12. Славин, Б. Люди или цифры – кто нужнее [Электронный ресурс] / Б. Славин. – Режим доступа: https://www.vedomosti.ru/opinion/articles/2017/01/17/673248-lyudi-tsifri-nuzhnee. – Дата доступа: 19.03.2023.

13. Михайленко, Н. В. Цифровое государственное управление / Н. В. Михайленко // Государственная служба и кадры. – 2020. – № 2. – С. 171–175.

14. Сливицкий, А. Б. Концепция оценки уровня готовности технологий, производств как механизм формирования единого инновационно-технологического пространства / А. Б. Сливицкий // Россия: тенденции и перспективы развития : Ежегодник / РАН, ИНИОН, Отд. науч. сотрудничества ; отв. ред. В. И. Герасимов. – М., 2017. – Вып. 12, ч. 1. – С. 618–624.

15. Сливицкий, А. Б. Система уровней готовности технологий как оптимальная модель организации и финансирования процесса создания научно-технического задела в российской промышленности / А. Б. Сливицкий // Россия: тенденции и перспективы развития : Ежегодник / РАН, ИНИОН, Отд. науч. сотрудничества ; отв. ред. В. И. Герасимов. – М., 2016. – Вып. 11, ч. 3. – С. 461–469.

16. Pause Giant AI Experiments: An Open Letter. Приостановить гигантские эксперименты с искусственным интеллектом: открытое письмо [Электронный ресурс]. – Режим доступа: https://futureoflife.org/open-letter/pause-giant-ai-experiments. – Дата доступа: 19.03.2023.

17. Все умрут, включая детей. Как искусственный интеллект изменит интернет и почему этого боится даже Илон Маск [Электронный ресурс]. – Режим доступа: https://lenta.ru/articles/2023/04/06/evilgpt. – Дата доступа: 08.04.2023.

18. Жилина, И. Ю. Теоретические основы социальной ответственности бизнеса: история, эволюция / И. Ю. Жилина // Экономические и социальные проблемы России. – 2016. – № 1. – С. 12–31.

Дата поступления в редакцию: 10.05.2023.

УДК 340.1

М. С. Абламейко
Научный сотрудник отдела исследований в области государственного строительства и международного права Института правовых исследований Национального центра законодательства и правовых исследований Республики Беларусь, доцент кафедры конституционного права юридического факультета Белорусского государственного университета, кандидат юридических наук, доцент

К ВОПРОСУ О РАЗРАБОТКЕ КОДЕКСА ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. В статье рассматривается содержание понятия этики искусственного интеллекта (ИИ) и вопросы этического применения систем ИИ. Выделяются

риски применения систем ИИ для человека, общества и государства. Проводится анализ подходов к регулированию ИИ на международном уровне. Предлагается разработка национального акта, носящего рекомендательный характер, – Кодекса этики искусственного интеллекта, направленного на формирование единых правовых и морально-этических норм, обеспечивающих безопасность личности, общества и государства и стимулирование развития технологий ИИ.

Abstract. The article discusses the content of the concept of ethics of artificial intelligence and the ethical application of AI systems. The risks of using AI systems for a person, society and the state are highlighted. An analysis of approaches to the regulation of AI at the international level is carried out. It is proposed to develop a national act that is advisory in nature – the Code of Ethics of Artificial Intelligence, aimed at the formation of uniform legal and moral and ethical standards that ensure the security of the individual, society and the state and stimulate the development of AI technologies.

Ключевые слова: искусственный интеллект, кодекс этики, системы искусственного интеллекта, этика искусственного интеллекта.

Keywords: artificial intelligence, code of ethics, artificial intelligence systems, artificial intelligence ethics.

Введение. За последние годы искусственный интеллект (ИИ) повсеместно интегрировался в нашу жизнь: беспилотные транспортные средства, банковские сервисы, медицинские системы и др. Большинство развитых стран мира в рамках государственных программ определяют приоритетность цифровизации в целом и стимулируют создание и развитие систем искусственного интеллекта. Однако функционирование систем без соответствующего правового регулирования может представлять угрозу для общества.

Значимость создания благоприятных правовых условий для разработчиков ИИ не раз подчеркивалась лидерами многих развитых стран, так как за его развитием стоит будущее всего мира. Вместе с тем мировое сообщество призывает к введению правил и норм применения ИИ в связи с возрастающими рисками для государства, общества и человека.

К вопросу развития технологий ИИ и их этического использования обращались такие ученые, как А. В. Попова, А. В. Минбалеев, И. А. Филипова, И. В. Понкин, О. Н. Толочко, А. Л. Савенок, Н. С. Минько и др.

Основная часть. *Этика ИИ*

Развитие систем ИИ привело к необходимости решения морально-этических вопросов использования таких систем в реальной жизни.

Этика ИИ является частью цифровой этики – областью прикладной этики, рассматривающей моральные проблемы современного общества в цифровой среде, в которой изучаются вопросы, связанные с разработкой, внедрением и использованием ИИ. Основная задача этики ИИ – определить, как ИИ может развиваться и какие проблемы, касающиеся благополучия человека (в том числе качества жизни, автономии и свободы, необходимой для существования демократического общества), могут возникнуть в связи с его развитием [1].

Этика ИИ относится к организационным конструкциям (корпоративные ценности, политика, этические кодексы и руководящие принципы), которые разграничивают «правильное» и «неправильное» в отношении применения

технологий ИИ. Эти конструкции устанавливают цели и рекомендации для ИИ на протяжении всего жизненного цикла продукта – от исследований и проектирования до создания и обучения, изменения и эксплуатации [2].

Этика ИИ – это комплекс ценностей, принципов и методов, основанных на общепринятых критериях «добра и зла», который определяет моральное поведение при разработке и использовании технологий ИИ [3].

Исходя из данных подходов, можно определить, что этика ИИ, рассматривая моральные нормы общества в цифровой среде, определяет границы дозволенного, разделяя поведение субъектов информационных отношений (разработчиков, пользователей и др.) на приемлемое и нет, при обязательной оценке рисков воздействия как на общество в целом, так и на каждого индивида в частности.

Этика ИИ соотносится с разными сферами (военной этикой; этикой права; этикой образования; медицинской этикой и др.), может расширяться и дополняться вместе с развитием технологий, в связи с чем носит междисциплинарный характер.

В эпоху внедрения ИИ во все сферы жизнедеятельности человека использование тех или иных технологий сопряжено с определенными опасностями и рисками и может угрожать существованию человечества [4].

Риски применения ИИ

За последние годы все больше внимания уделяется рискам применения систем ИИ. Связано это в первую очередь с тем, что, с одной стороны, внедрение ИИ в разных отраслях позволяет автоматизировать рутинные процессы, а человек становится контролером, а не исполнителем. С другой стороны, сбои работы таких систем могут привести к тяжелым последствиям, в связи с чем обеспечение безопасности функционирования таких систем является одной из первоочередных задач. В связи с этим классифицировать риски использования систем ИИ можно следующим образом:

1. Риск причинения вреда человеку.

Основным риском для человека принято считать вмешательство в частную жизнь. В связи с тем, что за последние десятилетия крупные компании накопили огромное количество данных, вопрос их правомерного использования стал особенно актуальным. Каждый раз, когда человек входит в сеть, он генерирует данные, и, соответственно, остается электронный след. Впоследствии интеллектуальные системы могут использовать эти данные в своих целях.

Вторым важным аспектом является использование систем ИИ при распознавании лиц с камер видеонаблюдения. Это позволяет отследить все перемещения человека, провести полную идентификацию, реиндетификацию, выявить девиантное поведение и др. Некоторые страны пошли по пути применения социального рейтинга, когда на основании собранных данных человеку присваивается определенный статус в обществе (человеку присваивается определенное количество баллов в зависимости от его поведения).

На основе анализа информации, которой интересуется человек в сети, системы ИИ формируют таргетированную рекламу. Социальные сети благодаря автоматическим алгоритмам очень эффективны в целевом маркетинге. В настоящее время уже происходит так называемое «предугадывание» желаний человека, когда системы ИИ на основании имеющихся данных предлагают тот или иной

продукт или услугу. В данном случае происходит навязывание и прогнозирование действий, что также несет угрозу для индивидуальности принятия решения.

Общение с ИИ. В последнее время стало весьма актуальной проблема, когда человек в процессе общения не может понять, общается он с ИИ или с человеком (тест Тьюринга). В случае несоблюдения ИИ морально-этических норм это может привести к серьезным последствиям для человека (склонение к самоубийству, дезориентация и др.). В связи с этим возникает проблема доверительного отношения ИИ и человека.

2. Риски для общества.

Основным риском в данной категории можно выделить манипулирование общественным мнением. Распространяя пропаганду лицам, искусственный интеллект может распространять любую информацию, которая потребуется в любом формате, который будет выглядеть наиболее убедительным, неважно, будет ли это правда или ложь. В качестве примера можно привести возможность прогнозирования результатов выборов или референдумов на основе данных пользователей социальных сетей.

Несовпадение целей ИИ и человека в случае постановки некорректной задачи. Если ИИ нацелен на достижение конкретной запрограммированной цели, а человек будет пытаться ему помешать, то это может привести к конфликту интересов.

3. Риски для государства.

Обеспечение безопасности функционирования цифровой экосистемы государства подразумевает бесперебойную работу с выявлением возможных внешних и внутренних угроз.

В большинстве своем ИИ – это «черный ящик», т.е. принимаемые им решения непрозрачны как для пользователя, так для оператора и разработчика. Отсюда возникает проблема интерпретируемости результатов и предвзятости систем ИИ. В настоящее время отсутствуют технические средства аудита систем ИИ, что позволяет разработчикам вносить в систему недокументированные функции при обучении (например, чужеродные объекты, которые система будет пропускать как разрешенные), обнаружить которые в работающей системе нельзя.

Использование данных для обучения нейронных сетей. Большинство стран придерживается политики защиты персональных данных, включая биометрические, визуальные, генетические, медицинские и др. Вместе с тем для стимулирования развития технологий ИИ требуется доступ к таким данным, что предполагает поиск баланса интересов компаний-разработчиков и государства. Также в случае недостатка данных при обучении ИИ повышается риск ошибок, что может привести к дестабилизации работы системы.

Выход системы из-под контроля человека. Существует вероятность, что автономная система ИИ с автоматическим принятием решений в критически важных, ответственных областях в результате ошибки или недостаточного обучения примет решение, которое нанесет ущерб людям или критической инфраструктуре.

Использование иностранных платформ. Большинство востребованных ИИ систем находятся на иностранных платформах, что дает возможность удаленного управления со стороны разработчика системы.

Рекомендации международного сообщества

В последние годы многие международные организации пошли по пути принятия различного рода документов так называемого «мягкого права», которые носят рекомендательный характер. Целью данных документов является установление основных принципов и норм, регулирующих разработку и применение систем ИИ как государственными, так и частными структурами.

Организация экономического сотрудничества и развития (ОЭСР) в 2019 г. утвердила «Руководящие принципы ИИ ОЭСР» (OECD AI Principles) в рамках документа «Рекомендации Совета ОЭСР по правовым инструментам ИИ» (Recommendation of the Council on OECD Legal Instruments AI) [5]. Целью данного документа является способствование развитию технологий ИИ и их этичному применению. Пять главных рекомендаций включают: ИИ должен служить людям и способствовать социально-экономическому развитию; технологии ИИ должны разрабатываться с учетом обеспечения и защиты прав человека; работа ИИ должна быть понятной и прозрачной для людей; технологии ИИ должны работать надежно и безопасно, а также постоянно оцениваться на предмет потенциального риска применения вреда человеку; разработчики и иные субъекты, причастные к управлению и внедрению систем ИИ, должны нести ответственность за их стабильное функционирование [6].

В 2020 г. Европейская комиссия опубликовала Белую книгу по ИИ – европейский подход к совершенству и доверию (A White paper on Artificial Intelligence – A European approach to excellence and trust) [7]. В Белой книге изложены варианты политики, позволяющие достичь двойной цели – стимулировать внедрение ИИ и устранить риски, связанные с определенными видами использования таких технологий.

В 2021 г. в ходе Генеральной конференции ЮНЕСКО 193 страны приняли первое глобальное соглашение по этике ИИ – Рекомендации по этическим аспектам ИИ [8], которое призвано служить этическим ориентиром и нормативной основой, позволяющей обеспечить строгое соблюдение принципа верховенства права в цифровом мире. Целью данного документа является, с одной стороны, стимулирование использования систем ИИ, а с другой – недопустимость причинения вреда как отдельному человеку, так и всей экосистеме. В Рекомендациях определены основные принципы, которых государствам следует придерживаться при формировании национального законодательства: соразмерность и непричинение вреда; безопасность и защищенность; справедливость и недискриминация; устойчивость; неприкосновенность частной жизни и защита данных; подконтрольность и подчиненность человеку; прозрачность и объяснимость; ответственность и подотчетность; осведомленность и грамотность; многостороннее и адаптивное управление и взаимодействие [9]. В рамках данного документа определены 11 стратегически важных направлений применения систем ИИ, включающих: оценку этического воздействия, ответственное управление и контроль, политику в области данных, развитие международного сотрудничества и др.

Содружество Независимых Государств также работает над проектом «Рекомендации по нормативному регулированию использования ИИ, включая этические стандарты для исследований и разработок», целью которого является разработка унифицированного подхода к правовому и этическому регулированию применения

технологий ИИ для государств-участников. Как и в выше представленных международных актах, с одной стороны, декларируется приоритет и стимулирование развития технологий, с другой – обеспечение безопасности личности, общества и государства, основанного на риск-ориентированном подходе. В дальнейших планах присутствует разработка Модельного кодекса этики в сфере ИИ для государств – участников Содружества Независимых Государств.

В Российской Федерации в 2021 г. разработан Кодекс этики в сфере ИИ. Это единая система рекомендательных принципов и правил, предназначенных для создания среды доверенного развития технологий искусственного интеллекта в России. Документ подписали Сбербанк, Яндекс, МТС, VK, «Газпром нефть» и Российский фонд прямых инвестиций, а также ведущие компании и научно-исследовательские организации. К настоящему времени кодекс подписала уже 181 компания [10].

Кодекс этики ИИ Республики Беларусь

В Республике Беларусь развитие технологий ИИ является приоритетным направлением научной, научно-технической и инновационной деятельности на 2021–2025 годы в рамках Указа Президента Республики Беларусь от 7 мая 2020 г. № 156.

Учитывая подходы к регулированию ИИ на международном уровне, считаем целесообразным разработать Кодекс этики ИИ в Республике Беларусь как документ рекомендательного характера применения систем ИИ в различных гражданских (не военных) целях. Кодекс должен быть направлен на формирование единых правовых и морально-этических норм, обеспечивающих безопасность личности, общества и государства и стимулирование развития технологий ИИ. Основное предназначение систем ИИ – работать на благо человечества. В процессе всего жизненного цикла системы ИИ необходимо применять там, где это принесет пользу людям и по назначению.

Основные принципы, которые следует заложить в Кодекс этики ИИ, исходя из международной практики, следующие:

– человекоориентированный подход, основанный на уважении прав и свобод человека;

– неприкосновенность частной жизни и защита данных, недискриминация;

– обеспечение безопасности, защищенность, непричинение вреда, минимизация рисков;

– устойчивость, прозрачность, подконтрольность и подчиненность человеку;

– ответственность и подотчетность.

Считаем целесообразным придерживаться риск-ориентированного подхода, что позволит провести классификацию систем ИИ в различных сферах: неприемлемый риск (удаленная биометрическая идентификация в режиме реального времени в общественных местах, за исключением случаев, когда такое использование необходимо для достижения строго поставленных целей); высокий риск (критическая инфраструктура, которая может подвергнуть риску жизнь и здоровье граждан); ограниченный риск (ИИ со специфическими требованиями к прозрачности); минимальный риск.

Следует определить идентификаторы воздействия ИИ на человека и общество и давать оценку с выявлением последствий и рисков, проводить мониторинг социального и экономического воздействия систем ИИ, аудит систем ИИ, давать оценку эффективности и действенности политики в области этики ИИ.

Заключение. Необходимость правового регулирования отмечается во многих стратегических документах разных стран мира. Многие развитые страны и международные организации активно работают над поиском баланса: стимулирования развития систем ИИ и минимизации рисков причинения вреда. Этические аспекты применения систем ИИ отнесены к так называемому «мягкому праву». Для решения этических вопросов принимаются Кодексы этики ИИ.

Современное национальное регулирование значительно отстает от скорости развития ИИ, в связи с чем считаем своевременным начать работу над Кодексом этики ИИ в нашей стране. Полагаем, инициатором и организатором разработки такого Кодекса мог бы выступить Парк высоких технологий Республики Беларусь или Объединенный институт проблем информатики Национальной академии наук Беларуси.

Необходимо определить путь создания Кодекса и зафиксировать терминологию в цифровом праве. Считаем целесообразным придерживаться риск-ориентированного подхода, что позволит провести классификацию систем ИИ в различных сферах. Следует разработать единые виды тестирования на безопасность, которые должны проводить компании-разработчики ИИ, и определить виды доступа к данным для проведения аудитов и оценок.

Важным шагом является принятие стандартов оценки и применения систем ИИ в отраслях. Они, естественно, будут различаться в зависимости от отрасли применения. Например, применение ИИ в отрасли здравоохранения требует самого пристального внимания.

Нормативные правовые акты, принимаемые в сфере ИИ, должны выстраивать баланс между инновациями и регулированием, чтобы развитие ИИ не имело негативных последствий от неконтролируемого применения.

Список цитированных источников:

1. Ethics guidelines for trustworthy AI [Electronic resource] // European Commission. – Mode of access: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. – Date of access: 12.05.2023.

2. AI ethics: A business imperative for boards and C-suites [Electronic resource] // Deloitte. – Mode of access: https://www2.deloitte.com/us/en/pages/regulatory/articles/ai-ethics-responsible-ai-governance.html. – Date of access: 12.05.2023.

3. Этика и «цифра»: от проблем к решениям / под ред. Е. Г. Потаповой, М. С. Шклярук. – М. : РАНХиГС, 2021. – 184 с.

4. Ройзензон, Г. В. Стандарты этики в искусственном интеллекте / Г. В. Ройзензон // Программная инженерия: методы и технологии разработки информационно-вычислительных систем (ПИИВС-2018) : сб. науч. тр. II Междунар. науч.-практ. конф., Донецк, 14–15 нояб. 2018 г. : в 2 т. / Донецк. нац. технич. ун-т. – Донецк : Донецк. нац. технич. ун-т, 2018. – Т. 1. – С. 227–236.

5. OECD AI Principles overview [Electronic resource]. – Mode of access: https://oecd.ai/en/ai-principles. – Date of access: 10.05.2023.

6. Системный подход к изучению государственных политик и процессов формирования этики применения технологий искусственного интеллекта: Глобальный атлас регулирования / Э. Г. Чаче, Р. И. Дремлюга, Н. А. Третьякова, А. В. Незнамов //

Докл. Рос. акад. наук. Математика, информатика, процессы управления. – 2022. – Т. 508, № 1. – С. 73–78.

7. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. COM/2021/206 final [Electronic resource] // An official web-site of the European Union. – 2021. – Mode of access: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. – Date of access: 15.03.2023.

8. Этические аспекты искусственного интеллекта [Электронный ресурс]. – Режим доступа: https://www.unesco.org/ru/artificial-intelligence/recommendation-ethics. – Дата доступа: 10.05.2023.

9. Лизикова, М. С. Этические и правовые вопросы развития искусственного интеллекта в предпринимательской и иной экономической деятельности / М. С. Лизикова // Труды Ин-та государства и права РАН. – 2022. – Т. 17, № 1. – С. 177–194.

10. Кодекс этики в сфере искусственного интеллекта [Электронный ресурс]. – Режим доступа: https://ethics.a-ai.ru/assets/ethics_files/2023/05/12/%D0%9A%D0%BE%D0%B4%D0%B5%D0%BA%D1%81_%D1%8D%D1%82%D0%B8%D0%BA%D0%B8_20_10_1.pdf. – Дата доступа: 10.05.2023.

Дата поступления в редакцию: 16.06.2023.

УДК 342.3

И. И. Костян
Научный сотрудник отдела исследований в области государственного строительства и международного права Института правовых исследований Национального центра законодательства и правовых исследований Республики Беларусь

ОСОБЕННОСТИ КУЛЬТУРЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПОЖИЛЫХ ЛЮДЕЙ

Аннотация. В статье анализируются вопросы формирования и развития культуры информационной безопасности особой категории субъектов – пожилых людей. Рассматриваются основные подходы к пониманию культуры информационной безопасности, предлагается авторская дефиниция последней. Обоснована необходимость выделения пожилых людей как отдельной группы субъектов правоотношений в контексте обеспечения их информационной безопасности. Рассматриваются основные меры по формированию и развитию культуры информационной безопасности пожилых людей, предлагаются соответствующие направления совершенствования законодательства Республики Беларусь.

Abstract. The article analyzes issues of the formation and development of the information security culture of a special category of subjects. The author considers