, стоит выделить сотрудников исследовательского центра в области ИИ при МГИМО А.В. Абрамову, А.Г. Игнатьева и А.А. Кулешова (Kuleshov et al., 2020), а также работы М.В. Федорова, А.В. Незнамова.

На наш взгляд, несмотря на определенный скептицизм общественности и отдельных экспертов в отношении эффективности «мягких» норм, этика ИИ играет самостоятельную роль. Поскольку технологические развитие, как и регулирование в сфере ИИ, – сложная нелинейно развивающаяся система, возникающие в этих сферах проблемы невозможно долгосрочно решить на уровне

конкретного кейса. При отсутствии фундаментальной основы, лежащей в гуманитарной плоскости, даже математически обоснованные методики оценки надежности систем ИИ могут оказаться неэффективными в мире сложных систем, поскольку в реальной жизни система ИИ действует не изолированно. При столкновении с проблемой велик соблазн решить ее, не принимая во внимание контекстов и окружающих ее взаимодействий (например, свести к ошибке кода или недостаточности базы данных), однако при этом упускаются из виду системные проблемы: недостаточность требований, недостатки управленческих и производственных процессов, отсутствие определяющих их этических рамок и т. д. (Lauer, 2021). Определение безопасности, вопросы ответственности за инциденты, соблюдение справедливости, прав граждан и принципа недискриминации при частичной автоматизации судебных решений – вопросы, требующие осмысления не только в рамках инженерии. Чем ближе технологии становятся непосредственно к человеку – к сфере образования, здоровья, развития, тем острее встают вопросы об их долгосрочном влиянии на жизнь каждого индивида, групп людей и общества в целом, о рисках для когнитивного развития, автономии и свободы выбора.

В условиях быстрого качественного развития ИИ и амбициозных планов цифровизации риск-ориентированный подход к регулированию нарождающихся отношений стал доминирующим и закреплен во многих стратегических документах.

Цель статьи – выявить и проанализировать общие черты риск-ориентированных подходов России и Европейского союза к регулированию ИИ на основании двух документов – Кодекса этики в сфере ИИ и проекта Регламента о гармонизации правил в сфере ИИ. Они же послужили основными источниками исследования.

Переходя к *методологии*, необходимо сразу оговориться, что сфера регулирования и тип рассматриваемых документов неодинаковы: Регламент – законопроект, в котором выстраивается многоуровневая система отношений и полномочий акторов ИИ, регистрации и мониторинга технологий до и после выхода на рынок ЕС; Кодекс – свод норм «мягкого права» и этичного поведения акторов ИИ.

С целью анализа текстов Кодекса[9] и Регламента[10] были выделены три основных критерия, характеризующих заложенный в документах риск-ориентированный подход: человеко-центрированность и гуманистичность; степени риска систем ИИ; система взаимодействия акторов ИИ и место саморегулирования в ней. Данные критерии были дополнены двумя другими, характеризующими процесс разработки этических норм в сфере ИИ (опора на экспертные знания) и возможные пути их дальнейшего развития (тенденция к эволюции «мягких» форм регулирования в более обязывающие).

---

[9] *Кодекс этики в сфере искусственного интеллекта* (2021, 26 октября). Взято 8 марта 2022, с https://www.aiethic.ru/code

[10] *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending certain Union Legislative Acts* (Artificial Intelligence Act) (2021, April 21). Retrieved March 8, 2022, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Проведенный анализ выявил ряд общих черт в российском и европейском подходах к этике ИИ. Поскольку формальные и содержательные различия данных документов достойны отдельного исследования, в данной статье основное внимание будет уделено именно сходствам в подходах России и ЕС к этике ИИ.

Результаты исследования

Человеко-центрированность и *гуманистичность* – первая характеристика, сближающая два подхода. В обоих документах делается акцент на защиту интересов, прав и свобод человека, непричинение вреда со стороны систем ИИ.

В Руководящих принципах на 2019–2024 гг. Еврокомиссия взяла на себя обязательство составить законопроект, который сформулировал бы европейский подход к человеко-центрированному ИИ и этическим последствиям его использования. В пояснительной записке к законопроекту отмечено, что он закладывает правовые основы развития экосистемы доверенного и надежного ИИ. Конечной целью развития ИИ провозглашается повышение благосостояния людей. Указано также, что создаваемые для ИИ правила должны быть ориентированы на человека, чтобы люди могли быть уверены, что технология используется безопасным образом, соответствует закону, учитывает основные права человека. Так, например, использование ИИ в целях манипуляций, эксплуатации и социального контроля названо в Регламенте вредоносным. Кроме того, в 2022 г. Комитет регионов предложил ряд поправок к законопроекту, в частности, уточнил требования к принципу недискриминации и добавил в число целей Регламента защиту основных прав граждан и немедленное уведомление их о том, что они подверглись воздействию высокорисковой системы ИИ[11].

Многие принципы российского Кодекса созвучны (хотя и не идентичны) принципам, обозначенным в Регламенте: он также провозглашает в качестве приоритета человеко-ориентированность технологий ИИ; уважение автономии и свободы воли человека; недискриминацию на всех этапах создания, обучения и использования ИИ, а также при формировании наборов данных; соответствие законодательству.

В Кодексе утверждается принцип ответственного отношения всех акторов ИИ к своей деятельности, приветствуется проведение оценки рисков и гуманитарного воздействия технологий ИИ на другие сферы жизни общества. Кроме того, акторам ИИ рекомендуется обеспечивать информационную безопасность и защиту персональных данных на всех этапах жизненного цикла ИИ, а также добровольно информировать пользователей о взаимодействии с такой системой, особенно в случае, если это затрагивает критически важные сферы жизни человека (здоровье, безопасность и т. д.). Обозначена ответственность разработчика за такое информирование. Кодекс также рекомендует обеспечить пользователю возможность отказаться от общения с машиной, отменить социально и юридически значимые решения и действия ИИ.

---

[11] Opinion of the European Committee of the Regions – European approach to artificial intelligence – Artificial Intelligence Act (revised opinion) (2022). *Official Journal of the European Union*. Retrieved March 8, 2022, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AJOC_2022_097_R_0012

Степени риска систем ИИ. В Регламенте предлагается запретить системы с неприемлемым риском: нацеленные на искажение поведения человека, формирование социального рейтинга, удаленную биометрическую идентификацию в режиме реального времени. Высокий риск связан с системами, затрагивающими здоровье, безопасность и основные права человека.

Кодекс не имеет такой классификации, но при этом в тексте можно найти положения, в которых содержатся ответы на ряд принципиальных вопросов, касающихся субъектности ИИ и очевидно относящихся к высокой степени риска. Так, Кодекс предполагает отсутствие правосубъектности ИИ, его поднадзорность и запрет на делегирование ИИ нравственного выбора. Отдельно говорится о необходимости контроля рекурсивного самосовершенствования систем ИИ и, в частности, контроле применения «сильного» ИИ со стороны государства.

Значительное внимание сегодня уделяется рискам, связанным с обработкой больших и персональных данных, их возможной утечкой, нецелевым использованием и т. д. Несмотря на огромный экономический потенциал технологий ИИ и амбициозные цели по их разработке и внедрению, *обеспечение безопасности данных пользователей* остается одной из основных задач государственных институтов как в России, так и в Евросоюзе. При этом защите конфиденциальности отдается приоритет перед возможными экономическими выгодами от более масштабного, но менее безопасного использования данных. Это также во многом объединяет подходы России и ЕС.

Система взаимодействия акторов ИИ и место саморегулирования в ней. Как уже было упомянуто, эффективность имплементации добровольных этических принципов зачастую вызывает скептицизм из-за отсутствия контроля и санкций за их нарушение со стороны самого мощного актора – государства. Отсутствие механизмов принуждения делает их трудными для практического применения и менее заслуживающими доверия в глазах общества. Они расцениваются как простая декларация, попытка отвлечь общественность от реальных проблем или как уловка компаний, стремящихся подменить ими законы или использовать этику в качестве маркетингового хода (Delacroix & Wagner, 2021). Во многих случаях даже четко сформулированным принципам не хватает механизмов реализации или четкого распределения ответственности, что оставляет их открытыми для различных интерпретаций и манипуляций. Поэтому стоит отдельно коснуться проблемы имплементации этики ИИ.

В обоих документах предпринимается попытка выстроить разноуровневую и разветвленную систему взаимоотношений государственных и негосударственных акторов ИИ.

Регламент выстраивает многоуровневую систему взаимодействий при направляющей и руководящей роли Еврокомиссии. Новая структура – Европейский совет по ИИ – призвана обеспечить сотрудничество национальных надзорных органов и ЕК, оказывать ей помощь и давать рекомендации. Государства смогут дополнять обозначенные в Регламенте правила и применять их с учетом национального законодательства. Контроль за соблюдением норм будут осуществлять национальные органы. При этом принципы и стандарты для систем ИИ с невысоким риском предлагается оставить на уровне саморегулирования – фиксировать в «кодексах поведения» компаний, разработка которых будет приветствоваться и поощряться.

Действительно, на уровне отдельной корпорации или отрасли нормы саморегулирования могут быть представлены в виде «кодексов поведения» или отраслевых стандартов. Имплементация же может обеспечиваться, например, через ведение реестров и «белых списков» профессиональными объединениями, ассоциациями и организациями; механизмы исключения из профессиональных объединений за несоответствие этическим нормам (Ибрагимов и др., 2021; Kuleshov et al., 2021).

В Кодексе также прописан механизм имплементации. Предполагается ведение Реестра уполномоченных представителей или комиссий по этике в структуре организаций-подписантов и создание «общей» администрирующей Комиссии. Документ ориентирован на поддержание общественного диалога, обмен опытом и развитие прикладного применения норм: составление «сводов практик», а в перспективе – «белой книги» акторов ИИ, к которым они могли бы обращаться при столкновении с конкретными этическими дилеммами.

Стоит отметить, что в Кодексе упоминаются две ситуации, в которых должны учитываться национальные приоритеты России. Первая связана с принятием значимых для общества и государства решений в сфере применения ИИ, которые могут вызвать изменения в ценностно-культурной парадигме развития общества. То есть в соответствии с российским риск-ориентированным подходом такие решения требуют предварительного междисциплинарного и научно выверенного исследования социально-экономических последствий и рисков. Вторая ситуация связана со стимулированием разработки, внедрения и развития безопасных и этичных решений в сфере технологий ИИ.

Опора на экспертные знания. Принимая во внимание сложный характер технологии ИИ и ее возможное социальное воздействие, этическое регулирование и разработка политики в этой сфере требуют многостороннего подхода. Опираясь на опыт из других сфер, в частности медицины, как российские (Ибрагимов и др., 2021), так и зарубежные (Delacroix & Wagner, 2021) исследователи указывают, что отношения между законодательным и этическим регулированием ИИ следует перестать рассматривать как антагонистические. Эксперты выступают за выстраивание системы взаимной поддержки государства и профессионального сообщества, интеграцию этики в высокотехнологичные отрасли не только «сверху вниз» (путем принятия законов и стандартов), но и «снизу вверх» за счет активной роли и социальной ответственности бизнеса, а государство видят скорее в роли фасилитатора, нежели «надсмотрщика».

Интересно, что процесс подготовки и обсуждения рассматриваемых документов в обоих случаях продемонстрировал определенные успехи на этом направлении. Так, упомянутые этические документы ЕС (Руководящие принципы по этике для надежного ИИ, методика оценки надежности ИИ и «Белая книга») разрабатывались при активном участии экспертов и корректировались с учетом общественных обсуждений. Только после этого был создан проект более обязывающего документа – Регламента. Российский Кодекс – тоже результат взаимодействия экспертов, бизнеса и государства: Альянса в сфере ИИ, Аналитического центра при Правительстве РФ, Минэкономразвития, ведущих вузов. В обсуждении проекта Кодекса участвовало более 500 экспертов. Не исключено, что некоторые этические положения также будут интегрированы в законодательство. Таким образом, в обоих случаях просматривается тенденция

к постепенной эволюции «мягких» этических норм в более обязывающие формы регулирования.

Заключение

Исходя из результатов исследования, можно заключить, что российский и европейский подходы к развитию этического регулирования ИИ имеют общие черты: заложенные в них гуманистические принципы направлены на сохранение автономии человека, защиту прав граждан и конфиденциальности их данных, развитие «доверенного» и безопасного ИИ. Кроме того, в обоих подходах делается попытка выстроить многостороннюю систему взаимодействия акторов ИИ, причем значительное внимание уделяется механизмам саморегулирования и дальнейшему развитию этических принципов по инициативе разработчиков и иных акторов. Значительное внимание и в России, и в ЕС уделяется научному обеспечению выработки политики и рекомендаций в сфере этики ИИ.

Таким образом, этическое регулирование – один из способов реагирования государственных и негосударственных акторов на глобальные технологические риски ИИ. Развитие этики ИИ гармонично дополняет развитие законодательного регулирования на различных уровнях взаимодействия. Это позволяет в целом снизить уровень неопределенности, связанный с цифровой трансформацией, по крайней мере в определении границ ее допустимых социальных и антропологических последствий. Сегодня существует уникальная возможность сформулировать их своевременно, а не постфактум, в порядке реакции на нежелательные последствия их применения.

Примечательна тенденция к эволюции «мягких» форм регулирования в более обязывающие, которая уже проявилась в ЕС. Кодекс тоже предполагает возможность постепенной интеграции отдельных положений в законодательство РФ. При этом вопрос ответственности и контроля выполнения этических принципов является ключевым для того, чтобы предотвратить их превращение в декларации или часть PR-кампаний. Кодификация норм на отраслевом и международном уровнях, их реализация и обновление на уровне профессиональных объединений и сообществ, предприятий и НКО, закрепление основополагающих этических принципов ИИ в законодательных нормах и стандартах для регулирования наиболее высокорисковых сфер развития систем ИИ могут лечь в основу эффективной системы имплементации этических правил в сфере ИИ.

В качестве направлений дальнейших исследований можно обозначить более подробный анализ подходов России и ЕС к разработке этических норм в сфере этики ИИ, рассмотрение формальных и содержательных различий российского и европейского риск-ориентированного подхода к развитию ИИ, выделение более разветвленной системы критериев для их сравнения.

Список литературы

1. Бек, У. (2000). *Общество риска. На пути к другому модерну*. М.: Прогресс-Традиция.

2. Гидденс, Э. (2011). *Последствия современности*. М.: Праксис.

3. Гришаев, В.В. (2002). *Риск и общество (дискуссия о понятии риска и библиография)*. М.: Социологический форум.
4. Ибрагимов, Р.С., Сурагина, Е.Д., Чурилова, Д.Ю. (2021). Этика и регулирование искусственного интеллекта. *Закон*, (8), 85–95.
5. Луман, Н. (2007). *Введение в системную теорию*. М.: Логос.
6. Сорокова, Е.Д. (2021). Глобальные риски военно-политического использования искусственного интеллекта в обществе модерна. В *Ломоносов-2021: материалы Международного молодежного научного форума (12–23 апреля 2021 г., Москва)*. М.: МАКС Пресс.
7. Хоружий, С.С. (2008). Проблема постчеловека, или трансформативная антропология глазами синергийной антропологии. *Философские науки*, (2), 10–31.
8. Хоружий, С.С. (2016). Проблематика рисков современности: концептуальные основания и ведущие подходы. В С.С. Хоружий, *Социум и синергия: колонизация интерфейса* (с. 135–165). Казань: Казанский инновационный университет имени В.Г. Тимирясова.
9. Bostrom, N. (2019). The vulnerable world hypothesis. *Global Policy*, *10*(4), 455–476. https://doi.org/10.1111/1758-5899.12718
10. Cotton-Barratt, O., Daniel, M., & Sandberg, A. (2020). Defence in depth against human extinction: Prevention, response, resilience, and why they all matter. *Global Policy*, *11*(3), 271–282. https://doi.org/10.1111/1758-5899.12786
11. Delacroix, S., & Wagner, B. (2021). Constructing a mutually supportive interface between ethics and regulation. *Computer Law & Security Review*, *40*. https://doi.org/10.1016/J.CLSR.2020.105520
12. Klinke, A., & Renn, O. (2002). A new approach to risk evaluation and management: Risk-based, precaution-based, and discourse-based strategies. *Risk Analysis*, *22*(6), 1071–1094. https://doi.org/10.1111/1539-6924.00274
13. Kuleshov, A., Ignatiev, A., Abramova, A., & Marshalko, G. (2020). Addressing AI ethics through codification. In *Proceedings of the 2020 International Conference Engineering Technologies and Computer Science* (EnT). https://doi.org/10.1109/EnT48576.2020.00011
14. Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics*, (1), 21–25. https://doi.org/10.1007/s43681-020-00013-4

References

1. Beck, U. (2000). *Obshhestvo riska. Na puti k drugomu modernu* [The risk society: Towards a new modernity]. Moscow: Progress-Tradiciya.
2. Bostrom, N. (2019). The vulnerable world hypothesis. *Global Policy*, *10*(4), 455–476. https://doi.org/10.1111/1758-5899.12718
3. Cotton-Barratt, O., Daniel, M., & Sandberg, A. (2020). Defence in depth against human extinction: Prevention, response, resilience, and why they all matter. *Global Policy*, *11*(3), 271–282. https://doi.org/10.1111/1758-5899.12786
4. Delacroix, S., & Wagner, B. (2021). Constructing a mutually supportive interface between ethics and regulation. *Computer Law & Security Review*, *40*. https://doi.org/10.1016/J.CLSR.2020.105520

5. Giddens, A. (2011). *Posledstviya sovremennosti* [The consequences of modernity]. Moscow: Praksis.

6. Grishaev, V.V. (2002). *Risk i obshhestvo (diskussiya o ponyatii riska i bibliografiya)* [Risk and society (discussion on the concept of risk and bibliography)]. Moscow: Sociologicheskij forum.

7. Ibragimov, R.S., Suragina, E.D., & Churilova, D. Yu. (2021). E'tika i regulirovanie iskusstvennogo intellekta [Ethics and AI regulation]. *Zakon*, (8), 85–95.

8. Khoruzhy, S.S. (2008). Problema postcheloveka, ili transformativnaya antropologiya glazami sinergijnoj antropologii [The problem of post-human, or Transformative anthropology in the light of synergetic anthropology]. *Filosofskie nauki*, (2), 10–31.

9. Khoruzhy, S.S. (2016). Problematika riskov sovremennosti: konceptual'nye osnovaniya i vedushhie podxody [The issues of modern risks: Conceptual basis and leading approaches]. V S.S. Khoruzhy, *Socium i sinergiya: kolonizaciya interfejsa* (pp. 135–165). Kazan: Kazanskij innovacionnyj universitet imeni V.G. Timiryasova.

10. Klinke, A., & Renn, O. (2002). A new approach to risk evaluation and management: Risk-based, precaution-based, and discourse-based strategies. *Risk Analysis, 22*(6), 1071–1094. https://doi.org/10.1111/1539-6924.00274

11. Kuleshov, A., Ignatiev, A., Abramova, A., & Marshalko, G. (2020). Addressing AI ethics through codification. In *Proceedings of the 2020 International Conference Engineering Technologies and Computer Science* (EnT). https://doi.org/10.1109/EnT48576.2020.00011

12. Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics*, (1), 21–25. https://doi.org/10.1007/s43681-020-00013-4

13. Luhmann, N. (2007). *Vvedenie v sistemnuyu teoriyu* [Introduction to systems theory]. Moscow: Logos.

14. Sorokova, E.D. (2021). Global'nye riski voenno-politicheskogo ispol'zovaniya iskusstvennogo intellekta v obshhestve moderna [Global risks of the military use of artificial intelligence in the modern society]. In *Lomonosov-2021: materialy Mezhdunarodnogo molodezhnogo nauchnogo foruma (12–23 aprelya 2021 g., Moskva)*. Moscow: MAKS Press.

Информация об авторе

Екатерина Дмитриевна Сорокова, аспирант, Московский государственный институт международных отношений (университет) МИД России, Москва, Россия, ORCID: https://orcid.org/0000-0002-4542-7767, e-mail: sorokova.e.d@my.mgimo.ru

Information about the author

Ekaterina Dmitrievna Sorokova, Post-graduate student, Moscow State Institute of International Relations (MGIMO University), Moscow, Russia, ORCID: https://orcid.org/0000-0002-4542-7767, e-mail: sorokova.e.d@my.mgimo.ru