Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Бурятский государственный университет
имени Доржи Банзарова»

На правах рукописи

Бадмаева Маина Харлановна

СОЦИАЛЬНО-ФИЛОСОФСКИЕ ПРОБЛЕМЫ И ПРИНЦИПЫ ПРИМЕНЕНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Специальность: 5.7.7 – Социальная и политическая философия
ДИССЕРТАЦИЯ
на соискание ученой степени
кандидата философских наук

Научный руководитель:
Золхоева М. В.,
д-р филос. наук, доцент

Улан-Удэ - 2023

Оглавление

ВВЕДЕНИЕ

Актуальность темы диссертационного исследования. В последние годы искусственный интеллект (далее – ИИ) прочно входит в нашу жизнь, раскрывая перед человеком многообразные возможности, улучшая качество человеческой жизни, расширяя горизонты самореализации человека. Однако ИИ таит в себе и новые опасности, оказывая значительное влияние на человека, общество, окружающую среду.

ИИ обладает способностью воздействовать на культуру, существенно трансформируя ее базовые составляющие. Можно с полным правом уже сегодня заявить о том, что ИИ – это сложное и многоплановое явление, источник масштабных социокультурных изменений как положительных, так и негативно воздействующих на человека и человеческую цивилизацию.

В этих условиях философия, будучи рациональным типом мировоззрения, обоснованно ставит перед собой задачу выработать аргументированное и взвешенное отношение к внедрению систем искусственного интеллекта, рассмотрев, какие риски сопровождают этот процесс, какое именно воздействие они оказывают на миропонимание современного индивида, на его отношение к окружающему миру.

Проблемы применения ИИ необходимо рассматривать сквозь призму подлинно философского толкования места и роли человека в мире, гуманистических целей и ценностей современного социума. Любые достижения в сфере ИИ имеют смысл только в том случае, если они соотносятся с идеалами процветания человека, человеческой цивилизации. Чрезмерное делегирование полномочий от человека системам ИИ, напротив, способно негативно повлиять на экзистенциальные основы бытия человека, с особой остротой поставив вопрос о смысле его существования и жизненном предназначении.

Социально-философский дискурс не раз обращался к теме искусственного интеллекта, но теоретическое осмысление происходящих внутри его систем изменений, современной специализации ИИ и воздействия результатов новейших разработок в этой сфере на человека и общество все еще существенным образом отстают от прогресса и внедрения этих технологий в жизнь современного социума. Теоретико-методологические и социально-философские основания планирования процесса развития и внедрения систем ИИ сегодня разработаны все еще недостаточно.

Эти недостатки могут быть преодолены разработкой философских принципов взаимодействия с ИИ, нацеленных на эффективное и безопасное использование искусственного разума человеком и обществом. Подобные исследования позволят сформулировать содержание основополагающих требований к разработчикам и пользователям систем ИИ на любых стадиях его развития и совершенствования, минимизировать негативные последствия, определить границы применения ИИ, разумно и гуманно реализовать его потенциал для решения глобальных проблем нашей цивилизации. В настоящее время на уровне государственных структур, научных организаций, бизнеса и гражданского общества многое сделано для выявления указанных принципов, но содержание их, как правило, не раскрыто, расплывчато, многозначно или вообще лишено каких-либо пояснений.

Таким образом, актуальность нашего исследования обусловлена необходимостью более глубокого и адекватного отражения в социально-философской рефлексии современных реалий воздействия искусственного интеллекта на общество и человека, потребностью выявить содержание базовых принципов взаимодействия человека с системами искусственного интеллекта посредством анализа его наиболее важных характеристик; выявления возможных социальных, этических последствий и перспектив взаимодействия человека с ИИ для достижения целей гуманистического, безопасного и эффективного развития современного социума.

Степень научной разработанности проблемы. В настоящее время существует значительное количество отечественных и зарубежных исследований технического, юридического, социологического и философского характера, в которых рассмотрен искусственный интеллект и проблемы его применения. Несмотря на это, все еще отмечается дефицит социально-философских исследований, посвященных рассмотрению разнообразных последствий применения искусственного интеллекта в современном обществе.

Рассмотрим круг работ, очерчивающих проблемное поле исследования, разделив их на несколько тематических блоков.

Исследования, посвященные влиянию техники на человека. В самом широком смысле, искусственный интеллект можно рассматривать как часть философии техники. К этому пулу исследователей относятся классические труды У. Бека[1], В. Г. Горохова[2], М. Кастельса[3], Г. Маркузе[4], М. Маклюэна[5], Л. Мэмфорда[6], Х. Ортеги-и-Гассета[7], В. М. Розина[8], Ф. Рело[9], А. Ридлера[10], М. А. Розова[11], С. А. Смирнова[12], В. С. Степина[13], П. Фло-

[1] Бек У. Общество риска. На пути к другому модерну / У. Бек. – М.: Прогресс-Традиция, 2000. – 383 с.
[2] Горохов В. Г. Место и роль философии техники и современной философии и её органическая связь с философией науки / В. Г. Горохов // Философия науки. – 2011. – № 1. – С. 181-199.
[3] Кастельс М. Информационная эпоха: экономика, общество и культура / М. Кастельс. – М.: ГУ ВШЭ, 2000. — 608 с.
[4] Маркузе Г. Одномерный человек / Г. Маркузе // Исследование идеологии Развитого Индустриального Общества. – М.: Reefl-book, 1994. – 368 с.
[5] Маклюэн М. Понимание медиа: внешнее расширение человека / М. Маклюэн. – М.: Жуковский: «Канон-пресс-Ц», 2003. – 464 с.
[6] Мэмфорд Л. Техника и природа человека / Л. Мэмфорд // Новая технократическая волна на Западе. – М.: Прогресс-Традиция, 1986. – С. 225-239.
[7] Ортега-и-Гассет Х. Размышления о технике / Х. Ортега-и-Гассет // Избранные труды пер. с исп.; сост., предисл. и общ. ред. А. М. Руткевича. – М.: Весь Мир, 1997. – С. 164-232.
[8] Розин В. М. Понятие и современные концепции техники / В. М. Розин. – М.: Институт философии РАН, 2006. - 255 с.
[9] Рело Ф. Техника и ее связь с задачею культуры /Ф. Рело. – СПб.: Типография министерства путей сообщения, 1885. – 27 с.
[10] Ридлер А. Германские высшие учебные заведения и запросы двадцатого столетия / А. Ридлер. – СПб.: типография Р. Голике, 1900. – 30 с.
[11] Степин В. С., Горохов В. Г., Розов М. А. Философия науки и техники / В. С. Степин, В. Г. Горохов, М. А Розов. – М., 1996. – 380 с.
[12] Смирнов С. А. Человек перехода / Отв. за вып. П. А. Носова. – Новосибирск, 2006. – 177 с.
[13] Степин В. С. Научное познание и ценности техногенной цивилизации / В. С. Степин // Вопросы философии. – 1989. – № 10. – С. 3-18.

ренского[14], Ф. Фукуямы[15], Ю. Хабермаса[16], М. Хайдеггера[17], И. В. Черниковой и Д. В. Черниковой[18] и др.

Так, Ж. Бодрийяр[19], Л. Мэмфорд, Д. Нейсбит[20], Ф. Фукуяма резко критиковали технологизацию и внедрение искусственного интеллекта без осмысления социальных и этических последствий. Отечественные исследователи В. А. Кутырев[21], И. В. Черникова и Д. В. Черникова[22] полагают, что человеческая идентичность подвержена опасности из-за необдуманного внедрения высоких технологий. А. Нордманн[23], Э. Тоффлер[24], Ф. Фукуяма, Ю. Хабермас исследуют использование новых технологий, задаваясь, главным образом, вопросом о вызванных им возможных изменениях человеческой природы. Д. В. Иванов[25], М. Кастельс, И. С. Мелюхин[26], А. И. Ракитов[27] анализируют социокультурные последствия использования информационных технологий. В. Г. Горохов рассматривает возможные результаты внедрения новых технологий с позиций этических норм и вводит новое понятие «наноэтика»[28].

---

[14] Флоренский П. Органопроекция / П. Флоренский // Русский космизм: антология философской мысли. – М.: Педагогика-пресс, 1993. - С. 149-162.

[15] Фукуяма Ф. Наше постчеловеческое будущее: Последствия биотехнологической революции; пер. с англ. М. Б. Левина / Ф. Фукуяма. – М.: АСТ : ЛЮКС, 2004. – 349 с.

[16] Хабермас Ю. Техника и наука как «идеология» / Ю. Хабермас. М.: Праксис, 2007. – 208 с.

[17] Хайдеггер М. Вопрос о технике / М. Хайдеггер // Время и бытие: статьи и выступления: пер. с нем. – М., 1993. – 49 с.

[18] Черникова Д. В., Черникова И. В. Образовательные и этические аспекты вызовов технонауки в пространстве университета / Д. В. Черникова, И. В. Черникова // Высшее образование в России. – 2021. – Т. 30, № 11. – С. 42-51.

[19] Бодрийяр Ж. Система вещей / Пер. с фр. С. Н. Зенкина. – М.: «Рудоми-но», 1999. – 224 с.

[20] Нейсбит Д. Высокая технология, глубокая гуманность: Технологии и наши поиски смысла / При участии Н. Нейсбит и Д. Филипса / пер. с англ. А. Н. Анваера. – М.: Транзит Книга, 2005. – 381 с.

[21] Кутырев В. А. Культура и технология: борьба миров / В. А. Кутырев. – М.: Прогресс-Традиция, 2001. – 240 с.

[22] Черникова Д. В., Черникова И. В. Образовательные и этические аспекты вызовов технонауки в пространстве университета / Д. В. Черникова, И. В. Черникова // Высшее образование в России. – 2021. – Т. 30, № 11. – С. 42-51.

[23] Nordmann A. Synthetic Biology at the Limits of Science / B. Giese, C. Pade, H. A. Wigger von Gleich // Synthetic Biology. Character and impact. – Heidelberg u. a.: Springer, 2015. – P. 3-7.

[24] Тоффлер Э. Шок будущего / Э. Тоффлер. – М.: АСТ, 2002. – 557 с.

[25] Иванов Д. В. Виртуализация общества / Д. В. Иванов. – СПб., 2002. – 96 с.

[26] Мелюхин И. С. Информационное общество: истоки, проблемы, тенденции развития / И. С. Мелюхин. – М.: Издательство Московского университета, 1999. – 206 с.

[27] Ракитов А. И. Наш путь к информационному обществу / А. И. Ракитов // Теория и практика общественно-научной информации. – М.: ИНИОН, 1989. – С. 50-68.

[28] Горохов В. Г. Социальные проблемы нанотехнологии / В. Г. Горохов // Высшее образование в России. – 2008. – № 3. – С. 84-98.

Исследования, посвященные определению и классификации ИИ. К этому пулу исследований относятся работы таких авторов, как А. Ю. Алексеев[29], В. Архипов[30], Р. Брукс[31], Н. Бостром[32], Р. Бродхэрст[33], В. В. Васильев[34], Н. Винер[35], Б. Герцель[36], П. Готовцев[37], Х. Де Гарис[38], И. Дубровский[39], А. Е. Евстратов[40], В. Карпов[41], Р. Курцвейл[42], Ш. Легг[43], Дж. Маккарти[44], К. Макниш[45], Р. Мерфи[46], М. Мински[47], П. М. Морхат[48], В.

[29] Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты / А. Ю. Алексеев. – М.: ИИнтеЛЛ, 2013. – 304 с.

[30] Архипов В. В., Наумов В. Б. О некоторых вопросах теоретических оснований развития законодательства о робототехнике: аспекты воли и правосубъектности / В. В. Архипов, В. Б. Наумов // Закон. – 2017. – № 5. – С. 157-170.

[31] Stone P., Rodney B. Artificial intelligence and life in 2030 / P. Stone, B. Rodney, B. Erik, C. Ryan, O. Etzioni // One-hundred-year study on artificial intelligence: Report of the 2015–2016. – Stanford, Stanford University. – URL: http://ai100.stanford.edu/2016-report (дата обращения: 13.10.2022)

[32] Бостром Н. Искусственный интеллект. Этапы. Угрозы. Стратегии / пер. с англ. С. Филина. – М.: Манн, Иванов и Фербер, 2016. – 496 с.

[33] Broadhurst R., Brown P et al. Artificial Intelligence and Crime / R. Broadhurst, P. Brown, D. Maxim, H. Trivedi, J. Wang // Research Paper, Korean Institute of Criminology and Australian National University Cybercrime Observatory, College of Asia and the Pacific. – Canberra, 2019. – Pp. 1-70.

[34] Васильев В. В. Трудная проблема сознания / В. В. Васильев. – М.: Прогресс-Традиция, 2009. – 272 с.

[35] Wiener N. The human use of human beings: cybernetics and society / N. Wiener. – Boston: Houghton Mifflin, Second Edition Revised, NY : Doubleday anchor, 1954. – 344 p.

[36] Goertzel B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects / B. Goertzel // Journal of Artificial General Intelligence. – 2014. – Vol. 5(1). – Pp. 1-46.

[37]Готовцев П. М., Ройзенсон Г. В. Характеристика проектов стандартов на этичный искусственный интеллект IEEE / П. М. Готовцев, Г. В. Ройзенсон // 390 Этика и «цифра». – 2020. – URL: https://ethics.cdto.center/ieee (дата обращения: 12.04.2022).

[38] Де Гарис Х. Искусственный мозг: подход с развитым модулем нейронной сети / Х. Де Гарис // World Scientific. – 2010. – 400 с.

[39] Дубровский Д. И. Искусственный интеллект и проблема сознания / Д. И. Дубровский // Философия искусственного интеллекта: материалы всерос. междисциплинар. конф., М., МИЭМ, 17-19 янв. 2005 г. – М.: ИФ РАН, 2005. – С. 26-31.

[40] Евстратов А. Э., Гученков И. Ю. Пределы применения искусственного интеллекта (правовые проблемы) / А. Э. Евстратов, И. Ю. Гученков // Правоприменение. – 2020. – Т. 4, № 4. – С.13-19.

[41] Карпов В. Э., Готовцев П. М., Ройзенсон Г. В. Машинная этика / В. Э. Карпов, П. М. Готовцев, Г. В. Ройзенсон // 390 Этика и «цифра». – 2020. – URL: https://ethics.cdto.center/3_4 (дата обращения: 23.04.2022).

[42] Kurzweil R. The Age of Intelligent Machines / R. Kurzweil. – Cambridge, MA : MIT Press, 1990. – 565 p.

[43] Legg S., Hutter M. A collection of definitions of intelligence / In B. Goertzel, P. Wang (Eds.) // Advances in artificial general intelligence: concept, architectures and algorithms. – Amsterdam : IOS Press., 2007. – Vol.157. – Pp. 17–24.

[44]McCarthy J. What is Artificial Intelligence? / J. McCarthy // Stanford University. – 2007. – URL: http://www-formal.stanford.edu/jmc/whatisai. (дата обращения: 21.09.2022)

[45] MacNish C., Pearce D., Pereira L. M. Logics in Artificial Intelligence / C. MacNish, D. Pearce, L. M. Pereira // European Workshop JELIA '94, York, UK, September 5-8, 1994. – 413 p.

[46] Murphy R. F. Artificial Intelligence Applications to Support K-12 Teachers and Teaching / R. F. Murphy // A Review of Promising Applications, Opportunities, and Challenges. RAND Corporation. – URL: https://www.rand.org/content/dam/rand/pubs/perspectives/PE300/PE315/RAND_PE315.pdf (дата обращения: 30.03.2021).

[47] Мински М. Фреймы для представления знаний / Мински М. – М.: Мир, 1979. – 151 с.

[48] Морхат П. М. Искусственный интеллект: правовой взгляд / П. М. Морхат // Институт государственно-конфессиональных отношений и права. – М.: Буки Веди, 2017. – 257 с.

Наумов[49], А. В. Незнамов[50], Н. Нильсон[51], П. Норвиг[52], Р. Пенроуз[53], М. Райан[54], С. Рассел[55], О. В. Ревинский[56], А. В. Резаев , В. Ручкина[57], Д. Серль[58], В. Н. Синельникова[59], Н. Д. Трегубова [60], А. Тьюринг[61], А. Турчин[62], В. Фулин[63], М. Хэнлэйн[64].

Впервые тема искусственного интеллекта в виде идеи «мыслящих машин» была введена в 1950 г. в работе британского математика А. Тьюринга «Вычислительные машины и разум»[65]. В 1956 г. Дж. Маккарти использовал термин «искусственный интеллект» для обозначения научных исследований, связанных с математическими, лингвистическими и алго-

[49] Наумов В. Б. Право в эпоху цифровой трансформации: в поисках решений / В. Б. Наумов // Российское право: образование, практика, наука. – 2018. – № 6 (108). – С. 4-11.
[50] Незнамов А. В. О концепции регулирования технологий искусственного интеллекта и робототехники в России / А. В. Незнамов // Закон. – 2020. – № 1. – С. 171-185.
[51] Нильсон Н. Искусственный интеллект: методы поиска решений / Пер. с англ. В. Л. Стефанюка; под редакцией С. В. Фомина. – М.: Мир, 1973. – 272 с.
[52] Norvig P. Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp / P. Norvig // Morgan Kaufmann. – 1991. – 948 p.
[53] Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики / Пер. с англ. под. ред. В.О. Малышенко. 3-е изд. – М.: Издательство ЛКИ, 2008. – С. 328.
[54] Ryan M. In AI we trust: Ethics, artificial intelligence, and reliability / M. Ryan // Science and Engineering Ethics. – 2020. – Vol. 26. – Pp. 2749-2767.
[55] Рассел С., Норвиг П. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. пер. с англ. 2-е изд. – М.: Вильямс, 2006. – 1408 с.
[56] Синельникова В. Н., Ревинский О. В. Права на результаты искусственного интеллекта / В. Н. Синельникова, О. В. Ревинский // Вестник Российской академии интеллектуальной собственности и Российского авторского общества. – 2017. – № 4. – С. 24-27.
[57] Ручкина Г. Ф. Искусственный интеллект, роботы и объекты робототехники: к вопросу о теории правового регулирования в Российской Федерации / Г. Ф. Ручкина // Банковское право. – 2020. – № 1. – С. 7-18.
[58] Серль Дж. Р. Сознание, мозг и программы / Дж. Р. Серль // Аналитическая философия: Становление и развитие: Антология / Общ. ред. и сост. А. Ф. Грязнов. – М., 1998. – 528 с.
[59] Синельникова В. Н., Ревинский О. В. Права на результаты искусственного интеллекта / В. Н. Синельникова, О. В. Ревинский // Вестник Российской академии интеллектуальной собственности и Российского авторского общества. – 2017. – № 4. – С. 24-27.
[60] Резаев А. В., Трегубова Н. Д. Искусственный интеллект и искусственная социальность: новые явления и проблемы для развития медицинских наук / А. В. Резаев, Н. Д. Трегубова // Эпистемология и философия науки. — М., 2019. — Т. 56, №4. – С. 183-199.
[61] Turing A. Computing machinery and intelligence / A. Turing // Mind. – 1950. – Vol. 59. – Pp. 433-460.
[62] Турчин А. В. Футурология. XXI век. Бессмертие или глобальная катастрофа? / А. В. Турчин, М. А. Бахтин. – Москва, 2013. – URL: https://libking.ru/books/nonf-/nonf-publicism/205876-aleksey-turchin-rossiyskaya-akademiya-nauk.html. (дата обращения: 12.09.2022)
[63] Фулин В. А. Универсальный искусственный интеллект и экспертные системы / В. А. Фулин, В. Н. Ручкин. – СПб.: БХВ-Петербург. – 2009. – 240 с.
[64] Kaplan A., Haenlein M. On the interpretations, illustrations, and implications of artificial intelligence / A. Kaplan, M. Haenlein. – URL: https://www.sciencedirect.com/science/article/abs/pii/ S0007681318301393 (дата обращения: 14.02.2021)
[65] Тьюринг А. Может ли машина мыслить / А. Тьюринг. – М.: Едиториал УРСС, Ленанд. – 2016. – 128 с.

ритмическими проблемами, необходимыми для имитации интеллекта человека с помощью компьютера[66].

Понятие «сильный искусственный интеллект» было введено американским философом Дж. Серлем и контекстуально определено следующим образом: «Более того, такая программа будет не просто моделью разума; она в буквальном смысле слова сама и будет разумом, в том же смысле, в котором человеческий разум — это разум»[67].

Определение «слабого (или узкого) ИИ» предложили, в частности, российские ученые В. Н. Синельникова и О. В. Ревинский, по мнению которых слабый ИИ – это компьютерная программа, спроектированная людьми и обладающая способностью, в соответствии с заложенной командной архитектурой, создавать новую информацию[68].

В дальнейшем представители аналитической философии (Н. Блок[69] , Д. Деннет[70], Т. Нагель[71], Х. Патнэм[72], Дж. Сёрл[73], Дж. Фодор[74], Д. Чалмерс[75] и др.) в своих трудах различали сильный ИИ и слабый ИИ как два основных подхода к искусственному интеллекту, сложившиеся в современной науке.

---

[66]McCarthy J. What is Artificial Intelligence? / J. McCarthy // Stanford University. – 2007. – URL: http://www-formal.stanford.edu/jmc/whatisai. (дата обращения: 21.09.2022)

[67] Серль Дж. Р. Разум мозга - компьютерная программа? / Дж. Р. Серль. – URL: https://psychosearch.ru/teoriya/psikhika/338-searle-john-razum-mozga-kompyuternaya-programma (дата обращения: 12.03.2022).

[68] Синельникова В. Н., Ревинский О. В. Права на результаты искусственного интеллекта / В. Н. Синельникова, О. В. Ревинский // Вестник Российской академии интеллектуальной собственности и Российского авторского общества. – 2017.– №4. – С. 17-27.

[69] Block N. Troubles with functionalism. Minnesota Studies in the Philosophy of Science / N. Block // Troubles with functionalism, Minnesota Studies in the Philosophy of Science. – 1978. – Pp.261-325.

[70] Деннет Д. Виды психики. На пути к пониманию сознания / пер. А. Веретенникова. – М.: Идея-Пресс, 2004. – 79 с.

[71] Нагель Т. Каково быть летучей мышью? // Хоф-штадтер Д., Деннет Д. Глаз разума / пер. с англ. М. А. Эскиной. – Самара : «Бахрах-М», 2003. – С. 349-360.

[72] Putnam H. The Project of Artificial Intelligence / H. Putnam // Renewing Philosophy. – Cambridge, MA : Harvard University Press, 1992. – P. 1-18.

[73] Серль Дж. Р. Сознание, мозг и программы / Дж. Р. Серль // Аналитическая философия: Становление и развитие: Антология / Общ. ред. и сост. А.Ф. Грязнов. – М., 1998. – 528 с.

[74] Fodor J. A. In critical condition: Polemical essays on cognitive science and the philosophy of mind / J. A. Fodor. – Cambridge, MA : MIT Press. – 1998. – P. 17

[75] Чалмерс Д. Сознающий ум. В поисках фундаментальной теории / Д. Чалмерс. – М.: Либроком, 2019. – 512 с.

Разнообразие многочисленных подходов к пониманию ИИ представлено в недавнем исследовании Ш. Легг и М. Хаттер «Коллекция определений ИИ»[76].

Исследования, посвященные проблемам и принципам применения искусственного интеллекта. Проблемы применения искусственного интеллекта обсуждались в трудах Р. Г. Апресяна[77], П. Готовцева[78], А. А. Гусейнова[79], В. Карпова[80], А. В. Разина[81], Г. В. Ройзензона[82], В. А. Цвык и И. В. Цвык[83], Б. Г. Юдина[84] и др. Эти авторы пришли к выводу о том, что проблемы разработки, внедрения и использования ИИ обладают ярко выраженной спецификой, отличающей их от комплекса вопросов, обсуждаемых в рамках биоэтики, генной инженерии, информатики и других областей научного знания. В. А. Цвык и И. В. Цвык отмечают, что последствия от внедрения ИИ оказывает глубокое влияние на развитие общества, науки, культуры и коммуникации. По их мнению, несмотря на то, что ИИ имеет потенциал изменения человечества в лучшую сторону, он порождает риски, угрожающие основным правам и свободам человека. П. Готовцев, В. Карпов, А. В. Разин, Г. В. Ройзензон в своих работах ставят

[76] Legg S., Hutter M. A collection of definitions of intelligence / In B. Goertzel, P. Wang (Eds.) // Advances in artificial general intelligence: concept, architectures and algorithms. – Amsterdam : IOS Press., 2007. – Vol.157. – Pp. 17-24.

[77] Апресян Р. Г. Этика и дискуссии об искусственном интеллекте / XI международная конференция «Теоретическая и прикладная этика: Традиции и перспективы - 2019. К грядущему цифровому обществу. Опыт этического прогнозирования (100 лет со дня рождения Д. Белла - 1919-2019)». Санкт-Петербургский Государственный Университет, 21-23 ноября 2019 г. Материалы конференции / Отв. ред. В. Ю. Перов. – СПб: ООО «Сборка», 2019. – С. 169-170.

[78] Готовцев П. М., Ройзенсон Г. В., Характеристика проектов стандартов на этичный искусственный интеллект IEEE / П. М. Готовцев, Г. В. Ройзенсон // 390 Этика и «цифра». – 2020. – URL: https://ethics.cdto.center/ieee (дата обращения: 12.04.2022).

[79] Гусейнов А. А. Размышления о прикладной этике / Доклад на основе статьи: Размышления о прикладной этике // Ведомости НИИПЭ, Вып. 25: Общепрофессиональная этика. – Тюмень : НИИПЭ, 2004. – 148 с.

[80] Карпов В. Э., Готовцев П. М., Ройзенсон Г. В. Машинная этика / В. Э. Карпов, П. М. Готовцев, Г. В. Ройзенсон //390 Этика и «цифра». – 2020. – URL: https://ethics.cdto.center/3_4 (дата обращения: 23.04.2022).

[81] Разин А. В. Этика искусственного интеллекта / Разин А. В. // Философия и общество. – 2019. – №1. – С. 57-73.

[82] Ройзензон Г. В. Проблемы формализации понятия этики в искусственном интеллекте / Г. В. Ройзензон // XVI национальная конференция по искусственному интеллекту с международным участием КИИ-2018. –М., 2018. – С. 245-252.

[83] Цвык В. А., Цвык И. В. Социальные проблемы развития и применения искусственного интеллекта / В. А. Цвык, И. В. Цвык // Вестник РУДН. – Серия: Социология. – 2022. – №1. – С. 58-69.

[84] Юдин Б. Г. Социальные технологии, их производство и потребление / Б. Г. Юдин // Эпистемология и философия науки. – 2012. – Вып. 31, № 1. – С. 55-64.

вопросы о том, какие этические нормы должны быть заложены в ИИ на этапе его разработки.

Целый ряд отечественных и зарубежных исследователей занимались исследованием проблем применения ИИ в конкретных сферах жизнедеятельности общества: Дж. Боссманн[85], А. Джобин, М. Йенка и Е. Вайена [86], В. Э. Карпов[87], М. Кэролан[88], В. А. Лаптев[89], П. М. Морхат[90], А. В. Попова[91], М. Райан[92], С. Рассел[93], Таддео[94], Э. Тополь[95], Л. Флориди[96], Т. Хагендорф[97], Э. Юдковски[98], Н. А. Ястреб[99].

Так, в трудах Л. Флориди обсуждаются вопросы, связанные с соблюдением конфиденциальности личных данных[100]. С. Паркенсон и Э. Харпер пытались выявить препятствия и трудности, возникающие в процессе включения конфиденциальных данных пользователей в наборы

[85] Bossmann Dzh. Top 9 Ethical Issues in Artificial Intelligence / Dzh. Bossmann. – URL: https://hr-portal.ru/article/9-glavnyh-eticheskih-problem-iskusstvennogo-intellekta (дата обращения: 22.03.2022).
[86] Jobin A., Ienca M., Vayena E. Artificial Intelligence: The Global Landscape of Ethics Guidelines / A. Jobin, M. Lenca, E. Vayena // Nature Machine Intelligence. – 2019. – Vol.1. – Pp. 389-399.
[87] Карпов В. Э, Готовцев П. М., Ройзензон Г. В. К вопросу об этике и системах искусственного интеллекта / В. Э. Карпов, П.М. Готовцев, Г. В. Ройзензон // Философия и общество. – 2018. – №2 (87). – С. 84-105.
[88] Carolan M. Automated agrifood futures: robotics, labor and the distributive politics of digital agriculture / M. Carolan // J. Peasant Stud. – 2020. – Vol. 47. – Pp.184-207.
[89] Лаптев В. А. Электронные доказательства в арбитражном процессе / В. А. Лаптев // Российская юстиция, № 2. – 2017. – С. 56-59.
[90] Морхат П. М. Искусственный интеллект: правовой взгляд / П. М. Морхат // Институт государственно-конфессиональных отношений и права. – М.: Буки Веди, 2017. – 257 с.
[91] Попова А. В. Этические принципы взаимодействия с искусственным интеллектом как основа правового регулирования / А. В. Попова // Правовое государство: теория и практика. – 2020. – № 3 (61). – С. 34-43.
[92] Ryan M. Ethics of using AI and big data in agriculture: the case of a large agriculture multinational / M. Ryan // ORBIT Journal. – 2019. – Vol.2(2). – 27 p.
[93] Russell S. Human-Compatible Artificial Intelligence. / S. Russel // Human-Like Machine Intelligence. – Oxford: Oxford University Press, 2021. – Pp 3-23.
[94] Taddeo M. Is cybersecutiry a public good? / M. Taddeo // Minds and machines. – 2019. – Vol. 29, № 3. - Pp. 349-354.
[95]Тополь Э. Будущее медицины: Ваше здоровье в ваших руках / Э. Тополь. – М.: Альпина нон-фикшн, 2016. – 491 с.
[96]Floridi L. The end of an era: from self-regulation to hard law for the digital industry / L. Floridi // Philosophy & Technology. – 2021. – Vol. 34, № 4. – Pp. 612-622.
[97] Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines / T. Hagendorff // Minds & Machines. – 2020. – Vol. 30. – Pp. 99-120.
[98] Юдковски Э. Систематические ошибки в рассуждениях, потенциально влияющие на оценку глобальных рисков. Новые технологии и продолжение эволюции человека? / Э. Юдковски // Трансгуманистический проект будущего. – М., 2008. – С. 182-225.
[99] Ястреб Н. А. Индустрия 4.0: киберфизические системы и интернет вещей / Н. А. Ястреб // Человек в технической среде: сборник научных статей / Под ред. доц. Н.А. Ястреб. – Вологда : ВолГУ, 2015. – С. 136-141.
[100]Mittelstadt B. D., Allo P. et al. The ethics of algorithms: Mapping the debate / B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi // Big Data and Society. – 2016. – Vol. 3(2). – Pp. 1-21.

больших данных[101]. Обоснование необходимости и проблем обеспечения прозрачности процессов, связанных с данными пользователей, рассматриваются в работах Дж. Баррел[102].

Преодоление и недопущение социальной несправедливости и предвзятого отношения – предмет исследований Т. Панч[103]. Российско-французский философ А. Гринбаум в труде «Машина-доносчица» поднимает вопросы, касающиеся «нравственности» узкого искусственного интеллекта, его ответственности перед человеком[104]. Автор называет узкий ИИ некой цифровой особой, которая сама по себе не может быть признана способной нести ответственность за совершенные действия и принятые решения. По-настоящему ответственной, нравственной ее может сделать только сам человек. Вообще, проблемы нравственности, соответствия критериям добра при использовании систем с ИИ поднимались еще во времена А. Тьюринга. По сей день в трудах многих ученых, исследователей обсуждается возможность причинения вреда человечеству со стороны сильного ИИ, звучит тревога и опасения за будущее, за безопасность человеческой цивилизации в целом.

Определенные результаты в данном контексте достигнуты в отечественном обществознании. Так, труды П. М. Морхата посвящены правовой регламентации процессов жизненного цикла систем ИИ[105]. Л. В. Баева, Храпов С. А. исследуют риски и последствия цифровых технологий, в том числе ИИ, в области образования[106].

---

[101]Harper E. M., Parkerson S. Powering Big Data for Nursing Through Partnership / E. M. Harper , S. Parkerson.– URL: https://pubmed.ncbi.nlm.nih.gov/26340243/ (дата обращения: 14.03.2022)

[102] Burrell J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms / J. Burrel // Big Data and Society. – 2016. – Vol. 3(1). – Pp. 1-12.

[103]Panch T., Mattie H. et al. Artificial intelligence and algorithmic bias: implications for health systems / T. Panch, H. Mattie, R. Atun // Journal of global health. – 2019. – Vol. 9(2). – Pp. 23-32.

[104]Гринбаум А. Машина-доносчица: как избавить искусственный интеллект от зла / А. Гринбаум. – М.: ТрансЛит, 2017. – 76 с.

[105] Морхат П. М. Искусственный интеллект: правовой взгляд / П. М. Морхат // Институт государственно-конфессиональных отношений и права. – М.: Буки Веди, 2017. – 257 с.

[106] Баева Л. В., Храпов С. А. Цифровизация образовательного пространства: эмоциональные риски и эффекты / Л. В. Баева, С. А. Храпов // Вопросы философии. – 2022. – №4. – С. 16-24.

В. Э. Карпов сосредоточил внимание на этических аспектах применения ИИ в жизни современного общества, базовых принципах взаимодействия ИИ с человеком.

Несмотря на внимание ученых к отдельным проблемам, возникающим в отношениях человека с ИИ, указанные трудности, как правило, не исследуются ими в комплексе, в целостности, на уровне социально-философского познания, что не позволяет выработать адекватную и безопасную стратегию взаимодействия общества с искусственным разумом и требует выработки философских принципов применения систем ИИ в современном мире.

Объектом исследования являются системы ИИ (слабого или узкого ИИ), созданные человеком для решения определенных практических задач.

Предмет исследования – проблемы применения систем ИИ, способы и пути их решения в современном обществе.

Цель данной работы – выявить и раскрыть содержание философских принципов взаимодействия систем ИИ с человеком для повышения безопасности и снижения рисков их использования в различных сферах общественной жизни.

Задачи исследования:

1. Выделить и охарактеризовать существующие в научной и философской литературе теоретические подходы к пониманию сущности искусственного интеллекта и классификации ИИ в контексте его практического применения в социальной реальности.

2. Определить наиболее значимые области использования ИИ в социальной практике.

3. Систематизировать существующие в настоящее время документы, регулирующие этические и правовые аспекты применения ИИ, выявить их содержание и основные идеи с целью установления возможных

рисков и угроз для безопасного и эффективного использования человеком систем ИИ.

4. Обосновать необходимость исследования проблем применения систем ИИ с позиций социальной философии.

5. Выделить проблемы, возникающие вследствие использования систем ИИ в различных сферах общественной жизни, отражающие основное содержание и отдельные стороны негативного воздействия указанных систем на человека и общество.

6. Определить совокупность проблем, вызванных применением систем ИИ, указывающих на причины их отрицательного воздействия на жизнь социума.

7. Раскрыть содержание и значение философских принципов, нацеленных на выявление условий эффективного и безопасного применения систем ИИ в современном обществе.

8. Выявить смысл и соотношение философских принципов, раскрывающих отдельные стороны, аспекты негативного влияния ИИ на жизнь человека и общества.

Теоретическая база и методология исследования. В качестве теоретической основы были использованы труды отечественных и зарубежных ученых: П. В. Алексеева[107], А. Гринбаума[108], А. Гринфилда[109], В. А. Кутырева[110], Х. Ортега-и-Гассет[111], М. Райана[112], Л. Флориди[113], М. Хайдеггера[114], Ф. Фукуямы[115], И. В. Черниковой и Д. В. Черниковой[116], Б.

[107] Алексеев П. В. Социальная философия / П. В. Алексеев. – М.: ООО «ТК Велби», 2003. – 256 с.
[108] Grinbaum A. et al. Ethics in Robotics Research/ A. Grinbaum, R. Chatila, L. Devillers, J. G. Ganascia, // IEEE Robotics and Automation Magazine. – 2017. – № 24. – Pp. 139-145.
[109] Гринфилд А. Радикальные технологии: устройство повседневной жизни / А. Гринфилд. – М.: Издательский дом «Дело» РАНХиГС, 2019. – 424 с.
[110] Кутырев В. А. Культура и технология: борьба миров / В. А. Кутырев – М.: Прогресс-Традиция, 2001. – 240 с.
[111] Ортега-и-Гассет Х. Размышления о технике / Х. Ортега-и-Гассет // Избранные труды пер. с исп.; сост., предисл. и общ. ред. А. М. Руткевича. – М.: Весь Мир, 1997. – С. 164-232.
[112] Ryan M. Ethics of using AI and big data in agriculture: the case of a large agriculture multinational / M. Ryan // ORBIT Journal. – 2019. – Vol. 2(2). – 27 p.
[113] Floridi L., Cowls J. et al. How to design AI for social good: Seven Essential factors / L. Floridi, J. Cowls, T.C. Kin, M. Taddeo // Sci. Eng. Ethics. – 2020. – Vol.26. – Pp. 1771-1796.
[114] Хайдеггер М. Вопрос о технике / М. Хайдеггер // Время и бытие : статьи и выступления : пер. с нем. – М., 1993. – 49 с.

Г. Юдина[117], К. Ясперса [118] и др. Исследование представленных в них положений позволило сформировать понятийный аппарат диссертации, выявить основные проблемы, тенденции и закономерности развития ИИ, вызовы и пути преодоления возникающих трудностей, обусловленных эволюцией и расширением сферы применения ИИ в различных областях социальной реальности.

Осмысление взаимодействия человека и техники в современном обществе стало предметом исследования работ А. Гринфилда, В. А. Кутырева, Х. Ортега-и-Гассета, Ф. Фукуямы, М. Хайдеггера, И. В. Черниковой, Д. В. Черниковой, Б. Г. Юдина, К. Ясперса, и др. М. Хайдеггер указывает на необходимость определения сущности техники, чтобы выявить, к каким последствиям должно быть готово общество при широком ее применении. При этом он рассматривает технику не в качестве простого, нейтрального и подчиненного человеку инструмента, предназначенного для достижения его целей и удовлетворения потребностей. По мнению философа, техника перестраивает самого человека, изменяя его сущность и подвергая риску его онтологический статус в мире. К. Ясперс видит в технизации угрозу утраты внутренней свободы и ценности личности. Он приходит к выводу, что именно за человеком должна оставаться целеполагающая функция, а техника является лишь средством достижения целей, поставленных человеком. Размышляя об экзистенциальных угрозах, вызванных тотальным распространением техники, он пишет, что техника – это «нечто такое, что подавляет, влияет на все их существование, противостоит им, не осознано ими, что словно бы происходит на заднем плане, не раскрыто»[119]. А. Гринфилд называет современные технологии «ра-

---

[115] Фукуяма Ф. Наше постчеловеческое будущее: последствия биотехнологической революции / Пер. с англ. М. Б. Левина. – М.: АСТ, 2004. – С. 364.
[116] Черникова Д. В., Черникова И. В. Образовательные и этические аспекты вызовов технонауки в пространстве университета / Д. В. Черникова, И. В. Черникова // Высшее образование в России. – 2021. – Т. 30, № 11. – С. 42-51.
[117] Юдин Б. Г. Социальные технологии, их производство и потребление // Эпистемология и философия науки. – 2012. – Вып. 31, № 1. – С. 55-64.
[118] Ясперс К. Смысл и назначение истории: пер.с нем / К. Ясперс. – М.: Политиздат, 1991. – 527 с.
[119] Ясперс К. Смысл и назначение истории: пер.с нем / К. Ясперс. – М.: Политиздат, 1991. – с. 137.

дикальными», поскольку они оказывают беспрецедентное в истории развития техники влияние на социальный порядок и повседневный уклад жизни общества. По его мнению, искусственный интеллект – это нечто, ни на что не похожее и способное кардинально изменить будущее человечества. Ф. Фукуяма в своих трудах выражает обеспокоенность о допустимости преобразований человека и общества с помощью применения современных технологий. Он предостерегает, что изменениям подвергнутся те существенные черты, которые делают человека человеком.

Социально-философскими проблемами, вызванными применением систем искусственного интеллекта, занимались А. Гринбаум[120], М. Райан, М. Форд[121], Л. Флориди[122] и др. В трудах М. Форда обсуждаются проблемы роста безработицы, профессиональной поляризации, необходимости перестройки структуры управления организации в связи с внедрением технологии ИИ.

Методологической базой исследования выступили системный, деятельностный подходы, герменевтический метод исследования и принцип историзма. Системный подход был применен при выделении комплекса основных социально-философских проблем применения ИИ и раскрытии содержания базовых принципов его использования в социальной практике. Также системный подход позволил подойти к выявлению сущности искусственного интеллекта как системного образования, составляющие которого образуют целостность, не сводимую к простой совокупности, сумме его частей.

Деятельностный подход и наследие философии техники применялись в оценке влияния систем ИИ на сущность человека, его идентичность как разумного существа, способного к целеполагающей, природо-

[120] Grinbaum A. et al. Ethics in Robotics Research / A. Grinbaum, R. Chatila, L. Devillers, J. G. Ganascia, // IEEE Robotics and Automation Magazine. – 2017. – № 24. – Pp. 139-145.
[121] М. Форд. Роботы наступают: развитие технологий и будущее без работы / пер.с англ. С. Чернина. – М.: Альпина нон-фикшн, 2016. – 572 с.
[122] Ryan M. Ethics of using AI and big data in agriculture: the case of a large agriculture multinational / M. Ryan // ORBIT Journal. – 2019. – Vol. 2 (2). – 27 p.

преобразующей, творческой деятельности, ориентированной на принятые индивидом ценности, нормы, идеалы.

Герменевтический метод был использован при интерпретации нормативно-правовых документов, регулирующих процессы разработки и внедрения систем ИИ. Также данный метод помог проследить процесс трансформации толкования сущности систем ИИ в сфере науки и практического применения этих систем.

Принцип историзма позволил рассмотреть системы ИИ в динамике их изменения, становления во времени, в связи с конкретно-историческими условиями их существования.

В диссертации также применяются традиционные общенаучные методы (анализ, синтез, дедукция, индукция и др.).

Новизна научной работы состоит в следующем:

1. Показано, что сущность искусственного интеллекта в контексте проблем его применения, отражающая существующий уровень его использования в современном социуме, характеризуется системной связью составляющих его элементов, а также нацеленностью на решение определенных, узконаправленных задач, поставленных человеком, что позволяет определить его, в соответствии с действующей классификацией ИИ, как «узкий» (или слабый) ИИ.

2. Определены основные сферы применения ИИ и аргументирован тезис о глубоком проникновении, встраивании его в основы, фундамент современного социума, что стало причиной возникновения риска новой экзистенциальной угрозы, обусловленной возможной утратой человеком привычного для него места в мире.

3. Выделены и систематизированы документы, регулирующие этико-правовые аспекты разработки, внедрения и применения ИИ, изданные различными социальными институтами и организациями современного общества, содержание которых позволяет выявить основные риски и угрозы для безопасного и эффективного ИИ.

4. Обоснована необходимость социально-философского исследования проблем применения систем ИИ, позволяющего подойти к их решению с позиций целостности социальной жизни и присущей философии общества нацеленности на индивида, его многогранные потребности и стремление к достижению социального идеала.

5. Выявлены социальные проблемы, возникающие вследствие применения систем ИИ, отражающие основные проявления отрицательного влияния ИИ на человека: проблемы причинения вреда, социальной несправедливости и нарушения автономии человека.

6. Определен круг проблем, раскрывающих причины возникновения нежелательных для человека и общества последствий взаимодействия с ИИ: проблемы непрозрачности, отсутствия ответственности и нарушения конфиденциальности.

7. Выявлено и раскрыто содержание философских принципов прозрачности, ответственности, конфиденциальности, отражающих условия безопасного и эффективного использования ИИ в различных областях жизни современного общества.

8. Определен смысл требований, составляющих содержание принципов социальной справедливости, автономии человека и непричинения вреда, представляющих отдельные стороны, грани безопасного, эффективного применения систем ИИ в различных областях жизни социума, а также обосновано основополагающее значение принципа непричинения вреда во всей совокупности рассмотренных выше философских принципов.

Основные положения, выносимые на защиту:

1. Существующий в настоящее время на уровне технологии и применяемый в различных областях общественной жизни искусственный интеллект представляет собой узкий (слабый) искусственный интеллект, созданный человеком для решения определенных узконаправленных задач и представляющий собой системы, включающие аппаратный ком-

плекс, программное обеспечение и набор данных. Особенностью ИИ является способность осуществлять сбор данных, их интерпретацию в виде рассуждений, принимать решения на основании имеющейся информационной базы практически в любой отрасли деятельности человека, что определяет постоянно расширяющийся спектр его применения.

2. Системы ИИ находят применение в ключевых сферах жизни социума (государственное управление, общественная безопасность, транспорт, сельское хозяйство, энергетика, здравоохранение, образование, правосудие, банковская, финансовая деятельность, военная сфера и др.), перечень которых непрерывно расширяется вследствие перманентного процесса их совершенствования, что привело к возникновению потенциальной угрозы пересмотра базовых мировоззренческих представлений о месте человека в мире, определяющем его роль и предназначение.

3. Экзистенциальная угроза, сформировавшаяся вследствие фундаментального встраивания систем ИИ в жизнь социума потребовала разработки и принятия целого комплекса документов, существующих в форме национальных стратегий, нормативно-правовых актов, этических руководств, рекомендаций и стандартов, изданных государственными институтами, коммерческими структурами, международными и неправительственными организациями и регулирующих этико-правовые аспекты использования ИИ.

4. Процесс применения ИИ требует глубокого и всестороннего социально-философского осмысления, фокусирующегося на человеке, его многогранных потребностях и стремлении к лучшей жизни. Исследование с теоретических позиций данной дисциплины необходимо для выявления рисков, вызываемых применением систем ИИ, для существования человека, сохранения его места и роли в современном мире.

5. Обобщение и осмысление существующего опыта взаимодействия человека с ИИ позволило выделить комплекс проблем их применения, отражающих основное содержание и различные стороны, аспекты отри-

цательного влияния ИИ на человека: проблемы причинения вреда, социальной несправедливости и нарушения автономии. Проблема причинения вреда приводит к риску отчуждения человека от самого себя, от собственной сущности и предназначения, искажению его особого статуса в мире как единственного существа, способного к интеллектуальной деятельности, нацеленной на изменение, преобразование окружающей реальности. Проблемы социальной несправедливости и нарушения автономии призваны раскрыть отдельные грани и описать разнообразные проявления причиняемого человеку ущерба, а также последствия, возникающие из-за некорректного внедрения, применения систем ИИ.

6. Отдельную группу трудностей, обусловленных применением ИИ, составили проблемы, раскрывающие причины формирования негативных последствий взаимодействия человека и ИИ: проблемы непрозрачности, отсутствия ответственности и нарушения конфиденциальности. Проблема непрозрачности заключается в том, что принцип работы систем ИИ становится в процессе его совершенствования все более непрослеживаемым, необъяснимым и неинтерпретируемым. Проблема отсутствия ответственности состоит в том, что природа ИИ не позволяет установить субъект ответственности, на которого однозначно можно было бы возложить вину в случае причинения вреда человеку системой ИИ. Проблема нарушения конфиденциальности возникает из-за угрозы утечки потока персональных данных или потери контроля над этими данными.

7. Для успешного разрешения существующих проблем применения систем ИИ и предотвращения возникновения новых затруднений раскрыто содержание философских принципов взаимодействия человека с системами ИИ, представляющих требования об открытости и доступности (принцип прозрачности), подотчетности систем ИИ человеку (принцип ответственности), о наложении запрета на нарушение личных границ (принцип конфиденциальности), выступающие условиями эффективного и безопасного применения систем ИИ.

8. Выявлен и сформулирован смысл требований об исключении любой предвзятости и дискриминации со стороны ИИ, о самостоятельности человека в принятии решений, составляющих содержание философских принципов социальной справедливости и автономии, в совокупности с вышеназванными принципами раскрывающих суть центрального философского принципа непричинения вреда человеку, призванного обеспечить безопасное для человека и общества, полезное и эффективное применение ИИ.

Теоретическое и прикладное значение. Диссертация является одним из первых в современном обществознании исследований, позволивших доказать необходимость рассмотрения проблем искусственного интеллекта с позиций социально-философского понимания места и роли человека в мире. Избранная теоретическая позиция, основанная на принципах гуманизма, позволила выявить и раскрыть содержание философских принципов применения систем ИИ, что составляет личный вклад соискателя в развитие концептуальных положений о безопасном, надежном и эффективном ИИ.

Полученные соискателем результаты основаны на широком круге источников, включающих помимо научной и философской литературы документы, регламентирующие этико-правовые аспекты использования ИИ, созданные и опубликованные государственными структурами, международными и неправительственными организациями, представителями бизнеса как в России, так и за рубежом. Тем самым автор диссертации вводит в научный оборот около ста работ (в том числе на английском и китайском языках), отражающих международный опыт применения систем ИИ, а также наиболее актуальные проблемы ИИ, стратегии и пути их преодоления.

Не менее важным теоретическим результатом является проведенная соискателем работа по выявлению и классификации основных проблем, обусловленных ИИ, осуществленная исходя из системного понимания

указанных проблем и их отношения к центральному принципу непричинения вреда, составляющему сущность и ядро концепции безопасного и надежного искусственного интеллекта.

Выводы работы могут быть использованы для создания документов, устанавливающих внутренние этические принципы компаний, занимающихся созданием и продвижением систем ИИ, а также послужить теоретической основой дальнейшего развития существующих общественных норм и правил в сфере применения ИИ, ориентированных на широкие социальные слои и группы.

Основные положения исследования могут найти применение при разработке просветительских программ для IT-специалистов, создающих новые программные продукты, для людей, которые применяют ИИ в промышленности, здравоохранении, транспорте, образовании и др. сферах общественной жизни, для сотрудников коммерческих структур, реализующих применение ИИ в малом и среднем бизнесе, для служащих государственных структур и общественных деятелей, поскольку фундаментальные философские принципы применения ИИ являются универсальными по своему характеру, дающими общее руководство по применению человеком любых систем ИИ.

Выводы исследования могут стать составной частью учебных курсов, посвященных актуальным проблемам социальной философии, спецкурсов, лекций и семинаров по машинной этике, философии техники, учебных и методических пособий по соответствующим дисциплинам, учебных курсов для средней школы в качестве источника информации и средства повышения общей осведомленности учащихся о проблемах и возможностям применения ИИ. Результаты диссертации также могут быть использованы при разработке широкого спектра научно-исследовательских программ социально-экономического развития.

Апробация работы. В процессе подготовки диссертации некоторые тезисы настоящего исследования были изложены и обсуждены в ряде публикаций: в монографии, научных сборниках и журналах.

Структура диссертационной работы. Структуру диссертации составляют введение, три главы, включающие восемь параграфов, заключение, список литературы и приложение.

ГЛАВА I. Искусственный интеллект: сущность, области применения, особенности исследования

1.1 Определение и классификация ИИ

Развитие науки и появление новых технологий, воспроизводивших те или иные возможности человеческого интеллекта и потому обозначавшихся термином «искусственный интеллект» привело к тому, что понятие «искусственный интеллект» стало чрезвычайно размытым и проблематичным. Но точное понимание смысла и значения термина «искусственный интеллект» является необходимым условием выявления и интерпретации социально-философских проблем, непосредственно связанных с внедрением систем искусственного интеллекта во все сферы жизни социума. Рефлексия над определением соответствующего ему понятия продолжается по сей день. Это понятие представляет собой спорную научную категорию, по-разному трактуемую представителями различных специальностей. Большая часть имеющихся определений не универсальны. Эти определения зачастую не покрывают тех значений понятия, которые уже существуют в современном обществе, трактуют ИИ упрощенно, искажают его сущность или делают акцент лишь на отдельных аспектах его функционирования (например, на системах машинного обучения или на способности к автоматическому принятию решений).

Несмотря на отсутствие общепринятого определения многочисленные дефиниции ИИ в своем содержании зачастую пересекаются, совпадают. Потому не удивительно, что Ш. Легг и М. Хаттер посвятили им свою отдельную работу, вышедшую в свет под названием «Коллекция определений ИИ»[123]. В ней авторы рассмотрели наиболее известные и упоминаемые в научных работах определения ИИ, выявив имеющиеся в них сходства и различия.

[123]Legg S., Hutter M. A collection of definitions of intelligence / In B. Goertzel, P. Wang (Eds.) // Advances in artificial general intelligence: concept, architectures and algorithms. – Amsterdam : IOS Press. – 2007. – Vol.157. – Pp. 17-24.

Пионером в области создания искусственного интеллекта по праву считается Н. Винер, автор работы «Кибернетика или управление и связь в животном и машине»[124]. Тогда понятия «искусственный интеллект» еще не существовало, однако ученым была высказана основная идея конструирования систем ИИ – универсальность алгоритмов решения любых, в том числе, практических задач, независимо от их воплощения. Н. Винер тем самым фактически начал поиск общего алгоритма решения любых задач.

В 1950 г. А. Тьюрингом в работе «Вычислительные машины и разум»[125] был разработан специальный тест, который, по замыслу разработчика, должен был отличить поведение объекта, наделенного ИИ, от поведения человека. Тест можно было считать успешно пройденным, если человек–экспериментатор, задающий ИИ вопросы в письменном виде, не сможет определить, кто на самом деле отвечает, человек или интеллектуальное устройство.

Впервые термин «искусственный интеллект» был введен в научный оборот информатиком Д. Маккарти, выступившим в 1956 г. на конференции в Дартмутском колледже, поэтому появление ИИ официально восходит именно к этой дате. Д. Маккарти понимал под ИИ возможность «обучить машины использовать различные навыки для решения задач, подвластных на данном этапе только людям»[126]. В 1958 г. он создал язык программирования LISP, разработанный для искусственного интеллекта.

В 1960 г. на свет появилась программа под названием GPS (General Problem Solver)[127], которая могла решать головоломки, заниматься вычислением неопределенных интегралов, решать такие интеллектуальные за-

[124] Wiener N. Cybernetics: Or Control and Communication in the Animal and the Machine / N. Wiener. – Paris : Hermann &Cie; Cambridge (MA) : MIT Press, 1948. – 229 p.
[125] Turing A. Computing machinery and intelligence / A. Turing // Mind. – 1950. – Vol. 59. – Pp. 433-460.
[126] McCarthy J. What is Artificial Intelligence? / J. McCarthy // Stanford University. – 2007. – URL: http://www-formal.stanford.edu/jmc/whatisai. (дата обращения: 21.09.2022)
[127] Newell A., Shaw J. C., Simon H. A. Report on a general problem-solving program / A. Newell, J. C. Shaw, H. A. Simon // Proceedings of the International Conference on Information Processing. – USA, 1959. – Pp. 256-264.

дачи, как доказательство теорем, игра в шашки или шахматы. В 1964 г. С. Маслов сформулировал идею метода автоматического поиска доказательства теорем в исчислении предикатов[128]. В 1966 году В. Турчин стал разработчиком алгоритмического метаязыка рекурсивных функций РЕФАЛ для описания языков и разнообразных способов их обработки[129]. В этом же году был создан первый анимационный робот «Shakey» в Стэнфордском университете[130].

Появление новых подобных программ и развитие существующих с течением времени столкнулось с проблемой недостаточности, предела имеющихся у ИИ знаний. В результате, в развитии ИИ произошли так называемые «зимы искусственного интеллекта», которые пришлись на начало 1970-х и конец 1980-х годов, когда надежды, возлагаемые на ИИ, потерпели крах, и финансирование их дальнейшего существования заметно сократилось[131]. В промежутке между периодами затишья запустили Интернет (1974 г.), а Г. Саймон получил Нобелевскую премию за теорию ограниченной рациональности, которая повлияла своим появлением на дальнейшее развитие области ИИ. После продолжительной «зимы» появляется первый персональный компьютер от IBM, человекоподобный робот от Cog, Deep Blue побеждает Г. Каспарова, появляются роботы Asimo, София и т.д. Поворотным пунктом в совершенствовании систем ИИ стало появление чат-бота GPT 4[132], новой генеративной языковой модели, способной обрабатывать как текстовые, так и визуальные данные.

Обобщая широкий круг научных работ, посвященных особенностям и эволюции ИИ, следует отметить, что упомянутая выше многозначность

---

[128]Маслов С. Ю. Обратный метод установления выводимости в классическом исчислении предикатов / С. Ю. Маслов // ДАН СССР. – 1964. – Т. 159, № 1. – С. 17-20.

[129] Турчин В. Ф. Алгоритмический язык рекурсивных функций (Рефал) / В. Ф. Турчин. – М.: Издательство ИПМ АН СССР, 1968.

[130] Shakey the Robot – SRI International. – URL: https://www.sri.com/hoi/shakey-the-robot/ (дата обращения: 12.05. 2022)

[131]Альманах «Искусственный интеллект». Анализ действующей нормативно-правовой базы // Центр компетенций НТИ на базе МФТИ по направлению «Искусственный интеллект». – 2020. – № 6. – URL: http://www.aiReport.ru (дата обращения: 12.09.2022)

[132] GPT 4 – OPEN AI. – URL: https://openai.com/product/gpt-4 (дата обращения: 02. 04. 2023)

толкования термина в них сохранилась. ИИ понимается как область знаний, технология, наука, искусственная система, совокупность метапроцедур представления знаний и т.д.

В понимании В. Архипова и В. Наумова, искусственный интеллект – это системное явление, обладающее рядом признаков: наличием встроенного технического устройства, которое способно воспринимать и передавать данные; наличием некоторой автономности; способностью анализировать, обобщать, выводить интеллектуальные решения; свойством обучаться, самостоятельно заниматься поиском информации и принимать на ее основе соответствующие решения[133].

В преамбуле Монреальской декларации искусственный интеллект характеризуется как «автономная система, способная выполнять сложные задачи, которые ранее совершались только «природным» интеллектом: обработку большого количества информации, вычисление, прогнозирование, обучение и адаптацию ответов к изменяющимся условиям, а также распознавание и классификацию объектов»[134].

Европейская комиссия по искусственному интеллекту определяет его как интеллектуальные системы, которые анализируют окружающую среду и действуют с некоторой степенью автономности[135]. Эти интеллектуальные системы могут быть представлены в виде программных продуктов, действующих в виртуальном мире (голосовые помощники, распознавание изображений, поисковые системы, распознавания речи и лиц). Кроме того, ИИ может быть встроен в аппаратные устройства (передовые

[133] Архипов В. В., Наумов В. Б. Искусственный интеллект и автономные устройства в контексте права: о разработке первого в России закона о робототехнике / В. В. Архипов, В. Б. Наумов // Труды СПИИ-РАН. – 2017. – № 6. – С. 46-62.

[134] Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).

[135] European Commission. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe. – URL: https://ec.europa.eu/digital-singlemarket/en/news/communication-artificial-intelligence-europe (дата обращения: 21.09.2022)

роботы, автономные автомобили, дроны или приложения Интернета вещей).

Достаточно обширное определение ИИ дала Л. С. Болотова. По ее мнению, ИИ можно интерпретировать как «совокупность метапроцедур — представления знаний, рассуждений, поиска релевантной информации в среде имеющихся знаний, логического вывода, пополнения знаний, их корректировки и т. д., то есть процедур, имитирующих мыслительную деятельность человека»[136]. Системы искусственного интеллекта — это «аппаратный и информационно-программный комплекс, действие которого аналогично действию механизмов человека»[137]. В данном определении утверждается, что система ИИ имеет внешнюю часть в виде аппарата и внутреннюю часть в качестве программного и информационного обеспечения. Также в определении отражено то, что системы ИИ обладают способностью к действиям и поведению, аналогичными человеческому. Но достигло ли развитие систем ИИ того уровня, когда их можно приравнять человеческому интеллекту? По нашему мнению, ошибочно представление тех разработчиков, которые пытаются при программировании имитировать человеческий разум, обойдя вопрос о сущности, природе этого разума.

Отдельные авторы пытаются определить ИИ посредством перечисления его наиболее существенных признаков. В работе И. В. Понкина и А. И. Редькиной среди характерных черт ИИ перечисляются высокая степень субстативности, автономности, способность самореферентно адаптировать собственное поведение и самообучаться, самостоятельно моделировать алгоритмы для решения проблем[138].

---

[136] Болотова Л. С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях: учебник / Л. С. Болотова. – М.: Финансы и статистика, 2012. – С.38-39.

[137] Болотова Л. С. Система поддержки принятия решений / Л.С. Болотова. – М.: Юрайт, 2019. – Часть 1. – С. 29.

[138] Понкин И. В. Искусственный интеллект с точки зрения права / И. В. Понкин, А. И. Редькина // Вестник Российского университета дружбы народов. Серия : Юридические науки. – 2018. – Т. 22, №1. – С. 91-109.

П. М. Морхат определяет ИИ как независимый комплекс программных или программно-аппаратных средств, обладающей целым рядом возможностей: способность этих систем осуществлять и демонстрировать антропоморфно-разумные мыслительные и когнитивные действия, такие, как распознавание, понимание, интерпретация и генерирование образов, символьных систем и языков, рефлексия, рассуждение, моделирование, образное (смысло-порождающее и смысло-воспринимающее) мышление, обобщение, анализ и оценка информации; способность самореферентности, саморегулирования, самоограничения, самоадаптирования под изменяющиеся условия, автономного самоподдержания себя в гомеостазе; способность самостоятельно накапливать информацию и опыт; способность автономно осуществлять генетический поиск и обработку информации; способность обучаться и самообучаться (также на своих ошибках и своём опыте); самостоятельно разрабатывать и самостоятельно применять алгоритмы самоомологации и др.[139]

Ассоциация развития искусственного интеллекта описывает ИИ как «научное понимание механизмов, лежащих в основе мышления и разумного поведения, и их воплощение в машинах»[140]. При этом уровень интеллекта в любой конкретной реализации ИИ может значительно варьироваться, и этот термин не обязательно подразумевает соответствие интеллекту человеческого уровня. ИИ включает в себя множество функций: а) глубокое обучение, трансферное обучение, обучение с подкреплением и их комбинации; б) понимание или представление глубоких знаний, необходимых для задач, специфичных для предметной области, таких как кардиология, бухгалтерский учет и юриспруденция; в) рассуждение, которое бывает нескольких разновидностей, таких как дедуктивное, индуктивное, временное, вероятностное и количественное; г) взаимодействие с

[139]Понкин И. В. Искусственный интеллект с точки зрения права / И. В. Понкин, А. И. Редькина // Вестник Российского университета дружбы народов. Серия : Юридические науки. – 2018. – Т. 22, №1. – С. 92-93.

[140] AI Overview: Broad Discussions of Artificial Intelligence, AI Topics. – URL: http://aitopics.org/topic/ai-overview. (дата обращения: 23.09.2022)

людьми или другими машинами для совместного выполнения задач и обучения в окружающей среде.

С. Рассел и П. Норвиг определяют ИИ как системы, созданные людьми, которые могут выполнять сложные задачи и обрабатывать информацию так же, как мы[141]. Этот подход был разработан в рамках функционализма, где познание представляет собой вычислительный процесс над символическими представлениями[142]. Основная цель исследований состояла в том, чтобы узнать больше о естественном интеллекте путем искусственного воспроизведения этих вычислений. Соответственно, системы ИИ были задуманы как запрограммированные для выполнения конкретных действий.

Однако, начиная с 1990-х годов, цели исследований ИИ изменились: целью стало создание интеллектуальных агентов, т. е. «сущностей, которые ощущают свое окружение и действуют в нем»[143]. В этой новой структуре ИИ не обязательно связан с алгоритмами для принятия решений. ИИ в этом случае выступает адаптивным процессом вычислений, взаимодействующим с окружающей средой для более эффективного принятия решений.[144] Другими словами, ИИ рассматривается не просто как процесс вычисления символов по определенным инструкциям, а как адаптивное и гибкое взаимодействие с окружающей средой.

Российское законодательство, а именно Указ Президента Российской Федерации от 10.09.2019 г. №490 «О развитии искусственного интеллекта в Российской Федерации», рассматривает ИИ как комплекс технологических решений, позволяющий имитировать когнитивные функции человека (в том числе самообучение и поиск решений без заранее за-

[141] Russell S., Norvig P. Artificial intelligence: International version: A modern approach / S. Russel, P. Norvig // The Knowledge Engineering Review. – Englewood Cliffs, NJ : Prentice Hall., 2010. – Vol. 11(1). – Pp.78-79.
[142] Brooks R. A. Elephants don´t play chess / R. A. Brooks // Robotics and Autonomous Systems. – 1990. – Vol. 6(1–2). – Pp. 3-15.
[143] Russell S. Rationality and intelligence: A brief update / In V. C. Müller (Ed.) // Fundamental issues of artificial intelligence. – Switzerland: Springer International Publishing, 2016.
[144]Там же.

данного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые как минимум с результатами интеллектуальной деятельности человека[145]. Этот комплекс охватывает информационно-коммуникационную инфраструктуру, программное обеспечение (в т. ч. использующее методы МО), процессы и сервисы по обработке данных, поиску решений).

Уже приведенный перечень дефиниций свидетельствует о чрезвычайно сложном, комплексном характере феномена, обозначаемого термином «искусственный интеллект», широком спектре мнений ученых о его сущности и предназначении, о принципиальных разногласиях в понимании основ и принципов взаимодействия искусственного интеллекта с сознанием, разумом, интеллектом человека. Не случайно специалисты в области ИИ, как правило, указывают на многоаспектность в толковании его сущности. Так, И. А. Филипова пишет: «Можно говорить об искусственном интеллекте как о явлении, как о группе технологий и как о научно-техническом направлении»[146].

Рассмотрев целый ряд теоретических подходов к выявлению сущности ИИ, мы пришли к выводу, что она не может быть познана, раскрыта без понимания разнообразия ИИ, без его классификации. Подобным образом рассуждали Д. Кастро и Д. Нью[147], которые, также рассмотрев различные попытки определения ИИ, пришли к мысли о том, что разница подходов к пониманию ИИ сводится к различию между «слабым» и «сильным» ИИ.

[145]Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). URL: https://base.garant.ru/72838946/ (дата обращения: 14.05.2022).

[146] Филипова И. А. Правовое регулирование искусственного интеллекта / И. А. Филипова // учебное пособие, 2-е издание, обновленное и дополненное – Нижний Новгород : Нижегородский госуниверситет, 2022. – С. 9.

[147] Castro D., New J. The Promise of Artificial Intelligence / D. Castro, J. New // Center for data innovation. – 2016. – P. 3. – URL: https://www2.datainnovation.org/2016-promise-of-ai.pdf (дата обращения: 23.09. 2022).

Действительно, большинство исследователей в области ИИ выделяют два типа ИИ[148]: слабая версия искусственного интеллекта и сильная версия искусственного интеллекта.

Слабый ИИ (или узкий ИИ) всеми специалистами определяется как система для выполнения конкретных, узкоспециализированных интеллектуальных задач. Например, А. Каплан и М. Хэнлайн характеризуют данный тип ИИ как тип аналитического интеллекта, который используется для определенных функций и приложений[149]. При этом слабый ИИ ограничен задачами, для выполнения которых обучен[150]. Но в выполнении узконаправленных задач он может превосходить присущие человеку интеллектуальные способности. Все современные системы ИИ и роботы с ИИ являются примерами того, что в современной литературе принято называть слабой версией ИИ.

Что касается сильного ИИ (или общего ИИ), то под этим термином, как правило, понимается программное обеспечение, обладающее такими же когнитивными способностями как у человека или даже превосходящими его в этом отношении. Он предназначен для решения огромного количества задач и проблем разными способами в автономном порядке. Так, по мнению П. Ванг, Б. Гёрцель и С. Франклин, общий ИИ (далее – ОИИ) – это настоящие мыслящие машины с интеллектом, подобным человеческому[151]. Сайт Search Enterprise AI определяет ОИИ как систему ИИ с обобщенными когнитивными способностями человека. При возникновении незнакомой задачи общий ИИ способен найти решение без вмешательства человека»[152].

---

[148] Шевченко А. И. К вопросу о создании искусственного интеллекта / А. И. Шевченко // Искусственный интеллект. – 2016. – № 2. – С. 7-15.

[149] Kaplan A., Haenlein M. Siri, Siri in my hand, who's the fairest in the land? On the interpretations, illustrations and implications of artificial intelligence / A. Kaplan, M. Heinlein // Business Horizons. – 2019. – Vol. 62(1). – Pp.15-25.

[150]Macnish K., Ryan M., Stahl B. Understanding ethics and human rights in smart information systems / K. Macnish, M. Ryan, B. Stahl // ORBIT Journal. – 2019. – Vol. 2(2). – Pp. 1-34.

[151] Wang P., Goertzel B., Franklin S. Artificial general intelligence, 2008: Proceedings of the first AGI conference / P. Wang, B. Goertzel, S. Franklin. – Washington DC : IOS Press, 2008. – 507 p.

[152]Search Enterprise AI. Artificial intelligence. – URL: https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence. (дата обращения: 15.05.2022)

В настоящее время общего ИИ не существует. Эту мысль в 1980 году сформулировал Дж. Сёрль в своей работе под названием «Сознание, мозг и программы». В ней, опираясь на созданный им мысленный эксперимент «Китайская комната», автор пришел к заключению, что создание общего ИИ невозможно, поскольку компьютер оперирует данными, не придавая им никакого значения и смысла, как это делает человек [153].

В то же время такие мыслители, как Р. Курцвейл[154], Б. Герцель[155] и Х. Де Гарис[156] считают, что мы вступаем в мир чрезвычайно интеллектуальных машин, в котором ОИИ будет способен непосредственно оказывать влияние на благополучие людей и сможет равным образом как помочь, так и навредить человеку. Например, представим, что ОИИ выходит из строя или попадает в руки небольших политически мотивированных террористических групп или крупных военных организаций. С помощью ОИИ они, например, обретут возможность шпионить, собирать конфиденциальную информацию о любых организациях, политиках и частных лицах, используя его в дальнейшем при решении собственных задач, нанося тем самым значительный ущерб мирному населению и государственным институтам.

Исследователь В. М. Маслов также делает вывод, что становится все больше аргументов за то, что сильный искусственный интеллект возможен[157].

К сожалению, нельзя однозначно утверждать, что эти опасения являются только плодом воображения. Сегодня мы становимся свидетелями того, что узкий ИИ уже используется странами первого мира в военных

---

[153]Серль Дж. Р. Сознание, мозг и программы / Дж. Р. Серль // Аналитическая философия: Становление и развитие: Антология / Общ. ред. и сост. А. Ф. Грязнов. – М., 1998. – 528 с.

[154] Kurzweil R. The Age of Intelligent Machines / R. Kurzweil. – Cambridge, MA: MIT Press, 1990. – 565 p.

[155] Goertzel B. Should humanity build a global AI nanny to delay the singularity until it's better understood? / B/ Goertzel // Journal of consciousness studies, 2012. – Vol.19. – Pp. 96-111.

[156] De Garis H. The Artilect War: A bitter controversy concerning whether humanity should build godlike massively intelligent machines / Eds. P. Wang, B. Goertzel, S. Franklin // Artificial general intelligence, 2008 : Proceedings of the first AGI conference. – Washington DC : ISO Press, 2008. – Pp. 362-373.

[157]Маслов В. М. Высокие технологии и феномен постчелочевеческого в современном обществе. автореферат диссертации на соискание ученой степени кандидата философских наук / В. М. Маслов. – Нижний Новгород, 2014. – 38 с.

целях. Речь идет, в частности, о беспилотных летательных аппаратах (дронах) Northrop Grumman X-47B, которые проходят испытания ВМС США[158]. Управление этой системой, выбор способа выполнения задачи полностью поручены не человеку, а узкому ИИ - дрону. В настоящее время неизвестно, будут ли подобные системы развиваться до уровня общего интеллекта. Тем не менее, по свидетельству Рональда К. Аркина, военные США уже проявили интерес к производству интеллектуальных систем, предназначенных для убийства[159].

Х. де Гарис в работе «Война за артилекты» задается вопросом о том, должны ли люди создавать ОИИ, и чем в итоге это может для них обернуться[160]. Данный вопрос будет доминировать в глобальной политике XXI века и, по мнению автора, положительный ответ на него может привести к крупной войне, в результате которой погибнут миллиарды людей»[161].

Р. Курцвейл, в свою очередь утверждает, что в связи с применением ИИ люди будут постепенно превращаться в киборгов, и четкая грань между машинами и людьми исчезнет. ОИИ может заменить врачей и других специалистов, выгодно отличаясь тем, что подобные системы не будут уставать, требовать длительной подготовки и при этом будут совершать меньше ошибок, чем живые люди. Как результат, в будущем во всех уголках мира доступная медицинская помощь станет нормой.

Экспоненциальный прогресс в технологиях обеспечит своеобразный интеллектуальный взрыв, результатом которого станет появление Отборного ИИ (Seed AI). Этот ИИ, прогнозирует Р. Курцвейл, будет способен модифицировать свою собственную программу, создавать более ра-

---

[158] Defense Tech. Navy's second stealthy X-47B drone flies. Defense Tech. org. – URL: http://defensetech.org/2011/11/28/second-x-47buav-flies/#more-15485 (дата обращения: 14.12. 2021)

[159] Arkin R. C. Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture / Eds. P. Wang, B. Goertzel, S. Franklin // Artificial general intelligence, 2008: Proceedings of the first AGI conference. – Washington DC : ISO Press, 2008. – Pp.51-62.

[160] De Garis H. The Artilect War: A bitter controversy concerning whether humanity should build godlike massively intelligent machines. Artificial general intelligence, 2008 // Proceedings of the first AGI conference / Eds. P. Wang, B. Goertzel, S. Franklin. – Washington DC : ISO Press, 2008. – Pp. 362- 373.

[161] Там же. С 440.

зумное собственное «я». Обновленная версия ИИ будет еще лучше программироваться, что, в свою очередь, позволит ей создавать еще более умные обновления[162].

Безусловно, развитие ИИ будет продолжаться, и результаты его уже сегодня являются одним из важнейших предметов обсуждения ученых во всем мире. Но в своей работе мы рассматриваем различные подходы к пониманию ИИ, учитывая существующий уровень развития его систем. Поскольку в настоящее время дебаты об отборном, универсальном, общем и т.п. интеллекте ведутся исключительно в сфере прогнозов и предсказаний, мы приходим к выводу о необходимости использовать самую общую классификацию ИИ, которая подразделяет множество существующих видов ИИ на сильный ИИ и слабый (узкий) ИИ. При этом сильный ИИ все еще остается спекулятивной темой, а объектом научного познания, по нашему мнению, сегодня может выступать лишь ИИ, созданный человеком для решения определенных практических задач, т.е. узкий ИИ.

Узкий ИИ может быть охарактеризован как группа технологий, как результат применения человеком научного знания для решения разнообразных практических задач. При этом большинство исследователей подчеркивают системный характер узкого ИИ. В частности, И.А. Филипова пишет: «Сегодня искусственным интеллектом признается полностью или частично автономная самоорганизующаяся система, обладающая возможностями мыслить, обучаться, самостоятельно принимать решения. …искусственная интеллектуальная система является программно-аппаратным комплексом, она включает аппаратное и программное обеспечение. Аппаратное обеспечение охватывает все физические части компьютера или машины – носителя искусственного интеллекта, то есть электронные и механические части, входящие в состав системы. Про-

[162]Singularity Institute for Artificial Intelligence, Seed AI. General Intelligence and Seed AI. – URL: http://singinst.org/ourresearch/publications/GISAI/paradigms/seedAI.html#glossary_crystalline (дата обращения: 12.09.2021)

граммное обеспечение включает программы, используемые для управления машиной. Компьютерщики нередко называют эти части «хард» и «софт». Проще говоря, искусственная интеллектуальная система – это компьютер, способный выполнять функции, ранее свойственные только человеку»[163].

Другими словами, узкий ИИ – это системы, элементами которых являются аппаратный комплекс, программное обеспечение, набор данных. Аппаратный комплекс — это физические устройства, которые обеспечивают работу программ. Программное обеспечение представляет собой набор из компьютерного кода и связей между его частями, который функционирует благодаря аппаратному комплексу. Набор данных включает в себя входные и выходные данные, при этом последние являются результатом обработки входных данных и существуют в виде решения и действия системы ИИ.

Системы ИИ могут быть представлены как в виде программ, работающих в виртуальном пространстве (например, программа для распознавания изображений или текста), так и в виде аппаратных устройств с программным обеспечением (например, беспилотный автомобиль, робот-уборщик). Подчеркнем еще раз, что все современные системы ИИ и роботы с ИИ являются примерами узкого ИИ.

На системный характер узкого ИИ указывает в своем определении экспертная группа высокого уровня Европейской комиссии, которая сформулировала следующую дефиницию ИИ. «Системы искусственного интеллекта (ИИ) – это программные (и возможно, также аппаратные) системы, разработанные людьми, которые, имея сложную цель, действуют в физическом или цифровом измерении, осуществляя сбор данных об окружающей их среде, интерпретируя собранные и структурированные или

[163] Филипова И. А. Правовое регулирование искусственного интеллекта / И. А. Филипова // учебное пособие, 2-е издание, обновленное и дополненное – Нижний Новгород: Нижегородский госуниверситет, 2022. – С. 9.

неструктурированные данные в виде рассуждений, основанных на ранее полученном знании или результатах обработки вновь поступившей информации и способные к принятию решения о наилучшем способе действия с точки зрения достижения поставленной человеком цели. Подобные системы искусственного интеллекта способны следовать заданным разработчиком правилам, сконструировать цифровую модель того или иного процесса либо явления, а также могут изменить своё собственное поведение, адаптируя его к изменившимся вследствие ранее совершенных действий условиям окружающей среды»[164]. Указанную дефиницию можно найти в отдельном документе, подготовленном экспертной группой и опубликованном под названием «Определение ИИ: основные возможности и дисциплины»[165].

Именно это определение, по нашему мнению, можно назвать наиболее корректным, поскольку оно содержит указание на основные свойства и способности ИИ, отличающие его от любых других предметов и явлений социальной реальности, отражая тем самым глубинную сущность этого феномена.

Отметим, что в отчете специального межправительственного комитета экспертов по искусственному интеллекту Совета Европы (CAHAI) именно это определение ИИ признано наиболее точным[166]. Однако данное определение нельзя считать окончательным, поскольку технические науки и технология ИИ в современном мире находятся в постоянном развитии.

Часто возникает путаница в понятийном аппарате из-за отождествления ИИ и методов, способов, алгоритмов, математических моделей, создаваемых для обеспечения его рабочих функций. Речь, в частности,

[164] High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.)
[165] A definition of Artificial Intelligence: main capabilities and scientific disciplines. – URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341 (дата обращения 26.08.2022)
[166] CAHAI. Draft analysis of the Multi-Stakeholder Consultation. – URL: https://rm.coe.int/cahai-cog2021-01-draft-report-msc-2763-7501-0051-v-1/1680 a2e2d5 (date of application 18.07.2022).

идет о машинном обучении и искусственной нейронной сети. Машинное обучение (далее – МО) — это группа методов (обучение с учителем, обучение без учителя, обучение с подкреплением), которые обеспечивают анализ данных программой для принятия ею решений. Посредством указанных методов осуществляется решение отдельных интеллектуальных задач, таких как сбор и анализ данных, построение выводов и др.

Искусственная нейронная сеть - это математическая модель, созданная для анализа обработки данных и принятия на их основе решений при использовании методов МО. Создатели искусственной нейронной сети вдохновлялись биологическими нейронными сетями, поэтому она построена из искусственных нейронов (процессоров). Они делают «интеллектуальным» программное обеспечение ИИ и выступают в роли методов и инструментов достижения системами ИИ желаемого результата. Сущность систем ИИ выражена в полной мере в представленным выше определении.

Таким образом, искусственный интеллект, получивший широкое распространение в современном обществе, существует в форме его слабой (или узкой) версии, предназначенной для выполнения строго определенных, поставленных человеком задач. Его отличает системный характер, обеспечивающий целостность взаимодействующих компонентов, представленных программным и информационным обеспечением, а также аппаратными устройствами. Он способен осуществлять сбор данных, их интерпретацию в виде рассуждений, принимать решение на основании имеющейся информационной базы о возможном и наилучшем способе решения поставленной перед ним задачи.

1.2 Основные сферы применения и особенности исследования ИИ

Изучение систем ИИ, уже нашедших применение в современном обществе, неизбежно ставит вопрос о наиболее актуальных сферах их внедрения и использования, результатах, проблемах, рисках и перспективах развития указанных систем. Ученые проводят самостоятельные ис-

следования или опираются на известные им и описанные в литературе эмпирические данные. Несмотря на обилие информации в этом вопросе в настоящее время по-прежнему не сложилось единого мнения, поскольку автоматизация производственных процессов, стремительный рост информационного оборота, инвестиций в эту сферу расширяют область применения ИИ практически ежедневно.

Так, в работе, подготовленной Специальным межправительственным комитетом экспертов по искусственному интеллекту Совета Европы (CAHAI)[167], обсуждаются следующие области социальной жизни, в которых применение систем ИИ получило широкое распространение: государственное управление, транспорт, сельское хозяйство, контроль окружающей среды и климатических изменений, здравоохранение, образование, правосудие, банковская, финансовая и страховая деятельность, военная сфера и др.

В статье Б. К. Шталя, Дж. Антониу, М. Райана[168], посвященных исследованию этических проблем применения ИИ, представлен следующий перечень отраслей жизни социума, отмеченных активным использованием систем ИИ: контроль и администрирование сотрудников, работа правительства, сельское хозяйство, обеспечение устойчивого развития окружающей среды, наука, страхование, энергетика и коммунальные услуги, коммуникации, СМИ и развлечения, розничная и оптовая торговля, производство и добыча природных ресурсов.

В другом источнике М. Райан, Дж. Антониу, Л. Брукс выделяют банковское дело, ценные бумаги, здравоохранение, страхование, розничную и оптовую торговлю, науку, образование, энергетику и коммунальные услуги, производство и добычу природных ресурсов, сельское хозяйство, коммуникации, СМИ и развлечения, транспорт, контроль и админи-

[167]Специальный межправительственный комитет экспертов по искусственному интеллекту (CAHAI). – URL: https://sk.ru/media/documents/CAHAI_AI_Research.pdf (дата обращения: 19. 11. 2022)
[168] Stahl B. C., Antoniou J. et al. Organisational responses to the ethical issues of artificial intelligence / B. C. Stahl, J. Antoniou, M. Ryan, K. Macnish, T. Jiya // AI & Society. – 2022. – Vol. 37. – Pp. 23-37.

стрирование сотрудников, обеспечение работы правительства, правоохранительных органов и правосудия, устойчивое развитие окружающей среды, оборону и национальную безопасность[169].

И. А. Асеева, также обсуждая этические проблемы применения ИИ, указывает на бизнес, здравоохранение, регулирование городской среды, государственные услуги, страхование и статистику, индустрию развлечений, маркетинг[170].

С. И. Конев, рассматривающий этико-правовые проблемы, выделяет такие области использования ИИ как медицина, сельское хозяйство, транспорт, юриспруденция, повседневный быт[171].

Проведенное нами исследование позволило, в частности, сделать вывод о том, что одной из важнейших сфер социальной жизни, в которой успешно находят применение системы ИИ, является транспорт. Здесь тестируют свои беспилотные автомобили крупные западные компании Google, General Motors, Tesla, BMW, Ford. Умное освещение дорожного покрытия, обеспеченное системами ИИ, может помимо выполнения основной функции анализировать состояние дорог, оказывать помощь в регулировании дорожного движения, предотвращая возникновение «пробок» на наиболее сложных и востребованных участках трассы. Беспилотные роботы-такси и автобусы разрабатываются и тестируются для удовлетворения нужд людей, использующих общественный транспорт. Уже доступны или разрабатываются процессоры, способные с помощью соответствующих датчиков следить за тем, что происходит рядом с автомобилем в режиме реального времени, определять местоположение автомоби-

[169] Ryan M., Antoniou J. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality / J. Antoniou, M. Ryan, L. Brooks, T. Jiya, K. Macnish, B. Stahl // Science and Engineering Ethics. – 2021. – Vol.27(2). – P. 16.

[170] Асеева И. А. Искусственный интеллект и большие данные: этические проблемы практического использования. (Аналитический обзор) / И. А. Асеева // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 8: Науковедение. – 2022. – № 2. – С. 89-98.

[171] Конев С. И. Этико-правовые проблемы регулирования искусственного интеллекта и робототехники в отечественном и зарубежном праве / С. И. Конев. – URL: https://cyberleninka.ru/article/n/etiko-pravovye-problemy-regulirovaniya-iskusstvennogo-intellekta-i-robototehniki-v-otechestvennom-i-zarubezhnom-prave (дата обращения: 19.09. 2022)

ля на карте и планировать маршрут, динамически адаптируясь к трафику. Программа European Truck Platooning Challenge запустила парк автономных грузовиков, который успешно преодолел более 2000 километров в режиме взвода. Они отслеживали скорость друг друга, поддерживали оптимальное расстояние между машинами с тем, чтобы максимально эффективно пройти по заданному маршруту движения[172].

В сфере образования большая часть технологий ИИ пока задействована для проверки посещаемости занятий и выполнения учениками заданий, оценки и анализа экзаменационных ответов, составления персональных планов обучения. IBM разработала Teacher Advisor, который помогает учителям математики третьего класса в США разрабатывать персонализированные планы уроков[173]. Эта система позволяет учителям адаптировать учебные материалы для учащихся одного класса, но с разным уровнем навыков. В будущем планируется увеличить количество предметных областей и классов, с которыми она может помочь в организации и проведении занятий.

Школьный округ Такома, штат Вашингтон совместно с Microsoft разработал модель машинного обучения для анализа данных учащихся. Система, используя данные об успеваемости, демографические и биографические сведения, строит прогнозы о том, кто из учащихся с большой долей вероятности может бросить учебу и предлагает профилактические меры воздействия на учеников и их окружение с целью недопущения негативного результата[174]. После многолетнего использования пилотной системы школьный округ Такома смог увеличить количество выпускников с 55% в 2010 году до 78% к 2014 году[175].

---

[172] Limer E. A Fleet of Self-Driving Trucks Just Completed a 1,000-Mile Trip Across Europe / E. A. Limer. – URL: http://www.popularmechanics.com/cars/trucks/a20310/european-platooning-challenge-self-driving-trucks-1000-miles/ (дата обращения: 11.07.2022)

[173] Harris E. Next Target for IBM's Watson? Third-Grade Math / E. Harris. – URL: http://www.nytimes.com/2016/09/28/nyregion/ibm-watson-common-core.html (дата обращения: 11.02.2022)

[174] ML Predicts School Dropout Risk & Boosts Graduation Rates. – URL: https://blogs.technet.microsoft.com/machinelearning/2015/06/04/ml-predicts-school-dropout-risk-boostsgraduation-rates/ (дата обращения: 16.08.2022)

[175] Там же.

Технологический институт Джорджии запустил автоматизированного помощника преподавателя Джилл, работающего на платформе когнитивных вычислений IBM Watson. Джилл помогает отвечать на различные запросы студентов в отношении онлайн-курсов (например, где найти материалы курса) и получает в среднем около 10 000 сообщений от студентов в течение семестра[176].

Компания Duolingo, занимающаяся разработкой программного обеспечения для освоения иностранных языков, использует МО для анализа активности пользователей и оценки прогресса в их обучении. Она разрабатывает персонализированные планы уроков, а также регулярно тестирует новые стратегии повышения их эффективности[177]. Duolingo структурирует планы уроков таким образом, чтобы каждый пользователь мог узнать, как ему улучшить собственные результаты и реализовать наиболее эффективные варианты обучения иностранным языкам.

В области энергетики IBM разработала систему машинного обучения SMT или «Самообучающаяся модель погоды и технология прогнозирования возобновляемых источников энергии». Этот инструмент предназначен для анализа данных с 1600 метеостанций, солнечных электростанций, ветряных электростанций и метеорологических спутников. На основе этих данных он генерирует прогнозы погоды на 30% точнее, чем Национальная Weather Service и прогнозирует доступность возобновляемой энергии на несколько недель вперед[178].

Окриджская национальная лаборатория Министерства энергетики США создала инструмент Autotune для суперкомпьютера Titan. Система использует МО для создания высокодетализированных моделей энерго-

[176] Maderer J. Artificial Intelligence Course Creates AI Teaching Assistant / J. Maderer. – URL: http://www.news.gatech.edu/2016/05/09/artificial-intelligence-course-creates-ai-teaching-assistant. (дата обращения: 12.03.2022)

[177] Goliani P. Duolingo Looks to Dominate the Mobile Education Market With New Flashcard App TinyCards / P. Goliani. – URL: http://www.forbes.com/sites/parulguliani/2016/07/22/duolingo-looks-to-dominate-the-mobile-education-market-with-new-flashcard-app/ (дата обращения: 17.09.2022)

[178] Mearian L. IBM’s Machine-Learning Crystal Ball Can Foresee Renewable Energy Availability / L. Mearian. – URL: http://www.computerworld.com/article/2948987/sustainable-it/ibms-machine-learning-crystal-ball-can-foresee-renewable-energy-availability.html. (дата обращения: 12.09.2022)

эффективного здания[179]. Инструмент контролирует около 150 параметров, влияющих на энергоэффективность, таких, например, как освещение и вентиляция, и анализирует, как оптимизировать эти факторы, чтобы использование электроэнергии в здании оставалось максимально эффективным.

Google внедрила подобное программное обеспечение искусственного интеллекта, разработанное Deep Mind, дочерней компанией Alphabet, для автоматической оптимизации энергоэффективности[180]. Система постоянно отслеживает 120 переменных и учится, как лучше настроить производительность оборудования и системы охлаждения, чтобы обеспечить высокий уровень эффективности работы центра обработки данных. В результате использования указанной системы энергопотребление центра обработки данных снизилось на 15 %.

Еще один успешный пример использования систем ИИ в сфере экологии – проект Orbital Insight, предназначенный для анализа спутниковых изображений лесов. Сравнивая снимки «до» и «после», система способна обнаруживать признаки незаконных рубок, отмечая те изменения, которые могут остаться вне поля зрения людей, отвечающих за сохранность лесного массива. Например, появление новых дорог для осуществления лесозаготовительных операций[181].

В мегаполисах системы, нацеленные на решение конкретных задач, также находят успешное применение. Исследователи IBM, например, спроектировали в Пекине систему машинного обучения для анализа данных об уровне загрязнения воздуха в городе. Уже сегодня система может прогнозировать изменение качества воздуха в течение ближайших трех

[179] ORNL Researchers Develop ‘Autotune’ Software to Make It Quicker, Easier, and Cheaper to Model Energy Use of Buildings. – URL: http://energy.gov/eere/buildings/articles/ornl-researchers-develop-autotune-software-make-it-quicker-easier-and. (дата обращения: 19.02.2022)

[180] Clark J. I’ll Be Back: The Return of Artificial intelligence / J. Clark. – URL: http://www.bloomberg.com/news/articles/2015-02-03/ill-be-back-the-return-of-artificial-intelligence (дата обращения: 12.01. 2022)

[181] Peters A. This AI Watches Satellite Photos and Says ‘It Looks Like You’re About to Cut Down a Forest. Could You Not? / A. Peters. – URL: https://www.fastcoexist.com/3046014/this-ai-watches-satellite-photos-and-says-it-looks-like-youre-about-to-cut-down-a-forest-cou. (дата обращения: 12.02. 2022)

суток на 30% точнее по сравнению с результатами, полученными традиционными способами[182]. Проект продолжают совершенствовать для прогнозирования изменения количества водителей на дорогах столицы, причем таким образом, что результаты будут сохранять свое значение на срок до 10 дней.

В области государственной службы и охраны правопорядка была протестирована программа Series Finder, выявляющая на основании анализа шаблона преступлений личности потенциальных преступников и прогнозирующая тем самым возможность возникновения противоправных деяний в будущем[183]. Так, для раскрытия уже совершенных краж и предсказания новых программа использовала данные отдела анализа преступности Кембриджского полицейского управления (КПУ), позволившие ей выявить своего рода modus operandi преступника, т.е. тип поведения, набор привычек, которые отличают правонарушителя. На основании полученных девяти поведенческих моделей, начав буквально с пары преступлений, Series Finder смогла восстановить большинство совершенных преступлений, зарегистрированных КПУ. Программа также выявила еще девять преступлений, укладывающихся в рамки созданных ею моделей, о которых КПУ ранее ничего не было известно.

В сфере государственной службы широкое применение получили системы распознавания изображений. Однако, как только точность распознавания лиц преодолела рубеж в 94%, в обществе возникли серьезные опасения в связи с опасностью нарушения конфиденциальности, вторжения в частную жизнь и возможной дискриминации прав граждан[184] как неизбежных последствий использования данных технологий.

---

[182] Knight W. How Artificial Intelligence Can Fight Air Pollution in China / W. Knight. – URL: https://www.technologyreview.com/s/540806/how-artificial-intelligence-can-fight-air-pollution-in-china/ (дата обращения: 11.03.2022)

[183] Rudin C., Sloan M. Predictive policing: using machine learning to detect patterns of crime / C. Rudin, M. Sloan. – URL: https://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime/ (дата обращения: 12.01.2022)

[184] Snow J. Google's New AI Smile Detector Shows How Embracing Race and Gender Can Reduce Bias / J. Snow // MIT Tecnology Review. – 2017. – P.189.

В США в сфере обеспечения общественной безопасности используется система Shot Spotter для обнаружения выстрелов из огнестрельного оружия. Система использует сетевые аудиосенсоры, разбросанные по городским кварталам, и МО для автоматического определения звуковых сигнатур выстрелов. Shot Spotter с высокой степенью точности сообщает полиции о стрельбе и районе, в котором она регистрируется[185]. Алгоритмы Shot Spotter могут различать выстрелы, определять координаты места, откуда произошел выстрел, а также устанавливать, было ли задействовано несколько видов огнестрельного оружия, и в каком направлении двигался стрелок[186].

Hitachi тестирует программное обеспечение в нескольких городах США, которое использует обработку естественного языка и МО для анализа сотен точек данных, таких как активность в социальных сетях, местоположение звонков в службу экстренной помощи, близость постов с геотегами к находящимся поблизости школам для создания тепловых карт районов с повышенным уровнем криминальной активности[187]. Данное программное обеспечение умеет, помимо прочего, анализировать общедоступную геотегированную активность в социальных сетях с тем, чтобы идентифицировать фразы, соответствующие им ключевые слова, которые с большой долей вероятности являются закодированными отсылками к наркотикам, процессам их распространения, к людям, занятым в этой противоправной деятельности.

Системы ИИ сегодня используются даже в творческих процессах. Например, при создании произведений живописи или художественной литературы. Так, проект Sketch-RNN[188] на самообучающейся нейросети

[185] Shot in the Dark: New Surveillance Tool Called ShotSpotter Traces and Records Incidents of Gunfire. – URL: https://www.sciencedaily.com/releases/2016/04/160416130850.html (дата обращения: 09.09.2022)
[186] Rozee M. The NYPD Can Now Pinpoint the Exact Location of a Gunshot / M. Rozee. – URL: http://gothamist.com/2015/03/17/nypd_pinpointing_gunshots.php (дата обращения: 12.09.2022)
[187] Captain S. Hitachi Says It Can Predict Crimes Before They Happen (September 28, 2015) / S. Captain. – URL: https://www.fastcompany.com/3051578/elasticity/hitachi-says-it-can-predict-crimes-before-they-happen (дата обращения: 12.04. 2022)
[188] David H., Douglas E. A Neural Representation of Sketch Drawings / H. David, E. A. Duglas. – URL: https://arxiv.org/pdf/1704.03477.pdf (дата обращения: 16.09.2022)

может рисовать эскизы различных предметов. Sketch-RNN распознает то, что нарисовано человеком и пытается повторить этот рисунок, но не как точную копию. Опираясь на эскиз, приложение создает свой оригинальный рисунок. Есть даже случаи, когда картины, написанные ИИ, было невозможно отличить от тех, которые создал человек[189]. Японский робот-писатель, в свою очередь, написал роман, который стал финалистом литературной премии имени Хоси Синъити[190]. Работа оказалась действительно хорошо структурированной, интересной и заслужила высокую оценку читателей.

Перечень подобных примеров использования систем ИИ может быть продолжен, поскольку область их применения в современном мире постоянно расширяется. Некоторые результаты рассмотрены нами в опубликованных по теме диссертации статьях.

Обширный перечень областей применения систем ИИ, свидетельствующий о фундаментальном встраивании их в жизнь современного социума, все возрастающем влиянии на общество и людей, их привычные формы бытия, сложившиеся мировоззренческие системы, ценностные установки, сам образ жизни человека в мире, обусловил невероятный интерес к исследованию систем ИИ со стороны самых разных специалистов, организаций, социальных и профессиональных групп.

Специфической особенностью исследования ИИ стало своеобразие источниковой базы его изучения. Наряду с научными статьями, монографиями, диссертационными исследованиями, проблемы разработки, внедрения и применения систем ИИ рассматриваются в документах, содержащих этические и правовые требования и принципы регулирования систем ИИ в самых разных областях жизни социума.

---

[189] Ефимова Е. Новое слово в живописи: искусственный интеллект «пишет» картины в уникальном стиле / Е. Ефимова. – URL: http://www.vesti.ru/doc.html?id=2905726 (дата обращения: 16.09.2022).

[190] Clark B. While Microsoft's Tay was being racist, an AI entered a writing contest — and nearly won/ B. Clark. – URL: https://thenextweb.com/insider/2016/03/24/while-microsofts-tay-was-being-racist-an-ai-entered-a-writing-contest-and-nearly-won/#gref (дата обращения: 16.09.2022).

Как объяснить появление такого источника, как многочисленные этические рекомендации, кодексы, стандарты, как нормативно-правовые акты, регламентирующие сегодня применение систем ИИ в мире?

По нашему мнению, ответ заключается в том, что, во-первых, системы ИИ строятся по примеру интеллекта человека. Они, по сути, представляют собой искусственный человеческий разум, т.е. своего рода аналог важнейшего, основного признака человека, его родовой сущности. С древних времен человек определялся как разумное существо, наделенное сознанием, способное к целенаправленной деятельности, основанной на процессах абстрактного мышления. Именно разум позволил человеку выделиться из природного мира и стать единственным существом на Земле, которое не просто приспосабливается к миру, но и приспосабливает мир к удовлетворению собственных потребностей, внося в мироздание порой принципиальные изменения.

В то же время разумность человека неотъемлема от его природно-биологической составляющей. Он подвержен многочисленным слабостям, страстям, болезням. Он устает, его дееспособность может снижаться, и он смертен. Кроме того, человек живет в социуме, также накладывающем на его жизнедеятельность многочисленные и разнообразные ограничения в виде этических, правовых, политических, социальных, ценностных требований, установок и законов, что также способно оказывать сдерживающее влияние на результаты его интеллектуальной и природопреобразующей деятельности.

Когда мы говорим об искусственном интеллекте, то все перечисленные выше препятствия отсутствуют, поскольку ИИ – это не естественный интеллект. Ему неведомы проблемы, с которыми сталкивается любой биологический организм. Он не болеет, не устает, не ошибается и не умирает. Для него не существует преград в виде призраков человеческого рода, пещеры, рыночной площади или театра, о которых писал Ф. Бэкон. Он быстро и эффективно обучается всему новому, уже сегодня

решает сложнейшие для обычного человека задачи (например, играет в шахматы на уровне гроссмейстера). Не удивительно, что его столь быстрый прогресс вызывает обеспокоенность и тревогу со стороны человека. Не станем ли мы по истечении непродолжительного отрезка времени неэффективными для систем ИИ? Не коснется ли проблема «лишних» людей всего человечества, всех представителей рода людского?

Иначе говоря, внедрение и совершенствование систем ИИ уже сегодня представляется не просто результатом эволюции, новым достижением человека. Оно вызывает трепет, беспокойство, страх, оно грозит стать уже в ближайшем будущем настоящей экзистенциальной угрозой человечеству, которая может означать пересмотр привычного для человека места в мире, смещение его на самый край, на периферию современного мира, освобождение человеком своего центрального положения в мире для быстро прогрессирующих «умных машин».

Возможно, стремясь избежать подобного финала, человечество заранее стремится не допустить столь нежелательного результата посредством создания многочисленных этических кодексов, требований, норм, руководств для разработчиков систем ИИ, для тех, кто занимается их внедрением и эксплуатацией. Человек надеется, что это поможет ему сохраниться как виду и сохранить привычное для него место в мире.

Ответом на обсуждение социально-философских проблем внедрения ИИ стали попытки установления неких правил, принципов, регулирующих его использование. С этой целью различные научно-исследовательские, политические, профессиональные организации, комитеты национального уровня, неправительственные и частные коммерческие компании разработали многочисленные законодательные акты, этические руководства и стандарты, решения этических комитетов и множество других аналогичных по смыслу документов.

Рассматривая эту группу источников, мы включили в нашу работу, в первую очередь, манифесты крупных политических образований (ЕС) и

профессиональных организаций (IEEE, Google), национальные стратегии государств, отдельные законодательные акты. Указанные документы, разработанные на основе научных исследований, носят предписывающий характер, созданы для специалистов, практически работающих с системами ИИ и лиц, принимающих решения и управляющих процессами создания и применения ИИ. Следовательно, их потенциальное влияние на дальнейшее развитие систем ИИ может быть чрезвычайно высоким. Источником рассмотренных документов являются правительственные структуры отдельных государств (например, Специальный комитет лордов Великобритании, Экспертная группа высокого уровня Европейской комиссии), авторитетные в научном сообществе организации (например, Future of Life Institute, IEEE, AI4People), компании, представляющие крупный международный бизнес (например, Google, IBM, Microsoft, Intel).

Источниковую базу исследования составили также нормативно-правовые акты, этические стандарты и руководства по регулированию систем ИИ, резолюции, международные декларации по внедрению этических систем ИИ, периодические издания.

Нормативно-правовые акты представлены международно-правовыми актами, законодательными актами отдельных государств, среди которых, например, «Закон о дорожном движении Германии»[191], Проект Федерального закона «О внесении изменений в Гражданский кодекс Российской Федерации в части совершенствования правового регулиро-

[191] Закон о дорожном движении Германии от 16.06.2017 (Strassen verkehrsgesetz). – URL: https://www.bundesregierung.de/breg-de/suche/automatisiertes-fahren-auf-dem-weg-326108 (дата обращения: 16.09.2022)

вания отношений в области робототехники»[192], «Закон об автоматизированных и электрических транспортных средствах»[193].

Отдельную группу документов, имеющих регламентирующий характер, составляют Резолюция о запрете применения автономных смертельных систем вооружения[194](Бельгия, 2018), Рекомендация по беспилотным автомобилям от Комиссии по этике при Министерстве транспорта и цифровой инфраструктуры Германии[195], Доклад об этике робототехники от ЮНЕСКО[196].

Появление такого рода источников, это ответ человеческого сообщества на еще одно важнейшее свойство систем искусственного интеллекта, отличающее их от всех ранее созданных человеком технологий: их способность действовать автономно.

Системы ИИ уже сегодня могут выполнять сложные интеллектуальные задачи (например, управление автомобилем) без активного вовлечения человека. Создание более универсальных продуктов ИИ, основанных на машинном обучении, со временем неизбежно приведет к появлению еще более автономных, а значит еще более неподконтрольных человеку систем. Эта особенность превращает системы ИИ в потенциальный источник экзистенциальной угрозы для человечества в целом, которая будет постоянно нарастать вследствие того, что системы ИИ в перспективе смогут совершенствовать собственное программирование и, как ре-

[192] Проект Федерального закона «О внесении изменений в Гражданский кодекс Российской Федерации в части совершенствования правового регулирования отношений в области робототехники» / Проект Д. С. Гришина (Grishin Robotics) (не вносился в Госдуму) // URL://https://www.dentons.com/ru/insights/alerts/2017/january/27/dentons-develops-first-robotics-draft-law-in-russia (дата обращения: 20.08.2022).

[193]Automated and Electric Vehicles Act. – URL: https://www.legislation.gov.uk/ukpga/2018/18/contents/enacted (дата обращения: 03.06.2022).

[194] Résolution visant à interdire l'utilisation, par la Défense belge, de robots tueurs et de drones armés // Chambre des représentants de Belgique. – URL: https://www.lachambre.be/FLWB/pdf/54/3203/54K3203005.pdf (дата обращения: 21.09.2022)

[195]Ethics Commission Automated and Connected Driving. URL: https: https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile (дата обращения:18.03.2022)

[196] Report of COMEST on robotics ethics. – URL: https://unesdoc.unesco.org/ark:/48223/pf0000253952 (дата обращения: 31.11.2022)

зультат, стать обладателями когнитивных способностей, превышающих познавательные возможности человека.

Безусловно, этот крайне негативный футуристический сценарий больше похож сегодня на сюжет фантастического рассказа. Однако и без такого рода прогнозов очевидно, что технологии ИИ требуют тщательного контроля и регулирования уже сейчас.

Поэтому не удивительно, что возможности ИИ и проблемы, создаваемые им, привели к появлению целого ряда нормативных и регламентирующих документов, этических руководств, стандартов, законодательных актов, деклараций и т.д. в самых разных отраслях и сферах общественной жизни. Мировое сообщество в лице правительств, международных организаций, научно-исследовательских институтов, бизнес-корпораций отреагировало таким образом на проблемы и риски, порождаемые ИИ, сформулировало свой ответ на вызов, брошенный ИИ глобальному человечеству.

Общее количество публикаций, посвященных регулированию проблем применения систем ИИ, постоянно растет. При этом большая часть существующих в настоящее время документов вышли в свет после 2016 года. В. Дигнум приводит результаты исследования, свидетельствующего о том, что в настоящее время выпущено уже около 600 политических рекомендаций, руководств или стратегических отчетов, связанных с этическим регулированием систем ИИ[197].

В настоящем параграфе своего исследования мы изучили обширный свод документов по этическому и правовому регулированию систем ИИ, существующих в мировой практике, и отобрали из них девяносто наиболее, на наш взгляд, значимых и содержательных. Перечень отобранных нами документов представлен в **Приложении** к диссертации**.** Документальные источники в этот список отбиралась на основе критери-

[197] Dignum V. Responsible Artificial Intelligence – from Principles to Practice / V. Dignum. – URL: file:///C:/Users/oem/Downloads/Responsible_Artificial_Intelligence_--_from_Princi.pdf (дата обращения: 12.09.2021)

ев надежности, новизны и разнообразия. Большинство документов было подготовлено частными компаниями и государственными учреждениями, за ними по мере убывания следуют академические организации, научно-исследовательские институты, межгосударственные или наднациональные организации, профессиональные ассоциации и т.д.

Попытки провести аналогичную работу предпринимались ранее другими исследователями. Отметим здесь обзор в Minds and Machines, в которых Хагендорф[198] изучил 22 руководящих принципа, и публикацию в Nature, в которой А. Джобин, М. Йенка и Е. Вайена[199] исследовали 84 источника руководящих этических принципов ИИ.

По нашему мнению, исследование принятых и действующих в мире документов, регламентирующих, упорядочивающих применение систем ИИ – необходимый шаг для понимания того, что в действительности происходит в сфере внедрения и применения ИИ, какие риски, опасности, трудности и препятствия возникают в процессах взаимодействия человека с ИИ, каковы перспективы развития систем ИИ в ближайшем и отдаленном будущем.

Подводя итоги параграфа, можно сделать вывод о перманентном расширении сфер применения систем ИИ. Последние уже сегодня активно используются в государственном управлении, на транспорте, в сельском хозяйстве, здравоохранении, образовании, правосудии, в банковской, финансовой и страховой деятельности, в военном деле и многих других областях деятельности человека. Столь значительный перечень говорит о глубоком проникновении ИИ в жизнь современного социума, все возрастающем влиянии его на общество и людей. В настоящее время не представляется возможным в силу перманентного совершенствования ИИ дать полный и исчерпывающий ответ о границах применения искус-

[198] Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines / T. Hagendorff // Minds & Machines. –2020. – Vol. 30. – Pp. 99-120.
[199] Jobin A., Ienca M., Vayena E. Artificial Intelligence: The Global Landscape of Ethics Guidelines / A. Jobin, M. Lenca, E. Vayena // Nature Machine Intelligence. – 2019. – Vol.1. – Pp. 389-399.

ственного разума в социальных процессах. Это породило тревогу, беспокойство и страх в душе человека, начинающего осознавать риски и угрозы, которые искусственный разум несет обществу вместе с очевидными преимуществами его использования. Социум в ответ на вызовы ИИ принялся формулировать многочисленные этические и правовые нормы и принципы регулирования систем ИИ с целью обезопасить свое взаимодействие с ними. Сегодня отражающие их документы технического, правового, этического характера составляют отдельный, самостоятельный источник актуальной информации о проблемах развития ИИ, само появление и формирование которого является специфической особенностью исследования проблем ИИ в современном обществознании.

1.3 Этическое и правовое регулирование ИИ

Начать аналитический обзор документов, в которых содержатся этические и правовые требования к регулированию ИИ, необходимо с национальных стратегий, которые выступают в роли документов программного характера, задают вектор будущего развития, способствуют формированию основ комплексного правового регулирования общественных отношений в рассматриваемой сфере социальной действительности. В большинстве своем современные программные документы содержат положения о необходимости этического аспекта при развитии индустрии ИИ. Речь, в частности идет о стратегиях, принятых в целом ряде государств, среди которых Россия, Китай, Франция, Финляндия, Швеция и др. Отметим, что в странах Южной Америки и Африки в настоящее время отсутствуют национальные документы рассматриваемой тематики, но эти государства сотрудничают по данным вопросам с Организацией экономического сотрудничества и развития (ОЭСР).

Российская национальная стратегия искусственного интеллекта представлена Указом Президента Российской Федерации от 10 октября 2019 года № 490 «О развитии искусственного интеллекта в Российской

Федерации»[200]. В стратегии установлены основные этические принципы развития и использования технологий ИИ, основные задачи развития ИИ, прописаны действия государственной власти для воплощения этих задач. В стратегии утверждается, что нормативно-правовую базу необходимо создавать поэтапно. Эта база должна быть способна обеспечить формирование и функционирование комплексной системы регулирования общественных отношений в сфере развития и использования технологий ИИ. В стратегии утверждается необходимость соблюдения основных этических принципов развития и применения ИИ, в частности, принципов безопасности, защиты прав и свобод человека, прозрачности, технологического суверенитета, целостности инновационного цикла, поддержки конкуренции и разумной бережливости.

Американская национальная стратегия развития ИИ «О сохранении американского лидерства в области искусственного интеллекта»[201] вступила в силу 11 февраля 2019 года. Стратегия содержит пять основных принципов развития ИИ: внедрение технологических прорывов; просвещение работников в области разработки и применения ИИ; соблюдение прав и свобод личности; укрепление доверия к системам ИИ. В стратегии утверждается, что формирование адекватной правовой базы напрямую зависит от уровня просвещенности у должностных лиц, принимающих решения в области ИИ.

Национальная стратегия КНР 2017 года «План развития искусственного интеллекта Нового поколения»[202] устанавливает этические принципы применения ИИ. Развитие ИИ является национальным приоритетом КНР. Национальная стратегия КНР преследует цель привести страну к

[200] Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). – URL: https://base.garant.ru/72838946/ (дата обращения: 14.05.2022).

[201] Executive Order on Maintaining American Leadership in Artificial Intelligence. – URL: https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/ (дата обращения: 20.06.2022).

[202] State Council, Notice of Issuing New Generation Artificial Intelligence Development Plan. – URL: http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm (на китайском языке) дата обращения: 16.05.2022 г.)

мировому лидерству в области ИИ к 2030 г. Она направлена на содействие научным исследованиям, просвещению специалистов в этой области, разработке стандартов, этических руководств, на решение вопросов безопасности. В стратегии определены основные этапы внедрения ИИ: удержание конкурентноспособной позиции в области ИИ на мировом рынке до 2020 г; активное использование ИИ на практике во всех секторах экономики до 2025 г; обеспечение мирового лидерства в области ИИ до 2030 г. Документ направлен на установление тесного взаимодействия между органами власти, частным сектором и академическими кругами. Разработке этических принципов Китай не уделяет первостепенного значения, однако в стратегии, тем не менее, они упоминаются.

Национальная стратегия Германии 2018 года утверждает, что внедрение систем ИИ должно сопровождаться соблюдением «европейских ценностей, таких как неприкосновенность человеческого достоинства, уважение частной жизни и принцип равенства»[203]. В Стратегии выделяется 12 сфер применения ИИ.

Национальная стратегия развития искусственного интеллекта Швеции ставит своей приоритетной целью лидерство в развитии новых цифровых технологий, в том числе ИИ[204]. Под лидерством следует понимать создание оптимальных условий для разработки и внедрения ИИ с целью процветания и благосостояния страны.

Национальная стратегия Дании дает определение ИИ, обозначает наиболее актуальные сферы (здравоохранение, энергетика и промышленность, сельское хозяйство, транспорт)[205]. В качестве одной из приоритет-

[203] Bundesregierung, Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz. – URL: https://www.bmbf.de/files/180718Eckpunkte_KI-StrategiefinalLayout.pdf (дата обращения: 13.08.2022.).

[204]Government offices of Sweden, National Approach to artificial intelligence. – URL: https://www.regeringen.se/4aa638/contentassets/a6488ccebc6f418e9ada18bae40bb71f/nationalapproach-to-artificial-intelligence.pdf (дата обращения: 21.03.2022 г)

[205]Ministry of Finance and Ministry of Industry, Business and Financial Affairs. National Strategy for Artificial Intelligence. – URL: https://eng.em.dk/media/13081/305755-gb-version_4k.pdf (дата обращения: 11.07.2022 г.).

ных задач в стратегии обозначено установление принципов ответственной разработки и использования ИИ.

Канада в настоящее время является одним из лидеров в области развития и внедрения искусственного интеллекта. При этом основной упор делается на научные исследования, а не на создание нормативной базы управления[206]. Панканадская стратегия ИИ[207] издана канадским Институтом перспективных исследований (CIFAR[208]), который частично финансируется государством. Если исходить из стратегии, то Канада стремится увеличить число исследований в области ИИ и создать взаимосвязанные и взаимодействующие друг с другом научные центры. Для того, чтобы добиться намеченных планов, Канада начала реализовывать программу AI&Society.

Эстонская Национальная стратегия развития ИИ[209] стремится придать статус правосубъектности практически всем алгоритмам, которые связаны с ИИ. Однако из содержания документа не ясно, по каким критериям может быть присвоен статус «цифрового субъекта», что затрудняет развитие систем ИИ, особенно учитывая, что новые технологии имеют полиморфный и непостоянный характер. Стратегия сосредоточивает свое внимание на роли ИИ в государственном управлении, а также на правовом регулировании процессов применения ИИ в отдельных секторах, среди которых разработка автоматизированных транспортных средств, роботов-доставщиков и т.д.

---

[206] Hirsh J. The Policy Deficit Behind Canadian Artificial Intelligence. Centre for International Governance Innovation. – URL: https://www.cigionline.org/articles/policy-deficit-behind-canadianartificial-intelligence (дата обращения: 19.09.2022 г.)

[207] Pan-Canadian Artificial Intelligence Strategy, Invest in Canada. – URL: https://www.investcanada.ca/why-invest/pancanadian-artificial-intelligence-strategy (дата обращения: 23.06.2022 г.)

[208] CIFAR is a Canadian-based global research organization. – URL: https://cifar.ca/ (дата обращения: 12.09.2022)

[209] Estonia accelerates artificial intelligence development. – URL: https://e-estonia.com/estonia-accelerates-artificial-intelligence/ (дата обращения: 19.10.2022 г.); Kratid Eesti heaks (на эстонском языке). – URL: https://www.kratid.ee/ (дата обращения: 19.10.2022 г.).

Национальная стратегия Франции[210] была представлена на конференции «Искусственный интеллект для человечества» (2018 г., Париж) президентом Франции Э. Макроном. Стратегия уделяет большое внимание этическим принципам, связанным с развитием технологий ИИ, в частности, принципу прозрачности технологий ИИ. С. Виллани, автор национальной стратегии и депутат парламента Франции, полагает, что принцип прозрачности должен быть поддержан на государственном уровне. Соблюдение данного принципа поможет, по его мнению, разработать более понятные системы ИИ. Стратегия содержит анализ актуальных проблем, которые касаются новых рисков, связанных с внедрением ИИ в производство. Также в документе утверждается, что ИИ должен быть изначально разработан с этическими нормами с целью предотвращения социально-экономических последствий его применения. До 2022 г. Франция определила следующие приоритетные направления деятельности в рассматриваемой сфере: усиление позиций Франции в ЕС; формирование политики открытых данных; адаптация национальных нормативных норм к европейским; определение основных этических проблем, связанных с разработкой и внедрением ИИ.

Национальная стратегия Великобритании по ИИ[211] для достижения лидерства в области развития ИИ планирует создание Совета по искусственному интеллекту, который будет целенаправленно руководить развитием ИИ, регулировать государственную политику в этой сфере, а также содействовать внедрению ИИ в промышленность. Важнейшими задачами указанный документ называет поддержку экспорта и привлечение иностранных инвестиций в область ИИ.

[210] Villani C. For a meaningful artificial intelligence towards. A French and European strategy. – URL: https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf (дата обращения 12.12.2022).
[211] Select Committee on Artificial Intelligence. AI in the UK: ready, willing and able. House of Lords, UK. – URL: https://www.scirp.org/(S(351jmbntvnsjt1aadkozje))/reference/referencespapers.aspx?referenceid=2884461 (дата обращения: 12.09.2022)

Японский Стратегический совет по технологиям искусственного интеллекта представил стратегию развития технологий ИИ, в которой разработаны дорожные карты индустриализации трех ведущих областей жизни общества, среди которых отмечаются медицина, социальное обеспечение, социальная мобильность[212]. Стратегия устанавливает этапы развития и основные приоритеты государства в области ИИ. В стратегии названы семь принципов ИИ для его этичного применения: принцип антропоцентричности, принцип просвещения (грамотности в области ИИ), принцип конфиденциальности, принцип безопасности, принцип честной конкуренции, принцип справедливости, подотчетности и прозрачности, принцип инноваций. В 2019 году были изданы два документа: «Социальные принципы искусственного интеллекта, ориентированные на человека»[213], «Стратегия искусственного интеллекта – 2019» (AI Strategy 2019. AI for everyone: People, Industries, Regions and Governments[214]).

Сингапурская национальная стратегия[215] вышла в свет в 2019 году. Она, как и многие подобные документы, определяет своей целью достижение мирового лидерства в области развития ИИ. В стратегии выделяются приоритетные области внедрения ИИ: логистика, медицина, образование, туризм и др. Особое внимание уделяется принципу доверия ИИ, защите прав и интересов граждан. Национальная стратегия Сингапура тесно работает с Консультативным советом по этичному использованию ИИ и данных (Advisory Council on the Ethical Use of AI and Data) с целью разработать необходимые этические кодексы в области ИИ. Сингапур также разработал Рамочную модель управления искусственным интел-

[212]Strategic Council for AI technology. – URL: http://www.nedo.go.jp/content/100865202.pdf (дата обращения: 13.10.2022 г.).
[213]Social Principles of Human-Centric AI. – URL: https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf (дата обращения: 12.09.2022)
[214]AI Strategy 2019. AI for Everyone: People, Industries, Regions and Governments. – URL: https://www8.cao.go.jp/cstp/english/humancentricai.pdf (дата обращения: 21.06.2021)
[215]National Artificial Intelligence Strategy. – URL: https://www.smartnation.gov.sg/why-SmartNation/NationalAIStrategy (дата обращения:12.12.2022)

лектом[216], в которой представлены принципы понятности, прозрачности и справедливости при внедрении и использования ИИ. В данном документе отмечается, что добиться идеальной объяснимости, прозрачности и справедливости трудно, однако необходимо стремиться к соблюдению названных требований. Также здесь утверждается, что решения ИИ должны быть антропоцентричными или ориентированными на человека. К 2025 году Сингапур поставил задачу достичь полной автоматизации иммиграционного контроля, а также повсеместно распространить датчики для систем ИИ.

Национальная стратегия Республики Корея под названием «Среднесрочный и долгосрочный генеральный план подготовки к интеллектуальному информационному сообществу: управление Четвертой промышленной революцией»[217] была издана еще в 2016 году. В этой стратегии искусственный интеллект был рассмотрен наряду с такими технологиями, как «интернет вещей», анализ больших данных и т.д. Также в документе утверждалось, что необходимо принять рамочное законодательство, устанавливающее необходимые этические и правовые нормы в сфере ИИ. В 2019 году была утверждена новая национальная стратегия[218] на ближайшие 10 лет, в которой позиция по отношению к разработке и внедрению ИИ существенно пересмотрена. Позиция сводится к тому, что прекращается жесткий контроль научно-исследовательских проектов в области ИИ, поскольку такое чрезмерный надзор тормозит развитие ИИ. Соответственно, было решено увеличить финансирование в этой области и дать больше свободы опытно-конструкторским работам. Также в стратегии

[216] Model Artificial Intelligence Governance Framework (Second Edition) 2020. Infocomm Media Development Authority Singapore. – URL: https://www.imda.gov.sg/-/media/Imda/Files/Infocomm-Media-Landscape/SG-Digital/Tech-Pillars/Artificial-Intelligence/Primer-for-second-edition-of-the-Model-Framework.pdf?la=en (дата обращения - 19.10.2022)

[217] Mid-to Long-term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution. – URL: https://english.msit.go.kr/cms/english/pl/policies2/__icsFiles/afieldfile/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf (дата обращения: 20.02.2022)

[218] Eun-jin K. Korean Government to Repeal Regulations in AI Industry December 18, 2019 / K. Eun-jin.– URL: http://www.businesskorea.co.kr/news/articleView.html?idxno=39324 (дата обращения: 10.03.2022)

утверждается, что военнослужащие и государственные работники должны быть просвещены в области ИИ и иметь соответствующее образование. Кроме того, планируется ввести в начальные и средние школы дисциплины по искусственному интеллекту.

Помимо принятия стратегий развития в названной сфере государственный сектор опубликовал ряд документов, посвященных выделению базовых принципов этичного применения ИИ. Например, правительство Канады разработало директиву об автоматизированном принятии решении[219], в котором особое внимание уделяется справедливости, прозрачности в процессе принятия решений с использованием искусственного интеллекта государственными органами власти. Однако данный документ не имеет никакой юридической силы в отношениях государства и частного бизнеса[220]. Австралия опубликовала концепцию этики ИИ[221], в которой на основе результатов соответствующих научных исследований предлагается набор инструментов для реализации на практике этичного ИИ. Счетная палата США в 2021 г. издала рекомендации о системе ответственности государственных структур, применяющих в своей деятельности технологию ИИ[222].

Изучая географию документов, устанавливающих этические принципы применения систем ИИ, можно сделать вывод о безусловном лидерстве в этом вопросе экономически развитых стран. Так, на США и Великобританию вместе приходится более трети всех документов по этическому регулированию систем ИИ, за ними следуют Япония, Германия, Франция и Финляндия. В Азии Китай находится на первом месте, следом

[219] Government of Canada, “Directive on Automated Decision-Making”, 2019. – URL: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 (дата обращения:12.09.2022)
[220] Government of Canada. Responsible use of artificial intelligence (AI). – URL: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html (дата обращения:21.07.2022)
[221] Dawson D et al. Artificial Intelligence - Australia’s Ethics Framework. Data61, CSIRO, Australia, 2019. – URL: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1 (дата обращения:14.10.2022)
[222] Artificial Intelligence. An Accountability Framework for Federal Agencies and Other Entities.GAO, 2021. – URL: https://www.gao.gov/assets/gao-21-519sp.pdf (дата обращения:25.05.2022)

за ним расположились Япония, Южная Корея и Сингапур. В Океании лидирует Австралия, за которой следует Новая Зеландия. В Африке Южная Африка стала первым государством, которое установило принципы этичного применения ИИ через Research ICT Africa[223], африканский аналитический центр, специально созданный для устранения стратегического пробела, имеющего место в развитии устойчивого информационного общества и цифровой экономики. Другие африканские страны решают похожие проблемы через институт членства в ОЭСР.

К похожим заключениям пришел в своем исследовании Л. Тиджон, который сформулировал и предложил несколько критериев, свидетельствующих, по его мнению, о готовности государств к внедрению ИИ[224]. Среди них, в частности, способность правительственных структур быстро адаптироваться к инновациям и перейти к их практическому внедрению в социальные процессы. Также имеет значение оценка самого процесса внедрения инструментов ИИ, а именно того, обучены ли они на высококачественных и репрезентативных данных, имеется ли соответствующая инфраструктура для предоставления решений ИИ и содействия их внедрению. В итоге рейтинг готовности к внедрению ИИ двадцати рассмотренных автором государств в 2021 году выглядит следующим образом: США, Сингапур, Великобритания, Финляндия, Нидерланды, Швеция, Канада, Германия, Дания, Республика Корея, Франция, Япония, Норвегия, Австралия, Китай, Люксембург, Ирландия, Тайвань, ОАЭ и Израиль. Кроме того, по мнению Л. Тиджона, США, Великобритания и Германия являются наиболее благоприятными государственными образованиями для развития ИИ, инвестиций, бизнеса и исследований, поскольку они лучше других подготовлены с точки зрения уровня технологий ИИ, управления ИИ, а также имеющихся баз данных и инфраструктуры.

[223] Research ICT Africa (RIA). – URL: https://researchictafrica.net/people/ (дата обращения: 21.09.2022)
[224] The Different Faces of AI Ethics Across the World: A Principle-Implementation Gap Analysis Lionel Nganyewou Tidjon and Foutse Khomh, Senior Member, IEEE. – URL: https://arxiv.org/pdf/2206.03225.pdf (дата обращения: 21.09.2022)

Анализ отобранных нами документов по целевой аудитории, к которой они обращены, позволило сделать вывод о том, что большинство из них не имеет универсального характера, но адресовано сразу нескольким заинтересованным сторонам (например, профессиональным сообществам, государственному сектору, частным компаниям, разработчикам, исследователям и т.д.). Есть примеры документов, принятых на национальном уровне и нацеленных на решение проблем в некоторых областях применения систем ИИ. Так, Австралия в 2019 г. сформулировала девять этических принципов ИИ в медицине[225], а Германия в 2017 г разработала этические принципы для эксплуатации беспилотного транспорта[226].

Обращаясь к юридическим документам, существующим в настоящее время в сфере регулирования ИИ, рассмотрим опыт Российской Федерации. В нашей стране Распоряжением Правительства от 19 августа 2020 г. № 2129-р[227] утверждена Концепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года. В данной концепции определен комплекс правовых перспектив в указанной области, а также основные направления работы в рамках создания многоаспектной правовой платформы в сфере робототехники и ИИ.

В контексте обсуждаемого вопроса интерес также представляет Федеральный закон от 24 апреля 2020 года № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта…»[228]. Основная цель документа - обеспечение выс-

[225] Ethical Principles for AI in Medicine. The Royal Australian and New Zealand College of Radiologists. – URL: https://www.ranzcr.com/documents/4952-ethical-principles-for-ai-in-medicine/file (дата обращения: 19.09.2022)

[226]Ethics Commission. Automated and Connected Driving. – URL: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connecteddriving.pdf?__blob=publicationFile (дата обращения: 14.01.2022)

[227] «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.». Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р // СЗ РФ. № 35, 2020 г. – Ст. 5593.

[228] «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Рос-

шим исполнительным органом власти Москвы необходимых условий для разработки, внедрения, реализации и оборота технологий ИИ. Благодаря ему в Москве был дан старт таким проектам, как создание и эксплуатация беспилотных автомобилей на дорогах общего пользования, использование роботов-курьеров службами сервиса и т.д.

21 мая 2021 года было издано Постановление Правительства РФ № 767[229], которым утверждены правила получения господдержки при реализации пилотных проектов по внедрению искусственного интеллекта в различных отраслях экономики. Это Постановление принято для реализации национальной стратегии России в сфере развития искусственного интеллекта. С этой же целью сегодня издаются дорожные карты развития технологий ИИ, которые содержат технологические задачи, сроки и этапы разработки продуктов в конкретной сфере, но эти документы уже не являются актами правового регулирования.

Необходимо упомянуть еще один закон «Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации» (Законопроект № 922869-7)[230], который готовится к рассмотрению. Согласно этому законопроекту, до 2024 года в области нормативного регулирования ИИ поставлены следующие задачи: обеспечение нормативных условий для доступа к данным; разработка более упрощенного административно-правового и нормативно-технического порядка тестирования и внедрения разработок в области ИИ и т.д.

В США очень активно проходит обсуждение различных вопросов правового регулирования ИИ. Так, уже принятый «Закон о защите кон-

---

сийской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных"». Федеральный закон от 24.04.2020 г. № 123-ФЗ. – URL: https://base.garant.ru/400797482/ (дата обращения: 12.09.2022)

[229] «Об утверждении Правил предоставления субсидии из федерального бюджета на поддержку некоммерческой организацией Фонд развития Центра разработки и коммерциализации новых технологий пилотных проектов апробации технологий искусственного интеллекта в приоритетных отраслях». Постановление Правительства РФ от 21 мая 2021 г. N 767 . – URL: https://base.garant.ru/400797482/ (дата обращения: 12.09.2022

[230] Законопроект № 922869-7. 2020. – URL: https://sozd.duma.gov.ru/bill/922869-7 (дата обращения: 20.04.2022)

фиденциальности»[231] (2016 г.) регламентирует передачу персональных данных между ЕС и США. Обсуждается целый ряд законопроектов в этой сфере, среди которых: проект закона о будущем ИИ (Future of AI Act, 2017); проект закона о Комиссии по национальной безопасности в области ИИ, одобренный Конгрессом США (National Security Commission Artificial Intelligence Act, 2018); проект закона о применении ИИ в государственном секторе (AI in Government Act, 2018) и др. В Канаде законодательное регулирование ИИ представлено «Законом о защите личной информации и электронных документов» (2000 г.)[232]. Во Франции принят закон «О цифровой Республике» (Loi pour une Republique numerique, 2016)[233]. В январе 2020 г. Национальное собрание Франции внесло предложение об издании Хартии искусственного интеллекта и алгоритмов (Charte de l'intelligence artificielle et des algorithmes)[234]. Это предложение рассматривалось парламентским комитетом по конституционному законодательству. Авторы законопроекта включили в его содержание ряд таких вопросов, так регулярный аудит систем ИИ, а также предотвращение злонамеренных манипуляций с ИИ. В Великобритании, ОАЭ, Германии существуют Законы о защите данных[235,236,237], Европейский Союз принял Общее положение о защите данных[238], Китай – Закон о защите личной

[231] Eu-U.S and swiss-U.S. privacy shield, 2016. – URL: https://www.privacyshield.gov/program-overview (дата обращения: 16.03.2022)

[232] Personal information protection and electronic documents act, 2000. – URL: https://lawslois.justice.gc.ca/ENG/ACTS/P-8.6/index.html (дата обращения: 21.02.2022)

[233] Loi pour une Republique numerique. No 2016-1321. 2016. – URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id (дата обращения:23.09.2022)

[234] Proposition de loi constitutionnelle relative à la Charte de l'intelligence artificielle et des algorithmes. – URL: http://www.assemblee-nationale.fr/dyn/15/textes/l15b2585_proposition-loi (дата обращения: 20.10.2022)

[235] Data protection act, 2018. – URL: https://www.hhs.gov/hipaa/for-professionals/index.html (дата обращения: 15.03.2022)

[236] Data protection law, 2021. – URL: https://u.ae/en/about-the-uae/digital-uae/data/dataprotection-laws (дата обращения: 23.07.2022)

[237] Bundesdatenschutzgesetz," 2009. – URL: https://www.datenschutz-wiki.de/BDSG (дата обращения: 12.05.2022)

[238] General data protection regulation, 2016. URL: https://gdpr-info.eu/ (дата обращения: 17.02.2022)

информации[239], Закон о кибербезопасности[240], Сингапур – Закон о защите от лжи и манипуляций в Интернете[241].

Б. П. Немиц в своем исследовании «Конституционная демократия и технология в эпоху ИИ»[242] задается вопросами о том, какие аспекты могут регулироваться законом, а какие – этическими кодексами. Важнейшим вопросом в области регулирования ИИ является проблема ответственности за действия систем ИИ. В этом отношении некоторые исследователи придерживаются позиции, что правовое регулирование становится необходимостью, в особенности, когда вопрос касается ответственности за действия систем ИИ и причиненный ими вред. Другим аспектом проблемы ответственности является вопрос о том, кто и как будет нести ответственность за использование ИИ во вредоносных целях. В настоящее время речь в большинстве случаев идет о привлечении к ответственности за вредоносное использование ИИ либо разработчика, либо исполнителя[243].

При этом ни в одном государстве мира не существует нормативно-правового регулирования, однозначно определяющего степень ответственности каждого из участников процессов взаимодействия общества с системами ИИ.

Определенные усилия в сфере регулирования ИИ предпринимаются бизнес-корпорациями, особенно теми из них, которые в своей деятельности активно используют системы ИИ.

---

[239]Personal information protection law, 2021. – URL: https://www.china-briefing.com/news/the-prc-personalinformation-protection-law-final-a-full-translation/ (дата обращения: 12.01.2022)

[240]Cyber security law, 2016. – URL: https://www.tradecommissioner.gc.ca/china-chine/cybersecuritycybersecuritechina-chine.aspx?lang=eng (дата обращения:30.04. 2021)

[241] Protection from online falsehoods and manipulation act, 2019. – URL: https://sso.agc.gov.sg/Acts-Supp/18-2019 (дата обращения: 11.03.2021)

[242] Nemitz P. Constitutional Democracy and Technology in the Age of Artificial Intelligence / P. Nemitz // Philosophical Transactions of the Royal Society. Mathematical, Physical and Engineering Sciences. – 2018. – Vol. 376. – Pp. 1-14.

[243] Аверинская С. А., Севостьянова А. А. Создание искусственного интеллекта с целью злонамеренного использования в уголовном праве Российской Федерации / С. А. Аверинская, А. А. Севостьянова // Закон и право. – 2019. – № 2. – С. 94-96.

В 2018 году компании Google и SAP опубликовали руководящие принципы в сфере ИИ. Google сформулировала семь этических принципов, которыми необходимо руководствоваться при разработке и внедрении систем ИИ: полезность; справедливость; безопасность; ответственность; конфиденциальность; поддержание стандартов высокого мастерства; ограничение применения вредоносных программ. Google также создала специальный инструмент, который позволяет влиять на работу тех инструментов, которые порождают несправедливые или дискриминационные решения для людей[244].

«Руководящий комитет SAP по этике ИИ» состоит из руководителей высшего звена, представляющих все подразделения организации. SAP гарантирует, что программное обеспечение, выпускаемое ею, построено в соответствии с принципами этики. Также в компании создали внешнюю «консультативную группу по этике ИИ», в которую вошли «эксперты из академических кругов, политики и бизнеса, чья специализация находится на стыке этики и технологий — в частности, технологии ИИ»[245]. SAP выступает за разработку глобального кодекса поведения в отношении этических методов ведения бизнеса с использованием ИИ, которым могли бы руководствоваться любые компании[246].

Этический комитет, созданный Microsoft, называется AETHER (AI Ethics in Engineering and Research). Он состоит из экспертов по прикладной этике, инженеров и представителей, назначенных руководителями основных подразделений»[247]. Комитет дает советы и разрабатывает рекомендации по инновациям в области ИИ для всех подразделений Microsoft. В частности, были созданы рабочие группы, занимающиеся такими про-

[244] Visually probe the behavior of trained machine learning models, with minimal coding. – URL: https://pair-code.github.io/what-if-tool/ (дата обращения: 21.09.2022)
[245] Там же.
[246] SAP. European Prosperity Through Human-Centric Artificial Intelligence; The Intelligent Enterprise: 2018, p.28. – URL://www.sap.com/documents/2018/01/3e67a134-ee7c-0010-82c7-eda71af511fa.html (дата обращения:12.04.2022)
[247] Putting principles into practice at Microsoft. – URL: https://www.microsoft.com/en-us/ai/our-approach?activetab=pivot1:primaryr5 (дата обращения:12.09.2022)

блемами, как «предвзятость и справедливость ИИ» и «прозрачность ИИ». Microsoft была первой крупной компанией, которая публично приняла государственное регулирование в отношении ИИ и одной из первых крупных компаний, выступивших за новое законодательство в области ИИ[248].

Amazon Web Services долгое время пыталась оставаться в стороне от дискуссий об этичности ИИ. По словам исполнительного директора AWS Питера Стански, «Клиенты сами должны решить, этично ли использовать инструменты AWS»[249]. Однако несколько спорных вопросов, связанных с ИИ, вынудили платформу изменить свою позицию. Сначала компании пришлось отказаться от использования созданного ею на основе технологии ИИ инструмента рекрутинга, поскольку он демонстрировал предвзятое отношение к претендентам на ту или иную позицию по расовому и половому признакам. В дальнейшем еще один инструмент, предназначенный для распознавания лиц (Rekognition), также подвергся критике по названной выше причине. Ведущие исследователи ИИ, представляющие промышленные предприятия и научное сообщество, осудили этот инструмент за гендерные и расовые предубеждения, а сотрудники и акционеры AWS призвали компанию прекратить продажу данного продукта правоохранительным органам. В конце концов, после двух лет осуждений Amazon неохотно согласился с критикой и остановил продажи[250]. Впоследствии AWS предложила рекомендации для будущего законодательного регулирования программных продуктов для распознавания лиц для правоохранительных органов, в числе которых обязательная проверка результатов работы систем ИИ человеком, уровень достоверности не ниже 99 %, обязательные отчеты о прозрачности, публичное уведомление

[248] Microsoft. The future computed: Artificial Intelligence and its role in society: chapter 2. – URL://news.microsoft.com/cloudforgood/_media/downloads/the-future-computed-english.pdf (дата обращения:17.09.2022).

[249] AWS is ethical about AI but 'we just don't talk about it' say APAC execs. – URL: http://cdn.computerworld.com.au/article/661203/aws-ethical-about-ai-we-just-don-t-talk-about-it-say-apac-execs/ (дата обращения: 11.10.2022)

[250] We are implementing a one-year moratorium on police use of Rekognition. – URL: https://blog.aboutamazon.com/policy/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition (дата обращения:12.09.2022)

всякий раз, когда видеонаблюдение и распознавание лиц используются совместно)[251]. Таким образом, AWS поддержала идею законодательного регулирования приложения ИИ.

IBM также была вовлечена в дебаты о необходимости регулирования ИИ из-за проблем с распознаванием лиц. Свой подход к государственному регулированию ИИ компания обозначила как «точное регулирование», что означает требование подвергать регулированию только определенные, конкретные приложения и всякий раз выяснять, где и по какой причине в процессе использования продукта возникли негативные для общества последствия.

Google была последней компанией, присоединившейся к группе тех, кто требовал нового законодательного регулирования инструментов ИИ. В их отчете об управлении ИИ говорится, что они «с нетерпением ожидают взаимодействия с правительствами, отраслевыми специалистами и гражданским обществом по вопросам, связанным с ИИ» — например, посвященным обсуждению новых видов вооружения или полицейского наблюдения[252].

Intel обсуждает вопросы регулирования ИИ в контексте обеспечения конфиденциальности. В 2017 году Intel заявила, что регулирующие органы должны осуществлять надзор и вмешиваться в случае необходимости, но в рамках существующих правовых норм. Компании, в свою очередь, должны в своей деятельности руководствоваться Принципами справедливой информационной практики (сформулированными ОЭСР) и иметь возможность продемонстрировать регулирующим органам, что эти требования ими соблюдаются[253].

---

[251] Some Thoughts on Facial Recognition Legislation. – URL: https://aws.amazon.com/ru/blogs/machine-learning/some-thoughts-on-facial-recognition-legislation/ (дата обращения:21.09.2022)

[252] Google. Perspectives on Issues in AI Governance. – URL: https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf. (дата обращения: 21.03.2022)

[253]Intel. Artificial Intelligence: The Public Policy Opportunity. – URL: https://blogs.intel.com/policy/files/2017/10/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf (дата обращения (16.02.2022)

Facebook компании Meta инициировал создание независимого наблюдательного совета, который наделен правом принимать решение в каждом конкретном случае оценки содержания того или иного контента (например, на предмет возможного разжигания ненависти, дискриминации, нарушения конфиденциальности и т.д)[254].

Международные организации (ОЭСР, ООН, Европейская Комиссия) вносят свой вклад в формирование системы регулирования ИИ. ООН заявила о целом ряде отдельных инициатив[255], однако до настоящего времени этой организацией все еще не принято ни одного полноценного договора. ЮНЕСКО в 2019 г. подготовила Рекомендации об этических аспектах искусственного интеллекта[256]. ОЭСР 22 мая 2019 г. приняла рекомендации, в которых были сформулированы пять принципов ответственного управления ИИ[257]. Среди этих основополагающих принципов ведущее место занимает принцип принесения пользы ИИ человечеству для его устойчивого развития и роста благосостояния. Остальные требования – это верховенство прав и свобод человека, демократические ценности, справедливость, прозрачность и ответственность. Помимо формулирования принципов, члены организации разработали общие рекомендации в области развития ИИ для правительств стран, входящих в организацию: а) необходимо содействовать государственным и частным инвестициям по вопросам исследований и разработки ИИ; б) необходимо обеспечить надлежащие политические условия для развертывания систем ИИ; в) важно обеспечить просвещение профессиональных кадров в области ИИ для развития навыков в новых условиях.

---

[254] Независимый надзорный совет. – URL: https://www.oversightboard.com/ (дата обращения: 12.04.2022)

[255] United Nations Activities on Artificial Intelligence (AI), 2019. – URL: https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2019-1-PDF-E.pdf (дата обращения: 01.12.2022 г.).

[256] Elaboration of a Recommendation on the ethics of artificial intelligence. – URL: https://en.unesco.org/artificial-intelligence/ethics#drafttext (дата обращения:27.01.2022)

[257] OECD moves forward on developing guidelines for artificial intelligence (AI). – URL: https://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm (дата обращения: 17.08.2022).

Указанные принципы развития ИИ одобрила так называемая «Большая двадцатка» в документе под названием «G20 Ministerial Statement on Trade and Digital Economy»[258], в котором, помимо прочего, страны-участницы G20 разработали рекомендации по улучшению внедрения систем ИИ в социальные процессы, среди которых необходимость разработки цифровой политики, направленной на уменьшение рисков, связанных с развитием цифровой экономики, верховенство принципа антропоцентричности при внедрении ИИ и учет рисков возникновения социальных проблем.

Руководство, созданное инженерной ассоциацией IEEE, возможно, является на сегодняшний день одним из наиболее полных и содержательных документов, посвященных регулированию ИИ. Оно нацелено на то, чтобы гарантировать заинтересованность всех сторон, участвующих в проектировании и разработке автономных и интеллектуальных систем, а также на то, чтобы они учитывали приоритет этических соображений для развития этих технологий на благо человечества»[259]. Отметим, что в указанном руководстве заявлено, что данный документ является исходной информацией для предстоящей «Серии P7000 стандартов» IEEE по этике ИИ. Этот проект, в свою очередь, обещает предоставить измеримый, поддающийся сертификации стандарт прозрачности автономных систем.

Следующий документ - Рекомендации от Группы экспертов высокого уровня Комиссии ЕС по искусственному интеллекту. Они призваны заложить основы нормативно-правовой базы регулирования ИИ в государствах ЕС. В изданных Группой экспертов высокого уровня «Этических рекомендациях для надежного ИИ»[260], подчеркивается, что ИИ должен быть «ориентированным на человека» и «заслуживающим доверия».

---

[258] G20 Ministerial Statement on Trade and Digital Economy. – URL: https://www.mofa.go.jp/files/000486596.pdf (дата обращения: 20.05.2022)

[259] IEEE. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. IEEE Standards Association. – URL: https://standards.ieee.org/industryconnections/ec/autonomous-systems.html (дата обращения: 20.04.2022)

[260] High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.).

Совет Европы также представил документы, направленные на регулирование применения ИИ. Отметим здесь Европейскую этическую хартию по использованию ИИ в судебных системах, которая утверждает, что ИИ способен повысить эффективность судебной системы, однако его внедрение не должно нарушать основные права и свободы человека[261]. Также указанный документ интересен тем, что он содержит всеобъемлющий обзор применения ИИ в судебных системах Европы.

В другом документе Совета Европы – Руководстве о защите данных при применении ИИ[262] – упоминаются следующие этические принципы применения ИИ: конфиденциальность, справедливость, ответственность, прозрачность, безопасность данных и управление рисками.

Помимо выше названных Европейский Союз разработал один из самых известных и обсуждаемых документов по применению ИИ - Резолюцию Европейского Парламента 2015/2103 «Нормы гражданского права о робототехнике»[263]. В нем прописано, что регулирование ИИ должно основываться на уважении прав человека, законах робототехники А. Азимова, нормах международного права. В Резолюции определены некоторые сферы социальной действительности, в которых регулирование ИИ чрезвычайно важно и способно привести к существенным позитивным изменениям: рынок труда, социальная сфера, государственное управление и др.

В Пекинских принципах ИИ[264], представляющих позицию китайского государства по вопросам этики ИИ, содержится ссылка на ключе-

[261] European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment. – URL: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (дата обращения: 19.11.2022г.)

[262]The Guidelines on Artificial Intelligence and Data Protection. – URL: https://rm.coe.int/lignesdirhttps:/rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8 (дата обращения: 18.11.2022 г.).

[263] Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) // OJ C. – 2018. – Vol. 252, 18.7. – Pp. 239-257. – URL: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN (дата обращения: 09.04.2022 г.).

[264]Beijing AI Principles. – URL: https://www.baai.ac.cn/news/beijing-ai-principles-en.html. (дата обращения: 12.09.2022)

вые концепции и ценности китайского общества. Например, здесь указано, что ИИ должен приносить пользу всему человечеству, уважать частную жизнь, достоинство, свободу, автономию и права человека, быть максимально справедливым, уменьшая возможную дискриминацию и предубеждения, обеспечивать прозрачность, объяснимость и предсказуемость принятия решений[265].

Рассматривая в их контексте такие проекты, как система социального кредитного рейтинга Китая, можно понять озабоченность общественности, СМИ, политических деятелей и ученых, которые усматривают в нем своеобразный аналог оруэлловского социального контроля, фактически уничтожающего любую конфиденциальность. Однако в реальности принцип защиты конфиденциальности личных данных четко сформулирован в законодательных актах Китая: более 100 приложений были запрещены правительством за нарушение конфиденциальности пользовательских данных, и еще десятки получили предписания о необходимости внесения изменений, связанных со сбором и хранением данных[266].

Исходя из анализа вышеперечисленных документов, можно сделать вывод, что в настоящее время в мире еще не выработан полноценный международный подход к регулированию ИИ. Можно говорить лишь об отдельных разрозненных документах, заявлениях, рекомендациях, формулирующих требования, принципы, правила, содержание которых расплывчато, многозначно и существует только на уровне деклараций.

Еще одна группа источников – этические стандарты, кодексы, декларации, источником которых выступают международные и неправительственные организации по всему миру.

---

[265] Там же.

[266] В 2019 году Министерство общественной безопасности КНР приостановило работу 100 приложений, не соответствующих стандартам конфиденциальности персональных данных. В этом же году Министерство общественной безопасности проверило 683 приложения. Также в 2019 году Министерство промышленности и информационных технологий КНР опубликовало перечень из 41 приложения, в которые к концу 2019 года должны быть перенастроены так, чтобы соответствовать требованиям к данным. В 2018 году в Шаньдуне возбудили крупное дело о нарушении прав на персональные данные против 11 компаний.

Как известно, Стандарты издаются на основе консенсуса, признания специалистами соответствия той или иной продукции установленным требованиям. Продукция, как правило, должна соответствовать условиям безопасности для природной среды и для здоровья человек.

Этические стандарты, как и этические кодексы, не являются законодательными документами и служат в качестве руководств или «мягких законов», не имеющих обязательной юридической силы. Их появление обусловлено стремлением действовать во благо человечества, в данном случае, в контексте взаимодействия человека с системами и технологиями ИИ.

Международная организация по стандартизации ISO выступает наиболее авторитетным в мировом сообществе разработчиком разного рода стандартов, технических отчетов, рекомендаций. ISO в 2017 году создала отдельный технический комитет по ИИ[267], членами которого стали 30 стран. Некоторые стандарты, изданные данным комитетом, уже вступили в силу: например, стандарт ISO/IEC 20546:2019, содержащий набор основных терминов и определений, имеющий целью улучшение коммуникации и понимания между людьми в области ИИ. В настоящее время в разработке у этой организации находится стандарт ISO/IEC AWI TR 24368 Information technology – Artificial intelligence – Overview of ethical and societal concerns, который описывает этические и социальные проблемы в области ИИ.

Этическое регулирование применения технологий ИИ находится в центре внимания многочисленных неправительственных организаций, подготовивших соответствующие документы. Наиболее известны сегодня Торонтская декларация[268], Универсальные рекомендации по ИИ[269], Ази-

[267]ISO/IEC JTC 1/SC 42 Artificial intelligence. – URL: https://www.iso.org/ru/committee/6794475.html?view=participation (дата обращения: 20.12.2022)
[268] The Toronto Declaration. 2018. – URL: https//www.torontodeclaration.org/declaration-text/english/ (дата обращения:21.01.2022)
[269] The Public Voice. Universal Guidelines for Artificial Intelligence, 2018. – URL://thepublicvoice.org/ai-universal-guidelines (дата обращения:12.01.2022)

ломарские принципы ИИ[270], Монреальская декларация[271] и 10 этических принципов ИИ Глобального союза UNI[272].

«Декларация Торонто»[273] уделяет особое внимание принципам равенства, недискриминации и инклюзивности. В декларации также затрагиваются такие принципы, как прозрачность и подотчетность систем ИИ. Документ был составлен в 2018 году Amnesty International (глобальное движение, выступающее за защиту прав человека) и Access Now (институт Нью-Йоркского университета, занимающийся исследованием социальных последствий ИИ).

«Универсальные рекомендации по ИИ»[274] включают те же принципы, но дополнительно обращают внимание на качество данных, общественную безопасность, запрет на секретное профилирование.

«Азиломарские принципы искусственного интеллекта»[275] были озвучены на конференции The Asilomar Conference on Beneficial AI, организованной бостонским некоммерческим исследовательским институтом Future of Life. Этот институт считает своей миссией поиск путей и способов смягчения экзистенциальных рисков, с которыми сталкивается человечество, в том числе, рисков, исходящих от ИИ. Около 100 исследователей, представляющих разные отрасли научного знания (экономику, право, философию и т.д.) стали участниками конференции, в ходе которой был принят документ, содержащий основные этические принципы применения ИИ, в числе которых доверие, открытость, ответственность, безопасность, защита данных, устойчивость, польза для всего человечества. Этот

[270] Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)
[271]Montréal Declaration: Responsible AI. 2018. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).
[272]UNI Global Union. Top 10 principles for ethical artificial intelligence. – URL://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (дата обращения: 12.08.2022)
[273] The Toronto Declaration. 2018. – URL: https//www.torontodeclaration.org/declaration-text/english/ (дата обращения: 21.01.2022)
[274] The Public Voice. Universal Guidelines for Artificial Intelligence, 2018. – URL://thepublicvoice.org/ai-universal-guidelines (дата обращения:12.01.2022)
[275] Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)

документ также содержит рекомендации для ученых, ведущих научные исследования в рассматриваемой сфере и государственных деятелей, определяющих политику в области развития ИИ и решения возникающих в связи с ним социальных проблем. На конференции, кроме того, обсуждались вопросы, связанные с правовыми рисками использования систем ИИ и их моральным статусом.

«Монреальская декларация ответственного развития ИИ»[276], составленная по итогам форума в Монреальском университете в 2017 г., помимо выделения этических принципов регулирования ИИ направлена на установление широкого диалога между общественностью, экспертами и лицами, принимающими решения на правительственном уровне.

Наконец, Глобальный союз UNI[277], глобальная федерация профсоюзов, сформулировал «10 основных принципов этического ИИ», которые, по мнению создателей документа, должны быть приняты во внимание профсоюзами работников в сфере материального производства в качестве требований к применению технологий ИИ.

Таким образом, проблема регулирования ИИ с позиций этики и права сегодня широко обсуждается и в сфере разработки и применения систем ИИ, и в науке, и на уровне государственных органов власти, крупных бизнес-корпораций, представителями международных, общественных, исследовательских, профессиональных и пр. объединений. Это свидетельствует о высокой степени актуальности рассматриваемой тематики, заинтересованности общества в развитии, распространении, совершенствовании систем ИИ, а также пристальном внимании социума к проблемам безопасности, ответственности, ограничений и запретов в процессах взаимодействия человека с искусственным разумом.

[276]Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).

[277] UNI Global Union (no year) Top 10 principles for ethical artificial intelligence. – URL://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (дата обращения: 12.08.2022)

Выводы по первой главе.

Имеющийся и представленный в разнообразных источниках информации опыт исследования искусственного интеллекта позволяет сделать вывод о том, что существующий в настоящее время на уровне технологии, применяемый в различных областях общественной жизни и требующий регулирования со стороны государственных, общественных организаций, общества в целом искусственный интеллект (ИИ) – это слабый (узкий) искусственный интеллект, созданный человеком для решения определенных, узконаправленных задач.

Узкий ИИ представлен системами, элементами которых являются аппаратные комплексы, программное обеспечение, наборы данных. Поэтому в рамках нашего исследования, посвященного узкому ИИ и связанным с его использованием рисков, мы используем понятия «искусственный интеллект (ИИ)» и «система искусственного интеллекта» как равнозначные, равнообъемные.

В своем исследовании мы рассмотрели ряд областей, в которых системы ИИ получили наибольшее распространение. Это транспорт, государственное управление, образование, оборона и национальная безопасность, сельское хозяйство, промышленность и энергетика. Безусловно, приведенный перечень не полон, поскольку в современном мире практически не осталось сфер деятельности человека, куда еще не проникли системы ИИ. Тем не менее, знакомство с указанными выше областями внедрения систем ИИ позволило выявить специфическую особенность исследования ИИ, которой, по нашему мнению, является своеобразие источниковой базы проблем ИИ. Помимо научных и философских трудов здесь обязательно должны рассматриваться документы регламентирующего характера, содержащие информацию об актуальных вызовах, с которыми человек сталкивается в процессе практического применения систем ИИ в различных сферах жизни общества, а также правовых нормах, моральных принципах регулирования ИИ, правилах и требованиях, уже

сформулированных различными социальными институтами и организациями в ответ на вызовы новой цифровой реальности.

Важнейшее значение в данной группе источников информации имеют национальные стратегии и нормативно-правовые документы, принятые государствами для регулирования отношений в сфере применения систем ИИ. Именно они формируют базовые условия существования и развития ИИ в социуме, указывают на главные проблемы и риски, возникающие в процессе взаимодействия человека и систем ИИ, а также обобщают известный опыт взаимодействия человека с искусственным разумом. Отметим, что, как правило, в них утверждается необходимость соблюдения основных этических принципов применения ИИ.

Исследование правового регулирования систем ИИ свидетельствует о том, что сегодня оно находится на стадии становления и все еще не способно дать адекватный ответ на уже существующие вопросы и трудности. В частности, речь идет о проблеме определения субъектов и степени ответственности участников процессов взаимодействия общества с системами ИИ.

Активность частного сектора в регулировании систем ИИ в настоящее время заметно снизилась, а результаты его деятельности в этой сфере вызывают большое количество вопросов в связи с тем, что проблемы применения систем ИИ, очевидно, не могут решаться только техническими методами или только в частном порядке, применительно к отдельным случаям использования конкретных систем ИИ. Масштабы распространения устройств на основе ИИ в социуме, их разнообразие и сложность сегодня настоятельно требуют правового регулирования не на уровне частных компаний, а на уровне государств, законодательных органов власти.

Этическое регулирование применения технологий ИИ со стороны международных неправительственных организаций, как правило, сводится к принятию рамочных, декларативных документов, которые обознача-

ют лишь общие подходы и цели участников взаимодействия общества, его отдельных социальных и профессиональных групп с системами ИИ. Содержащиеся в них этические принципы разработки, внедрения и развития систем ИИ, как правило, играют роль своеобразного предостережения человечеству от возможных негативных последствий применения технологий ИИ. Однако, несмотря на значительный объем совпадений в предлагаемых организациями принципах, содержание этих норм интерпретируется ими по-разному, либо вообще остается нераскрытым.

Все вышеперечисленное подтверждает важность и значимость избранной цели исследования и указывает на необходимость социально-философского осмысления наиболее актуальных проблем применения систем ИИ в современном обществе.

Таким образом, при подготовке параграфа нами исследовано содержание значительного пула документов по этическому и правовому регулированию систем ИИ, отобранных на основе критериев надежности, новизны и разнообразия. Источником большинства документов выступили частные компании, государственные учреждения, академические организации, научно-исследовательские институты, межгосударственные или наднациональные организации, профессиональные ассоциации (полный перечень приведен в Приложении к диссертации). Рассмотренные документы, заявляя о необходимости учета этического аспекта при развитии ИИ, в то же время лишены целостности, единства в оценке и интерпретации уровня и перспектив развития ИИ, не учитывают содержание других источников и потому не могут служить общим, универсальным руководством для безопасного взаимодействия систем ИИ с человеком и обществом. Так, национальные стратегии определяют самые общие теоретические подходы и самые общие требования к процессу внедрения ИИ. Этические стандарты и кодексы, адресованные отдельным заинтересованным сторонам (например, профессиональным сообществам, государственному сектору, частным компаниям, разработчикам, исследователям и

т.д.), напротив, построены на результатах осмысления практики использования ИИ в некоторых узких и специфических отраслях деятельности.

Попытки выработать стратегию по безопасному использованию ИИ на уровне академической науки в настоящее время ограничено выявлением основополагающих принципов и проблем, тогда как их содержание все еще остается не раскрытым, что также затрудняет применение выделенных принципов на практике.

Правовое регулирование, по оценкам специалистов, существенно отстает от процесса внедрения ИИ в различные области социальной реальности. Существующие документы позволяют сделать обоснованный вывод об отсутствии в настоящее время полноценного подхода к регулированию ИИ с позиции права. Фрагментарный, разрозненный характер принятых документов, содержание которых расплывчато, многозначно и декларативно, свидетельствует о необходимости дальнейшего исследования и обобщения опыта применения ИИ в современном мире.

ГЛАВА II. СОЦИАЛЬНО-ФИЛОСОФСКИЕ ПРОБЛЕМЫ ПРИМЕНЕНИЯ ИИ В СОВРЕМЕННОМ ОБЩЕСТВЕ

2.1 Разнообразие и комплексный, социально-философский характер проблем применения ИИ

Обсуждение проблем применения систем ИИ необходимо предварить обращением к классике, к первому закону роботехники, сформулированному писателем-фантастом А. Азимовым в рассказе «Хоровод», вышедшем в свет в далеком 1942 году: робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.

Распространяя действие первого закона на системы ИИ, подчеркнем, что последние создаются человеком для решения важных для него задач, достижения поставленных им целей, но не для продуцирования дополнительных трудностей, препятствий и проблем. Тем более, они ни при каких условиях не должны превращаться в угрозу для человека, его здоровья, безопасности, жизни, для его настоящего и будущего. Напротив, они призваны облегчить жизнь человека, освободив его от рутинных и неинтересных, либо чрезвычайно трудоемких операций, для созидательных и раскрывающих его безграничный творческий потенциал занятий.

Тем не менее, исследуя наиболее актуальные области применения ИИ, мы выяснили, что в процессе проектирования, внедрения и распространения систем ИИ неизбежно возникает ряд специфических проблем, перечень которых по мере развития сферы применения систем ИИ постоянно расширяется.

Традиционно эти проблемы принято обозначать термином «этические» проблемы, однако, на наш взгляд, они далеко выходят за пределы собственно этической проблематики. Мы характеризуем их как социально-философские, поскольку они охватывают широкий комплекс вопросов, затрагивающих в целом взаимодействие человека с искусственным интеллектом в различных сферах общественной жизни, порождая помимо

собственно этических, нравственных проблем, экзистенциальные, аксиологические и др. вопросы. Кроме того, их действие ухудшает состояние и положение человека в мире. Следовательно, они требуют осмысления и разрешения не только из этических соображений, но, главным образом, с точки зрения социальной философии.

Так, экзистенциальная проблема утраты человеком смысла бытия может возникнуть, например, в процессе применения систем ИИ в современном производстве. Человеку, потерявшему работу из-за замены его роботом, зачастую сложно найти другие формы успешной социализации, и он ощущает свою ненужность, невостребованность в социуме, порождающую вопросы о предназначении, ценности и смысле его жизни. Аксиологический аспект воздействия технологий искусственного интеллекта на миропонимание современного человека связан с тем, что искусственный интеллект воспринимается в обществе в качестве абсолютной, безусловной ценности, вытесняя на обочину бытия традиционные ценности человеческой жизни.

Нельзя не отметить особый класс проблем, порожденных относительной новизной феномена ИИ, все еще существующим недостатком изученности опыта применения его систем в социальной практике и, как следствие, недостаточной осведомленностью человека о реальных и потенциальных возможностях ИИ.

Проблемы применения ИИ не раз становились предметом научной дискуссии. По ее итогам учеными получены значимые результаты и опубликованы обзоры наиболее важных, по их мнению, вопросов.

Так, Боссманн выделил девять основных проблем в области ИИ: безработица, неравенство, человечность, искусственная глупость, расистские роботы, безопасность, злые джинны, сингулярность и права роботов[278]. М. Райан, Дж. Антониу говорят уже о семнадцати подобных вызо-

[278] Bossmann Dzh. Top 9 Ethical Issues in Artificial Intelligence. – URL: https://hr-portal.ru/article/9-glavnyh-eticheskih-problem-iskusstvennogo-intellekta (дата обращения: 22.03.2022).

вах[279]. Т. Хагендорф в своем фундаментальном исследовании различных этических руководств[280], выпущенных к 2020 г., обнаружил проблемы непрозрачности, нарушения конфиденциальности, дискриминации, угрозы безопасности, отсутствия подотчетности (ответственности). Доклад «Искусственный интеллект и жизнь в 2030 г.»[281], выпущенный Стэнфордским университетом в 2016 г., в числе насущных проблем применения систем ИИ называет следующие: угроза безопасности, конфиденциальности личных данных, проблема нарушения уголовной и гражданской ответственности, безработица, сертификация систем ИИ. Эти же проблемы упоминаются в Резолюции Европейского парламента «Нормы гражданского права»[282]. В Национальной стратегии РФ обозначена проблема причинения вреда, а также указано, каким образом и вследствие каких причин системы ИИ могут причинить вред человеку[283]. И. Н. Мосечкин утверждает, что самообучающиеся системы ИИ представляют опасность для общества и рассматривает некоторые инциденты с участием ИИ, требующие, по его мнению, законодательных мер регулирования[284].

Институтом инженеров электротехники и электроники (Institute of Electrical and Electronics Engineers, IEEE) были опубликованы «Рекомендации по этически обоснованному проектированию» (Ethically Aligned

---

[279] Ryan M., Antoniou J. et al. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality / J. Antoniou, M. Ryan, L. Brooks, T. Jiya, K. Macnish, B. Stahl // Science and Engineering Ethics. – 2021. – Vol. 27**(**2). – P. 16.

[280] Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines / T. Hagendorff // Minds & Machines. – 2020. – Vol. 30. – Pp. 99-120.

[281] Stone P., Brooks R. et al. Artificial Intelligence and Life in 2030 / P. Stone, R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager // One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University. – Stanford, CA, September 2016. – 50 p.

[282] Delvaux M. Draft report with recommendation to the Commission on Civil Law Rules on Robotics. Committee on Legal Affairs of European Parliament. – URL: http://robotrends.ru/images/1702/853648/Draftreport_with_recommendations_to_the_commission_on_civil_Law_Rules_on_Robotics.pdf (дата обращения: 12.01.2022)

[283] Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). – URL: https://base.garant.ru/72838946/ (дата обращения: 14.05.2022).

[284] Мосечкин И. Н. Искусственный интеллект и уголовная ответственность: проблемы становления нового вида субъекта преступления / И. Н. Мосечкин // Вестник Санкт-Петербургского университета. Право. – № 10 (3). – С. 461-476.

Design)[285], в которых содержится перечень социальных и этических проблем внедрения систем ИИ: проблема определения ответственности, непрозрачности, недостаточной осведомленности; проблема неуниверсальности и конфликтности этических норм, зависящая от характера задач и различий между сферами применения; проблема достижения доверия между людьми и системами ИИ; проблема непредвиденного поведения систем ИИ; трудность в обеспечении безопасности работы систем ИИ; проблема конфиденциальности персональных данных; проблема обеспечения прозрачности в работе систем ИИ; проблема юридической ответственности в случае нанесения системами ИИ вреда человеку.

Проблему распределения ответственности между людьми и системами ИИ исследовал Д. Балкин[286]. Он размышляет над тем, что усложнение и распространение систем ИИ приводит к вопросу о наделении этих систем юридической ответственностью, субъектностью. Иначе невозможно понять, кто будет отвечать перед законом, если в результате использования системы ИИ возникнут негативные последствия и будет нанесен вред, ущерб кому-либо из участников инцидента.

Системы ИИ активно внедряются в так называемые «умные города», и «широкое использование технологий в повседневной жизни приводит к значительным изменениям в материальной и духовной жизни человека»[287]. М. Шрехер[288] и С. Рид[289] обратили внимание специалистов на проблему распределения ответственности в контексте применения систем ИИ в умных городах. Они пишут о том, что системы ИИ должны быть всегда регулируемыми, подотчетными и снабженными механизмами

[285] Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2/ The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. – URL: http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html (дата обращения: 01.01.2022)
[286] Balkin J. B. The Path of Robotics Law / J. B. Balkin // California Law Review. – 2015. – Vol. 6. – Pp. 45-60.
[287] Бадмаева М. Х. Повседневная жизнь человека в умном городе / М. Х. Бадмаева // Вестник Бурятского государственного университета. Философия. – 2020. – Вып. 4. – С. 34.
[288] Scherer M.U. Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. / Harv. J. // Law Technol. – 2015. – Vol. 29. – P. 353.
[289] Reed C. How should we regulate artificial intelligence? / C. Reed // Philos. Trans. R. Soc. – 2008. – Vol. 376. – Pp. 201.

компенсации сбоев. Речь идет именно о таком виде регулирования, которое стоит на страже человеческих ценностей, обеспечивая непричинение вреда жизни человека, устойчивое развитие и защиту окружающей среды, соблюдение требования социальной справедливости. Эти общественные ценности, на их взгляд, должны иметь безусловный приоритет перед экономическими ценностями, соображениями прибыли и выгоды.

О том, что экономические цели корпораций и интересы политических элит не должны быть приоритетной задачей для ИИ, также писали Л. Флориди, Дж. Коулс, Т. Кин и М. Таддео[290]. Они убеждены, что ИИ является общим ресурсом, призванным работать исключительно на благо каждого человека и общества в целом.

В центре внимания В. Майер-Шёнбергер и К. Кукьер[291] – проблема непрозрачности, ставшая ощутимой, по мнению авторов, в связи с анализом больших данных, поскольку операции с ними невозможно контролировать так, как это было возможно в случае с простым компьютерным кодом.

Дж. Баррелл, исследуя проблему непрозрачности, попытался выяснить причины ее появления[292]. При этом он различает преднамеренную непрозрачность, возникающую, например, в тех случаях, когда правительство или корпорации, преследуя некие собственные цели, принимают решение оставить последовательность действий системы ИИ недоступной для общественности, общественного мнения и непрозрачность, обусловленную масштабами применения системы ИИ, когда принципы машинного обучения или количество задействованных акторов делают алгоритм ее действий непрозрачным даже для экспертов.

---

[290] Floridi L., Cowls J. et al. How to design AI for social good: Seven Essential factors / L. Floridi, J. Cowls, T.C. Kin, M. Taddeo // Sci. Eng. Ethics. – 2020. – Vol.26. – Pp. 1771-1796.

[291] Майер-Шёнбергер В., Кукьер К. Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим / Пер. с англ. – М.: Издательство «Манн, Иванов и Фербер», 2014. – 240 с.

[292] Etzioni A., Etzioni O. Should artificial intelligence be regulated. / A. Etzioni, O. Etzioni. – URL: https: //www.issues.org/334/perspective-should-artificial-intelligence-be-regulated/ (дата обращения:11.01.2022)

А. Гринфилд[293], Т. Ахмад, А. Тередесей и К. Эккерт[294] рассмотрели проблему непрозрачности систем ИИ в связи с вопросом о недоверии к ИИ. Они считают, что системы ИИ уже стали неотъемлемой частью повседневной жизни человека, но при этом принцип работы их остается окутанным завесой тайны. Таким образом, непрозрачный характер принятия решений системами ИИ подрывает доверие к ним со стороны общества и его институтов.

Особую остроту указанные проблемы приобретают в процессе использования сравнительно недавно появившегося чата GPT 4, основанного на ИИ. Новый чат-бот с искусственным интеллектом GPT 4 - это генеративная языковая модель, запущенная коммерческой компанией Open AI 14 марта 2023 г. в развитие ранее разработанных чат-ботов GPT. Генеративной она называется, поскольку способна создавать новые данные, а не только анализировать существующие. Chat GPT 4, основанный на МО, не только распознает шаблоны, но и использует их для создания новых данных.

В настоящее время чат-бот пишет стихи и песни, готовит слайды в определенном стиле, успешно проходит собеседование по поводу приема на работу в Google, пишет маркетинговые кампании для определенной демографической группы, комментарии к онлайн-играм и создает изображения с высоким разрешением, выстраивает прогнозы, финансовую аналитику, генерирует простейший код, общается на многих языках мира. Продукты, полученные GPT 4, почти неотличимы от контента, созданного человеком, так как эта система обучается на базе всей информации, доступной в настоящее время в Интернете. Поэтому, на первый взгляд,

[293] Гринфилд А. Радикальные технологии: устройство повседневной жизни / А. Гринфилд. – М.: Издательский дом «Дело» РАНХиГС, 2019. – 424 с.
[294] Ahmad M. A., Teredesai A., Eckert C. Fairness, accountability, transparency in AI at scale: Lessons from national programs / M. A. Ahmad, A. Teredesai, C. Eckert // Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. – Barcelona, Spain, 2020. – Vol. 27 (30). .– Pp. 690-699.

GPT 4 выглядит полноценным партнером человека в творческой работе по созданию инновационных продуктов.

Преимущества этого нового инструмента искусственного интеллекта по сравнению с его предшественниками очевидны, однако важно видеть и понимать оборотную сторону этого явления и те проблемы, которые оно порождает. Так, в недавнем мультидисциплинарном исследовании, объединившем 43 статьи экспертов из разных областей, утверждается, что применение Chat GPT может привести к репутационным и юридическим рискам, использованию оскорбительного контента или контента, защищенного авторским правом, потере конфиденциальности, мошенничеству, транзакции и распространению ложной информации. Угрозы, исходящие от Chat GPT 4 и подобных ИИ-ботов, связаны с проблемой непрозрачности, могут способствовать дискриминации и предубеждениям, нарушению авторских прав, плагиату или сфабрикованному ботами неаутентичному текстовому контенту, проблеме распространения поддельных медиа[295].

В сложных видах деятельности, имеющих творческий и инновационный характер, GPT 4 сегодня не способен создавать действительно оригинальный, новый продукт. Музыкант Ник Кейв недавно получил от одного из своих поклонников текст песни, который был написан Chat-GPT на основе конкретной подсказки: «в стиле Ника Кейва». Разочарованный Кейв ответил поклоннику: «Написание хорошей песни — это не мимикрия, не репликация и не стилизация…»[296]. Л. К. Джена, С. Гоял

[295] Dwivedi, Y, Kshetri, N et al .Hughes, Laurie & Slade, Emma & Jeyaraj, Anand & Kar, Arpan & Baabdullah. “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy / Y. Dwivedi, N. Kshetri, L. Hughes, E. Slade, A. Jeyaraj // International Journal of Information Management – 2023. – URL: https://www.researchgate.net/publication/369170284_So_what_if_ChatGPT_wrote_it_Multidisciplinary_perspectives_on_opportunities_challenges_and_implications_of_generative_conversational_AI_for_research_practice_and_policy (дата обращения: 03.04. 2023)

[296] Cain S. «This song sucks»: Nick Cave responds to ChatGPT song written in style of Nick Cave” / S. Cain // The Guardian. – 2023. – URL: https://www.theguardian.com/m usic/2023/jan/17/this-song-sucks-nick-cave-responds-to-chatgpt-song-written-in-style-of-nick-cave (дата обращения: 01.04.2023)

считают, что стать по-настоящему творческими гораздо больше шансов имеют системы с формами эмоционального интеллекта[297].

Бум распространения чат-бота вновь актуализировал опасения, касающиеся страха растущей зависимости людей от подобных программ и постепенного делегирования им человеком решения все более сложных и значимых задач[298]. Например, в развивающихся странах, где не достает экспертов в тех или иных отраслях знаний, Chat GPT 4 может заменить их, генерируя необходимую пользователям информацию. Широкое использование чат-бота в образовании уже в настоящее время требует перестройки и адаптации учебного процесса, работы преподавателей, научных сотрудников и руководства университетов к условиям быстро меняющейся цифровой среды. По мнению ряда исследователей, подобные технологии вполне могут использоваться и используются студентами для написания научных работ, что порождает проблемы с плагиатом и академической честностью[299]. Помимо очевидных, лежащих на поверхности результатов применения указанной технологии, необходимо учитывать риск возникновения более глобальных и отдаленных во времени негативных последствий. Речь идет, например, о возможном отказе учащихся от углубленного изучения и критического анализа различных аспектов изучаемого предмета вследствие использования ими Chat GPT 4, позволяющего получить быстрый результат без реального погружения и овладения новыми навыками и знаниями. С. Коннор пишет, что это может привести к подавлению критического мышления, творческого потенциала, абсолютно необходимых для развития человека[300].

[297] Jena L. K., Goyal S. Emotional intelligence and employee innovation: Sequential mediating effect of person-group fit and adaptive performance/ L. K. Jena, S. Goyal // European Review of Applied Psychology. – 2022. – Vol. 72(1). – Article 100729.

[298] Baird A., Maruping L. M. The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts / A. Baird, L. M. Mapuring // MIS Quarterly. – 2021. – Vol. 45(1).

[299] Stokel-Walker C. AI bot ChatGPT writes smart essays - should professors worry? / C. Stoker-Walker // Nature. – London. – 2022.

[300] O'Connor S. ChatGPT. Editorial: Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? / S. O' Conner // Nurse Education In Practice. – 2023. – Vol. 66. - Article 103537.

Проблема непрозрачности в связи с применением Chat GPT 4 выражается в том, что от разработчика и тем более пользователя скрытыми остаются механизмы работы бота, системы ранжирования и рекомендации. GPT 4, как и многие другие системы ИИ, основанные на МО, может использовать данные, представляющие собой сложившиеся в обществе предубеждения и заблуждения, способствуя усугублению проблемы социальной несправедливости. Д. Винбергер писал, что «предвзятость — это первородный грех машинного обучения»[301]. Как известно, знания мы обретаем вследствие формулирования объяснений и проверки их на соответствие реальности, тогда как система опирается на всю совокупность данных, перенесенных человеком в сеть. Следовательно, избежать в полной мере предвзятости на данном этапе развития технологии ИИ не представляется возможным. Chat GPT 4 фактически строит свои рассуждения на основе информации, которая далеко не всегда надежна и достоверна. Не будучи способным к подлинному пониманию, он занимается сопоставлением, компилированием больших объемов данных, находящихся в открытых для обучения источниках. Сгенерированная GPT 4 информация требует тщательной проверки, поскольку его программное обеспечение создает неточный и ложный контент, который может выглядеть вполне достоверным, но на деле является ложным результатом логически некорректных выводов.

Данная система, помимо прочего, нивелирует значимость фактов и выводит на поверхность глобальных информационных потоков любую информацию, в том числе и конфиденциальную, что порождает проблемы нарушения конфиденциальности и управления данными, а также проблему подотчетности в том случае, когда организация не знает и не может определить, кто должен нести ответственность за недостоверную

[301] Weinberger D. How Machine Learning Pushes us to Define Fairness. URL: https://hbr.org/2019/11/how-machine-learning-pushesus-to-define-fairness. – 2019.

информацию, полученную от Chat GPT 4 и активно используемую в работе сотрудниками компании.

Доступность данной технология для массового потребителя и разработчиков через языки программирования (например, Python), облачные сервисы (например, Amazon Web Services, Microsoft Azure, Google) позволяет применять ее так же широко, как, например, программы Excel и Access. Эта легкость и простота кратно увеличивают риск нанесения вреда, ущерба человеку вследствие использования информации, сгенерированной данной системой, которая не имеет ничего общего с действительным решением поставленных человеком задач.

По нашему мнению, указанные проблемы, порожденные применением ИИ, требуют социально-философского осмысления, поскольку по своему смыслу и содержанию они соответствуют основной интенции социально-философского исследования. Социальная философия традиционно изучает с позиций целостности и системности универсальные, сущностные черты общества, наиболее общие законы его динамики, глубинные причины тех или иных событий и процессов, перспективы развития социума. Но все эти знания ей необходимы для уяснения места и роли человека, выступающего в качестве творца мира культуры, мира ценностей и норм, необходимых для совместного бытия больших групп людей, функционирования социальных институтов, существования социальной реальности в целом. Иначе говоря, в фокусе социально-философского исследования всегда находится человек как центральный элемент мироздания, создатель и носитель всех свойств и качеств социального, без которого говорить об обществе не имеет смысла и не представляется возможным. Следовательно, проблемы, возникающие в процессе взаимодействия человека с системами ИИ должны быть введены в круг вопросов, исследуемых социальной философией, поскольку обсуждение и решение этих вопросов затрагивает основы бытия человека в мире, его положение по отношению к другим составляющим социальной реальности, а также

ближайшие и отдаленные перспективы его развития. Как отмечает П. В. Алексеев, «…подлинно философские трактовки социальной философии, ее задач и предмета фокусируются на индивиде, на его многогранных запросах и обеспечении лучшей жизни человека. Именно эти интересы…должны просвечивать все исследования по социальной философии. Научность социально-философского познания должна сливаться с гуманистичностью – таков ведущий принцип познания в сфере социальной философии»[302].

Социально-философское исследование также необходимо для выявления комплексной природы проблем ИИ, для поиска решений этих проблем не только с теоретических позиций этики, аксиологии, философской антропологии и других отраслей знаний о человеке и обществе, а из понимания целостности социальной жизни людей, реализации гуманистических идеалов, деятельностной природы, целеполагающей функции человека, его способности к природопреобразующей деятельности и обусловленного этим центрального положения человека в социальной действительности.

Обобщая вышесказанное, отметим в завершение параграфа 2.1, что проблема угрозы безопасности или причинения вреда, сформулированная А. Азимовым в качестве первого закона роботехники, выступает в роли центральной, системообразующей социально-философской проблемы применения технологии ИИ, по отношению к которой остальные представляются раскрывающими возможный источник, причину возникновения вреда, либо его разновидности, варианты реализации.

Как следует из рассмотренных нами и упомянутых выше источников, к наиболее часто обсуждаемым в них вопросам и трудностям применения ИИ авторы и исследователи относят проблемы нарушения автономии человека, социальной несправедливости как представляющие аспекты, формы, грани, разновидности этого вреда и проблемы нарушения

[302] Алексеев П.В. Социальная философия. – М.: ООО «ТК Велби», 2003. – С. 8.

конфиденциальности, отсутствия ответственности, непрозрачности как порождающие, вызывающие возможное причинение ущерба, вреда человеку вследствие использования технологии ИИ.

Безусловно, предложенный перечень проблем применения ИИ не является полным, окончательным, исчерпывающим. В силу повсеместного проникновения систем ИИ в различные сферы общественной жизни, перманентного совершенствования технологии и постоянно растущего влияния их на жизнь и деятельность человека этот перечень может изменяться, пополняться, уточняться. Кроме того, очевидно, что социальные проблемы могут быть вызваны не только применением ИИ. Иные технологии, не связанные с ИИ, также могут выступать в роли источника этих проблем. Однако использование ИИ способно усугубить действие других причин, а также стать самостоятельной причиной ухудшения положения человека в мире.

Рассмотренный в данном параграфе пример Chat GPT 4 подтверждает тезис о принципиально «узком» характере современного ИИ, который, несмотря на свои выдающиеся характеристики, все же не может стать полноценной заменой человеку. Человек остается по отношению к искусственному разуму оригиналом, творцом, создателем действительно нового. Его сущность заключается в неисчерпаемости, открытости, бесконечности, загадочности и необъятности. Технологии воспроизводят и используют результаты деятельности человека, но превзойти его непостижимую с позиций рациональности природу они не в состоянии. При этом созданный GPT 4 контент в виде фейковых новостей, пропаганды, дезинформации уже сегодня может вводить людей в заблуждение, усугубляя существующие в обществе предубеждения, разжигая вражду, ненависть, подрывая социальную сплоченность, доверие людей друг к другу и к результатам развития новых технологий, технологий будущего. Все это обосновывает необходимость системного, целостного, взвешенного исследования воздействия ИИ на общество и человека средствами и ме-

тодами гуманистически ориентированной социально-философской интерпретации процессов взаимодействия человека с искусственным разумом, открывающей возможность отыскания путей решения, способов преодоления уже существующих и только формирующихся в данной области проблем и препятствий.

2.2 Основные проявления отрицательного влияния ИИ на человека

Причинение вреда. Проблема причинения вреда вызывает беспокойство и самую серьезную озабоченность ученых, общественных деятелей, мирового сообщества в целом. Британский физик С. Хокинг в 2014 году сказал: «Успешное создание искусственного интеллекта станет самым большим событием в истории человечества. К сожалению, оно может оказаться последним, если мы не научимся избегать рисков... Когда искусственный интеллект начнет управлять финансовыми рынками, научными исследованиями, людьми и разработкой оружия, это будут вещи, недоступные нашему пониманию. Если краткосрочный эффект искусственного интеллекта зависит от того, кто им управляет, то долгосрочный эффект - от того, можно ли будет им управлять вообще» [303].

Обсуждаемая проблема зачастую сводится к намеренному использованию ИИ в целях, представляющих угрозу для жизни, здоровья и благополучия человека и общества. Так, в системе обеспечения правопорядка применение ИИ зачастую усугубляет существующие негативные тенденции.

Некоторые системы ИИ используются в разработке автономного оружия, специально создаются для решения военных задач. В 2015 г. на Международной конференции по искусственному интеллекту исследователи в области робототехники и ИИ, общественные деятели, ученые призвали прекратить дальнейшее развитие вооружения, основанного на ИИ,

[303] Хокинг С. Искусственный интеллект - величайшая ошибка человечества / С. Хокинг // Техномания. - 2014. – URL: https://texnomaniya.ru/other-interesting-news/hoking-iskusstvenniiy-intellekt-velichaiyshaya-oshibka-chelovechestva.html (дата обращения 11.04.2021).

которое может действовать без строгого человеческого контроля. Открытое письмо подписали более 20 тыс. человек[304].

Тем не менее, в настоящее время существуют разные виды автономного вооружения на основе ИИ. Помимо тех из них, которые подконтрольны человеку и предоставляют ему возможность отмены решения системы, обсуждается идея создания полностью автономного оружия, принципиально неподвластного человеческому контролю[305]. Есть попытки регулирования процесса введения автономного и полуавтономного оружия: в одном из документов министерства обороны США указано, что командиры и операторы должны иметь возможность осуществлять необходимый уровень человеческого суждения относительно использования силы[306]. Но что именно подразумевается под выражениями «необходимый уровень человеческого суждения» остается неясным.

Автономное вооружение, таким образом, представляется еще более опасным в сравнении с традиционным оружием, поскольку его применение существенно сокращает время на принятие и осмысление важных стратегических решений, минимизирует контроль над ситуацией со стороны человека и лишает человека возможности самостоятельно принимать решения, сокращая его автономию, ограничивая его свободу. Непрозрачный характер приводит к отсутствию взаимопонимания между машиной и человеком, а логика принятия решений ИИ может оставаться непонятной для последнего.

Угрозы постепенно возникают во всех сферах: в медицине – ненадлежащая деятельность автономного медицинского оборудования; в агросекторе – разрушения, причиненные беспилотной сельскохозяйственной

[304] Открытое письмо в ООН с призывом запретить разработку вооружений с ИИ. – URL: https://robogeek.ru/iskusstvennyi-intellekt/otkrytoe-pismo-v-oon-s-prizyvom-zapretit-razrabotku-vooruzhenii-s-ii# (дата обращения: 12.04.2022)

[305] Etzioni A., Etzioni O. Should artificial intelligence be regulated. / A. Etzioni, O. Etzioni. – URL: https: //www.issues.org/334/perspective-should-artificial-intelligence-be-regulated/ (дата обращения:11.01.2022)

[306] Sharkey N. Towards a principle for the human supervisory control of robot weapons / N. Sharkey // Politica&Società. – 2014. – № 2. – Pp. 305-324

техникой (например, трактор Spirit[307] или устройство Lettuce Bot[308]); в обороне — ошибка в системе «свой — чужой» военных устройств для подавления противника.

Международные стандарты безопасности ИИ в основном находятся в стадии разработки и обычно касаются только таких ее отдельных аспектов, как объяснимость[309] или управляемость[310] систем ИИ. Они не рассматривают и не предлагают структурированный процесс для полной оценки рисков в области ИИ или полную таксономию источников рисков, которые были бы необходимы для разработки соответствующих стандартов, либо дают лишь краткое описание источников риска, что затрудняет общее понимание связанных с ними трудностей. Существует всего несколько исследований, посвященных определению и описанию конкретных источников рисков, исходящих от ИИ, но и они содержат лишь поверхностное и неполное описание[311].

К сожалению, применение систем ИИ в военных и преступных целях не исчерпывает весь потенциальный вред, наносимый человеку. Проблема причинения вреда имеет более сложный, комплексный характер, проявляющийся посредством множества социально-философских проблем, раскрывающих отдельные грани причиняемого человеку вреда и породившие их причины.

До недавнего времени считалось, что технологии ИИ заменят человека в выполнении относительно простых, рутинных работ, но в действительности оказалось, что они могут заменить человека, выполняя практически любую работу, алгоритм которой можно вычислить и предсказать.

---

[307] AT400 Spirit. – URL: http://robotrends.ru/robopedia/at400-spirit (дата обращения: 17.01.2022)

[308] Lettuce Bot: Roomba for Weeds. – URL: https://modernfarmer.com/2013/05/lettuce-bot-roomba-for-weeds/ (дата обращения:15.03.2022)

[309] ISO/IEC AWI TS 6254; Information Technology-Artificial Intelligence-Objectives and Approaches for Explainability of ML Models and AI Systems. International Electrotechnical Commission; International Organization for Standardization: Geneva, Switzerland, 2021.

[310] Там же.

[311] Steimers A., Bömer T. Sources of Risk and Design Principles of Trustworthy Artificial Intelligence / A. Steimers, T. Bömer // Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. AI, Product and Service. HCII 2021. Lecture Notes in Computer Science; Duffy, V.G., Ed.; Springer. – Berlin/Heidelberg, Germany, 2021. – Vol. 12778. – Pp. 239-251

Не зря еще на заре технологий ИИ Н. Винер предостерегал, что цифровые вычислительные машины – это нечто, что коренным образом отличается от всех предыдущих механических средств и способно разрушить привычный социальный порядок[312]. Конечно, следует учитывать, что разные технологии, автоматизирующие работу, имеют разный потенциальный эффект. Тем не менее, некоторые из них способны привести к потере человеком работы, которую он ценит и считает важной и значимой, лишить его рабочего места, законного заработка и средства обеспечить себя и своих близких всем необходимым. Более того, технологии ИИ могут сделать это очень тонко, убрав определенные задачи в алгоритмах действий работников, сделав участие человека в них не обязательным, не выгодным и излишним для работодателя.

Уникальная особенность ИИ как технологии заключается в том, что он обладает уровнем интеллекта, беспрецедентным по сравнению с другими предшествующими ему в истории развития технологиями (информационными и коммуникационными), проявляющемся в способности ИИ принимать решения, решать проблемы и даже думать в узкоспециализированном понимании этого слова. Системы ИИ, подробно изучая описание наших действий и повторяя за нами, могут овладеть необходимым навыком, качественно выполняя гораздо больше когнитивных задач, чем среднестатистический работник-человек. Тем самым системы ИИ обретают возможность влиять на характер происходящих изменений, порождая новые, неожиданные, противоречивые и негативные последствия для занятости людей, для их самооценки, самореализации, понимания своего места и роли в мире.

Можно предположить, что если системы ИИ берут на себя более сложные и важные задачи, то последствия могут быть непредсказуемыми. Пример с GPT 4 демонстрирует, что профессии, которые ранее считались

[312] Dyson G. Turing's cathedral: The origins of the Digital Universe / G. Dyson. – New-York : Vintage. 2012. – 464 p.

прерогативой высококвалифицированных кадров, становятся теперь уязвимыми для автоматизации системами ИИ и создают риски трудоустройства, профессиональной востребованности для специалистов, занятых в творческих и интеллектуальных видах деятельности.

Системы ИИ активно внедряются в области живописи и музыки. Известен пример, когда ИИ сочинил целое музыкальное произведение[313]. Тогда как относиться к тому творческому продукту, который осуществляется системами ИИ? Можно ли результат их «творчества» воспринимать как нечто, заслуживающее внимания? Гораздо больше опасений вызывает то, что подобные программные продукты основываются на генетическом программировании, которое нацеливается на постоянное обучение до тех пор, пока не начнет решать задачу близко к ее идеальному варианту выполнения. Тогда возникает вопрос, возможно ли, что алгоритмы научатся в ближайшем будущем создавать за нас, например, правовые концепции или решать задачу в области государственного управления?

Обратимся к еще одному примеру – системе ИИ под названием AlphaFold, предназначенная для автоматизации и ускорения процесса предсказания структуры белка в разработке новых методов лечения болезней человека[314]. Эта система может решить быстро и точно задачу, которую обычно выполняли ученые, обучавшиеся для ее выполнения годами. Подобная ситуация создает значительные риски для способности этих специалистов использовать имеющиеся у них знания и умения, демонстрировать свое мастерство и использовать имеющиеся у них навыки, а значит, быть и ощущать себя профессионалами, нужными, востребованными в обществе людьми, следующими своему предназначению, своей миссии.

[313]Iamus, a music-making computer, could be the next Mozart. – URL: https://www.vice.com/en/article/pgg8yy/iamus-a-music-making-computer-could-be-the-next-mozart (дата обращения: 21.01.2023)

[314] Hassabis D., Revell T. With AI, you might unlock some of the secrets about how life works / D. Hassabis, T. Revell. – New Scientist. – 2021. – Vol. 249(3315). – Pp. 44-49.

Проблема деквалификации актуализирует социально-политический аспект применения систем ИИ, вопросы о том, каким образом общество и государство должно реагировать, если большие массы людей окажутся внезапно лишенными источника дохода без перспективы найти себе другое занятие, чтобы удовлетворять свои базовые потребности, иметь возможности и условия для личной и социальной реализации. М. Форд отмечает, что в настоящее время политическое регулирование не способно минимизировать последствия структурных изменений, вызванных внедрением цифровых технологий, уменьшить растущее неравенство в контексте сокращения доли труда в национальном доходе. При этом он подчеркивает повсеместное присутствие ИИ, тот факт, что с каждым днем остается все меньше профессий, где человека не заменила бы машина[315].

Кроме очевидной экономической угрозы материальному благополучию людей, существует еще один не менее важный вопрос, связывающий потерю работы с экзистенциальной угрозой потери смысла жизни[316], о котором, в частности, пишут Ким и Шеллер-Вольф. Более того, зафиксировав данную взаимосвязь, авторы указывают на сохранение актуальности проблемы смыслоутраты даже при компенсации потери дохода, возникшей вследствие автоматизации. Безработные страдают от потери навыков, социальной изоляции и отсутствия цели в жизни, утрата работы оказывает долгосрочное негативное воздействие на их социальное самочувствие [317]. В более поздних исследованиях появился такой феномен как «смерть от отчаяния», возникающий в связи с отсутствием оплачиваемой работы[318]. Это означает, что автоматизация может нести потенциальную экзистенциальную угрозу для человечества, поскольку без ра-

[315] М. Форд. Роботы наступают: развитие технологий и будущее без работы / пер.с англ. С. Чернина. – М.: Альпина нон-фикшн, 2016. – 572 с.
[316] Kim T. W., Scheller-Wolf A. Technological unemployment, meaning in life, purpose of business, and the future of stakeholders/ T. Kim, A. Scheller-Wolf // Journal of Business Ethics. – 2019. – Vol. 160(2). – Pp. 319-337.
[317] Clark A., Oswald A. Unhappiness and unemployment / A. Clark, A. Oswald // Economic Journal. – 1994. – Vol. 104(424). – Pp. 648-659.
[318] Case A., Deaton A. Deaths of despair and the future of capitalism / A. Case, A. Deaton // Princeton University Press. – 2020.

боты гораздо больше людей имеют шанс столкнуться с хроническими заболеваниями и, в крайнем случае, утратив надежду, сделать непоправимый и неправильный выбор против самой жизни.

Несколько иная ситуация складывается, если человек занимает роль оператора, осуществляющего контроль за эффективностью работы систем ИИ[319]. На первый взгляд, такой вид работы может показаться более щадящим вариантом по сравнению с первым сценарием. Однако человек, постоянно выполняя повторяющиеся и фрагментарные по характеру действия, отчуждается от результатов своей работы и утрачивает возможность видения целостной картины. Такой характер работы дает мало возможностей для развития разнообразных навыков, для раскрытия и реализации собственного потенциала.

Таким образом, мы видим, что системы ИИ, выполняя значимую, интересную, творческую часть работы человека, способны отчуждать его от непосредственной производительной деятельности, что негативно влияет на родовую сущность человека как производящего существа, единственного в мире субъекта природопреобразующей деятельности, направленной на адаптацию, приспособление мира к удовлетворению потребностей человека, занимающего главное, центральное положение в известной и доступной ему реальности. Согласно Марксу, рабочие с появлением машин сталкиваются с технологическим отчуждением и борются с технологической безработицей. В контексте применения систем ИИ работники отчуждаются от продукта собственной деятельности, который вступает с ними в жесткую конкуренцию на рынке труда, все больше превосходя их способности и возможности. Человек также отчуждается от самого себя, от своей родовой сущности, поскольку все меньше ему удается решать задачи, соответствующие его творческому общественно-производственному потенциалу. Он перестает быть целью и становится

[319] Langlois R. N. Cognitive comparative advantage and the organization of work: Lessons from Herbert Simon's vision of the future / R. N. Langlois // Journal of Economic Psychology. – 2020. – Vol. 24(2). – P. 174.

средством достижения целей, поставленных не им. Системы ИИ усугубляют процесс отчуждения, сталкивая человека с чем-то чуждым ему по своему происхождению и субстрату, но в то же время сверхъестественно знакомым по своим способностям. Если бы системы ИИ соотносились с механическим ткацким станком или паровой машиной, то экономическое положение человека не подвергалось бы такой серьезной угрозе. Если раньше рабочий беспокоился об утрате работы в своем профессиональном секторе, то у него всегда была возможность переобучиться и снова обрести свою безусловную ценность с экономической точки зрения. Использование систем ИИ в современном мире чрезвычайно усложняет, проблематизирует эту возможность, стремительно сокращая перечень областей деятельности человека, не подлежащих автоматизации.

Таким образом, можно сделать вывод, что системы ИИ, в отличие от других технологий, в перспективе могут стать не ограниченными в своих возможностях и способными обучаться и осуществлять производственный процесс во всех областях жизни социума. Соответственно, возникает риск отчуждения человека от самого себя вследствие его взаимодействия с системами ИИ, наделенными человеком его самой существенной человеческой особенностью, способностью к интеллектуальной деятельности и построенной на ее основе целенаправленной материальной деятельности, нацеленной на изменение, преобразование окружающей реальности, исходя из целей и задач человека, человеческой цивилизации. Именно это составляет главное отличие технологии ИИ от всех иных разработанных ранее технологий, которое обусловливает широкий перечень проблем, рисков и опасностей, возникающих в процессе развития человека и общества, развития их взаимодействия с системами ИИ. В краткосрочной перспективе, как было установлено выше, применение систем ИИ грозит безработицей, а в долгосрочной перспективе возникает угроза не только полной безработицы, но и настоятельной необходимости защищать особый статус человека в мире.

По отношению к проблеме причинения вреда все остальные трудности выступают в роли вторичных, подчиненных, раскрывающих и уточняющих смысл главной проблемы. Они призваны выявить источник, причины возникновения нежелательных для человека последствий, а также описать разнообразные проявления возможного ущерба для человека, и последствия, возникающие в случае некорректного внедрения, применения систем ИИ или ошибок, допущенных создателями таких систем в процессе их разработки.

Социальная несправедливость. Помимо указанных выше проблем, в процессе исследования результатов применения ИИ выяснилось, что системы ИИ могут быть предвзятыми в отношении определенных индивидов или групп, подвергая их несправедливой дискриминации по этническим, гендерным или религиозным основаниям, что позволяет сделать вывод о наличии проблемы социальной несправедливости**.** Ее обсуждению посвящен целый ряд работ, появившихся в последнее время.

Лагерь исследователей по теме справедливости делится на тех, кто считает, что справедливость возможно алгоритмизировать и ввести в системы как измеряемый показатель[320] и тех, кто уверен, что справедливость нельзя толковать с точки зрения четко определенных количественных показателей[321].

Несправедливость в контексте применения ИИ принято обозначать в литературе как алгоритмическую предвзятость[322]. Справедливость выражается главным образом в предотвращении или смягчении последствий нежелательной предвзятости и дискриминации. Источники определяют

[320] Michael V., Van Kleek M., Binns R. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making / V. Michael, M. Van Kleek, R. Binns // Proceedings of the 2018 chi conference on human factors in computing systems. – 2018. – Pp. 1-14.

[321] Holstein K., Wortman J. et al. Improving fairness in machine learning systems: What do industry practitioners need? / K. Holstein, V. Wortman, J. Daumé, M. Dudík, H. Wallach // Proceedings of the 2019 CHI conference on human factors in computing systems. – 2019. – Pp.1-16.

[322] Ryan M., Antoniou J. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality. – URL: https://edepot.wur.nl/543861 (дата обращения: 12.09.2021)

несправедливость систем ИИ как отсутствие разнообразия[323], интеграции[324] и равенства[325]. Представители государственного сектора по проблеме несправедливости систем ИИ уделяют особое внимание их влиянию на рынок труда[326].

Принято считать, что системы ИИ могут помочь при справедливом принятии решений и способствовать предотвращению предвзятости[327]. Однако практика показала, что принятие решений системами ИИ зачастую воспроизводит существующие в обществе системные, институциональные и социальные предубеждения. Такие предубеждения могут привести к дискриминации, несправедливости и проблемам нарушения конфиденциальности, нарушению прав человека**.**

Так, в большинстве европейских государств в сфере обеспечения правопорядка применяются системы ИИ для профилирования людей, попыток предсказать вероятное преступное поведение людей или места будущих преступлений, а также для оценки предполагаемого риска причастности отдельных лиц к совершению преступлений (например, список преступников 600 (Нидерланды))[328]. Результаты этих прогнозов используются для слежки, проведения обысков или допросов лиц, которым системой ИИ был присвоен высокий рисковый рейтинг.

Оценивая работу подобных систем, специалисты пришли к неожиданным результатам. Так, Дж. Ангвин, Дж. Ларсон, С. Матту в своем ис-

---

[323] Green Digital Working Group. Position on Robotics and Artificial Intelligence. – URL: https://felixreda.eu/wp-content/uploads/2017/02/Green-Digital-Working-Group-Position-on-Robotics-and-Artificial-Intelligence-2016-11-22.pdf (дата обращения: 12.01.2022)

[324] SAP's guiding principles for artificial intelligence. – URL: https://www.sap.com/products/leonardo/machine-learning/ai-ethics.html#guidingprinciples. (дата обращения:12.09.2021)

[325] The Future Society. Science, Law and Society (SLS) Initiative. The Future Society. – URL: https://web.archive.org/web/20180621203843/http://thefuturesociety.org/science-lawsociety-sls-initiative/ (дата обращения: 19.06.2022)

[326] UNI Global. 10 Principles for Ethical AI. – URL: http://www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/ (дата обращения: 10.12.2021)

[327] Kleinberg J., Mullainathan S., Manish R. Inherent Trade-Offs in the Fair Determination of Risk Scores / J. Kleinberg, S. Mullainathan, R. Manish // 8th Innovations in Theoretical Computer Science Conference. Schloss Dagstuhl-Leibniz-Zentrumfuer Informatik. – 2018. – Pp. 1-23.

[328]Pushback against AI policing in Europe heats up over. – URL: https://www.globaltimes.cn/page/202110/1237232.shtml (дата обращения: 09.08.2022)

следовании выяснили, что системы ИИ предвзяты в автоматических оценках возможности повторения преступления осужденными[329].

Системы искусственного интеллекта применяются также в сфере образования и трудоустройства для отбора кандидатов при поступлении в ВУЗы, для найма работников, например, с просмотром резюме и для целевой рекламы вакансий[330]. М. Боген и А. Рик отметили, что риск систематической ошибки был зафиксирован на каждом из этих этапов процесса найма[331]. Применение систем ИИ в указанном направлении, безусловно, имеет преимущества в виде автоматизации рутинной работы и сокращения временных затрат. Однако возникают опасения по поводу того, насколько адекватной является оценка системами ИИ навыков соискателей, на каких основаниях и фактах принимается решение, не подвергается ли угрозе безопасность социума и отдельного человека.

В финансах и банковском деле алгоритмы ИИ составляют основу множества различных приложений, таких как прогнозирование рынка для торговли, управление рисками для оценки кредитоспособности, распределение кредитов и определение ставок по ипотечным кредитам[332]. Учеными выявлены многочисленные случаи, когда решения по кредитным заявкам были несправедливыми и предвзятыми по отношению к заемщикам из числа национальных и гендерных меньшинств. Речь идет, в частности, о более высоких процентах отказов в ипотеке и кредите для латиноамериканских и чернокожих заемщиков в США[333] или более низких

[329]Angwin J., Larson J. et al. Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks / J. Angwin., J. Larson, S. Mattu, L. Kirchner. – URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (дата обращения: 13.08.2022)

[330] Tristram A. E. AI in talent acquisition: A review of AI-applications used in recruitment and selection / A. E. Tristram // Strategic HR Review 18, 2019. – Vol.5. – Pp. 215-221.

[331] Bogen M., Rieke A. Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias / M. Bogen, A. Rieke. – URL: https://apo.org.au/node/210071 (дата обращения:09.07.2022)

[332]Longbing C. AI in finance: Challenges, techniques, and opportunities / C. Longbing // Comput Surveys. – 2022. – № 55. – Pp.1-38.

[333]Bartlett R., Morse A., Stanton R., Wallace N. Consumer-lending discrimination in the FinTech era. – URL: https://ideas.repec.org/p/nbr/nberwo/25943.html (дата обращения: 21.09.2022)

кредитных лимитах для женщин по сравнению с мужчинами, имеющими равные кредитные характеристики[334].

Рекомендательные системы в поисковиках, основанные на ИИ, помогают пользователям ориентироваться в сети, предлагают интересующую их информацию. Было установлено, что рекомендательные системы усиливают различные виды предвзятости, усиленно предлагая материалы с чрезмерным участием мужчин, представителей европеоидной расы и молодых пользователей[335], или представляя в качестве лучших только определенные предприятия и маркетплейсы[336]. Это увеличивает дисбаланс сил между доминирующими на рынке крупными платформами и более мелкими, которые не имеют доступа к равным объемам высококачественных потребительских данных, которые жизненно важны для выхода на рынок.

На наш взгляд, такая концентрация власти в руках очень небольшого числа компаний, которые разрабатывают большинство приложений ИИ и действуют при этом из соображений получения ими максимальной прибыли, а не в интересах развития общества, является дополнительной угрозой для институтов демократии. Поскольку рекомендательные системы поиска нацелены на привлечение внимания и потому определяют большую часть информации, которая становится известна пользователям, высок риск искаженного восприятия последними информации, что в перспективе способно лишить людей возможности осознанно участвовать в здоровом политическом и социальном дискурсе. Так, системы искусственного интеллекта крайне негативно влияют на участие избирателей в демократических выборах, подрывая сами основы демократического ми-

[334]Haridasani G. A. Are Algorithms Sexist? / G. A. Haridasari. – URL: https://www.nytimes.com/2019/11/15/us/apple-card-goldman-sachs.html (дата обращения: 21.05. 2022)
[335]Ribeiro F. N. Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale / Filipe N. Ribeiro, L. Henrique, F. Benevito et al. – URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17878 (дата обращения: 06.09.2022)
[336]Dash A., Chakraborty A. et al. When the umpire is also a player: Bias in private label product recommendations on e-commerce marketplaces / A. Dash, A. Chakrabatov, S. Ghosh, A. Mukherjee, P. Krishna // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. – 2021. – Pp. 873-884.

роустройства, посредством назойливо предлагаемой теми или иными группировками недобросовестной целевой рекламы.

Особенно очевидной становится социальная несправедливость в контексте европейского законодательства, согласно которому мужчины и женщины должны пользоваться равными правами на рынке труда и иметь равный доступ в сфере предложения товаров и услуг[337]. Также должно быть гарантировано равное отношение к людям, независимо от их расового или этнического происхождения[338]. Аналогичным образом должно соблюдаться равенство в сфере труда и занятости, если речь идет о людях с инвалидностью, о людях с различными религиозными предпочтениями, убеждениями, о людях различных возрастных групп и сексуальной ориентации[339].

С. Вахтер, Б. Митльштат, С. Рассел [340], изучив целый ряд европейских документов, пришли к выводу, что, в отношении ИИ чрезвычайно сложно определить, были ли в действительности нарушены права людей в том или ином случае, поскольку алгоритмы ИИ осуществляют свою деятельность с такой скоростью, масштабом и уровнем сложности, которые в принципе не сопоставимы с уровнем человеческого интеллекта, не поддаются человеческому пониманию[341]. Например, потребителям сложно оценить, была ли им предложена лучшая возможная цена. Они даже не имеют возможности узнать о том, что определенные рекламные объявления им вовсе не были предложены. Масштабы и эффекты автоматизированного принятия решений не позволяют людям понять, что они оказались в обделенном, ущербном положении, и, как результат, у них не бу-

[337] Council of the European Union. Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services. – 2004.
[338]Там же.
[339]Там же.
[340] Wachter S., Mittelstadt B. Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI / S. Wachter, B. Mittelstadt, C. Russel // Computer Law & Security Review. – 2021. – Vol. 41. – Pp.105-567.
[341] Burrell J., How the Machine “Thinks:” Understanding Opacity in Machine Learning Algorithms / J. Burrell // BIG DATA SOC. – 2016. – Pp.1-12.

дет явных оснований для предъявления иска в соответствии с законом о недискриминации. Использование систем ИИ, таким образом, поднимает вопрос о справедливости на фундаментально новый уровень, предполагающий выполнение требований о методах обнаружения возможной дискриминации и обязательном представлении доказательств, свидетельствующих о том, что права людей не были ущемлены.

Перед разработчиками и пользователями ИИ встает вопрос, может ли справедливость быть автоматизирована? ИИ дискриминирует людей так, что становится очень сложно, практически невозможно обнаружить, расследовать и предотвратить случаи дискриминации. Справедливость часто зависит от контекста и не может (и, возможно, не должна) быть принципиально автоматизирована. Поэтому традиционные (например, правовые) процедуры обеспечения справедливости при принятии решений должны быть более последовательными, постоянными и четкими в своих определениях. Даже если появятся стандартные показатели и пороговые значения, позволяющие в каких-либо ситуациях отличить справедливое от несправедливого, проблема несправедливости останется, ибо она никогда не сможет охватить все возможные случаи нарушения нормы.

Еще один аспект рассматриваемой проблемы возникает из-за сложного характера современных систем машинного обучения, действия которых осуществляются по принципу «черного ящика», когда проследить за целым рядом факторов, составляющих алгоритм деятельности ИИ, не представляется возможным.

Проблема заключается в том, что человек не знает, каким образом системы ИИ, основанные на нейронных сетях, приходят к тем или иным результатам. Например, вышеупомянутый алгоритм COMPAS, оценивающий вероятность совершения насильственного преступления, как оказалось, дискриминирует преступников с темным цветом кожи. Однако, как показало исследование, система в действительности не учитывала расовые признаки в качестве входных данных. Вместо этого она получила

конфиденциальные данные о расовой принадлежности в качестве следствия из другой доступной ей информации о месте жительства потенциального преступника.

Аналогичным образом две программы ИИ, которые независимо друг от друга научились распознавать изображения лошадей, использовали при этом совершенно разные подходы[342]. Первая программа ИИ правильно сосредоточилась на особенностях животного, вторая строила выводы на основании набора пикселей в нижнем левом углу каждого изображения лошади. ИИ приходил к верным заключениям, исходя из совершенно ошибочных посылок.

В сфере здравоохранения существует специфический риск применения технологии ИИ, заключающийся в том, что они могут усугубить существующее неравенство, обусловленное, в частности, этнической принадлежностью, социально-экономическим статусом, возрастом пациента. Сбор данных для системы ИИ при этом может привести к получению некорректных сведений вследствие языковых барьеров, а также того обстоятельства, что люди будут предоставлять неверную или неполную информацию из-за испытываемого ими недоверия к автоматизированным системам. Ошибки приложений ИИ, в свою очередь, в конечном итоге могут негативно сказаться на клинических результатах пациентов, на состоянии их здоровья. Человеческое измерение здравоохранения предполагает уникальные отношения между профессионалом и пациентом, пронизанные особыми ценностями и обязанностями. Это отношение рассматривается учеными как требующее особого подхода, ориентированного на пациента, предполагающего автономию пациента и способствующую его осознанному выбору, который человек имеет право и может сделать в соответствии с разделяемой им системой ценностных установок.

[342]Lapuschkin S., Wäldchen S. Unmasking Clever Hans predictors and assessing what machines really learn / S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek and K.R. Müller // Nature Communications. – 2019. – Vol.10 (1). – P.1096.

В сфере сельского хозяйства внедрение ИИ приводит к увеличению цифрового разрыва между отдельными фермами и странами. Одни активно используют ИИ и получают представляемые им преимущества, другие, в силу отставания в технологическом развитии, лишены такой возможности. Задача общества, государства состоит в том, чтобы обеспечить справедливую и инклюзивную пользу от ИИ для всех производителей и пользователей. С другой стороны, как отмечает М. Райан, проблема заключается в том, что ИИ приносит экономическую выгоду агробизнесу и технологическим компаниям, но не самим фермерам, которые зачастую получают экологический и социальный ущерб от применения систем ИИ[343]. Цвык В. А и И. В. Цвык отмечает, что, несмотря на явные преимущества применения ИИ в сельском хозяйстве, он может способствовать отчуждению человека от животного мира и окружающей среды[344].

Справедливость связана также с солидарностью. Последняя обсуждается чаще всего в связи с последствиями использования систем ИИ на рынке труда, в контексте обеспечения социальной защиты и равного отношения ко всем претендентам на вакантные рабочие места. Солидарность выражается в том, чтобы добиться справедливого перераспределения преимуществ, предоставляемых ИИ, осуществляющим подбор сотрудников для имеющихся вакансий между претендентами из всех без исключения социальных групп. А. Джобин, М. Йенка и Е. Вайена, в частности, описывают существующие на рынке труда практики сбора данных, ориентированные на отдельных, соответствующих неким определенных требованиям лиц, которые своим избирательным механизмом фактически наносят ущерб социальной солидарности и действуют по принципу «радикального индивидуализма»[345].

---

[343] Ryan M. Agricultural big data analytics and the ethics of power / M. Ryan // Journal Agric Environ Ethics. – 2020. – Vol. 33 (1). – Pp.49-69.

[344] Цвык В. А., Цвык И. В. Социальные проблемы развития и применения искусственного интеллекта / В. А. Цвык, И. В. Цвык // Вестник РУДН. – Серия: Социология. – 2022. – №1. – С. 58-69.

[345] Jobin A., Ienca M., Vayena E. Artificial Intelligence: The Global Landscape of Ethics Guidelines / A. Jobin, M. Lenca, E. Vayena // Nature Machine Intelligence. – 2019. – Vol.1. – Pp. 389-399.

По нашему мнению, внедрение и использование систем ИИ во всех сферах общественной жизни должно строиться на основе справедливого, равного, беспристрастного отношения ко всем участникам разнообразных процессов социального взаимодействия. Подобные требования выдвигают сегодня многочисленные международные организации, комитеты по этике, научные институты, разрабатывающие собственные этические кодексы, а также ученые, участвующие в работе научных конференций, подобных конференции ACM «Справедливость, подотчетность и прозрачность».[346] По мере того, как все больше и больше решений делегируется человеком системам ИИ, общество, его институты должны обеспечить, чтобы эти решения были свободны от любой предвзятости и дискриминации.

Нарушение автономии. Данная проблема, специально выделяемая исследователями систем ИИ, заключается в том, что последние могут посягать на автономию человека, ущемляя свободу принятия им решения и уязвляя таким образом человеческое достоинство. У людей отнимают саму возможность самостоятельно, без помощи ИИ принимать осознанные и независимые решения, что можно трактовать как утрату человеком его права на самоопределение. Л. Н. Мешкова утверждает, что ИИ и подобные ему технологии способны существенно влиять на свободу человека, порождая новые зависимости и ограничения[347]. Системы ИИ, например, могут прямо или косвенно навязывать людям определенный образ жизни, применяя репрессивные механизмы наблюдения или стимулирования[348]. В результате, страдает процесс целеполагания, теряются навыки самостоятельности, совершается менее аутентичный выбор, принцип человеческой автономии фактически нарушается. Привлекательные на первый

[346] ACM FAccT Conference. – URL: https://facctconference.org// (дата обращения: 04.01.2022)
[347]Мешкова Л. Н. Цифровые технологии как фактор трансформации культуры / Л. Н. Мешкова // Вестник Бурятского государственного университета. Философия. – 2020. – Вып. 3. – С. 53-60.
[348]Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).

взгляд перспективы использования систем ИИ (расширение когнитивных способностей человека, освобождение человека от монотонного труда, повышение качества его жизни и обслуживания и др.) сопровождаются прокламациями о неминуемых угрозах и рисках для самостоятельности и автономии человека, неизбежно возникающих вследствие внедрения ИИ в жизнь человеческого общества.

Проблема нарушения автономии особенно наглядна в сфере медицины, где системы ИИ преобразуют базисные структуры врачевания, меняют социальный контекст оказания медицинской помощи. Преобразования эти имеют системный характер, поскольку кардинально затрагивают все аспекты деятельности врача, изменяя диагностические процедуры[349], практики принятия врачебных решений[350], порядок проведения необходимых манипуляций (например, хирургических операций[351], послеоперационного ухода[352]).

Системы ИИ в области медицины уже сегодня претендуют на роль автономных, свободно действующих субъектов, выполняя ассистирующую и консультирующую функции. Роботы-ассистенты в некоторых клиниках занимаются уходом за больными после тяжелых операций[353], а программы на основе ИИ консультируют врачей, помогая им в принятии решений[354]. В области инновационной медицины уже ведутся дискуссии о том, как разработать и внедрить полностью автономную систему ИИ. Так, на заседании президиума РАН по развитию робототехники в меди-

[349] Abdulkareem M., Leiner T., Petersen S.E. Artificial intelligence will transform cardiac imaging – opportunities and challenges / Abdulkareem M., Leiner T., S. E. Petersen // Front Cardiovasc Med. – 2019. – Vol.6. – 133 p.

[350]Shortliffe E. H., Sepúlveda M. J. Clinical decision support in the era of artificial intelligence / E. H. Shortliffe, M. J. Sepúlveda // JAMA. – 2018. – Vol. 320 (21). – Pp. 2199-2200.

[351]Там же.

[352] Hashimoto D.A et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy / D. A. Hasimoto, G. Rosman, E. R. Witkowski // Ann Surg. – 2019. – Vol. 270 (3). – P.21.

[353]Bian Y.et al. Artificial intelligence–assisted system in postoperative follow-up of orthopedic patients: exploratory quantitative and qualitative study / Y. Bian, Xiang, B. Tong, B. Feng, X. Weng // J. Med. Internet Res. – 2020. – Vol. 22 (5). – 23 p.

[354]Обзор российских систем принятия врачебных решений (СППВР) // WEBIOMED 30 января 2021 г. – URL: https://webiomed.ai/blog/obzor-rossiiskikh-sistempodderzhki-priniatiia-vrachebnykh-reshenii/ (дата обращения: 12.06. 2022)

цине, в частности, было заявлено о необходимости создания роботов, полностью заменяющих хирургов[355]. О. Р. Чепьюк выражает свое беспокойство тем, что четвертая промышленная революция может стать непосредственной угрозой автономии человека, когда на «умных заводах» и в «умных» отраслях решения будут приниматься ИИ[356].

Проблема нарушения автономии, в свою очередь, порождает проблему деквалификации сотрудников и проблему доверия. У врачей, которые все больше опираются на ассистирующие системы ИИ, со временем меняются методы принятия решений (имеется в виду обоснованное принятие решения), стиль обучения, снижается качество медицинских знаний, вырабатывается стереотипное, небрежное, обезличенное отношение к пациентам. Этот феномен деквалификации (постепенной утраты навыков) выявил Т. Хофф в ходе исследования работы врачей первичного звена[357]. У начинающих специалистов навыки и вовсе не будут развиваться, считает ученый, и врач не сможет адекватно выполнять свои профессиональные обязанности. Опасения по проблеме деквалификации и изменении природы человека выражает М. В. Заладина, утверждая, что чем внушительнее становится влияние ИИ на человеческую природу и окружающую среду, тем труднее его контролировать, тем больше опасность для человека[358]. Л. В. Баева утверждает, что роботизация и замещение человека приведут к значительной трансформации ценностных приоритетов человека, изменятся его воззрения о долге, справедливости, необходимости и т.д. Автор озадачен вопросом о том, меняют ли эти технологии саму

[355] Заседание Президиума Российской академии наук «О внедрении робототехники в отечественную медицину – проблемы и пути решения». – URL: http:// www.ras.ru/news/news_release.aspx?ID=2eaccb0cd887-4d02-bb10-7caae48b083b&print=1 (дата обращения: 23.02.2022)

[356]Чепьюк О. Р. Экономическая бессубъектность как фактор дегуманизации социальных отношений: диссертация на соискание ученой степени доктора философских наук / О. Р. Чепьюк. – Нижний Новгород, 2020. – 371 с.

[357]J. D. Karpicke, J. R. Blunt. Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping / D. Jeffrey // Science, 2011. – Vol.331. – Pp. 772-775.

[358] Заладина М.В. Социально-философский анализ духовного отчуждения: автореферат диссертации на соискание кандидата философских наук. / М. В. Заладина. – Нижний Новгород, 2020. – 25 с.

генетическую и духовную сущность личности[359]. К тому же, системы на основе глубокого обучения не вполне прозрачны, и действия их принципиально непонятны для врача, не говоря уже о том, что они еще более непонятны пациентам. Можно сказать, что подобные системы в определенном смысле обладают субъектностью, поскольку им делегируется часть врачебных полномочий. Это, в свою очередь, вновь поднимает вопрос о том, кто несет ответственность: человек или машина? Некоторые компании даже пытаются запустить проекты, подобные «Moral Machine», имеющие целью наделить ИИ полноценной субъектностью[360].

Внедрение систем ИИ, все больше посягающих на автономию врачей, порождает проблему доверия. Она сводится к необходимости поиска ответа на вопросы о том, будет ли пациент доверять ИИ так же, как врачу, будут ли врачи доверять системам ИИ, принцип работы которых может быть непрозрачен для человека. Непрозрачность решений систем ИИ угрожает автономии пациента, который становится ограниченным в принятии свободных и осознанных решений по поводу своего здоровья на основе неполного информирования. К тому же, если пациент остается неосведомленным о том, с кем он взаимодействует, с врачом или искусственным разумом, то выходит, что система обманывает пациента и нарушает его права. Недоверие системам ИИ выразили врачи в IBM во время клинического испытания Watson Oncology, посчитав рекомендации проекта недостаточно компетентными и даже опасными[361]. Таким образом, чрезмерное доверие системам ИИ может привести к нерешаемым проблемам в сфере оказания медицинских услуг. В связи с этим возрастает необходимость серьезной оценки последствий использования таких систем.

---

[359] Баева Л. В. Социокультурные изменения в условиях развития высоких технологий / Л. В. Баева // Инноватика и экспертиза: научный журнал. - 2012. - №2 (11). - С. 110-119.

[360] Whitby B. Automating medicine the ethical way / B. Whitby // Machine Medical Ethics. – Cham: Springer, 2015. – P. 223-233.

[361] Bioethics briefing note: artificial intelligence (AI) in healthcare and research. – URL: https://www.abhi.org.uk/resource-hub/file/9335 (дата обращения: 12.09.2022)

Проблема нарушения автономии человека в значительной степени проистекает из самообучаемости систем ИИ, их способности независимо от действий производителя/пользователя привести к неправильной последовательности действий (или бездействию) системы. Дело в том, что системы с высокой степенью собственной автономности могут демонстрировать неожиданное поведение и принять неправильное решение, причинив тем самым ущерб субъекту, ограничив его автономию, его самостоятельность в принятии решений. Более того, риски, создаваемые высокоавтоматизированными системами ИИ, в свою очередь, могут усугубляться человеческим фактором (например, свойственным среднестатистическому человеку временем реакции на ситуацию, ее истолкование). Не случайно И. В. Понкин и А. И. Редькина утверждают, что правовое положение систем ИИ должно определяться мерой и природой их автономности от человека[362]. Как правило, высокая степень самостоятельности ИИ ограничивает возможности контроля и влияния на него со стороны человека, а значит и автономию человека. Поэтому разработчикам систем ИИ следует предусмотреть, чтобы действия человека всегда имели приоритет при их использовании, то есть чтобы человек всегда был в центре приложения. Лучший способ обеспечить необходимый уровень автономии человека — вовлечь будущих пользователей, а также экспертов в предметной области в сам процесс разработки системы ИИ. Степень автоматизации должна соответствовать контексту приложения и предоставлять пользователям все необходимые возможности управления. В конечном итоге, это приведет к созданию искусственного интеллекта, ориентированного на человека. Системы ИИ уже сегодня широко используются во многих приложениях, связанных с безопасностью, причем в таких отраслях деятельности, которые сопряжены с чрезвычайно высоким уровнем риска для человека, его жизни и здоровья (например, авиация

[362] Понкин И. В., Редькина А. И. Искусственный интеллект с точки зрения права/ И. В. Понкин, А. И. Редькина // Вестник Российского университета дружбы народов. – Серия: Юридические науки №1, 2018. – С. 91-109.

или работа атомных электростанций). Здесь особенно важно обеспечить, чтобы средства управления системой были понятны и подконтрольны людям, вели себя в работе так же, как на этапе проектирования.

Проблема нарушения автономии проявляется и в невозможности для человека выбирать ту работу, которая способствует саморазвитию, нравственному совершенствованию и позволяет придерживаться собственных ценностей, жить в соответствии с теми принципами, которые они действительно разделяют. В этом контексте Л. Флориди считает, что постепенно разрастающееся делегирование полномочий системам ИИ снижает человеческую автономию[363]. Это означает, что в случае активного участия ИИ в процессе решения задачи, последний может ограничивать человека в его стремлении следовать высоким духовным целям своей деятельности, что приводит к потере важных, существенных человеческих качеств и ограничивает возможности для развития человека. ИИ также может сократить нашу автономию и в том случае, если доступ к передовой технологии будет закрыт для большинства членов общества. Автономия работника, выполняющего функции оператора систем ИИ, может быть нарушена посредством неправомерного ограничения искусственным разумом объема информации, доступной для просмотра и использования[364]. К тому же, если работа оператора сама по себе является обыденной и скучной, то это заставляет человека чувствовать себя «рабом машины» и, по этой причине также испытывать чувство несвободы.

По нашему мнению, куда больший риск для автономии человека возникает из-за слежки и манипуляций со стороны систем ИИ. М. Фуко обозначал данную проблему ростом «общества слежки», которое заставляет людей постоянно чувствовать себя под наблюдением[365]. Когда люди

[363] Floridi L. et al. AI4People - An ethical framework for a good AI society / L. Floridi, J. Cowls, M. Beltrametti et al. // Minds and Machines. – 2018. – Vol. 28(4). – Pp. 689-707.
[364] Kellogg K. C., Valentine M. A. Algorithms at work: The new contested terrain of control / K. Kellogg, M. A. Valentine, A. Christin // Academy of Management Annals. – 2020. – Vol. 14(1). – Pp.366-410.
[365] Abrams J. J. Pragmatism, artificial intelligence, and posthuman bioethics: Shusterman, Rorty, Foucault / J. J. Abrams // Human Studies. – 2020. – Vol. 27(3). – Pp. 241-258.

находятся под наблюдением, они, как правило, ощущают скованность и действуют менее свободно. Использование систем ИИ для наблюдения за работниками, как мы предполагаем, будет иметь аналогичные последствия и может стать для работодателей средством усиления контроля за работниками. Например, использование камер с искусственным интеллектом для наблюдения за водителями службы доставки Amazon заставляет их чувствовать себя зажатыми, ограниченными и неспособными действовать автономно[366]. Некоторые системы ИИ внедряются для мониторинга онлайн-совещаний с целью осуществления наблюдения за вовлеченностью работников в процесс обсуждения[367], что может привести к стрессу, депрессии и пр. негативным последствиям ощущения собственной несвободы.

Слежка за рабочими с помощью систем ИИ и работа в качестве оператора на службе у машины может привести также к профессиональной поляризации, к разделению людей на неквалифицированных и привилегированных работников. Последним при этом достается более интересная и увлекательная работа. Эти опасения отсылают к проблеме социальной несправедливости, возникающей при распределении доли преимуществ и объема тяжелой, рутинной работы на каждом рабочем месте, что подрывает солидарность в коллективе, раскол между теми, кто получает выгоду от внедрения ИИ, и теми, кто ее не получает. Например, в работе любого колл-центра ИИ может использоваться для мониторинга и оценки звонков, осуществляемых и поступивших каждому оператору в отдельности. Такое усиленное наблюдение может быть воспринято операторами как навязчивое и ограничивающее их автономию. Но, с другой стороны, подобное использование систем ИИ может обеспечить успех организации, помогая ей лучше обучать и управлять деятельностью опера-

[366]Asher-Schapiro A. Amazon AI van cameras spark surveillance concerns / A. Asher-Schapiro // News.Trust.Org. – 2021. – URL: https://news.trust.org/item/20210205132207-c0mz7/ (дата обращения: 12.09.2022)

[367] Pardes A. AI can run your work meetings now / A. Pardes // Wired. – URL: https://www.wired.com/story/ai-can-run-work-meetings-now-headroom-clockwise (14.09.2022)

торов. Однако неизбежно будут формироваться микрогруппы внутри организации, по-разному оценивающие роль ИИ: одни будут видеть в ИИ ненужное и мешающее работе устройство, а другие — удачный способ повышения результативности работы исполнителей. Люди, принадлежащие к первой группе и испытывающее дискомфорт от давления со стороны систем ИИ, могут даже принять решение об увольнении, либо утратить чувство солидарности, ответственности и принадлежности к единому коллективу, сопричастности его успехам и результатам.

Области применения ИИ в будущем будут только расширяться, и человеку необходимо позаботиться о том, чтобы при разработке интеллектуальных систем любого уровня сложности человек не утратил своего центрального положения в мироздании, чтобы в основе любых технологий неизменным оставался подход, ориентированный на человека, на достижение им целей и задач своего развития, на реализацию его подлинного предназначения в мире.

Таким образом, в параграфе мы анализировали основные проявления негативного влияния ИИ на человека, представленные проблемами причинения вреда, социальной несправедливости и нарушения автономии человека. Указанные трудности исследованы нами во взаимосвязи, что позволило выявить роль и значение каждой из названных проблем.

Выделенные в данном параграфе проблемы упоминаются в целом ряде работ, посвященных определению конкретных источников рисков, исходящих от ИИ, однако указанные источники содержат лишь поверхностное и неполное их описание, не вскрывают их взаимосвязь и взаимообусловленность. Так, международные стандарты безопасности ИИ касаются лишь отдельных аспектов причинения вреда (например, необъяснимости или неуправляемости систем ИИ). При этом в них отсутствует обоснованная оценка выделенных рисков и необходимая для практического применения таксономия источников рисков, без чего разработка соответствующих стандартов не может привести к достижению целей

безопасного применения систем ИИ. Поскольку документы дают лишь краткое описание источников рисков, практически невозможно сформировать общее понимание потенциальных трудностей и нежелательных для человека последствий, обусловленных существующими угрозами.

На наш взгляд, проблема причинения вреда, возникающая в процессе взаимодействия человека с системами ИИ, приводит к риску отчуждения человека от самого себя, от собственной сущности и предназначения, возможной утрате и искажению его особого статуса в мире как единственного существа, способного к интеллектуальной деятельности, нацеленной на изменение, преобразование окружающей реальности. Проблемы социальной несправедливости и нарушения автономии призваны уточнить смысл, раскрыть отдельные грани и описать разнообразные проявления причиняемого человеку ущерба, а также последствия, возникающие из-за некорректного внедрения, применения систем ИИ.

2.3 Причины формирования негативных последствий взаимодействия человека и ИИ

Непрозрачность. Первой в списке подобных проблем, по нашему мнению, стоит проблема, которая сегодня привлекает все большее внимание специалистов – это проблема непрозрачности систем ИИ. Она указывает на одну из важнейших причин возникновения негативных для человека последствий применения систем ИИ.

Проблема непрозрачности заключается в том, что системы ИИ из-за постепенной эволюции и усложнения превращаются в своего рода «черные ящики»[368]. Принцип работы этих систем становится все более непрослеживаемым, необъяснимым и неинтерпретируемым. Понять действия или мотивы принятия решения такими непрозрачными системами ИИ становится практически невозможно. Система не демонстрирует ход своей работы, последовательность выстраиваемой ею цепочки рассуждений.

[368] Isak M., Lee A. B. The Right Not to Be Subject to Automated Decisions Based on Profiling / M. Isak, A. B. Lee // EU Internet Law: Regulation and Enforcement / ed. T. Synodinou, P. Jougleux, C. Markou, T. Prastitou. — Springer, 2017. – Pp. 77-98.

Для человека (в том числе разработчика системы) остается неизвестным то, на основании каких данных был совершен тот или иной вывод, получено то или иное заключение. М. Ананни и К. Кроуфорд подготовили подробный обзор множества случаев, связанных с проблемой непрозрачности[369].

Ежегодно в мире проходят международные конференции, посвященные обсуждению прозрачности, справедливости, подотчетности систем ИИ (например, FAT-ML[370], семинар ICML[371]).

В настоящее время известны многочисленные прецеденты нарушения прав людей, возникшие в результате непрозрачной работы систем ИИ. Непрозрачной может быть либо система в целом, либо отдельные решения, принятые системой. Некоторые исследователи обеспокоены тем, что требование к прозрачности систем ИИ находится на нереально высоком уровне, поскольку даже решения, принимаемые человеком, далеко не всегда могут быть понятны и прозрачны для окружающих[372]. Они полагают, что требовать прозрачности всего алгоритма бесполезно, поскольку сам процесс трудозатратен, дорог и не всегда может раскрыть логику принятия решений. Авторы приходят к выводу, что важнее подумать о том, в каких случаях объяснения действий системы полезны для человека. Они считают, что целесообразнее будет установление требований, направленных на раскрытие сведений о логике принятия конкретных решений, касающихся входных данных, статистического анализа, проверки кода и т.д.

[369] Ananny M., Crawford K.. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability / M. Ananny, K. Crawford // New Media & Society, 2018. – Vol. 20(3). – Pp. 973–989.

[370]Fairness Accountability and Transparency in Machine Learning. – URL: http://www.fatml.org/ (дата обращения:19.01.2022)

[371]Workshop on Human Interpretability in Machine Learning. – URL: https://sites.google.com/view/whi2018/. (дата обращения: 12.09.2022)

[372] Zerilli J., Knott A. et al. Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? / J. Zerilli, A. Knott, J. Maclaurin, C. Gavaghan //Philosophy and Technology, 2019. – Vol. 32 (4). – Pp. 661-683.

На наш взгляд, непрозрачность в работе систем ИИ может быть продиктована сложностью алгоритмов ИИ, обусловливающей низкий уровень интерпретируемости результатов деятельности системы, а также вероятностью возникновения ошибок в ее работе.

При этом если от системы требуется получение информации высокой степени точности, то проблема непрозрачности должна быть максимально устранена. Если же некоторые погрешности не оказывают значительного негативного влияния на результат работы системы, то требование прозрачности можно считать излишним. Например, для системы ИИ, подбирающей целевую рекламу для того или иного пользователя социальной сети, достаточно будет относительно низкого уровня интерпретируемости, поскольку последствия ее неправильной работы не столь значительны, они не способны нанести заметный ущерб человеку, потребителю этой рекламы. В то же время интерпретируемость работы системы на основе ИИ, осуществляющей диагностику состояния здоровья человека, должна быть предельно высокой, поскольку любые ошибки могут навредить здоровью пациента, и здесь непрозрачность в действиях и решениях системы недопустима.

В середине 1990-х система ИИ, внедренная в алгоритмы работы одной из клиник и обученная предсказывать, каких пациентов клиники следует госпитализировать, а каких лечить амбулаторно, сделала вывод, что пациенты с пневмонией и астмой имеют более низкий риск смерти и поэтому не подлежат госпитализации. С медицинской точки зрения это решение абсолютно неверно, поскольку пациентов с астмой и пневмонией требуется не просто госпитализировать, но и поместить в отделение интенсивной терапии[373]. Клиника, столкнувшись с подобной проблемой, приняла решение отказаться от данной системы ИИ, и в настоящее время

[373] Caruana R, Lou Y. et al. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission / R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, N. Elhadad // Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. – Sydney, NSW, Australia: ACM, 2015. – Vol. 30. – P. 17-21.

исследователи проводят работу, направленную на то, чтобы сделать клинические системы на основе ИИ более прозрачными.

Беспилотные транспортные средства, создание которых в значительной степени было ответом на стремление снизить количество смертей, происходящих в результате дорожно-транспортных происшествий, из-за непрозрачности принимаемых ими решений могут, как свидетельствует существующий опыт их применения, оказаться еще одной угрозой жизни и здоровью человека. Так, беспилотный автомобиль Uber убил женщину в Аризоне[374]. Это был первый известный случай со смертельным исходом, связанный с полностью автономным транспортным средством. Беспилотные автомобили должны верно оценивать дорожную ситуацию и принимать правильные решения. Однако в данном случае оказалось, что программное обеспечение автомобиля ошибочно распознало возникший перед ним объект как пластиковый пакет, а не человека, переходящего дорогу. Непрозрачность работы алгоритма, управляющего движением автомобиля, послужила причиной гибели человека. Только прозрачная система, решения которой известны и понятны человеку, одобрены человеком, может предотвратить возникновение подобных ситуаций. Работы по исследованию поведения беспилотных транспортных средств с целью объяснить алгоритм их действия и сделать его понятным для людей, уже ведутся[375,376], но каковы будут его результаты, сегодня предсказать практически невозможно.

В правосудии создаваемые системы ИИ могут помочь, например, в оценке риска возникновения рецидивов преступлений. Но здесь важно получить гарантии того, что прогнозирование риска рецидива осуществ-

[374]McFarland M. Uber Shuts Down Self-Driving Operations in Arizona CNN / M. McFarland. – URL: http://money.cnn.com/2018/05/23/technology/uber-arizona-self-driving/index.html. (дата обращения:12.01.2021)

[375] Bojarski M. del Testa D., Dworakowski D.et al. End to end learning for self-driving cars / M. Bojarski. D. del Testa, D. Dworakowski, B. Firner, B. Flepp et al. – 2016. – Pp. 1-9. – URL: http://arxiv.org/abs/1604.07316. (дата обращения: 16.09. 2022)

[376] Haspiel J. Explanations and Expectations: Trust Building in Automated Vehicles / J. Haspiel, J. Meyerson, L. P. Robert Jr. et al. – URL: https://deepblue.lib.umich.edu/handle/2027.42/140746 (дата обращения:18.03.2022)

ляется с позиций справедливости и не ущемляет права человека. Известный пример дела 2013 года «Лумис против Висконсина»[377], когда обвиняемый получил максимальное наказание, свидетельствует о существующей в сфере правосудия опасности принятия необъективных решений вследствие применения систем ИИ. Задействованная в данном случае система ИИ COMPAS, принимая во внимание пол, расу, место жительства и т.п. критерии, присвоила Лумису высокий рисковый рейтинг повторения преступления. Но используемые ею алгоритмы составляли коммерческую тайну, а сам процесс установления причинно-следственных связей между данными анкеты, обрабатываемой COMPAS и предлагаемым ею решением не был очевиден для судьи. Непрозрачность в данной области имела высокие шансы обернуться крайне негативными последствиями для осужденного. О возможности принятия необъективного решения Верховный суд заявил в принятом им решении. Но в настоящее время ситуация практически не изменилась, и автоматизированное принятие решений в правовой системе не стало более прозрачным[378].

В сфере финансовых услуг системы ИИ могут найти применение, например, в процедуре оценки кредитных рейтингов. Трудности начинаются тогда, когда система не объясняет, почему клиенту было отказано в кредите. Такие кредитные бюро, как Equifax и Experian, работают над многообещающими исследовательскими проектами, нацеленными на то, чтобы генерировать автоматические коды причин и сделать решения о кредитных рейтингах на основе ИИ более прозрачными[379].

Принципиально непрозрачная природа систем ИИ обусловлена тем, что они построены на основе искусственных нейронных сетей методом глубокого обучения и способны к самостоятельному извлечению инфор-

[377] Lightbourne J. Damned lies & criminal sentencing using evidence-based tools / J. Lightbourne. – 2017. – Pp. 327-343.
[378] Detecting bias in black-box models using transparent model distillation. – URL: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_96.pdf (дата обращения: 04.05.2022)
[379] Equifax Launches NeuroDecision Technology. – URL: https://investor.equifax.com/news-events/press-releases/detail/203/equifax-launches-neurodecision-technology (дата обращения: 12.09.2022)

мации из массива данных. Результат работы такой системы может оказаться совершенно непредсказуемым. Например, чат-бот Microsoft под названием Tay, сконструированный для общения с людьми в возрасте от 18 до 24 лет, уже через сутки начал проявлять признаки некорректного поведения, допуская нацистские и расистские высказывания[380].

В Национальной стратегии развития искусственного интеллекта, принятой в РФ на период до 2030 года, указано, что алгоритмы работы нейронных сетей могут быть чрезмерно сложными для интерпретации, и, соответственно, результаты их работы могут быть даже отменены человеком[381]. Следовательно, данные, которые получены с помощью систем ИИ, должны быть проверены человеком, а механизм принятия решения должен быть прозрачным.

Учитывая, что системы ИИ будут со временем выполнять гораздо более сложные задачи и оказывать большее влияние на наш мир, чем технологии предыдущих поколений, следует признать, что проблема непрозрачности приобретает еще большую остроту.

Помимо прочего, непрозрачность работы систем ИИ порождает проблему доверия и угрозы достоинству личности, которая может, в конечном счете, остановить дальнейшее внедрение систем ИИ в различные сферы жизни социума и нивелировать объективные преимущества, которыми они в действительности обладают. Трудно и опасно доверить важные решения системе, принцип работы которой скрыт и недоступен.

В школьном округе Хьюстон (США) с помощью системы ИИ было произведено оценивание эффективности работы преподавателей и влияния их на обучающихся. Результаты оценивания впоследствии стали основанием для увольнения ряда педагогов, признанных недостаточно ква-

[380]Чат-бот от Microsoft за сутки научился ругаться и стал расистом. – URL: https://www.interfax.ru/world/500152 (дата обращения: 16.09.2022)
[381]Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). – URL: https://base.garant.ru/72838946/ (дата обращения: 14.05.2022).

лифицированными[382]. В данном случае угроза достоинству личности заключается в том, что оценивание компетентности специалистов производилось на основе данных, обработанных системой ИИ без раскрытия аргументов, послуживших основанием для принятия негативных для карьеры преподавателей решений. Люди должны понимать и знать, на основании чего принимаются такие решения. Но алгоритм работы систем ИИ не позволяет получить эту информацию. В настоящее время проблема интерпретации непрозрачных решений широко обсуждается в самых разных областях, в том числе специалистами в области права и социальной философии, однако консенсус все еще не достигнут.

Отсутствие ответственности. Проблема отсутствия ответственности является одним из наиболее распространенных следствий некорректного применения систем ИИ. В современной литературе неоднократно обсуждались случаи причинения вреда человеку системами ИИ, в которых виновные не были установлены и наказаны, т.е. никто не понес ответственность за гибель и страдания людей. Так, на заводе Volkswagen робот-манипулятор прижал к плите работника, последний не смог выбраться и скончался на месте[383]. Boeing 737 попал в аварию из-за неверных показаний датчика и системы контроля полета[384]. Беспилотный автомобиль не снизил скорость и сбил человека[385].

Проблема отсутствия ответственности состоит, прежде всего, в том, что в подобных описанным выше ситуациях не представляется возможным установить ответственное лицо, субъекта ответственности, на кото-

[382] Gacuta J., Selvadurai N. A statutory right to explanation for decisions generated using artificial intelligence / J. Gacuta, N. Selvadurai // International journal of law and information technology. – Oxford, 2019. – Vol. 28 (3). – Pp. 193-216.
[383] Попова Н. Ф. Основные направления развития правового регулирования использования искусственного интеллекта, роботов и объектов робототехники в сфере гражданских правоотношений / Н. Ф. Попова // Современное право. – 2019. – № 10. – С. 69-73.
[384] Авария Boeing 737 Max глазами разработчика ПО. – URL: https://habr.com/ru/post/449564/ (дата обращения: 3.02. 2022)
[385] Чем опасно внедрение технологий с искусственным интеллектом. Рядом с нами появились электронные личности // Российская газета. – 2020. – № 18 (8072).

рого однозначно можно было бы возложить вину за произошедшие негативные для человека события.

Ю. Тихомиров считает, что в подобных ситуациях «…ответственность, прежде всего, лежит на создателе (изготовителе), отдельно выделяется лицо, осуществляющее обучение ИИ, оператор, собственник и третьи лица, повлиявшие на решение ИИ»[386]. О. Я. Ястребов, сформулировавший концепцию опережающего регулирования, считает проблему отсутствия ответственности одной из основных[387].

Вопрос об ответственности систем ИИ в настоящее время нашел отражение в специальных документах компаний-разработчиков систем ИИ. Например, Google выпустила руководства для регулирования систем ИИ[388], предусматривающие ограничение доступа к отдельным приложениям и технологиям.

В то же время очевидно, что слишком жесткое или непродуманное регулирование может замедлить развитие систем ИИ. Р. Кларк предлагает в этой связи альтернативные подходы к регулированию ИИ, такие как саморегулирование, отраслевое регулирование, механизмы совместного регулирования и формальное право[389]. Он убежден, что регулирование должно быть тщательно разработано и оценено с учетом технических и политических сложностей. К. Форбс отмечает, что некоторые из предложенных видов регулирования активно используются сегодня для стимулирования процессов внедрения ИИ, особенно это касается практики саморегулирования[390].

---

[386] Юридическая концепция роботизации: монография /отв. ред. Ю.А. Тихомиров, С.Б. Нанба. – М.: Проспект, 2019. – С. 89.

[387] Ястребов О. Искусственный интеллект в правовом пространстве: концептуальные и теоретические подходы / О. Ястребов // Правосубъектность: общетеоретический, отраслевой и международно-правовой анализ: сб. материалов к XII Ежегодным науч. чтениям памяти проф. С.Н. Братуся. – М.: Статут, 2017. – С. 280.

[388] Google. AI at Google: Our Principles. – URL: https://ai.google/principles/ (дата обращения: 18.04.2022)

[389] Roger C. Regulatory alternatives for AI / C. Roger // Computer Law & Security Review. – 2019. – Vol. 35. – Pp. 398-409.

[390] Forbes K. Opening the path to ethics in artificial intelligence / K. Forbes // AI Ethics 1. – 2021. – Pp. 297-300.

Вопрос об ответственности систем ИИ особенно остро встает, когда они применяются там, где деятельность их может напрямую влиять на жизнь, здоровье и благополучие людей. Речь идет, например, о применении ИИ в сфере здравоохранения, где ИИ сегодня используется в качестве консультантов (в системах поддержки принятия решений), выступающих в виде самостоятельных автономных субъектов (роботов). Недавно в Китае компанией iFlytekCo. Ltd., был разработан робот под названием Xiaoyi (Сяо И), который успешно прошел экзамен (456 баллов из 600 возможных), позволяющий получить лицензию врача. Врачом он пока не работает, но находит и анализируют информацию о пациентах. Компания IBM разработала суперкомпьютер Watson, который располагает данными около 100 млн. пациентов[391]. Эта система проанализировала 15 млн. страниц биомедицинской литературы, более 200 учебников по медицине, 300 медицинских журналов. Компания-разработчик уверена, что система на деле улучшает качество медицинской диагностики и гарантирует лучшие результаты лечения, будучи способной в каждом конкретном случае дать наиболее точный медицинский диагноз. Однако даже на такую суперамашину нельзя переложить ответственность за ошибочное и, возможно, опасное для жизни и здоровья пациента решение. Онкологические центры, сотрудничающие на этапе запуска с Watson, признали несостоятельность проекта и пришли к выводу, что поглощение такого объема данных — это не то же самое, что их осмысление и разумное использование.

Кто же будет нести моральную и правовую ответственность в случае принятия неправильного решения: лечащий врач, использующий систему ИИ, или эксплуатирующая ее медицинская компания? Может быть, ответственность также ложится на производителя системы ИИ, из-за недостатков которой произошел инцидент? Как распределить ответственность между различными субъектами оказания конкретной медицинской

[391]IBM Pitched Its Watson Supercomputer as a Revolution in Cancer Care: It's Nowhere Close. – URL: https://www.statnews.com/2017/09/05/watson-ibm-cancer/ (дата обращения: 17.09. 2022)

услуги? В настоящее время специалисты ведут бурные дискуссии о статусе ИИ, в том числе юридическом[392], и в ходе обсуждений уже очевидно стремление участников дискуссий к переформатированию этико-правовых оснований взаимодействия врачей и пациентов в случае использования в процессе диагностики и лечения систем ИИ.

Проблема отсутствия ответственности и необходимость ее решения очевидны при исследовании применения беспилотных транспортных средств. На ком лежит ответственность, если произойдет аварийная ситуация: на владельце беспилотного транспортного средства, разработчике технологических основ беспилотного транспортного средства или его изготовителе? Многие считают, что внутри беспилотного транспортного средства всегда должен присутствовать человек, который в любую секунду может и готов взять на себя управление и ответственность, особенно при аварийной ситуации. Этой позиции придерживается большинство стран, которые разрабатывают беспилотные транспортные средства. Она закреплена в положениях Венской конвенции о дорожном движении (1968 г), которая устанавливает, что каждое транспортное средство или состав транспортных средств, находящихся в движении, обязаны иметь водителя[393]. Однако в этом случае транспортное средство перестает считаться беспилотным.

В настоящее время ни одно государство не готово в полной мере использовать для транспортных перевозок беспилотные автомобили, управляемые ИИ. Одна из причин – отсутствие необходимого законодательного регулировании применения беспилотных транспортных средств. Нормативные правовые акты, существующие сегодня, разрознены, слабо коррелируют с существующим законодательством отдельных государств и международными правовыми актами.

---

[392]Nechkin A.V. Constitutional and legal status of artificial intelligence in Russia: present and future / A. V. Nechkin // Lex Russica. – 2020. – Vol. 73 (8). – Pp. 78-85.

[393]Конвенция о дорожном движении (в приложении также технические условия, касающиеся автомобилей и прицепов) (г. Вена 8.11. 1968 г.) // СПС «Консультант Плюс». – С.8. – URL: https://docs.cntd.ru/document/1901133 (дата обращения: 12.09.2022)

В Германии в середине 2017 года отдельным законодательным актом были внесены изменения в действующий Закон о дорожном движении[394], предусматривающие допуск высокоавтоматизированных автомобилей на дороги общего пользования. В документе имеется 20 требований к беспилотному транспорту, их производителям и водителям. Например, требование к ценности человеческой жизни означает, что система автопилота в любой внештатной ситуации должна быть нацелена на сохранение жизни людей. При этом подчеркивается, что автопилот не должен совершать определенный выбор, чья именно жизнь должна быть спасена. Напротив, он должен действовать в интересах всех участников аварии в равной степени. В случае аварии встроенный в автомобиль «черный ящик» определяет, под чьим управлением находилось транспортное средство и, следовательно, уточняет, несет ли ответственность за аварийную ситуацию водитель или производитель беспилотного транспортного средства.

Также в законе утверждается, что водитель может отвлекаться во время движения, но при этом система гарантирует ему «адекватный запас времени» в том случае, если человеку придется взять управление на себя. Очевидно, что водитель не будет иметь права спать, но остается неясным, что законодатель считает «адекватным». Во время юридических прений по данному документу были упомянуты временные промежутки продолжительностью 2-5 секунд. Однако следует иметь в виду, что указанное положение будет применяться как к городскому движению со скоростью 30 км/ч, так и к движению по шоссе со скоростью 130 км/ч.

Настораживает и то, что производители автомобилей не несут прямой ответственности в случае возникновения аварийной ситуации. Эта ответственность целиком ложится на водителя, который, по сути, непосредственно не управляет автомобилем. Подчеркнуто, что немецкий за-

[394]Восьмой закон о внесении изменений в Закон о дорожном движении от 16 июня 2017 года. – URL: http://robopravo.ru/uploads/s/z/6/g/z6gj0wkwhv1o/file/5MZOclyT.pdf (дата обращения: 11.05.2022)

конодатель на данном этапе технологического развития не считает возможным менять действующие правила дорожного движения в части ответственности водителя автотранспортного средства. Подобные изменения, по мнению законодателя, возможно, будет внести только после достижения последней стадии автоматизации.

В Норвегии в январе 2018 года также был принят Закон «Об испытаниях высокоавтоматизированных автотранспортных средств»[395], позволяющий проводить испытание машинам с разным уровнем автоматизации на дорогах общего пользования (даже при полном отсутствии в машине водителя).

В России первый акт регулирования беспилотного транспортного средства в виде Постановления Правительства РФ «О проведении эксперимента по опытной эксплуатации на автомобильных дорогах общего пользования высокоавтоматизированных транспортных средств» появился в ноябре 2018 года[396]. В нормативном акте указывается, что в России планируется проведение правового эксперимента, в рамках которого высокоавтоматизированные автомобили получат возможность двигаться по дорогам общего пользования наравне с другими участниками движения (на территории г. Москвы и Республики Татарстан). В документе также указано, что разрешение могут получить только юридические лица, а ответственность за причинение вреда каждым автомобилем должна быть застрахована на 10 миллионов рублей. Водители этих транспортных средств должны иметь стаж не менее трех лет. При этом в их биографии должен отсутствовать факт лишения их прав на вождение по какой-либо причине. Но самым примечательным и спорным положением является пункт об ответственности, в котором отмечено, что полную ответствен-

[395]Об испытаниях высокоавтоматизированных автотранспортных средств. – URL:https://lovdata.no/dokument/NL/lov/20171215112?q=Lov%20om%20utpr%C3%B8ving%20av%20selvkj%C3%B8rende (дата обращения: 12.09.2022).

[396] О проведении эксперимента по опытной эксплуатации на автомобильных дорогах общего пользования высокоавтоматизированных транспортных средств»: Постановление Правительства РФ от 26 ноября 2018 г. № 1415. – URL: http://publication.pravo.gov.ru/File/GetFile/0001201811270008?type=pdf (дата обращения: 01.10.2012).

ность за все происшествия (кроме тех, что произошли по вине других участников) несет собственник автомобиля. При этом Постановление указывает, что тестируемые беспилотники должны управляться водителем, который находится во время проведения эксперимента на месте водителя. Таким образом, регулирование беспилотных автомобилей ничем не отличается от регулирования обычных серийных автомобилей. Видимо, водитель нужен для того, чтобы его в любой ситуации можно было бы назначить субъектом ответственности за любой возможный вред, который причинит беспилотник.

Мы полагаем, что деятельность, связанная с использованием ИИ, нуждается в специальном правовом регулировании, поскольку речь идет о жизни и здоровье человека. А поскольку системы ИИ могут быть потенциально опасны для людей, необходимо строго лицензировать деятельность, связанную с их производством и эксплуатацией. Для этого необходимо на законодательном уровне четко определить, за кем должна быть закреплена ответственность в случае возникновения нештатных ситуаций, а также ответить на вопрос, возможен или правомерен переход ИИ из статуса объекта в статус субъекта правоотношений.

Европейский парламент рассматривает законопроект о присуждении роботизированным системам ИИ статуса «электронной особы»[397]. Авторы законопроекта полагают, что возрастающая автономия позволяет считать их не вещью, а самостоятельными субъектами правоотношений. Соответственно, этот статус позволит возложить на них и юридическую ответственность за совершаемые ими действия. В документе указано, что «обычные нормы ответственности становятся неэффективными, и появляется необходимость создать новые нормы, которые в первую очередь определят, как аппараты могут быть привлечены к ответственности —

[397]Европарламентарий могут предоставить роботам юридический статус. – URL: https://ria.ru/world/20170114/1485715425.html?inj=1 (дата обращения: 19.08.2022).

полной или частичной — за свои действия или бездействие»[398]. Документ касается роботов, автономных транспортных средств, дронов и т.п. устройств. Данный законопроект многие раскритиковали, около 150 экспертов из 14 стран написали открытое письмо Европейскому парламенту, осуждающее содержание документа[399]. По их мнению, машина не может нести ответственность, поскольку нанесенный ею вред возник вследствие ошибок, допущенных человеком, создавшим алгоритм ее действий или в результате сбоя в работе этого алгоритма. В настоящее время не существует системы ИИ, которая была бы полностью автономной.

Однако попытки превратить системы ИИ в автономного агента можно встретить в других отраслях современной экономики. Например, в 2016 г. в российских и зарубежных СМИ появилась новость, что система ИИ, принадлежащая Сбербанку, в ближайшие годы научится принимать около 80 % всех своих решений с помощью ИИ[400].

Таким образом, проблема отсутствия ответственности состоит в невозможности однозначно определить субъекта и меру ответственности в инцидентах, произошедших с участием систем ИИ, в силу несовершенства существующей законодательной базы, относительной новизны феномена ИИ, поспешного и зачастую непродуманного внедрения указанных систем в различные сферы жизни современного общества.

Правовая регламентация ответственности в случае применения систем ИИ, на наш взгляд, должна быть распределена между всеми участниками процесса разработки, внедрения и использования этих систем. Следовательно, необходимо установить границы ответственности изобретателей, разработчиков, производителей технологий ИИ. Также определенную долю ответственности должны нести пользователи систем ИИ. Речь

[398]Там же.

[399] Ксенофонтова А. Бесправный механизм: почему ученые выступили против присвоения роботам статуса «электронной личности» / А. Ксенофонтова. – URL: https://russian.rt.com/science/article/504118-roboty-evroparlament-yuridicheskoye-lico (дата обращения: 17.05.2022)

[400]Греф: Сбербанк сможет принимать 80 % решений искусственным интеллектом. – URL: https://ria.ru/economy/20160908/1476449735.html (дата обращения: 19.07.2022)

идет об ответственности человека за его деятельность по эксплуатации систем ИИ, которая может привести к созданию ситуаций повышенной опасности, возникших вследствие специфических свойств этих технологий и низкого уровня контроля за процессом со стороны человека. Применение мер уголовной ответственности непосредственно по отношению к системам ИИ нецелесообразно, так как они не обладают должным уровнем самосознания. Добавим также, что вопрос об ответственности в каждом конкретном случае, конечно, должен решаться в судебном порядке.

Нарушение конфиденциальности. Не менее значимой является проблема нарушения конфиденциальности, также возникающая в случае применения систем ИИ. Сегодня они широко используются для сбора, обработки и защиты личных данных. Это необходимо, например, для поиска и отслеживания людей в больших и густонаселенных городах, биометрической аутентификации по голосу и лицу человека, для поиска преступников по поведенческому шаблону и т.д. Указанная проблема возникает из угрозы утечки потока персональных данных или потери контроля над этими данными при использовании систем ИИ. Эти данные могут быть скомпрометированы, опубликованы в открытых источниках и в дальнейшем использоваться мошенниками.

Проблема нарушения конфиденциальности имеет комплексный характер. Во-первых, она проявляется в том, что данные могут собираться или передаваться без явного и осознанного согласия пользователя. Например, во время вспышки лихорадки Эбола в Западной Африке в 2014 г. пришлось экстренно рассекречивать данные и анализировать записи звонков мобильных телефонов пациентов. Эти меры позволили приостановить эпидемию, поскольку эпидемиологи сумели отследить распространение болезни по полученным данным. Важно отметить, что дальнейшая публикация данных была приостановлена из-за опасений по поводу нарушения конфиденциальности и использования рассекреченных данных промышленными конкурентами.

Другим примером является проект социального кредитования в Китае, введенный в 2014 г[401]. Он направлен на регулирование делового и частного поведения граждан, обеспечение объективности в принятии решений о мерах их наказания или поощрения. Главная цель – укрепление доверия граждан к органам власти. В документе «Пекинские принципы ИИ» специально подчеркивается, что «акторы должны иметь достаточное информированное согласие о влиянии системы на их права и интересы»[402]. Казалось бы, речь идет о защите конфиденциальных данных. Но реализация проекта социального кредитования противоречит принципу конфиденциальности. Проект собирает сведения о повседневной жизни человека: совершенных им покупках, поведении в соцсетях, выбираемых им компьютерных играх и т.д., затем переводит полученную информацию в числовые данные по утвержденным правительством правилам. Это делается для того, чтобы произвести оценку положительных или отрицательных сторон личности и выстроить рейтинг социального доверия, который в дальнейшем учитывается при совершении индивидом самых разнообразных действий, вплоть до бронирования отелей или поездок на отдельных видах транспорта, которые становятся недоступны пользователю из-за введенных ограничений. Результатом проекта должно стать повышение прозрачности деятельности всех организаций, органов власти, усиление социального контроля, вовлечение широкой общественности в процесс управления государством, и наконец, достижение главной цели — повышение уровня доверия со стороны общества к органам власти. Однако уже в 2020 году стало очевидно, что проект оценивается в обществе

[401]Chorzempa M. China’s social credit system: a mark of progress or a threat to privacy? / M. Chorzempa, P. Triolo, S. Sacks // Policy Briefs PB18–14, Peterson Institute for International Economics. – 2018. – № PB18-14.

[402]Beijing AI Principles. URL: https://www.baai.ac.cn/news/beijing-aiprinciples-en.html.(дата обращения:19.01.2022)

как нежелательный синтез, «брак между коммунистическим надзором и капиталистическим потенциалом»[403].

Проблема в том, что на практике информированное согласие сводится к тому, что пользователем достаточно формально принимаются условия системы. Он в действительности не осознает, какие именно его персональные данные будут использоваться, в каких ситуациях они будут применены. Помимо этого, использование данных пользователя, сведений о его активности успешно монетизируется различными бизнес-структурами, но он сам об этом практически ничего не знает и никакой компенсации не удостаивается. К тому же, ответственность за сохранность персональных данных несет сам же пользователь. Соглашаясь на обработку персональной информации, он соглашается нести ответственность и за возможные негативные последствия. Многие пользователи думают, что без предоставления ими неких данных о себе система не сможет полноценно функционировать[404], но, на самом деле, приложениями используется лишь малая часть данных, собранных без ведома человека. В результате человек оказывается совершенно беззащитен перед алгоритмами ИИ. Без специального просвещения, образования в этой области он не имеет представления, на что именно соглашается, принимая условия системы. Он не может спрогнозировать, какие данные о нем получит система ИИ, и как они будут в дальнейшем использованы. Он не знает, не будут ли нарушены в данном случае его права.

Практика передачи от человека к системам ИИ конфиденциальных личных данных в современном мире продолжается, принимает новые формы, но последствия и риски этого совершенно не просчитаны и слабо осознаются.

---

[403]Botsman R. Big data meets Big Brother as China moves to rate its citizens. URL: https://www.wired.co.uk/article/chinesegovernment-social-credit-score-privacy-invasion (дата обращения: 14.01.2022)

[404] Большие данные в социальных и гуманитарных науках. Сб. обзоров и рефератов / РАН. ИНИОН. Центр науч.-информ. исслед. по науке, образованию и технологиям / отв. ред. Гребенщикова Е. Г. – М., 2019. –193 с.

Второй аспект проблемы нарушения конфиденциальности заключается в появлении риска утечки собранных данных внешним агентам или попыток деанонимизации[405]. Отдельные пользователи систем ИИ могут стать мишенью для разнообразных вредоносных агентов, что неизбежно приведет к нарушению их прав (например, в том случае, если личная информация используется таким образом, что под угрозой оказывается индивидуальная автономия человека).

Особенно это актуально, когда человек вынужден взаимодействовать с системой ИИ, например, в сфере оказания неотложной медицинской помощи. Здесь существует риск того, что данные пользователей могут быть переданы компаниям, которые используют технологии ИИ для маркетинга товаров и услуг, для извлечения прибыли, для создания и реализации каких-либо продуктов. Эти сведения могут использоваться страховыми фирмами или крупными технологическими компаниями.

Факт подобного использования данных о здоровье людей, собранных компанией Google, зафиксированный в мае 2017 года, имел огромный резонанс и с особой остротой поставил вопрос о том, как, в каких целях может быть вообще использована персональная информация. Google тогда объявила о стратегическом партнерстве с Чикагским университетом и Медицинским университетом Чикаго в США[406]. Целью партнерства стала разработка новых инструментов машинного обучения для прогнозирования медицинских событий (например, срочной госпитализации). Для реализации этой цели университет предоставил Google доступ к сотням тысяч историй болезни «обезличенных» пациентов. Спустя некоторое время, в июне 2019 года один из пациентов Университета М. Динерштейн подал групповой иск против Университета и Google от имени всех

[405] Narayanan A., Shmatikov V. Robust de-anonymization of large sparse datasets / A. Narayanan, V. Shmatikov // SP ’08: Proceedings of the 2008 IEEE Symposium on Security and Privacy. – Washington, DC, USA, 2008. – Pp.111-125.

[406] Wood M. UChicago Medicine collaborates with Google to use machine learning for better health care. (May 2017). – URL: https://www.uchicagomedicine.org/forefront/research-and-discoveries-articles/uchicago-medicine-collaborates-with-google-to-use-machine-learning-for-better-health-care (дата обращения: 17.01.2022)

пациентов, чья конфиденциальная информация была раскрыта[407]. В результате расследования стало очевидно, что технологии ИИ, предсказывающие появление различных заболеваний, могут порождать излишнюю стигматизацию отдельных лиц, подвергать людей агрессивному маркетингу со стороны фармацевтических компаний и других фирм, представляющих широкий спектр коммерческих медицинских услуг.

Можно сделать вывод о том, что сегодня существует риск переоценки преимуществ использования ИИ, появления завышенных ожиданий от его потенциальных возможностей и, как следствие, внедрение непроверенных продуктов и услуг, не прошедших тщательную оценку безопасности и эффективности. Вполне может быть, что сама технология искусственного интеллекта может не соответствовать стандартам научной достоверности и точности, которые в настоящее время применяются к медицинским технологиям. На наш взгляд, даже чрезвычайная ситуация не может быть оправданием для развертывания непроверенных технологий. Например, цифровые технологии, разработанные и применяемые на ранних этапах пандемии COVID-19, не всегда соответствовали каким-либо объективным стандартам эффективности, оправдывающим их использование[408]. Технологии ИИ были внедрены в качестве ответных мер на пандемию без достаточных доказательств и клинических испытаний. Указанные примеры позволяют сделать вывод о необходимости правового регулирования подобной деятельности и тщательного анализа возникающих проблем.

И, наконец, третий аспект проблемы нарушения конфиденциальности, дающий о себе знать на этапе получения выводов, которые система ИИ может сделать из имеющихся в ее распоряжении дан-

[407]Minssen T, Gerke S, Shachar C. Is data sharing caring enough about patient privacy? Part I: The background / T. Minssen, S. Gerke, C. Shachar. – URL: https://blog.petrieflom.law.harvard.edu/2019/07/26/is-data-sharing-caring-enough-about-patient-privacy-part-i-the-background/ (дата обращения: 07.04.2022)

[408] Gasser U, Lenca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: Taxonomy, ethical challenges, and navigation aid / U. Gasser, M. Lenca, J. Scheibner, J. Sleigh, E. Vayena // Healthpolicy. –2020. – Vol. 2. – Pp. 425-434.

ных. Пользователи, как правило, не знают о содержании этих выводов и могли бы возражать против такого использования их личных данных в том случае, если бы их проинформировали.

Например, в Китае система QR-кодов была создана на основе цифровой платежной системы Alipay, платформы для мобильных и онлайн-платежей[409], которая была значительно расширена за счет создания инфраструктуры цифровой идентификации для хранения данных о здоровье граждан. В результате возник риск превращения этой системы в форму дополнительного контроля за действиями людей и даже применения в отношении отдельных лиц карательных мер со стороны государства в случае передачи собранных данных его учреждениям.

Подтверждением данного опасения стало, например, признание правительством Сингапура в начале 2021 года о том, что данные, полученные из приложения для отслеживания контактных лиц COVID-19 (Trace Together), могут быть доступны «для целей уголовного расследования», несмотря на предыдущие заверения в том, что этого не произойдет[410].

Похожая ситуация складывается в сельскохозяйственном секторе, где фермеры беспокоятся, какие именно типы данных извлекает ИИ, и как они будут распространяться и использоваться[411]. Многие из фермеров жертвуют своей конфиденциальностью, соблазняясь преимуществами, которые сулит внедрение ИИ в процесс производства. Однако около 78% фермеров выражают обеспокоенность тем, что их данные передаются и продаются корпорациями[412]. Агробизнес может недобросовестно и неответственно использовать эти данные для перепродажи продукции ферме-

[409]Mozur P., Zhong R., Krolik A. In coronavirus fight, China gives citizens a color code, with red flags / P. Mozur, R. Zhong, A. Krolik– URL: https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html (дата обращения: 19.01.2021)

[410]Illmer A. Singapore reveals COVID privacy data available to police. – URL: https://www.bbc.com/news/world-asia-55541001 (дата обращения: 09.09.2021)

[411]Ryan M. Ethics of using AI and big data in agriculture: the case of a large agriculture multinational / M. Ryan // ORBIT Journal. – 2019. – Vol. 2(2). – 27 p.

[412] Stock R, Gardezi M. Make bloom and let wither: Biopolitics of precision agriculture at the dawn of surveillance capitalism / R. Stock, M. Gardezi // Geoforum. – 2021. – Vol.122. – Pp.193-203.

рам на невыгодных для них условиях для того, чтобы привязать применение последними систем ИИ к поставкам семян и техники, а также использовать эти данные для скупки земли по очень низким ценам.

Неправомерный доступ к базе данных систем ИИ поднимает вопрос об ответственности за возможный вред, причиненный этими системами, который вновь отсылает нас к правовому регулирования процесса получения, хранения и передачи персональных данных в соответствии с уголовным, административным, трудовым и гражданским правом.

В заключение отметим, что в параграфе 2.3 нами раскрыто содержание проблем непрозрачности, нарушения конфиденциальности и отсутствия ответственности. Существующие источники, в которых упоминается о существовании названных проблем, как правило, фиксируют их, не исследуя подробно их суть и значение. Герменевтический анализ и сопоставление целого ряда документов правового и этического характера, а также результатов научных исследований последних лет позволили нам выявить сущность и взаимосвязь данных проблем с центральной проблемой причинения вреда человеку вследствие применения им систем ИИ.

Проблема непрозрачности заключается в том, что принцип работы систем ИИ становится в процессе его совершенствования все более непрослеживаемым, необъяснимым и неинтерпретируемым. Человеку оказывается принципиально недоступен порядок его работы, а также последовательность выстраиваемых им цепочки рассуждений. Непрозрачность в работе систем ИИ чаще всего обусловлена сложностью этих систем и использованием метода глубокого обучения, а также высокой степенью вероятности возникновения ошибок в алгоритмах их деятельности.

Проблема отсутствия ответственности состоит в том, что природа ИИ не позволяет установить ответственное лицо, субъекта ответственности, на которого однозначно можно было бы возложить вину в случае причинения вреда человеку системой ИИ. Также данная проблема заключается в невозможности определить меру ответственности тех или

иных лиц в инцидентах, произошедших с участием систем ИИ, в силу недостаточности и несовершенства правового регулирования развития и применения ИИ, поспешного и зачастую непродуманного внедрения его систем в различные сферы жизни современного общества.

Проблема нарушения конфиденциальности возникает из-за угрозы утечки потока персональных данных или потери контроля над этими данными, которые могут быть скомпрометированы, вопреки воле и желанию человека опубликованы в открытых источниках и в дальнейшем использоваться для совершения мошеннических действий.

Указанные проблемы, по нашему мнению, выступают потенциальными источниками, причинами причинения вреда человеку, угрожающими его безопасности и порождающими, в свою очередь, другие негативные последствия взаимодействия человека с системами ИИ.

Выводы по второй главе

Во второй главе исследования мы рассмотрели ряд социально-философских проблем, которые, по нашему мнению, отражают уникальный, специфический характер систем ИИ, отличающий их от любых других созданных человеком технических систем и технологий. Указанные проблемы образуют систему взаимосвязанных, взаимообусловленных трудностей, выступающих различными формами проявления центральной проблемы использования систем ИИ: проблемы причинения вреда человеку. Для достижения цели нашего исследования они разделены нами на две группы, одна из которых посвящена рассмотрению возможных проявлений вреда, наносимого человеку системами ИИ, вторая объединяет причины возникновения негативных эффектов взаимодействия человека с ИИ.

Изучение проявлений, условий, последствий действия рассмотренных проблем позволяет сделать вывод об исключительном потенциале их воздействия на человека. Этот потенциал уже сегодня сопоставим с потенциалом самых передовых ядерных технологий или биомедицины, что

обусловливает необходимость особого подхода и отдельного философского исследования систем ИИ с точки зрения оценки их возможного влияния на будущее человека и человечества. Очевидно, что массовое внедрение систем ИИ, способных к самообучению и бесконечному совершенствованию, не обремененных ограничивающими их требованиями и нормами нравственности, составляющими одну из сущностных особенностей человека, многократно увеличивает опасность того, что в какой-то момент человек, создавший указанные системы, перестанет понимать смысл принимаемых ИИ решений, его цепочку «рассуждений», ход его «мыслей». Это в будущем грозит ему утратой контроля над искусственным разумом, его возможными стремлениями и действиями.

Неуклонный рост числа официальных документов, декларирующих этические принципы разработки и применения систем ИИ, свидетельствует о стремлении разных социальных групп и участников процесса создания и внедрения ИИ достичь консенсуса в вопросах использования систем ИИ мировым сообществом. Основной вопрос заключается в том, чтобы понять, что системы ИИ существуют для людей, а не люди для них. Главное - не возможности технологий, а возможность людей понимать самих себя, понимать, что им действительно нужно, в чем состоит их подлинное благо. Запретить внедрение этих систем, как и отменить их развитие, невозможно, техно-пессимизм тоже бесполезен. Значит, человеку необходимо понять, что угроза исходит не из самих технологий, способных нанести непоправимый вред, а из веры человека в свое безоговорочное превосходство в мире, его недооценки оборотной стороны внедрения в жизнь систем ИИ, имеющих в том числе и негативное значение для настоящего и будущего созданной человеком цивилизации. Наша позиция заключается в том, что в разработке и развитии технологий ИИ важно выделять критические точки этого развития с позиций социальной, этической, аксиологической разрешимости, допустимости рассматриваемых технологий. Необходимо сосредоточиться на том, каких результатов

мы ожидаем от технологий, а не на том, что технологии ожидают от нас. Разработка, внедрение систем ИИ – это междисциплинарная задача, которая требует подлинного объединения, синтеза технических, технологических, социологических, философских знаний. Она требует иных форм проектной работы, иного уровня осознания стоящих перед человеком проблем и задач. При этом ведущую, направляющую роль в подобных исследованиях должна играть социальная философия, предметом постижения которой является общество, рассмотренное с позиций целостности и системности, наиболее общих законов его динамики, осмысления фундаментальных причин событий и процессов, основных направлений развития социума, необходимых для выявления места и роли человека в мире, реализации заложенного в нем творческого, созидательного потенциала.

Именно социально-философское познание, научность которого, по словам П.В. Алексеева «…должно сливаться с гуманистичностью», позволяет в процесс обсуждения проблем применения ИИ подойти к пониманию, выявлению сущности базовых принципов, требований, норм, регулирующих и обеспечивающих безопасное и эффективное применение искусственного разума в любых отраслях и сферах жизни социума.

ГЛАВА III. ФИЛОСОФСКИЕ ПРИНЦИПЫ РАЗРАБОТКИ, ВНЕДРЕНИЯ И ПРИМЕНЕНИЯ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

3.1 Условия эффективного и безопасного применения систем ИИ

Системы ИИ, широко применяемые в современном мире, помимо благ и преимуществ, приносят в жизнь человека новые риски, поскольку их бездумное внедрение потенциально может увеличить количество несчастных случаев или привести к новым видам несчастных случаев, жертвы которых, помимо прочего, рискуют остаться без компенсации из-за сложности применения существующих режимов юридической ответственности, из-за нерешенности множества вопросов о статусе систем ИИ в современном социуме.

Иначе говоря, взаимодействие человека с этими системами нельзя назвать нейтральным, принципиально не способным вызвать негативные последствия. Напротив, оно таит в себе угрозы безопасности как личности, так и обществу, миру в целом.

Реакцией социума на возникшую и постоянно возрастающую угрозу со стороны ИИ стали многочисленные попытки выработать перечень основополагающих принципов взаимодействия человека с искусственным разумом, необходимых для того, чтобы наиболее эффективным способом предотвращать и снижать возможность возникновения и эскалации социально-философских, этических проблем в процессе разработки, внедрения и применении систем ИИ во всех отраслях жизни современного общества. Эти принципы обусловлены рассмотренными нами в предыдущем разделе проблемами применения систем ИИ и потому имеют коррелирующее с ними наименование: принципы непричинения вреда, прозрачности, ответственности, конфиденциальности, справедливости и автономии. Необходимость решения задач нашего исследования обусловила разделение этой совокупности на две группы:

Безусловно, каждый из названных принципов имеет множество собственных нерешенных проблем (например, проблема с определением самого понятия, реализацией принципа на практике и т.д.). Не вызывает сомнения и тот факт, что содержание каждого из них подлежит дальнейшему исследованию и уточнению, а перечень – пополнению.

Тем не менее, выбор указанных шести принципов определен выводами и содержанием значительного количества источников разного происхождения, посвященных регулированию и применению систем ИИ, подтверждающих своим существованием актуальность и несомненную важность этих фундаментальных принципов взаимодействия человека с системами ИИ. Эти источники зачастую используют различные термины для обозначения одних и тех же по содержанию норм и требований, поэтому для удобства анализа в представленных ниже таблицах, предваряющих рассмотрение основных принципов применения систем ИИ, нами собраны близкие по значению термины, встречающиеся в такого рода документах. Отметим, что в собранном нами своде документов их авторы, как правило, лишь формулируют названные принципы, но не раскрывают их содержание, выявление и прояснение которого стало целью предпринятого нами исследования.

Мы предполагаем, что могут возникнуть вопросы, касающиеся неполноты перечня философских принципов. Например, почему устойчивость и экологическая ответственность не упомянуты в представленном списке. Мы обеспокоены экологическими проблемами и влиянием, которое системы ИИ оказывает и будут оказывать на окружающую среду, поэтому решили включить принцип «экологической ответственности» или «устойчивости» в принцип непричинения вреда. Подобным образом мы объединяли ряд принципов в один общий, если обнаруживали тождественность смыслов, общность содержания. Так, «подотчетность» была включена в принцип «ответственность». По той же логике японский принцип ИИ о добросовестной конкуренции, утверждающий, что не дол-

жен осуществляться недобросовестный сбор данных и посягательство на суверенитет, был включен нами в «принцип конфиденциальности».

В данном параграфе предметом исследования стали принципы, раскрывающие условия безопасного применения систем ИИ. Первый в названной группе – принцип прозрачности.

Принцип прозрачности — это требование, согласно которому основа конкретного решения системы ИИ всегда должна быть наглядна для пользователя этой системы. При этом должна быть обеспечена возможность выяснения причин в случае возникновения в ее работе какого-либо нарушения, сбоя. Принцип «прозрачности» заключается в утверждении, что системы ИИ должны разрабатываться и внедряться таким образом, чтобы можно было осуществлять надзор за их деятельностью.

Прозрачность	Объяснимость, открытость, отслеживаемость, предсказуемость, отчетность, понятность, интерпретируемость, открытые данные и алгоритмы, право на информацию, уведомление при принятии ИИ решения о личности, регулярная отчётность, проверяемость, инспектируемость.

Прозрачность характеризует то, что представляют собой данные, которыми располагает система ИИ, где осуществляется сбор этих данных, что с ними происходит в процессе работы системы ИИ, а также то, как они ею используются[413]. Международный форум по борьбе с мошенничеством в государственном секторе обозначил принцип прозрачности как требование обеспечения объяснимости как процесса работы систем ИИ, так и полученных в результате этой работы результатов, причем как для самих экспертов, так и для широкой общественности[414]. Участники фору-

[413] Digital Curation Centre. The University of Edinburgh The Role of Data in AI: Report for the Data Governance Working Group of the Global Partnership of AI. – URL: https://www.research.ed.ac.uk/en/publications/the-role-of-data-in-ai (дата обращения:12.12.2022)

[414] Office S. F. The use of Artificial Intelligence to Combat Public Sector Fraud. Professional Guidance. International Public Sector Fraud Forum / S. F. Office. – 2020.

ма пришли к выводу, что открытость алгоритма дает возможность отследить, по какому принципу, на каком основании было принято то или иное решение.

В Национальной стратегии развития искусственного интеллекта РФ на период до 2030 года[415] принцип прозрачности является одним из основных условий развития и внедрения систем ИИ. Принцип прозрачности в данном документе трактуется как объяснимость работы ИИ, процесса достижения им результатов, отсутствие дискриминации в доступе пользователей к продуктам, которые созданы с помощью ИИ.

Принцип прозрачности был также упомянут во многих этических рекомендациях, в том числе в этических рекомендациях Европейской комиссии по надежному ИИ[416], Европейского парламента (2021 г.)[417], ЮНЕСКО в отношении проблем образования (Ф. Мияо и др., 2021 г.)[418], в Пекинском консенсусе ЮНЕСКО (2019 г.)[419] и «Принципах ответственного управления надежным ИИ» ОЭСР (2021 г.)[420]. Однако описания принципа прозрачности в этих отчетах и руководствах различаются. Например, если Европейская комиссия[421] понимает под прозрачностью объяснимость и прозрачность элементов, относящихся к системе ИИ, то проект ЮНЕСКО 2020 г. под прозрачностью имеет в виду связь с ответственностью и подотчетностью. В документе прозрачность признается необходимым условием выполнения принципа ответственности. В Европей-

[415]Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). – URL: https://base.garant.ru/72838946/ (дата обращения: 14.05.2022).

[416] High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.).

[417] European Parliament Report on artificial intelligence in education, culture and the audiovisual sector Committee on Culture and Education. – URL: https://www.europarl.europa.eu/ doceo/document/A-9-2021-0127_EN.html (дата обращения: 12.07.2022)

[418] Miao F., Holmes W. et al. AI and education: Guidance for policy-makers. United Nations Educational, Scientific and Cultural Organization / F. Miao, W. Holmes, R. Huang, H. Zhang. – URL: https://unesdoc.unesco.org/ ark:/48223/pf0000376709 (дата обращения: 10.04.2022)

[419]Beijing Consensus on Artificial Intelligence and Education // International Conference on Artificial Intelligence and Education, Planning Education in the AI Era: Lead the Leap, Beijing. – 2019. – 70 p.

[420]OECD's Principles for responsible stewardship of trustworthy AI. URL: https://oecd.ai/en/ai-principles (дата обращения: 18.05.2022)

[421] High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.).

ской этической хартии использования искусственного интеллекта в судебных системах и окружающих их реалиях принцип прозрачности считается важнейшим принципом применения искусственного интеллекта в судебной и правоохранительной системах, а основными составляющими данного принципа объявлены объяснимость, понятность, доступность[422].

Что касается доступности, то здесь неизбежно возникает целый ряд проблем. Даже если удастся объяснить содержательную составляющую алгоритмов ИИ, то по экономическим, юридическим или политическим причинам система ИИ все равно может стать непрозрачной.

Кроме того, продукты глубокой нейронной сети зависят от количества и уровней вычислений, которые связаны между собой столь сложным образом, что ни один ввод или вычисление не может по определению являться доминирующим фактором[423]. Иначе говоря, надзор за деятельностью систем ИИ в отдельных случаях сложно осуществлять, поскольку эти системы могут принимать в значительной степени автономные решения. Такие системы ИИ представлены в самых разных областях, например, в медицине в качестве диагностических инструментов для обнаружения диабетической ретинопатии[424], в сети Интернет в виде рекомендательных алгоритмов YouTube[425], в предиктивном полицейском управлении и при вынесении уголовных приговоров[426].

Высшая экспертная группа по искусственному интеллекту Европейской комиссии признает, что существуют технические ограничения для того, чтобы система была прозрачной, и что иногда невозможно дать

[422]Европейская этическая хартия об использовании искусственного интеллекта в судебных системах и окружающих их реалиях. URL: https://rm.coe.int/ru-ethical-charter-en-version-17-12-2018-mdl-06092019-2-/16809860f4 (дата обращения: 12.09.2022)

[423] Miller T. Explanation in Artificial Intelligence: Insights from the Social Sciences / T. Miller // Artificial Intelligence. – 2019. – Vol. 267. – P. 1-38.

[424] Abramoff M. D., Lavin P. T.et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices / M. D. Abramoff., P. T. Lavin, M. Birch, N. Shah, J. C. // NPJ Digit Med. – 2018. – Vol. № 1. – P. 39.

[425] Bishop S. Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm / S. Bishop // Convergence, 2018. – Vol. № 24(1). – Pp. 69-84.

[426] Brayne S., Christin A. Technologies of crime prediction: The reception of algorithms in policing and criminal courts / S. Brayne, A. Christin // Social Problems. – 2020. – Vol. 68 (3). – Pp. 608-624.

объяснение того, как система пришла к тому или иному решению. Вследствие этого реализация принципа прозрачности упирается в так называемую проблему «черного ящика», которая делает системы неинтерпретируемыми и необъяснимыми. Эта же проблема делает системы непредсказуемыми, следовательно, они могут нанести вред людям, например, вследствие предвзятости решений. Предвзятость вытекает из анализа больших данных, который осуществляется порой по непонятным и недоступным человеку закономерностям.

Все эти особенности способны подорвать общественное доверие к ИИ, которое, в свою очередь, может поставить под сомнение легитимность государственного сектора в целом.

Иначе говоря, понятие прозрачности означает, помимо прочего, юридическую прозрачность, т.е. открытость кода программы, несмотря на утвержденные в законодательстве режимы интеллектуальной собственности, коммерческой тайны и т.п. В соответствии с указанными правовыми нормами, никто не может иметь право доступа к исходному коду и технической документации продукта[427]. Даже оператор алгоритма не знает, как данные обрабатываются, и как принимаются те или иные решения. Отсюда возникает необходимость в принудительном раскрытии принципов работы алгоритма в тех случаях, когда принятое ИИ решение оказывается способно оказывать существенное влияние на жизни людей (назначение штрафов, диагностика болезней, назначение лечения)[428]. Однако даже при наличии доступа к исходному коду алгоритма и коэффициентам, при высокой точности и надёжности применяемых алгоритмов все еще отсутствует гарантия объяснимости. Это проистекает из особенностей работы некоторых систем ИИ (в частности, основанных на искусственных нейронных сетях), а именно из того, как они генерируют решения

[427] Burrell J. How the machine 'thinks': Understanding / J.Burrell // Big Data & Society. – 2016. – Vol. 1-2. – P. 1-12.
[428]Kaminski M. E. Binary governance: Lessons from the GDPR'S approach to algorithmic accountability / M. E. Kaminski // Southern California Law Review, 2019. – Vol. 92 (6). – Pp. 1529-1616.

или прогнозы на основе имеющихся в их распоряжении входных данных, о чем мы уже упоминали выше.

По нашему мнению, надзор за деятельностью системы ИИ должен обеспечиваться в форме публичного раскрытия информации о рассматриваемой системе, ее процессах, прямом и косвенном влиянии на права человека и мерах, принимаемых для выявления и смягчения неблагоприятного воздействия системы на субъекта этих прав в случае возникновения негативной для пользователя ситуации. Кроме того, полученная информация должна быть полной и достоверной, позволяющей специалистам проводить содержательную оценку работы системы ИИ.

Например, если человеку отказано в получении им кредита, то он вправе затребовать от кредитной организации информацию о причине отказа в удовлетворении кредитной заявки. Иначе говоря, человек не должен становиться жертвой автоматизированной обработки данных с использованием ИИ. Автоматизированная обработка данных должна осуществляться с полным информированием субъекта данных, возможностью получить объяснения в отношении принятого системой решения и вытекающим отсюда правом человека оспаривать принятое ИИ решение.

Особое место принцип прозрачности занимает в Кодексе этики для разработчиков робототехники, который считается сегодня одним из наиболее комплексных и проработанных документов в области ИИ[429]. В данном документе, говоря о принципе прозрачности, авторы различают техническое и организационное направление. Организационное направление касается разработчиков и обязанности их предоставлять необходимую информацию пользователям, оповещать об угрозах и рисках, которые могут нести системы ИИ. Техническое направление задается вопросом о предоставлении доступа к алгоритму в случае причинения вреда систе-

[429]Civil Law Rules on Robotics: Resolution with recommendations to the Commission N 2015/2103(INL): adopted by European Parliament 16.02.2017 // EURO-PARL.EUROPA.EU - the official website of the European Parliament. – URL: https://www.europarl.europa.eu/ doceo/document/TA-8-2017-0051_EN.html (дата обращения: 22.12.2020).

мой ИИ с тем, чтобы обеспечить возможность отследить причины возникновения сбоя в работе системы.

Общий регламент по защите данных ЕС[430] к вопросу прозрачности и подотчетности алгоритмов относит положение о том, что субъект вправе получать информацию «о наличии системы автоматизированного принятия решения, включая профилирование, а также о значимости и предполагаемых последствиях такой обработки для субъекта данных»[431]. Это положение интерпретируют по-разному: либо пользователю раскрывается вся логика принятия решения[432], либо пользователя уведомляют в самых общих чертах о процессе принятия решений алгоритмом[433].

Некоторые авторы полагают, что прозрачность может относиться либо в целом к системе, либо к принятию отдельных решений[434]. При этом они утверждают, что обеспечение полной прозрачности работы алгоритма не имеет смысла и слишком ресурсозатратно, к тому же не ведет к значимому объяснению логики принятия решений. В литературе даже упоминается термин «принцип относительной прозрачности»[435].

В документе «Этически обусловленное проектирование» (IEEE)[436] и Белой книге ЕС[437] прозрачность определяется как один из восьми всеобъемлющих принципов, как способность обнаруживать основу решения,

---

[430] General Data Protection Regulation (GDPR). – URL: http://www.gdpr.su/doc/gdpr%20reglament.odt (дата обращения: 30.08.2022).

[431]Там же. (ст. 13-15 Регламента)

[432]Malgieri G., Comandé G. Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation / G. Malgeri, G. Comandé // International Data Privacy Law. – 2017. – Vol. 7 (4) – Pp. 243-265.

[433]Wachter S., Mittelstadt B., Floridi L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation / S. Wachter, B. Mittelstadt, L. Floridi // International Data Privacy Law. – 2017. – Vol. 7(2). – Pp. 76-99.

[434]Koene A., Clifton C., Hatada Y., Webb H., Richardson R. A governance framework for algorithmic accountability and transparency / A. Koene, C. Clifton, Y. Hatada, H. Webb, R. Richardson // Brussels: European Parliamentary Research Service, 2019. – 124 p.

[435]Бахтеев Д. В., Тарасова Л. В. Применение искусственного интеллекта в деятельности арбитражных судов РФ: перспективные направления и проблемы / Д. В. Бахтеев, Л. В. Тарасова // Вестник Костромского государственного университета. – 2020. – Т. 26, № 4. – С. 249-254.

[436] Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version 2 / The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. – URL: http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html (дата обращения: 01.01.2022)

[437]European Commission. White paper on artificial intelligence: A European approach to excellence and trust, 2020. – URL:https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. (дата обращения:12.03.2022)

принятого системами ИИ. IEEE рекомендуют разработать «новые стандарты, описывающие измеримые, поддающиеся проверке уровни прозрачности, позволяющие объективно оценивать работу системы».

В «Руководящих принципах по этике для надежного ИИ»[438] прозрачность имеет решающее значение для создания и поддержания доверия пользователей к системам ИИ. Это означает, что процессы должны быть прозрачными, возможности и назначение систем искусственного интеллекта — открытыми, а решения — насколько это возможно — объяснимыми для тех, кого они прямо или косвенно затрагивают. Без такой информации решение не может быть должным образом обжаловано.

В Азиломарских принципах под прозрачностью понимается возможность выяснения причин при сбоях в системе, требование того, чтобы процесс работы должен быть виден специалисту, если он желает увидеть подробности[439]. Иными словами, любой процесс принятия решений системами ИИ, оказывающий существенное влияние на права человека, должен быть идентифицируемым. Люди должны иметь возможность понять, как принимаются решения, и как эти решения проверяются. Это, конечно, не значит, что пользователю должны выдавать информацию о коэффициентах нейронной сети, на которой построен ИИ. Это означает, что общие принципы работы должны быть понятными и ясными, чтобы была открыта возможность осуществления общественного контроля за работой этих систем.

Интересно, что рассматриваемый документ отдельно формулирует требование к языку описания работы систем ИИ, использование которых, по мнению авторов, должно быть обнародовано в четких и доступных терминах.

[438] High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.).
[439] Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)

Разъяснение смысла этого требования, по нашему мнению, содержат «Принципы подотчетных алгоритмов» FAT-ML[440], в которых прозрачность означает обеспечение того, чтобы решения систем ИИ могли быть объяснены конечным пользователям и другим заинтересованным сторонам в нетехнических терминах. При этом заинтересованным сторонам должна быть предоставлена информация, которая позволяет объяснить общее функционирование системы, конкретное использование данных в системе и отдельные решения, принимаемые системой.

На наш взгляд, необходимые технические и организационные меры должны быть направлены на обеспечение прозрачности с самого начала разработки системы ИИ, а не только после завершения ее проектирования. Проектировщики систем и заинтересованные стороны должны с самого начала подумать о том, как сделать информацию о стандартах принятия решений доступной и понятной для самых различных категорий пользователей, а также, например, для регулирующих и контролирующих органов власти.

При разработке систем ИИ следует учитывать сложность требований к прозрачности, сложность процесса принятия решений. В зависимости от информационной цели и аудитории стандарты принятия решений могут быть представлены в виде общих принципов высокого уровня или более конкретных и подробных стандартов. Даже если добиться соответствующего уровня прозрачности не удается, необходимо стремиться к возможному уменьшению непрозрачности в работе системы ИИ. Так, IBM в 2018 г. признал, что, хотя предубеждения и дискриминация никогда не могут быть полностью устранены в системах ИИ, необходимо приложить усилия для их уменьшения или смягчения[441].

[440] Fairness, Accountability, and Transparency in Machine Learning. – URL: https://www.fatml.org/ (дата обращения: 12.01.2023)

[441] IBM. Principles for trust and transparency. – URL: https://www.ibm.com/blogs/policy/wp-content/uploads/2018/05/IBM_Principles_OnePage.pdf (дата обращения: 16.09.2022)

Особенно важно сделать прозрачными риски, связанные с работой системы ИИ. Риски, связанные с автоматизированным принятием решений, обсуждались в трудах Т. Араужо[442] и Дж. Бахнера[443] (например, нарушения конфиденциальности, дискриминация и предвзятость при принятии решений). Отметим здесь, что степень обеспечения прозрачности в данном контексте в значительной степени зависит от сферы применения систем ИИ. Например, в сфере медицины требование прозрачности представляется важным и необходимым. Пациент должен обладать максимально подробной и полной информацией о том, какие операции будут производиться с его организмом, и к каким последствиям он должен быть готов. Это чрезвычайно важно, чтобы оценить возможные риски, возникающие вследствие использования ИИ.

С юридической точки зрения до сих пор не ясно, какие средства правовой защиты будут у людей против непрозрачных систем ИИ. Дискриминационная практика, возникающая в результате использования таких технологий, недостаточно описана в законе, например, в законе о защите данных[444].

В целях обеспечения прозрачности в работе систем ИИ важно, чтобы деятельность их разработчиков тоже была прозрачной и доступной для заказчиков и потенциальных пользователей посредством предоставления подробных отчетов, в которых содержится описательная и сводная информация о системе ИИ с точки зрения ее внедрения и использования. Например, разработчики алгоритмов уголовного правосудия и инструментов оценки криминального риска должны публиковать сведения о том, где используются эти инструменты (т.е. в каких юрисдикциях),

[442]Araujo T., Helberger N. et al. In AI we trust? Perceptions about automated decision-making by artificial intelligence / T. Araujo, N. Helberger, S. Kruikemeier, C. H. De Vreese // AI & SOCIETY, 2020. – Vol.35. – Pp. 611-623.
[443]Bahner J. E., Hüper A. D., Manzey D. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience / J. E. Bahner, A. D. Hüper, D. Manzey // International Journal of Human-Computer Studies. – 2008. – Vol. 66(9). – Pp.688-699.
[444]Wachter S., Mittelstadt B. A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. Columbia Business Law Review, 2019. – Vol. 7(2). – Pp.494-620.

сколько решений было принято системой, и какими они были. Там, где это возможно, следует оценивать качество принятия решений системами ИИ путем сравнения их с результатами, полученными вследствие применения других подходов к принятию решений в данной области. Такие отчеты могут публиковаться с разной периодичностью в зависимости от важности и своевременности информации.

Возможно, решению проблемы прозрачности могло бы содействовать закрепление требования к разработчикам предоставлять информацию о последовательности операций, которые осуществляются системой ИИ и результатах, к которым они приводят. Описательную схему операций можно сделать в виде блок-схемы, которая показывает главные вариации решений, к которым система ИИ подведет на определенных стадиях своих вычислений. Важно здесь не опуститься до предельного упрощения в объяснении технического содержания работы системы, или напротив, не дойти до предельного усложнения, когда даже экспертам будет затруднительно понять принцип работы ИИ. Такое обеспечение будет труднодостижимо в случае систем, основанных на методах машинного и глубокого обучения, так как алгоритм может обучаться и выявлять при обработке данных такие закономерности, которые могут быть неизвестны нам.

Интеграция прозрачности в процесс проектирования и внедрения систем искусственного интеллекта — непростая задача. Скорость технологического развития, многочисленные аспекты концепции прозрачности, неопределенность в отношении того, где требуется прозрачность, как лучше всего подходить к общению с различными заинтересованными сторонами и как встроить меры прозрачности в значимые и организационно реалистичные меры подотчетности — все это существующие проблемы, ожидающие своей практической реализации.

Вероятно, можно было бы увеличить прозрачность в работе систем ИИ, если бы принцип прозрачности был положен в основу работы любых

систем ИИ, если бы он был закреплен в виде нормативных правовых актов, разного рода конвенций, требований и т.д. Людям тогда, например, было бы проще оспаривать решения, принятые ИИ и причинившие им какой-либо ущерб. Мы считаем, что в любом случае, особенно в сферах, где жизнь человека зависит от принятого системой решения, окончательный вердикт всегда должен оставаться за человеком. Проектирование и разработка систем ИИ высокого риска должны осуществляться таким образом, чтобы изначально можно было обеспечить необходимую прозрачность их функционирования, позволяющую пользователю интерпретировать результаты работы систем ИИ на любом этапе ее деятельности. Для достижения этой цели в процессе разработки систем ИИ необходимо создавать междисциплинарные группы специалистов в области программирования, философии, права, психологии и т.д., что позволит принимать более продуманные, этически выверенные, а также соответствующие существующим в обществе юридическим, политическим, культурным нормам и требованиям решения.

Важно здесь отметить, что не от машин мы должны требовать прозрачности, а от людей, ответственных за внедрение этих систем и способных обеспечить прозрачность. Только люди могут и должны объяснять, что они решают и делают. Принцип ответственности в данном случае коррелирует с понятием подотчетности, когда разработчик должен предоставить необходимую для пользователя информацию. Развитие технологий и технологическое образование должны быть скорректированы таким образом, чтобы лучше помочь пользователям и разработчикам ИИ ответить на эти вызовы. Кроме того, обществу нужно задуматься о том, следует ли допускать высокую степень автоматизации в случаях, когда высокая скорость, важность и объем решений (например, в военном деле), требуют мгновенного ответа, поскольку для взвешенного решения может не хватить временного ресурса.

Философский принцип прозрачности служит предложением того, как можно конкретизировать требование прозрачности. Указанный принцип подразумевает, что по мере возможности во всех случаях применения систем ИИ человеку, пользователю:

– должна быть предоставлена адекватная информация о характере и функциональности системы ИИ, о возможных вариантах изменения функций системы, а также сведения о данных, которые обрабатывает система ИИ;
– должна быть обеспечена открытость в отношении логической схемы принятия системой ИИ решений;
– известно о потенциальных рисках и угрозах принимаемых системой ИИ решений;
– эти данные должны быть донесены на доступном и понятном языке;
– должна быть предоставлена информация о механизме подачи жалоб и претензий, этапах прохождения этой юридической процедуры, точных полномочиях контактных лиц, ориентировочных сроках и ожидаемых результатах.

Без реализации принципа прозрачности невозможно претворить ведущий принцип – принцип непричинения вреда. Соответственно, от реализации принципа прозрачности зависит дальнейшая судьба систем ИИ, их развития и внедрения.

Несоблюдение принципа прозрачности и отсутствие достаточной степени объяснимости, в свою очередь, создает огромную проблему, связанную с ответственным использования ИИ. Ведь для того, чтобы действовать ответственно, важно уметь объяснить кому-то принципы принятия решения. Прозрачность делает возможным реализацию требования ответственности, ответственного поведения. Эта взаимосвязь прозрачности и ответственности подводит нас к следующему принципу – принципу ответственности.

*Принцип ответственности***.**

Ответственность	Подотчётность, моральная и правовая ответственность, добросовестное принятие решений.

По мере совершенствования систем ИИ и увеличения их влияния на жизнь социума возникает необходимость переосмысления понятия ответственности. Часто обсуждение «ответственности» в научной литературе сводится к обвинению кого-либо в безответственности, однако мы полагаем, что это не способствует прояснению содержания термина. Некоторые исследователи поставили перед собой задачу сформулировать новую этику, соответствующую технологическим реалиям современности (Р.Г. Рополь, А. Маттиас, Г. Йонас). Они полагают, что новая этика ответственности должна быть нацелена на сохранение будущего человечества, на осмысление отдаленных последствий осуществляемой им технической, технологической деятельности, на предотвращение, недопущение непредсказуемых негативных ее результатов.

Ответственность, рассматриваемая сквозь призму последствий, приводит к изменению временного горизонта, а также объекта и субъекта ответственности. Г. Рополь[445], в частности, пришел к выводу, что человек, совершающий действие, оказывающее определенное воздействие на нравственно значимые субъекты, может быть привлечен к ответственности, может быть субъектом ответственности. «Быть ответственным» для ученого означает, что субъекты ответственности должны оправдывать свои действия перед пострадавшей стороной. При этом, по его мнению, ответственность может быть наложена ретроспективно (после) или перспективно (заранее), поэтому необходимо в обязательном порядке учитывать временной аспект.

С ним согласна Х. Арендт, которая также считает, что человек, меняя мир, должен нести (не только принимать) ответственность за измене-

[445]Ropohl G. Neuewege, die technik zuverantworten / G. Ropohl // H. Lenk& G. Ropohl (Eds.), Technik und ethik 2nd ed., 1993. – Vol. 8395. – Pp. 149-176.

ния, которые стали следствием его деяний[446]. Иными словами, новатор несет ответственность за перемены, которые он производит, даже в том случае, если эти последствия наступят значительно позже, то есть ответственность должна быть пролонгирована в будущее, на весь период существования последствий деятельности новатора.

Еще одна проблема, возникающая в связи с толкованием ответственности – это проблема субъекта ответственности. Например, в генной инженерии, где широко обсуждается вопрос кризиса человеческой идентичности вследствие биологических трансформаций человека, сложно определить, кто выступает субъектом ответственности, на кого она должна быть возложена. Более того, возникает трудность в отношении понимания морали как одного из сущностных признаков человека, своего рода моральной способности человека[447].

Ответственность в профессиональной этике часто обсуждается применительно к коллективу, как групповая ответственность. При этом Х. Ленк, в частности, полагает, что ответственность здесь зачастую слишком абстрактно описана. Фактически вся ответственность падает на разработчика, инженера, непосредственного исполнителя, имеет место чрезмерное перекладывание ответственности на инженера[448]. Решая эту проблему, некоторые исследователи считают, что в данном случае ответственность инженера становится ролевой. Он отвечает за свои действия только перед начальством, организацией, но не перед своей совестью или обществом в целом. Коллективный характер ставит под сомнение саму реализацию моральной ответственности, провоцирует безнаказанность, вседозволенность, что едва ли может способствовать достижению социального благополучия и отвечать критериям прогрессивного развития общества.

---

[446] Arendt H. Collective Responsibility / H. Arendt // Responsibility and Judgment. – New York : Schoken books, 2003. – 295 p.
[447] Хабермас Ю. Будущее человеческой природы / Ю. Хабермас. – М.: Весь мир, 2002. – 144 с.
[448] Ленк Х. Размышления о современной технике / Х. Ленк. – М.: Аспект-Пресс, 1996. – 183 с.

Ряд теоретиков в отношении этого вопроса придерживаются позиции, согласно которой приписать коллективу моральную ответственность можно, так как, взаимодействуя с коллективом, мы больше не взаимодействуем с отдельным индивидом (Э. Штрекер[449], Х. Ленк[450]).

С ними не согласны другие исследователи, полагающие, что невозможно возложить ответственность на коллектив, поскольку нельзя игнорировать личностный аспект, решения и поступки каждого отдельного индивида[451]. Возможно, решением может послужить то, что возложение моральной ответственности должно быть реализовано, но без разделения на части. Другими словами, ответственность приписывается не целому коллективу, а именно тем лицам, которые причастны.

Вопрос о том, в какой степени люди могут или должны нести ответственность за поведение ИИ, стал одним из самых обсуждаемых в сфере ИИ[452,453,454]. В настоящее время существуют самые разные подходы к установлению ответственности за вред, причиненный системами ИИ. Например, по мнению Р. В. Душкина, автономная система ИИ, причинившая вред ее пользователю, должна сама нести за это ответственность[455]. Обычно провинившийся человек возмещает свой ущерб в виде выплаты штрафа или отбывания в тюрьме. Соответственно система тоже должна понести наказание в виде выплаты штрафа. Например, беспилотный автомобиль имеет свои потребности в обслуживании, которое стоит денег. Система зарабатывает эти деньги своей функциональностью, своим трудом, собирая с пассажиров деньги и накапливая их. Эти деньги яв-

[449] Stroker E. Ich und anderen: die Frage der Mitverantwortung / E. Stroker. – Frankfurt am Main : Klostermann, 1984. – P.64.
[450] Ленк Х. Размышления о современной технике / Х. Ленк. – М.: Аспект-Пресс, 1996. – 183 с.
[451] Neimaier O. Wofur sind wir verantworlic/ O. Neimaier // Conceptus 24. – 1990. – Vol.63. – Pp. 43-54.
[452]Special Interest Group on Artificial Intelligence. Dutch Artificial Intelligence Manifesto. – 2018.
[453]MI Garage Ethics Framework - Responsible AI. MI Garage. – URL: https://www.migarage.ai/ethics-framework/ (дата обращения: 18.03.2022)
[454]Accenture UK. Responsible AI and robotics. An ethical framework. – URL: https://www.accenture.com/gb-en/company-responsible-ai-robotics (дата обращения: 12.02.2022)
[455] Этичное применение искусственного интеллекта. – URL:https://ethics.cdto.center/3_3 (дата обращения: 19.12.2022)

ляются для нее ценностью, поэтому она должна ими выплачивать штраф, если причинит вред использующему ее человеку.

А.А. Щитова также считает, что система ИИ потенциально может быть ответственной и возмещать ущерб материально. Следовательно, развивает свою мысль ученый, ее можно наделить определенным правом собственности в отношении материального имущества[456]. Но если система ИИ не владеет этим имуществом, то следует ее привлечь к обязательным работам, наложив запрет на занятие определенными видами деятельности на установленный срок.

А.В. Незнамов считает наложение ответственности непосредственно на систему ИИ своего рода крайним вариантом, которого следует избегать, равно как и другой крайности, когда ответственность не несет никто, когда виновника обнаружить не удается в принципе[457].

В 2004 году в свет вышла работа А. Маттиаса[458], в которой автор обсуждает «разрыв ответственности» во взаимодействии с обучаемыми системами ИИ. Он указал на то, что ни один человек не может быть законно обвинен или признан виновным в нежелательных результатах действий, опосредованных системами ИИ. Проблема заключается в том, что поведение системы ИИ, использующей сложные алгоритмы, слишком автономно и слишком непредсказуемо, чтобы кто-либо из людей, внесших в ее разработку и создание свой вклад, мог в дальнейшем нести ответственность за результаты работы этой системы. В этом случае выявить субъекта ответственности не представляется возможным.

[456]Щитова А. А. Правовое регулирование информационных отношений по использованию систем искусственного интеллекта: диссертация на соискание ученой степени кандидата юридических наук. /А. А. Щитова. – М., 2021. – 225 с.

[457] Незнамов А. В. О концепции регулирования технологий искусственного интеллекта и робототехники в России / А. В. Незнамов // Закон. - 2020. - № 1. - С. 171-185.

[458] Matthias A. The responsibility gap: Ascribing responsibility for the actions of learning automata / A. Mathias // Ethics and Information Technology. – 2004. – Vol. 6(3). – Pp.175-183.

Также существуют работы, в которых делаются попытки уточнить распределение ответственности между разными субъектами[459]. Речь здесь идет и о юридической ответственности[460].

Ответственность, связанная с применением ИИ, часто рассматривается не только в научных работах, но и в различных кодексах и руководствах по этике ИИ и других видах мягкого права. Однако в большинстве случаев трактовка ответственности здесь страдает чрезмерной абстрактностью и потому оказывается неприменима к конкретным ситуациям. Кроме того, содержание понятия ответственности в такого рода документах, как правило, включает ее нормативный или профессиональный аспекты, что сужает понимание ответственности и исключает из ее толкования любые другие смыслы. Например, в Азиломарских принципах про ответственность написано буквально одно предложение, которое не раскрывает содержания принципа ответственности: «разработчики продвинутых систем ИИ играют ключевую роль в формировании нравственных последствий использования ИИ, неправильного использования ИИ и действий ИИ; они имеют возможностью и несут обязанность влиять на такие последствия»[461].

Отвечая на вопрос, может ли субъект ответственности быть представлен системой ИИ, мы считаем, что вспомогательные системы ИИ (те, которые основаны не на машинном обучении) не могут быть привлечены к ответственности, поскольку не обладают ни свободой воли, ни сознанием. Соответственно, ответственность в этом случае должна быть возложена на людей, которые разрабатывают и используют технологии ИИ.

Трудности с возложением ответственности возникают в случае с самообучающимися системами ИИ, которые обладают наибольшей авто-

[459]Guidelines for artificial intelligence. Deutsche Telekom. – URL: https://www.telekom.com/en/company/digital-responsibility/details/artificial-intelligence-aiguideline-524366 (дата обращения: 13.02.2022)

[460] Responsible bots: 10 guidelines for developers of conversational AI. – URL: https://www.microsoft.com/en-us/research/publication/responsible-bots/ (дата обращения: 06.09.2022)

[461]Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)

номностью по сравнению с предыдущими, принцип работы которых не до конца ясен даже самим разработчикам. Более того, даже разработчики не имеют достаточного контроля над действиями таких систем ИИ.

Рассмотрим аварию с беспилотным автомобилем, в которой машина сбивает пешехода. На кого возложить ответственность? Кто виноват в случившемся?

Другой пример с чат-ботом Tay от Microsoft Twitter[462], который после взаимодействия с пользователями начал высказывать расистские и женоненавистнические комментарии. В разбирательстве этого дела участвовало много сторон: разработчики, дизайнеры, компании и пользователи. Суароз-Гонзало, анализируя этот инцидент, указывает, что ответственность должна быть возложена в основном на дизайнеров, разработчиков, а не на пользователей Twitter, взаимодействовавших с ботом[463].

Р. Спарроу уверен, что в подобных ситуациях необходимо вовсе препятствовать использованию любых автономных систем, особенно автономных систем вооружения[464]. По его мнению, разработка все более автономных систем вооружения неэтична еще и потому, что в данном случае трудно найти лицо, ответственное за гибель людей в ходе войны[465]. Исчезает сама возможность привлекать к ответственности конкретных лиц, виновных в совершении преступлений, противоправных деяний.

А. Гринбаум предлагает рассматривать системы ИИ в качестве «цифровых особ», которые, в отличие от всех других видов техники, алгоритмически наделены способностью к машинному обучению, соответственно, обладают определенной степенью автономии и способностью

[462]Робот, общаясь в Твиттере, за сутки стал расистом и женоненавистником. – URL: https://www.techinsider.ru/technologies/237098-robot-obshchayas-v-tvittere-za-sutki-stal-rasistom-i-chelovekonenavistnikom/ (дата обращения: 21.02. 2022)
[463]Suárez-Gonzalo S., Mas-Manchón L., Guerrero-Solé F. Tay is you. The attribution of responsibility in the algorithmic culture/ S. Suárez-Gonzalo, L. Mas-Manchón, F. Guerrero-Solé // Observatorio, 2019. – Vol.13(2). – Pp.1-14.
[464]Sparrow R. Killer robots / R. Sparrow // Journal of Applied Philosophy. – Vol.24(1). – Pp. 62-77.
[465] Sparrow R. Predators or plowshares? Arms control of robotic weapons / R. Sparrow // IEEE Technology and Society, 2009. – Vol. 28(1). – Pp. 25-29.

принимать самостоятельные решения о совершаемых ими действиях[466]. Указанные системы не обладают сознанием, но при этом не находятся под полным контролем человека и совершают действия, не вполне прозрачные для него. Эти системы основаны на многослойных технологиях обучения (например, на глубоком обучении), где выдача строгой математической модели невозможна. Даже эксперт не способен объяснить связь между исходными данными и конечным результатом. Следовательно, такого рода системы могут представлять особую опасность, например, в области медицины, когда жизнь и здоровье человека оказываются в руках у машины. Работая с подобными системами, люди должны быть хорошо осведомлены о том, что эти системы не подконтрольны человеку. Задачей разработчика подобных систем является необходимость избавить других и себя от иллюзии детерминизма в процессе разработки и взаимодействия с такого рода системами ИИ и четко определить те виды деятельности, которые человек готов делегировать указанным системам, не беспокоясь о возможных последствиях принятых ими решений.

Обобщая вышесказанное, можно сделать вывод, что философский принцип ответственности устанавливает требование, согласно которому вовлеченные в создание ИИ-систем стороны обязаны нести моральную и юридическую ответственность за все действия, совершенные такими устройствами на всех этапах их жизненного цикла. При этом необходимо понимать, что цель исследования, проектирования, создания, развертывания и применения ИИ-систем состоит в том, чтобы помочь людям принимать решения и выполнять определенные функции, но не заменить человека, на котором всегда в конечном итоге будет лежать правовая и моральная ответственность за принятые решения и совершенные либо несовершенные деяния.

[466] Гринбаум А. Машина-доносчица: как избавить искусственный интеллект от зла / А. Гринбаум. – М.: ТрансЛит, 2017. – 76 с.

Так, в Азиломарских принципах утверждается, что разработчики и создатели продвинутых систем ИИ должны нести полную ответственность за действия, за последствия использования и за возможные злоупотребления, допущенные этими системами. Аналогичное требование можно найти в Монреальской декларации[467].

На наш взгляд, окончательное решение всегда должен принимать человек, и это решение должно быть свободным и осознанным. Так, решение жизни и смерти других людей всегда должно приниматься самими людьми, и ответственность за это решение ни в коем случае не должна перекладываться на ИИ. Ответственность за преступление или правонарушение несут лица, которые уполномочивают ИИ на совершение преступления или правонарушения или проявляют халатность, позволяя ИИ совершать противоправные деяния.

Таким образом, содержание **философского принципа ответственности** может быть сформулировано следующим образом:

– Заинтересованные стороны (разработчики, пользователи) должны нести правовую и моральную ответственность за то, чтобы системы ИИ использовались в надлежащих условиях и четко выполняли ряд поставленных им задач. При этом должно быть определено конкретное лицо, на которого может быть возложена ответственность в случае возникновения сбоев и неполадок в работе систем ИИ.
– Оценка и испытание систем ИИ на этапе разработки и внедрения должна быть произведена человеком. Иными словами, именно человек должен в конечном итоге подтвердить безопасность и эффективность систем ИИ для пользователей.
– Организации, внедряющие системы ИИ, должны предусмотреть в своих регламентах выполнение обязательных работ по разбору, анализу и

[467]Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).

выплате компенсации пострадавшим лицам и организациям в случае, если система ИИ причинила вред человеку.

Принцип конфиденциальности.

Конфиденциальность	Приватность, защита данных, личная, частная информация, защита от утечки персональных данных.

Системы ИИ аккумулируют и используют конфиденциальные данные о человеке, такие как имя, адрес пребывания, цифровые следы человека в сети, а также информацию, которая собирается через различные приложения. Здесь встает вопрос — каким образом использование систем ИИ отразится на нас: приведет ли это к лучшему пониманию общественных процессов, более взвешенному принятию решений или же закончится эксплуатацией, контролем со стороны властных структур?

К. Керри заметил, что использование ИИ открывает возможности, которые могут нарушать интересы конфиденциальности, выводя анализ личных данных на новый уровень мощности и скорости[468]. Шеннон Валлор пришел к выводу, что технологии наблюдения, которые нарушают конфиденциальность, в долгосрочной перспективе могут препятствовать нашему моральному и культурному росту[469].

Реган утверждает, что конфиденциальность присуща индивидууму как личности и важна для его саморазвития и установления человеческих отношений с другими членами социума[470]. Э. Блаустейн характеризует конфиденциальность как сохранение независимости, достоинства и целостности человека[471].

[468] Kerry C. Protecting Privacy in an AI-Driven World / C. Kerry. – Brookings, 2020. – 239 p.
[469] Shannon V. Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting / V. Shannon. – Oxford, UK: Oxford University Press, 2016. – 328 p.
[470] Priscilla R. M.. Legislating Privacy: Technology, Social Values, and Public Policy. Chapel Hill: University of North Carolina Press. – URL: https://www.jstor.org/stable/10.5149/9780807864050_regan (дата обращения: 12.04.2022)
[471] Bloustein E. J. Privacy as an aspect of human dignity: An answer to Dean Prosser / E. J. Bloustein // Philosophical Dimensions of Privacy: An Anthology. – 1984. – Pp. 156-202.

Таким образом, конфиденциальность является важнейшим условием сохранения личного достоинства человека, предполагающего осознание им собственной ценности, уникальности, уважения к себе.

Попытки более точно определить конфиденциальность встречаются редко, но чаще всего она связывается с защитой данных[472,473,474] и их безопасностью[475,476].

В настоящее время консенсус в обществе достигнут относительно необходимости контроля за соблюдением конфиденциальности на законодательном уровне, что выгодно отличает рассматриваемый принцип от всех остальных. В то же время в сфере бизнеса многие компании выступают за смягчение правового регулирования в области конфиденциальности, поскольку излишне жесткие требования защиты персональных данных ограничивают рост и развитие этих компаний. Многие страны, напротив, настаивают на сохранении существующих требований по обезличенности общедоступных данных. Но проблема здесь заключается в том, что обезличенность данных далеко не всегда гарантирует их конфиденциальность. Так, в сфере действия правоохранительных органов и правосудия нередко закрытые в соответствии с законом персональные данные оказываются легко доступны для всех заинтересованных сторон. По сетевым следам специалистам не трудно получить информацию о пользователе и его действиях в сети. Это означает, что юридически конфиденциальность данных подлежит защите, однако на деле эти нормы постоянно нарушаются.

---

[472] Internet Society. Artificial Intelligence&Machine Learning: Policy Paper. Internet Society. – URL: https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-andmachine-learning-policy-paper/ (дата обращения: 12.02.2023)

[473] Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)

[474] Microsoft. Responsible bots: 10 guidelines for developers of conversational AI. – URL: https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf (дата обращения: 12.03.2022)

[475] Dawson D et al. Artificial Intelligence - Australia's Ethics Framework. Data61, CSIRO, Australia. – URL: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1 (датаобращения:14.10.2022)

[476] Sony. Sony Group AI Ethics Guidelines. – URL: https://www.sony.com/en/SonyInfo/sony_ai/guidelines.html (дата обращения:15.02.2022)

На наш взгляд, юридический контроль должен быть направлен непосредственно на то, в каких конкретных случаях будут использоваться данные. В. Майер-Шёнбергер, философ в области больших данных, утверждает: «Разумнее было бы отменить практику индивидуального управления конфиденциальностью и заменить ее расширенной подотчетностью, которая предъявлялась бы к пользователям данных, повышая их ответственность за свои действия. Компании, работающие с данными, больше не смогли бы приводить в свое оправдание то, что человек разрешил их использовать. Напротив, им пришлось бы оценивать потенциальные опасности, с которыми могут столкнуться люди при вторичном применении их данных»[477].

Особенно важно получать согласие пользователей на обработку их личных данных и не преступать при этом так называемый порог согласия. Социальное напряжение часто возрастает при возникновении экстренных ситуаций: когда люди находятся на грани жизни и смерти, в условиях катастрофы или пандемии. Например, во время вспышки лихорадки Эбола в 2014 г. срочное рассекречивание и анализ записей звонков пользователей мобильных телефонов в регионе позволили эпидемиологам остановить быстрое распространение опасной болезни. Но публиковать данные запретили из-за риска нарушения конфиденциальности пользователей, а также по причине потенциальной ценности данных для промышленных конкурентов.

В Монреальской декларации принцип конфиденциальности является одним из ключевых, а разработка систем ИИ, по мысли создателей декларации, должна предотвращать риски неправомерного использования персональных данных и защищать их целостность и конфиденциаль-

[477]Майер-Шёнбергер В., Кукьер К. Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим / Пер. с англ. – М.: Издательство «Манн, Иванов и Фербер», 2014. – 240 с.

ность[478]. Согласно документу, люди всегда должны иметь право на «цифровое отключение» в своей личной жизни, и ИИ должен сам предлагать возможность отключения через регулярные промежутки времени, не побуждая людей постоянно оставаться на связи. Мысли и эмоции людей, их переживания должны быть строго защищены от ИИ, способного не только помочь, но и причинить вред. Люди должны иметь полный контроль над информацией о себе и своих предпочтениях. ИИ не должен создавать профили индивидуальных предпочтений, чтобы влиять на поведение людей без их свободного и информированного согласия.

Положение об информированном согласии входит в перечень основных принципов программы «Beijing AI Principles». В базовом документе указанной программы говорится о том, что «стороны должны иметь достаточное информированное согласие о влиянии системы на их права и интересы»[479]. Однако, несмотря на это, сегодня в Китае правительство очень активно использует данные граждан, собирая информацию об их повседневной жизни: досуге, покупательской активности, активности в сети[480]. На практике пользователь, предоставивший информированное согласие, оказывается не осведомлен о том, как именно его данные будут использоваться, для каких задач и в каких ситуациях. Возможно, используя данные, одни зарабатывают и получают незаконную прибыль, а другие не получают никакой компенсации, а, напротив, еще и платят за установку приложений и устройств. Добавим к сказанному и то обстоятельство, что информированное согласие на обработку данных автоматически возлагает всю ответственность за дальнейшее использование информации на самого пользователя. Например, подобное происходит в

[478]Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).
[479]Beijing AI Principles. – URL: https://www.baai.ac.cn/news/beijing-ai-principles-en.html. (дата обращения: 12.09.2022)
[480] Chorzempa M. China’s social credit system: a mark of progress or a threat to privacy? / M. Chorzempa, P. Triolo, S. Sacks // Policy Briefs PB18–14, Peterson Institute for International Economics. – 2018. – № PB18-14.

тех случаях, когда пользователи, предоставляющие такое согласие, считают, что оно является условием работы приложения, без которых нужный им процесс не запустится[481].

В Азиломарских принципах под конфиденциальностью подразумевается, что люди должны иметь доступ к персональным данным и их изменению, а также иметь возможность отслеживать их и контролировать[482].

Последнее становится возможным, благодаря механизму «динамического согласия»[483] (в некоторых источниках «дифференцированного согласия»[484]), предоставляющего пользователям возможности отслеживать и корректировать свои предпочтения в отношении конфиденциальности более детально. Здесь человек вправе разрешать или запрещать сбор отдельных видов данных. К примеру, он получает возможность расплатиться за пользование программой и при этом не предоставляет системе свои данные. Может показаться, что в данном случае пользователь оказывается не в лучшем положении, лишается каких-либо преимуществ, но, на самом деле, именно такой тип отношений свидетельствует о признании ценности личных данных и праве пользователя распоряжаться ими по собственному усмотрению.

Таким образом, в процессе работы системы ИИ с персональными данными в настоящее время существуют риски нарушения конфиденциального характера этих данных. Однако при должном регулировании эти риски могут быть успешно устранены. Важно усилить ответственность

---

[481] Большие данные в социальных и гуманитарных науках. Сб. обзоров и рефератов / РАН. ИНИОН. Центр науч.-информ. исслед. по науке, образованию и технологиям / отв. ред. Гребенщикова Е. Г. – М., 2019. – 193с.

[482] Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)

[483] Dynamic consent: A patient interface for twenty-first century research networks / Kaye J., Whitley E. A., Lund D., Morrison M., Teare H., Melham K. // European journal of human genetics. – 2015. – Vol. 23, № 2. – Pp. 141-146.

[484] Contract for the Web. – URL: https://9nrane41lq4966uwmljcfggvwpengine.netdna-ssl.com/wp-content/uploads/Contract-for-the-Web-3.pdf (дата обращения: 17.09.2022)

перед законом всех лиц, отвечающих за разработку и дальнейшее использование систем ИИ в различных отраслях и сферах общественной жизни.

ЕС рассматривает в качестве главного нормативного акта в сфере защиты личных данных Регламент Европейского Парламента и Совета Европейского Союза 2016 г. «О защите физических лиц при обработке персональных данных и о свободном обращении таких данных, а также об отмене Директивы 95/46/ЕС». Он регламентирует процесс обработки личных данных, который производится полностью либо частично при помощи автоматизированных средств. Регламентом предусмотрен комплекс мер, направленных на защиту пользователей от каких-либо действий с их биометрическими данными, предпринимаемых третьими лицами или организациями без их прямого согласия.

Российское законодательство еще не имеет законодательных актов, которые регулировали бы деятельность по организации и осуществлению видеонаблюдения, использованию биометрических данных. Существуют разрозненные рамочные упоминания о защите данных в отдельных законах, но только малая часть их имеет отношение к системам ИИ (например, ст. 152.1 ГК РФ, ст. 3 ФЗ РФ от 27.07.2006 г. № 152-ФЗ «О персональных данных»).

Принцип конфиденциальности, на наш взгляд, является одним из самых труднореализуемых на практике, поскольку осуществление его может порой противоречить действию других принципов. Например, повышение прозрачности автоматических решений за счет постфактум объяснений может привести к невозможности реализовать принцип конфиденциальности.

Кроме того, реализация данного принципа затруднена тем обстоятельством, что в каждом отдельном случае требуется выработать конкретные правила профилирования, предусмотреть гарантии эффективной защиты данных, определить перечень и меру ответственности тех лиц, кто отвечает за работу с персональными данными пользователей. Здесь

важно осуществить правильный выбор в пользу сотрудничества, а не чрезмерной специализации систем ИИ. Например, эксперт по обеспечению справедливости алгоритмов ИИ и эксперт по конфиденциальности данных должны иметь возможность сотрудничать для разработки модели машинного обучения, которая одновременно является и справедливой, и обеспечивающей конфиденциальность личных данных пользователя. Не менее важно, чтобы решения ИИ контролировались человеком, когда они могут затронуть интересы отдельных лиц, как группы, так и индивидуально. Особой проблемой является проверка качества и эффективности человеческого контроля, а также выбора момента его вмешательства (до того, как решение будет принято машиной или после завершения рабочего процесса).

Подводя итог вышесказанному, можно **сформулировать философский принцип конфиденциальности** в процессах взаимодействия с системами ИИ как обращение к разработчикам и пользователям с требованием:

– предотвращать риски неправомерного сбора, хранения и использования пользовательских данных и защищать их целостность и конфиденциальность на протяжении всего жизненного цикла: от начала и до конца ее обработки;
– обеспечить в процессе разработки систем ИИ надлежащие механизмы управления данными на всех этапах жизненного цикла проектируемых ИИ-систем, в том числе касающиеся сбора данных, контроля использования данных на основе информированного согласия и разрешений, обеспечения личных прав пользователей на владение собственными данными и получение доступа к ним, а также процедуры раскрытия информации о применении и использовании данных;
– усилить юридическую ответственность лиц, отвечающих за разработку и внедрение в социальную практику систем ИИ, в частности, несущих ответственность за работу с персональной информацией пользователей,

установить определенные правила профилирования, чтобы гарантировать надежную защиту данных человека;

– обеспечить максимальную анонимизацию или неидентифицируемость данных, которую возможно осуществить в процессе сбора данных, при обработке информации или ее постобработке.

Таким образом, обобщая в параграфе 3.1 причины возникновения любого возможного ущерба человеку со стороны систем ИИ, мы выделили три философских принципа: принципы прозрачности, ответственности и конфиденциальности, отражающие соответственно требования об открытости и доступности, подотчетности систем ИИ человеку, о наложении запрета на нарушение его личных границ. Указанные принципы раскрывают возможные условия безопасного и эффективного применения ИИ человеком.

Открытость, доступность систем ИИ конкретизирует содержание принципа прозрачности. Указанный принцип подразумевает необходимость предоставления пользователю всей полноты информации о системе ИИ, принципах и методах ее работы, используемых ею данных, возможных негативных последствиях и угрозах, возникающих вследствие ее применения и присущих ей ограничениях.

Требование подотчетности систем ИИ раскрывает сущность принципа ответственности, согласно которому устанавливается обязательность выполнениях всех необходимых условий корректного использования ИИ для любых лиц или организаций, причастных к разработке и эксплуатации систем ИИ на всем протяжении их жизненного цикла. Причем обязательность сопровождается определением конкретного лица, на которое возлагается вся, в том числе, правовая ответственность за недочеты, неисправности, сбои, возникающие в работе системы.

Наконец, философский принцип конфиденциальности раскрывается в хорошо известных требованиях соблюдения правил работы с личными, персональными данными пользователей систем ИИ, что необходимо для

защиты и соблюдения границ личного пространства человека, обеспечения безопасности и защищенности его от нежелательного вторжения третьих лиц.

3.2 Принципы справедливости, автономии и непричинения вреда человеку в контексте использования ИИ

Названные принципы конкретизируют различные грани, аспекты, стороны безопасного и эффективного применения ИИ. Их рассмотрение мы начнем с принципа справедливости.

Принцип справедливости.

Справедливость	Недискриминация, равенство, солидарность, разнообразие, инклюзивность, честность, исключение предрассудков, беспристрастность, стандарты для всех, минимизация предвзятости

Философский энциклопедический словарь определяет справедливость как «понятие о должном, соответствующее определениям о сущности человека и его неотъемлемых правах… содержит в себе требования соответствия между практической ролью различных индивидов (социальных групп) в жизни общества и их социальным положением, между их правами и обязанностями, между деянием и воздаянием, трудом и вознаграждением, преступлением и наказанием, заслугами людей и их общественным признанием[485]. Справедливость в отношении систем ИИ означает предотвращение и устранение предвзятости, равное отношение к группам и индивидам. Философский принцип справедливости призывает к тому, что системы ИИ должны быть разработаны и обучены таким образом, чтобы не создавать, не усиливать и не воспроизводить дискриминацию, основанную на социальных, половых, этнических, культурных, религиозных или иных различиях, а, напротив, способствовать ее устранению либо сокращению.

[485]Философский энциклопедический словарь / Л. Ф. Ильичев, П. Н. Федосеев, С. М. Ковалев, В. Г. Панов. М.: Советская энциклопедия, 1983. – С.622.

Внедрение систем ИИ должно помочь устранить отношения доминирования между группами и людьми, основанные на различиях во власти, богатстве или знаниях. Развитие ИИ должно приносить социальные и экономические выгоды для всех посредством сокращения существующего социального неравенства и социальной уязвимости отдельных лиц, социальных групп и слоев общества. ИИ не должен отдавать предпочтение, предоставлять исключительные права, преференции каким-либо группам, подгруппам или индивидам. Недавние исследования показывают, что алгоритмы машинного обучения способны различать расу и пол человека. Это касается алгоритмов автоматического анализа лица и наборах данных в отношении фенотипических подгрупп[486]. Учитывая это, необходимо исключить, в частности, любую дискриминацию по демографическим показателям (по возрасту, полу, этнической принадлежности и т. д.).

В Стандарте рассмотрения алгоритмических предубеждений (IEEE p7003)[487] отдельным лицам или организациям, создающим алгоритмические системы, предписывается обязательное соблюдение правил, помогающих избежать непреднамеренных, необоснованных и необъяснимо различающихся результатов работы ИИ с данными пользователей. Это практическое руководство для разработчиков позволяет определить, когда им следует отступить, остановиться, чтобы оценить внезапно возникшие в работе их продукте проявления предвзятости, социальной несправедливости.

Авторы Монреальской декларации[488] призывают к тому, что ИИ должен избегать использования полученных данных для блокировки отдельных лиц в профиле пользователя, фиксации личной идентичности

[486]Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification / J. Buolamwini, T. Gebru. – URL: https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf (дата обращения:12.09.2022).
[487]Koene A., Dowthwaite L., Seth S. IEEE P7003™ Standard for Algorithmic Bias Considerations: Work in Progress Paper / A. Koene, L. Dowthwaite, S. Seth // Proceedings of the International Workshop on Software Fairness. – 2018. – Pp. 38-41.
[488]Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).

или ограничения фильтром, что препятствовало бы их личностному росту и развитию, особенно в таких областях, как образование, правосудие или бизнес. Для соблюдения принципа справедливости, считают авторы документа, важно особую осторожность проявлять как на этапах разработки, так и на этапе развертывания ИИ-систем. Иными словами, еще на этапе предварительной обработки данных важно так сбалансировать данные, чтобы избежать в дальнейшем какой-либо дискриминации. В том случае, если признаки дискриминации все же проявятся, на этапе постобработки следует переназначить метки, изначально предсказанные моделью непрозрачной системы ИИ с тем, чтобы осуществить перевод ее в более справедливое состояние. При выявлении дискриминации необходимо рассмотреть вопрос о корректирующих мерах по ограничению или нейтрализации этих рисков, а также задуматься о повышении осведомленности заинтересованных сторон. ИИ не должен разрабатываться или использоваться с целью ограничения свободного выражения идей или возможности услышать различные мнения, поскольку и то, и другое являются неотъемлемыми условиями демократического общества.

Учитывая быстро меняющийся характер и темп смены технологий, методов ИИ, общество обязано знать и обсуждать возникающие в связи с этим проблемы. Это особенно актуально для поддержки наиболее незащищенных слоев населения и групп, для которых в силу определенных социальных особенностей ИИ, как и любые другие технологии эпохи информатизации, представляет серьезную опасность.

ИИ может привести как к лучшему, более справедливому и более эффективному обществу, так и, напротив, способствовать обогащению небольшой группы цифровой элиты. В научном обзоре Б. Митльштат, П. Алло, М. Таддео отмечено, что большая часть рассмотренной ими литературы посвящена тому, как дискриминация возникает вследствие необъективной, предвзятой обработки данных и столь же предвзятого принятия

решений[489]. Ф. Паскуале, анализируя дискриминационную практику применения систем ИИ, тоже выражает тревогу и озабоченность: «если общество не будет привержено принципам справедливой обработки данных, то цифровая дискриминация будет только усиливаться»[490]. Поэтому в области ИИ инклюзивная политика должна быть направлена на достижение двух целей: гарантировать, что развитие этих технологий не приведет к увеличению социального и экономического равенства, и использовать ИИ для уменьшения уже существующего неравенства.

Принцип справедливости неразрывно связан с принципом прозрачности, поскольку обеспечение прозрачности систем ИИ может помочь избежать несправедливости. Ф. Паскуале подчеркивает, что отсутствие прозрачности, характерное для многих систем ИИ, затрудняет оценку его внутренней работы[491]. Нам нужно знать, какие именно параметры были заданы в качестве главных, основных, как данные были собраны, и как они повлияли на результат. Непрозрачность в работе систем ИИ не позволяет получить эту информацию, что впоследствии неизбежно приводит к обострению социальной дискриминации. В ряде случаев в системах ИИ используются достаточно простые модели, поэтому реализовать принцип прозрачности сравнительно легко, если только сами компании или организации не блокируют раскрытие используемых ими алгоритмов. Если речь идет о более сложных моделях (в частности, алгоритмах глубокого обучения), то обеспечить прозрачность процедур становится гораздо сложнее. В данном случае обеспечению прозрачности препятствует не нежелание компании или сложность корреляций, а природа и структура самих алгоритмов, недоступных для понимания, для корректной интерпретации.

[489]Mittelstadt B. D., Allo P. et al. The ethics of algorithms: Mapping the debate / B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi // Big Data and Society. – 2016. – Vol. 3(2). – Pp. 1- 21.
[490] Pasquale F. The Black Box Society. The secret algorithms that control money and information / F. Pascuale // Harvard University Press, Cambridge, 2015. – P. 145.
[491] Там же.

Поучительным примером является использование инструмента под названием «профилирование правонарушителя для осуществления альтернативных санкций (COMPAS)», используемого судами в нескольких штатах США. Данная система используется также для оценки риска совершения нового преступления обвиняемым. ProPublica подвергла резкой критике COMPAS за предвзятое отношение к подсудимым с темным цветом кожи[492]. Однако владельцы системы ИИ не признали свою вину и не раскрыли в полной мере алгоритм, используемый для расчета оценки риска совершения повторного преступления[493]. На данный момент остается недоказанным, действительно ли COMPAS является дискриминационным или нет. Учитывая значимость той сферы деятельности, в которой система COMPAS используется, едва ли можно считать сложившуюся ситуацию нормальной или приемлемой для общества и его институтов.

Концепция «объяснимого ИИ» пытается решить эту проблему, приняв такие меры, как отслеживаемость, аудит и прозрачное информирование о возможностях системы[494]. И это все, что можно было бы сделать человеку для увеличения доступности результатов работы сложных алгоритмов систем ИИ.

Для реализации принципа справедливости необходимо обеспечить прозрачность в отношении использования личных данных человека. Члены общества должны твердо знать о том, какие именно особенности, признаки, свойства людей не должны служить критерием различения их, разделения на иерархические подклассы, подмножества в том или ином контексте, поскольку применение их в данном качестве может оказаться

[492]Angwin J., Larson J. et al. Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks / J. Angwin., J. Larson, S. Mattu, L. Kirchner. – URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (дата обращения: 13.08.2022)

[493]Dieterich W, Mendoza C, Brennan T. COMPAS risk scales: demonstrating accuracy equity and predictive parity performance of the COMPAS risk scales in Broward County / W. Dieterich, C. Mendoza, T. Brennan. – URL: http://www.go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf. (дата обращения:12.09.2022)

[494]Wierzynski C. The challenges and opportunities of explainable AI / C. Wierzynski. – URL: https://www.ai.intel.com/the-challenges-and-opportunities-of-explainable-ai/. (дата обращения: 04.09.2022)

несовместимым с существующим в общественном сознании представлением о справедливом обществе. Другими словами, члены общества должны быть осведомлены о том, какие параметры учитывает алгоритм ИИ, и как они влияют на результат его работы.

Обобщая вышесказанное, сформулируем основные положения **философского принципа справедливости**:

- системы ИИ изначально должны разрабатываться таким образом, чтобы исключить предвзятое, несправедливое отношение по признакам гендерной принадлежности, уровня заработной платы, возраста, расовой принадлежности и других характеристик;
- должно быть обеспечено широкое использование или равный доступ к технологиям ИИ, которые приносят благо обществу, вне зависимости от уровня благосостояния и социального благополучия пользователей;
- внедрение и использование систем ИИ не должно усугублять существующие в обществе формы дискриминации и предвзятости. Напротив, применение систем ИИ призвано создать условия для нивелирования и искоренения подобных явлений;
- необходимо исключить передачу полномочий от человека системам ИИ в ситуациях ответственного нравственного выбора, поскольку только человек как физическое или юридическое лицо может быть полноправным субъектом ответственности;
- в процессе разработки, внедрения и использования систем ИИ следует четко определить субъекта ответственности на каждом из этапов жизненного цикла этих систем ИИ, учитывая специфику каждого из названных периодов.

Принцип автономии.

Автономия	Человеческий контроль над технологией, управляемость, право человека на самооопределение, расширение возможностей человека, право выбора за человеком, а не за машиной.

Прогресс в развитии искусственного интеллекта, открывающий новые, ранее не существовавшие перспективы развития для человека, одновременно создает значительные риски для его автономии, его права и возможности совершать поступки в соответствии с собственными целями и ценностями. Учитывая распространенность технологий ИИ — от поисковых систем до автоматических сканеров лица и тела — и их влияние на жизнь людей, мы делаем вывод, что человеческая идентичность и достоинство сегодня формируются не только в социальном, но и в социотехническом аспекте.

Сохранение автономии человеком во взаимодействии с системами ИИ – важнейшая составляющая благополучия современного человека. Обман, манипуляции или принуждения с помощью систем ИИ, получившие в современном обществе небывалый размах и невиданное ранее разнообразие, позволяют предположить, что технологии ИИ могут серьезно навредить автономии человека и уже сейчас служат орудием совершения мошеннических действий, введения в заблуждение с целью нанесения вреда, ущерба людям во всем мире.

Попытка Cambridge Analytica манипулировать избирателями — лишь один из примеров подобного воздействия[495]. Другой пример – эксперимент Facebook по «эмоциональному заражению», в ходе которого пользователей склоняли к принятию определенных решений под воздействием эмоций[496]. Попытки мошенников похитить финансовые средства

[495] Susser D., Roessler B., Nissenbaum H. Technology, Autonomy, and Manipulation / D. Susser, B. Roessler, H. Nissenbaum // Technical report, Social Science Research Network. – Rochester, NY, June 2019. – 82 p.
[496]Experimental evidence of massive-scale emotional contagion through social networks. – URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4066473/ (дата обращения: 12.09.2022)

клиентов банковской системы, телекоммуникационных компаний посредством технологий ИИ, стали настоящим испытанием и проклятием для абсолютного большинства граждан современных государств.

Таким образом, возникает абсолютная уверенность в том, что системы ИИ могут нанести ущерб человеческой автономии, препятствовать достижению его целей и совершению менее аутентичного выбора, снижать компетенции человека в области автономии, подрывать автономию личности, делая ее более уязвимой для цифровых мошенников, грабителей, преступников. В литературе достаточно часто в контексте обсуждения автономии человека встречается проблема правосубъектности и автономности систем ИИ. По нашему мнению, системы ИИ могут способствовать или препятствовать автономии человека, но при этом сами они не обладают подлинной субъектностью, не являются моральными агентами, и потому не могут в буквальном смысле слова самостоятельно угрожать человеческой автономии.

Тем не менее, активное распространение систем ИИ и их использование в различных мошеннических действиях чрезвычайно актуализируют вопрос, можно ли в процессе взаимодействия с системами ИИ считать выбор пользователя подлинным, и не страдает ли, не ущемлена ли в этом случае способность человека к самоопределению?

Ответ на вопрос начнем с теоретического обсуждения проблемы автономии. Под «автономией» мы понимаем именно автономию человека, а не систем ИИ и исследуем в своей работе вопрос о потенциальном воздействии систем ИИ на человеческую автономию.

В этом же контексте мыслят специалисты, представляющие наиболее авторитетные организации, занимающиеся развитием и управлением процессов создания и использования систем ИИ. Так, группа экспертов высокого уровня (HLEG) Европейской комиссии перечисляет «автономию» в качестве первого из четырех своих ключевых этических принци-

пов в документе под названием «Руководство по надежному ИИ»[497]. Несколько других программных документов, в том числе Монреальская декларация по ответственному развитию искусственного интеллекта[498] и Белая книга Европейской комиссии по искусственному интеллекту[499], в равной степени подчеркивают необходимость защиты и уважения автономии. Организация экономического сотрудничества и развития (ОЭСР) называет автономию одной из основных ценностных установок, ориентированных на человека[500].

В определении автономии мы полагаемся на трактовку, данную М. Райаном и Л. Деси[501], которая согласуется как с аналитической (Франкфурт, 1971[502]; Фридман, 2003[503]), так и с феноменологической точкой зрения (А. Пфандер, 1967[504]; П. Рикёр, 1966[505]). М. Райан и Л. Деси рассматривают автономию как чувство готовности и воли в действиях, которые «одобряются или будут одобрены самим человеком». Иными словами, автономия предполагает действие в соответствии со своими целями и ценностями, а не просто независимость или контроль. Это положение необходимо подчеркнуть, поскольку в процессе взаимодействия с ИИ люди могут одобрять или не одобрять передачу контроля системе ИИ.

Иначе говоря, человеческая автономия в технологических системах требует чувства готовности, желания и стремления совершить действие;

---

[497]High Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.)

[498]Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).

[499]ECWP. On Artificial Intelligence - A European approach to excellence and trust. White Paper COM. – URL: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (дата обращения: 21. 06. 2022)

[500]OECD. Recommendation of the Council on Artificial Intelligence. – URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (дата обращения:17.10.2022)

[501]Ryan M., Deci L. Self-Determination Theory and the Facilitation of Intrinsic Motivation Social Development, and Well-Being / M. Ryan, L. Deci // The American Psychologist, 2000. – Vol.55. – Pp. 68–78.

[502] Frankfurt H. G. Freedom of the Will and the Concept of a Person / H. G. Frankfurt // Journal of Philosophy, 1971. – Vol. 68. – Pp. 5-20.

[503] Friedman M. Autonomy, Gender, Politics / M. Friedman. – New York : Oxford University Press, 2003. – 263 p.

[504]Pfander A. Motive and Motivation. Munich: Barth, Translation in Phenomenology of Willing and Motivation / ed. H. Spiegelberg // Evanston: Northwestern University Press, 1967. – 98 p.

[505]Ricoeur P. Freedom and Nature: The Voluntary and Involuntary / trans: Kohák, E.V. // Evanston: Northwestern University Press, 1966. – Vol.1.

отсутствия внешнего давления или принуждения; возможности контролировать действия, предпринимаемые системой ИИ; отсутствия обмана или преднамеренной дезинформации.

Безусловно, это далеко не полная и не достаточная концептуализация автономии человека в системах ИИ, но она формирует необходимую основу, обеспечивающую условия для устранения большого количества ключевых противоречий, возникающих в контексте проблемы автономии.

Рассмотрим, как эту проблему решают другие исследователи. А. Лайтинен выделяет следующие условия реализации автономии человека: способность к самоопределению, уважение и поддержание человеческой автономии, признание и уважение со стороны других членов общества и т.д. Ученый отмечает, что различные материальные, экономические, правовые, культурные и информационные ресурсы также могут рассматриваться в качестве необходимых составляющих автономии человека[506]. Так, способность человека к самоопределению позволяет выработать такое мировоззрение, при котором человек формирует свои собственные оценки, ценности и действует в соответствии с ними. Эта же способность порождает самоуважение и уважение права на самоопределение других индивидов. Социальное признание и социальное взаимодействие являются важными факторами, способствующими развитию и реализации человеческих способностей. Далее А. Лайтинен выделяет несколько видов препятствий, которые подрывают или ограничивают автономию человека: прямое вмешательство, принуждение, угрозы, неприкрытая сила, манипуляции, идеологическая обработка, обман, подталкивание, патернализм и т.д.

Понятие «самоуважение» Джона Ролза, на наш взгляд, также отражает сущностную сторону автономии человека. Самоуважение, согласно воззрениям Ролза, является неотъемлемой частью развития социальной

[506]Laitinen A, Sahlgren O. AI Systems and Respect for Human Autonomy / A. Laitinen, O. Sahlgren // Front. Artif. Intell. – 2021. – Vol.4.

справедливости, относится к первичным благам, распространять которые призвана его теория справедливости[507]. Самоуважение, по Ролзу, обеспечивает индивиду как «чувство собственной ценности», так и «твердую убежденность в том, что его представление о своем благе, его план жизни заслуживают осуществления»[508]. Самоуважение у Ролза основано на ощущении себя равноправным членом общества, на одинаковых условиях разделяющим с другими ответственность за вынесение фундаментальных суждений по социальным и политическим вопросам.

Действительно, без твердой убежденности в себе и своем жизненном плане, без ощущения себя равным среди равных жить в мире и достигать своих целей намного труднее. Самоуважение обеспечивает равенство и независимость личности по отношению к другим членам общества, обеспечивая и активно формируя в процессах взаимодействия с системами ИИ ее моральное, политическое и культурное развитие. И наоборот, если паттерны лишения прав, дискриминации встроены (сознательно или случайно) в системы ИИ, то последние оказываются способны препятствовать автономии человека, лишая его самоуважения, уверенности в себе и своем праве.

Признавая значимость автономии человека и обсуждая проблему автономии в контексте взаимодействия человека с системами ИИ, создатели отдельных регулирующих эти взаимодействия документов провозглашают принцип автономии как утверждение так называемой позитивной свободы, свободы процветать, права на самоопределение демократическими средствами, права устанавливать и развивать отношения с другими людьми, свободы отзывать согласие или свободы использовать предпочтительную платформу/технологию[509].

---

[507] Rawls J. A Theory of Justice / J. Rawls. – Cambridge, Massachusetts : Belknap Press, 1971. – 562 p.
[508] Там же.
[509] Internet Society. Artificial Intelligence & Machine Learning: Policy Paper. – URL: https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-andmachine-learning-policy-paper/ (дата обращения: 23.09.2022)

Другие источники в трактовке принципа автономии указывают также свободу от технологических экспериментов, манипуляций или наблюдения[510]. Обращаясь к вопросу о человеческой автономии, AI HLEG[511] указывает, что люди должны иметь возможность принимать обоснованные автономные решения в отношении систем ИИ. Необходимо, чтобы автономия не влекла за собой неоправданного принуждения, обмана или манипуляций со стороны систем ИИ. Люди вправе обладать необходимой информацией и инструментами для того, чтобы адекватно и без ущерба для себя и других взаимодействовать с такими системами. Компания Google, которая сформулировала семь этических принципов применения ИИ, считает, что технологии ИИ должны быть подотчетными человеку, подлежать надлежащему руководству и контролю со стороны человека[512].

В области здравоохранения принцип сохранения автономии человека при использовании систем ИИ может быть признан одним из наиболее часто обсуждаемых этических принципов[513]. Проблемы, связанные с этим принципом, заключаются в отсутствии ориентированности на пациента[514] и в отсутствии совместного принятия решений человеком и системой ИИ[515].

Принцип автономии особенно важен при разработке и внедрении автономных систем вооружения, которые могут выбирать и атаковать цели без вмешательства человека или самостоятельно инициировать атаки.

[510]European Group on Ethics in Science and New Technologies. Statement on Artificial Intelligence, Robotics and Autonomous Systems. 2018.
[511] High Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.)
[512] Google. AI at Google: Our Principles. – URL: https://ai.google/principles/ (дата обращения: 18.04.2022)
[513]Karimian G., Petelos E., Evers S. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review / G. Karimian, E. Petelos, S. Evers // AI Ethics, 2022. – Vol. 2. – Pp. 539–551.
[514]Blease C., Kaptchuk T.J. et al. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views / C. Blease, T. J. Kaptchuk, M. H. Bernstein, K. D. Mandl, J. D. Halamka et al. // Journal of Med. Internet Research. – 2019. – Vol. 21 (3). – Pp. 1-24.
[515]Liyanage H., Liaw S.T. et al. Artificial intelligence in primary health care: perceptions, issues, and challenges / H. Liyanage, S. T. Liaw, J. Jonnagaddala, R. Schreiber, C. Kuziemsky C. et al. // Yearb. Med. Inform. – 2019. – Vol. 28. – Pp. 41-46.

Эти системы вызывают опасения по поводу потери человеком контроля над вооружением.

В Национальной стратегии Дании[516] принцип автономии подразумевает возможность принимать осознанные и независимые решения без искусственного интеллекта, лишающего человека права на самоопределение. В других документах подчеркивается, что необходим «контроль человека над системами ИИ и знания об автономных системах»[517], что принципы автономии человека трансформируются в «полномочия человека принимать решения»[518]. Эти выводы близки к идеям Дж. Фьелд, Н. Ахтен, Х. Хиллигосс, А. Надь, М. Срикумар, утверждающих в своей работе, что автономия обычно обеспечивает теоретическое обоснование принципов «человеческого контроля над технологиями»[519].

Несмотря на оживленную дискуссию об автономии человека и во многом совпадающие позиции ее участников, какие-либо единые рекомендации для разработчиков или пользователей, способствующие устранению потенциальных рисков, угрожающих человеческой автономии, выработаны не были. Отсутствие единства в данном вопросе усугубляется сложностью технического ландшафта и большой неопределенностью социальных последствий внедрения ИИ. В совокупности все это способно нивелировать уже предпринятые усилия ученых и специалистов в области ИИ для успешного управления системами ИИ и с еще большей остротой поставить вопрос о целесообразности и безопасности использования систем ИИ в различных сферах общественной жизни.

На наш взгляд, задача систем ИИ в рассматриваемом аспекте – это поддержка людей в принятии взвешенных и информированных решений

[516]The Danish National Strategy for artificial intelligence. – URL:https://en.digst.dk/strategy/the-danish-national-strategy-for-artificial-intelligence/ (дата обращения:17.09.2022)
[517]ECAA. Statement on Artificial Intelligence, Robotics, and ‘Autonomous’ Systems. Technical report. – 2018. – 34 p.
[518] Floridi L., Cowls J. A Unified Framework of Five Principles for AI in Society / L. Floridi, J. Cowls // Harvard Data Science Review, 2022. – Vol. 1 (1). – Pp. 535-545.
[519] Fjeld J., Achten N. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI / J. Fjeld, N. Achten, H. Hilligoss, A. Nagy, M. Srikumar. – 2020. № 2020-1. – URL https://papers.ssrn.com/abstract=3518482 (дата обращения: 21.08.2022)

в соответствии с их собственными целями. Сегодня существует реальная угроза, когда системы ИИ потенциально могут быть использованы для влияния на поведение человека в интересах третьих лиц или организаций посредством скрытых манипуляций, обмана, слежки и понуждения[520], нарушающих индивидуальную автономию человека. Потеря способности принимать самостоятельно решения возникает, если системам ИИ регулярно передается все больше и больше задач, включая принятие решений в социальных, медицинских или финансовых вопросах. Такой патернализм предполагает «благонамеренные» посягательства на автономию человека против его воли, своего рода насильственное «причинение добра». Подобное использование систем ИИ может привести к необоснованному искажению подлинных убеждений, мотивов, ценностных установок и целей человека.

Философский принцип автономии пользователей может быть сформулирован как требование к разработчикам систем ИИ обеспечить в процессах взаимодействия с ИИ:

- реализацию права человека самостоятельно принимать осознанные и независимые решения по любым вопросам в соответствии с собственными целями и ценностями,
- реализацию права человека обладать всей необходимой информацией для успешного взаимодействия с ИИ- системами,
- возможность для человека на любом этапе взаимодействия полностью контролировать процесс,
- защиту человека от влияния третьих лиц или организаций, оказываемого посредством психологических или иных манипуляций, обмана, шантажа, угроз, слежки, понуждения и т.п. действий.

В свете вышесказанного принцип автономии, по нашему мнению, должен стать одним из основных принципов при разработке, внедрении и

[520] EU Member States sign up to cooperate on Artificial Intelligence//European Commission. – URL: https://ec.europa.eu/jrc/communities/en/community/humaint/news/eu-member-statessign-cooperate-artificial-intelligence (дата обращения: 22.04.2022).

оценке функциональности систем ИИ. Ключом к выполнению этого принципа станет право не подчиняться решению, основанному исключительно на автоматизированной обработке. Тем более, если это решение создает правовые последствия для пользователей. Защита людей должна быть важнее всех других соображений полезности. Цель состоит в том, чтобы уменьшить ущерб до такой степени, чтобы избежать его вообще. Государственный сектор должен нести ответственность за обеспечение безопасного для автономии человека внедрения и распространения систем ИИ. Автономное принятие решений людьми является выражением позиции общества, согласно которой в центре внимания при решении любых социальных вопросов и проблем всегда должен быть человек с его правом на развитие и потребностью в защите. При этом речь не идет о вседозволенности и отсутствии общественных норм и регуляторов, поскольку максимум личной свободы выбора в общем порядке развития всех членов общества уравновешивается свободой других и их безопасностью.

Принцип автономии призван помочь людям, пользователям систем ИИ, стать более независимыми, самостоятельными, а не пытаться манипулировать, деквалифицировать, незаконно контролировать и неуважительно относиться к автономии других. Окажется ли влияние систем ИИ на автономию человека положительным или отрицательным, покажет будущее. Ясно то, что потенциальное влияние систем ИИ на человеческую автономию чрезвычайно глубоко и многогранно.

Принцип непричинения вреда, объединивший несколько близких по смыслу требований: безопасность, защита жизни, благодеяние, устойчивость и т.д означает, что деятельность, связанная с исследованием, проектированием, созданием, развертыванием и применением систем ИИ, должна исключать любое непреднамеренное причинение вреда человеку.

Непричинение вреда	Защита жизни, безопасность, физическая и психическая неприкосновенность человека, устойчивость (экологическая безопасность), защищенность человека, благодеяние, благосостояние человечества, обратимость, минимизация ущерба, предотвращение несчастных случаев, управление рисками.

Системы ИИ должны проектироваться, разрабатываться и развертываться, во-первых, с целью предотвращения или решения проблем, отрицательно влияющих на жизнь человека и благополучие окружающей среды, а во-вторых, должны представлять собой идеи и разработки, одобренные обществом и признанные экологически безопасными.

По нашему мнению, принцип непричинения вреда играет ведущую, основную роль во всей совокупности представленных выше принципов. Он является системообразующим по отношению к ним, поскольку, взятые вместе, они раскрывают сущность этого принципа, представляют собой его различные стороны, аспекты, грани. Несоблюдение любого из пяти оставшихся требований автоматически делает невозможным безопасное для человека применение систем ИИ. И наоборот, только через соблюдение принципов прозрачности, конфиденциальности, справедливости, автономии и ответственности можно добиться осуществления принципа непричинения вреда человеку системами ИИ.

Для соблюдения рассматриваемого принципа необходимо, чтобы разработчики гарантировали фальсифицируемость наиболее важных требований и гипотез при запуске систем ИИ в безопасных, защищенных контекстах. Нужно также учитывать необходимость исключительно постепенного развертывания систем ИИ, предоставляющего возможность пошаговой, поэтапной проверки их деятельности, учета возникающих критических замечаний и создания условий для анализа их полной работоспособности. Должна быть готовность со стороны пользователей систем ИИ к быстрой остановке или изменению курса развертывания данных систем, в случае обнаружения опасных или нежелательных последствий.

Принцип непричинения вреда, отсылающий к первому закону робототехники А. Азимова, не исключает того, что вред человеку все-таки может быть причинен вследствие применения системы ИИ, однако основная идея данного принципа заключается в том, что непричинение вреда человеку должно быть заложено разработчиком в саму систему ИИ в качестве целеполагания, в качестве основной цели ее деятельности. И в том случае, если ИИ случайным образом нанес ущерб человеку, разработчики, и пользователи должны иметь возможность найти оптимальное решение для исключения подобных негативных последствий в будущем.

Здесь речь не идет об использовании систем ИИ для каких-либо видов современного оружия. Данная тема требует отдельных исследований и далеко выходит за пределы обсуждения проблемы, рассматриваемой в нашей работе.

В Азиломарских принципах[521] непричинение вреда интерпретируется посредством следующих положений: 1) должна быть обеспечена безопасность и защищенность систем ИИ на протяжении всего срока эксплуатации; 2) необходимо обязательно выявлять причины причинения вреда системой искусственного интеллекта; 3) цели и поведение систем ИИ и человека должны быть в обязательном порядке синхронизированы.

В Рекомендациях по искусственному интеллекту, созданных Организацией экономического сотрудничества и развития (ОЭСР)[522], системы ИИ должны быть нацелены на принесение пользы людям, действие в рамках закона, безопасное функционирование в течение всего срока эксплуатации.

Принцип «предотвращения вреда» в AIHLEG[523] специально оговаривает, что «хотя ИИ может принести много пользы, он также может

[521]Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)
[522]OECD's Principles for responsible stewardship of trustworthy AI. – URL: https://oecd.ai/en/ai-principles (дата обращения: 18.05.2022)
[523]High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.).

причинить вред». Этот вред может быть как материальным (безопасность и здоровье людей, в том числе гибель людей, материальный ущерб), так и нематериальным (утрата права на неприкосновенность частной жизни, ограничение права на свободу выражения мнений, человеческое достоинство, дискриминация, например, при приеме на работу), и может относиться к широкому спектру рисков.

ECWP[524] комментирует эти положения: «нормативная база должна концентрироваться на том, как свести к минимуму различные риски потенциального вреда». Риски могут быть вызваны недостатками в разработке технологии ИИ, быть связаны с проблемами непрозрачности и качества данных.

Существует позиция, согласно которой системы ИИ способствуют повышению общественного благосостояния, справляясь со многими задачами гораздо лучше, чем человек, могут даже взаимодействовать с живыми существами, составлять им компанию и заботиться о них. Однако, по нашему мнению, следует задуматься над тем, что эффект обеспечиваемых системами ИИ «преимуществ» в значительной степени сомнителен. Беспрецедентные изменения, которые производят в нашем мире системы ИИ, как было показано выше, неизбежно порождают разнообразные социально-философские проблемы.

Философский принцип непричинения вреда, по нашему мнению:

- требует разработать тщательную основу для понимания и проверки любых социальных изменений, вызванных использованием ИИ прежде, чем они будут приняты в виде безусловной ценности и названы «преимуществами».

Важно сосредоточиться на более целостном изучении последствий применения технологий ИИ в нынешнем глобальном контексте. Кроме того, следует задаться вопросом, кто определяет, что в каждом конкрет-

[524]European Commission. White paper on artificial intelligence: A European approach to excellence and trust. – . URL: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf. (дата обращения:12.03.2022)

ном случае будет полезно для человека, применяющего системы ИИ, а что нанесет ему вред;

- требует, чтобы системы ИИ проходили строгий процесс проверки безопасности и производительности.

В настоящее время существует множество неопровержимых доказательств того, что эти системы не являются объективными или ценностно нейтральными[525] или даже надежными и безопасными[526];

- призывает разработчиков при создании систем ИИ с целью обеспечения их безопасного использования консультироваться с пользователями, которые будут непосредственно взаимодействовать с этими технологиями, подвергаться их влиянию.

Возможно, понимание целевых установок пользователя, условий эксплуатации систем ИИ, применяемых пользователем методов управления и оценка последствий взаимодействия пользователя с системой ИИ позволят в перспективе сделать указанные системы более безопасными. Быть может, своевременное обеспечение этих условий помогло бы избежать трагической катастрофы двух авиалайнеров Boeing 737 Max, когда пилоты всеми силами пытались ликвидировать сбой неисправных датчиков системы ИИ, но все их попытки были тщетны из-за того, что отсутствовали так называемые «дополнительные функции безопасности», которые данная компания не предоставляет в исходном комплекте, а продает отдельно[527]. Таким образом, оценка уровня рисков должна учитывать известные и потенциальные риски в краткосрочной и долгосрочной перспективе;

[525] O'Neil C. Weapons of math destruction: How big data increases inequality and threatens democracy / C. O'Neil – New York : Crown, 2016. – 272 p.
[526] Qayyum A., Qadir J., Bilal M., Al-Fuqaha A. Secure and robust machine learning for healthcare: a survey / A. Qayyum, J. Qadir, M. Bilal, A. Al-Fuqaha // IEEE Rev Biomed Eng. 2020. – Vol. 14. – Pp.156-80.
[527] Boeing во время аудита обнаружила ошибки в обновленном ПО для самолетов модели 737 MAX. – URL: https://habr.com/ru/news/t/484434/ (дата обращения: 12.09.2022)

- означает, что системы ИИ должны использоваться для расширения человеческих возможностей, защиты благополучия и безопасности каждого члена общества.

Необходимо разрабатывать и внедрять решения, которые не создают препятствий процессу реализации всех потенциальных творческих возможностей человека для достижения гармонии во всех сферах жизни общества: экономической, социальной, духовной и политической;

- формулирует к разработчикам систем ИИ требование учета сложившихся и принятых в данном обществе ценностей, прав и свобод граждан, существующих механизмов и способов раскрытия творческого и интеллектуального потенциала людей, а также имеющихся культурно-исторических особенностей коллективного бытия.

Человек и его потребности должны быть единицей измерения допустимости использования ИИ. Человек, его права и свободы должны рассматриваться как наивысшая ценность;

- предполагает оценку влияния применения ИИ на его пользователей, общество и окружающую среду.

Ответственность акторов ИИ должна учитывать влияние этих систем на общество и граждан на каждом этапе их жизненного цикла, включая неприкосновенность частной жизни, этичное, безопасное и ответственное использование персональных данных. В случае, если развертывание систем ИИ может привести к неприемлемым для человека и общества последствиям, должны быть приняты меры для предотвращения возникновения подобных негативных эффектов в будущем. Использование систем ИИ, способных целенаправленно причинять вред окружающей среде, жизни и благополучию человека, принципиально недопустимо.

Завершая рассмотрение принципов применения систем ИИ, в параграфе 3.2 мы сосредоточили внимание на требованиях, обеспечивающих исключение возможной предвзятости и дискриминации человека со стороны ИИ, а также гарантирующих самостоятельность, независимость че-

ловека в принятии решений, что составляет суть принципов справедливости и автономии, выступающих важнейшими проявлениями принципа непричинения вреда человеку, как центрального, ведущего во всей совокупности правил и норм взаимодействия человека и ИИ. Эти принципы представляют важнейшие признаки безопасного и эффективного применения ИИ.

Социальная несправедливость в контексте применения систем ИИ может возникать из-за объективно существующих различий индивидов, разницы их экономических, политических статусов, принадлежности к разным группам, выделенным по возрасту, полу, национальной принадлежности и т.д. Системы ИИ не должны усугублять уже существующие в обществе предрассудки и заблуждения. Напротив, их применение должно быть организовано таким образом, чтобы все группы населения имели свободный и равный доступ к преимуществами выгодам, предоставляемым новыми технологиями.

Автономия в рассматриваемом контексте предполагает защиту тезиса о главенствующем, центральном положении человека во всей совокупности его отношений с искусственным разумом, поддержку людей в принятии ими взвешенных и обоснованных решений в соответствии исключительно с их собственными целями и задачами, позитивную свободу человека, его неотъемлемое право защищать себя от неоправданного принуждения, обмана или манипуляций со стороны систем ИИ.

Указанные принципы в совокупности с требованиями, рассмотренными выше, позволяют подойти к пониманию центрального, основополагающего принципа применения ИИ, принципа непричинения вреда. Его содержание обобщает призыв о недопущении нежелательных, негативных следствий во взаимодействии человека и ИИ, а также дает современную интерпретацию, толкование, понимание того, что собой представляет корректный, безопасный, эффективный ИИ, осуществляющий по-

мощь и поддержку человеку в решении им собственных проблем, в достижении им собственных целей развития.

Выводы по третьей главе

Рассмотренные в третьей главе диссертации философские принципы прозрачности, ответственности, автономии, конфиденциальности, справедливости и непричинения вреда необходимо отличать от регламентаций другого уровня, таких, как международные нормы в области прав человека, внутренние или региональные нормы, профессиональные нормы, с которыми они, безусловно, взаимосвязаны, однако выступают по отношению к ним в качестве универсальных, всеобщих принципов, имеющих значение для всех без исключения индивидов и групп, взаимодействующих с системами ИИ во всех областях применения данных систем.

Указанные принципы позволяют подойти к пониманию центрального, основополагающего во всей совокупности правил и норм взаимодействия человека и ИИ принципа непричинения вреда, содержание которого сводится к недопущению нежелательных, негативных следствий во взаимодействии человека и ИИ. Кроме того, принцип непричинения вреда выступает современной интерпретацией того, что собой представляет действительно надежный и безопасный, приносящий пользу человеку в решении им проблем и в достижении им собственных целей развития искусственный интеллект. Принцип социальной справедливости как требование, исключающее возможную предвзятость и дискриминацию человека со стороны ИИ, а также принцип автономии, гарантирующий самостоятельное, независимое принятие решений человеком, выступают важнейшими проявлениями принципа непричинения вреда человеку.

Принципы прозрачности, ответственности и конфиденциальности, отражающие требования об открытости и доступности систем ИИ, подотчетности их человеку, о наложении запрета на нарушение его личных

границ, формулируют условия реализации ключевого принципа непричинения вреда.

Философские принципы предназначены служить основой нормотворческой деятельности в сфере применения ИИ. Мы убеждены, что использование ИИ должно регулироваться на законодательном уровне, однако деятельность в указанном направлении сегодня практически отсутствует или крайне редко встречается.

Выявляя сущность и раскрывая содержание основных философских принципов применения систем ИИ, мы параллельно рассматривали и формулировали рекомендации, необходимые для успешной реализации указанных принципов. Среди них инклюзивность и разнообразие командных ролей, обучение и осведомленность об этических ценностях, непрерывное планирование, выполнение и мониторинг фундаментальных принципов в жизненном цикле систем ИИ, начиная с разработки и заканчивая применением на практике. Не менее важной рекомендацией является стандартизация, поскольку требование единого стандарта предназначено для достижения функциональной совместимости и совместной работы между производителями, недопущения отраслевой монополии и ограничения прав пользователей. Нельзя забывать о постоянном контроле со стороны общественности, информировании ее о реальных трудностях и проблемах, возникающих в процессе использования систем ИИ, поскольку именно общество, его социальные группы и институты формулируют цели и задачи той деятельности, для которой проектируются, создаются и используются системы узкого ИИ.

ЗАКЛЮЧЕНИЕ

В современном мире технологии искусственного интеллекта существенным образом влияют на происходящие в обществе процессы, на самые разные аспекты его жизнедеятельности. Их влияние выражается, в частности, в трансформации экономических, политических или иных социальных стандартов и норм.

В нашем исследовании были рассмотрены существующие определения понятий «искусственный интеллект», «системы искусственного интеллекта», выбрана наиболее всеобъемлющая, отражающая современный уровень развития систем ИИ дефиниция; представлены существующие в научной литературе подходы к классификации искусственного интеллекта; установлена специфическая особенность источниковой базы исследования проблем ИИ; выявлены основные социально-философские проблемы, риски и угрозы применения систем искусственного интеллекта, возникающие в ходе их разработки и внедрения. В диссертации подробно рассмотрена реакция мирового сообщества на возникшие трудности, исследован имеющийся опыт этического и правового регулирования систем ИИ, систематизирован обширный свод документов, содержащих правовые и этические принципы регулирования отношений в сфере ИИ, раскрыто содержание основных социально-философских принципов применения систем ИИ, необходимых как для их разработки, внедрения, так и успешного, безопасного применения в любых сферах общественной жизни.

В результате проведенного исследования были сделаны следующие основные выводы.

1. В настоящее время, характеризующееся бурным ростом и развитием, широким распространением ИИ-систем в социальной практике, все еще не выработано единого общепринятого определения понятия «искусственный интеллект». Учитывая имеющийся уровень технологических достижений в области информационных систем и искусственного интел-

лекта, позволяющий с позиции научного исследования обсуждать искусственный интеллект только в виде «узкого» ИИ, а также содержание целого комплекса документов, посвященных проблемам их этического и правового регулирования, наиболее приемлемым, раскрывающим суть изучаемого феномена, представляется понятие «системы искусственного интеллекта», определенное экспертной группой высокого уровня Европейской комиссии как «…программные (и возможно, также аппаратные) системы, разработанные людьми, которые, имея сложную цель, действуют в физическом или цифровом измерении, воспринимая свою среду посредством сбора данных, интерпретируя собранные структурированные или неструктурированные данные, рассуждений на знаниях или обработке информации, полученной из этих данных, и принятия решения о наилучших действиях для достижения поставленной цели. Системы искусственного интеллекта могут либо использовать символические правила, либо изучать цифровую модель, а также могут адаптировать своё поведение, анализируя влияние предыдущих действий на окружающую среду»[528].

2. Результатом обобщения опыта применения систем ИИ стало выделение имеющих наиболее универсальный характер социально-философских проблем, с которыми сталкиваются разработчики и пользователи систем ИИ в процессе их внедрения в разнообразные социальные и производственные процессы. Это проблемы непрозрачности, нарушения конфиденциальности, социальной несправедливости, нарушения автономии, отсутствия ответственности, причинения физического или морального вреда. Эти проблемы мы рассматриваем как социально-философские, поскольку они охватывают широкий комплекс вопросов взаимодействия человека и искусственного интеллекта, содержание которых выходит далеко за пределы предмета науки этики.

[528]High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. 8 April 2019. URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.).

3. Возникновение вышеназванных социально-философских проблем потребовало необходимой реакции со стороны государственных структур, промышленных и деловых кругов, ученых и общественников, предпринявших многочисленные попытки упорядочить отношения человека и искусственного интеллекта посредством публикации документов, содержащих этические и правовые требования, нормы, принципы, предназначенные для разработчиков и пользователей систем ИИ. Большинство указанных документов носит декларативный характер, представляет собой стратегические документы, определяющие направления развития и финансового обеспечения проектов по внедрению и использованию систем ИИ, этические руководства, которые своей целью имеют регулирование отдельных систем, нашедших применение в различных сферах общественной жизни. Сформулированные в них этические и правовые принципы разрознены, не в полной мере обоснованы, их содержание не раскрыто или имеет многозначную интерпретацию. Наиболее значимыми в рамках данного исследования видятся требования, касающиеся непричинения вреда, прозрачности, ответственности, конфиденциальности, справедливости и автономии человека.

4. Нами были выявлены основные философские принципы регулирования разработки и применения систем ИИ, раскрыто их содержание и предназначение. Мы считаем, что они необходимы для предотвращения, снижения риска возникновения и эскалации социально-философских, этических проблем применении систем ИИ во всех отраслях жизни современного социума. Это принципы прозрачности, ответственности, автономии, конфиденциальности, социальной справедливости и непричинения вреда человеку. Указанные принципы, понимаемые нами как наиболее общие требования, предъявляемые к разработчикам и пользователям систем ИИ, необходимо отличать от регламентаций другого уровня, таких, как международные нормы в области прав человека, внутренние или региональные нормы, профессиональные нормы и т.д., поскольку

они выступают по отношению к последним в качестве универсальных, всеобщих требований, имеющих значение для всех без исключения индивидов и групп, взаимодействующих с системами ИИ во всех областях их применения. Эти фундаментальные, философские принципы, по нашему мнению, могут стать основой нормотворческой деятельности в сфере ИИ и руководством для дальнейшего этического регулирования процессов взаимодействия человека с искусственным интеллектом.

Рассмотрение проблем ИИ на уровне социально-философского исследования позволяет сделать вывод об актуальности избранной темы, ее масштабности, сложности и многогранности. Сам объект исследования находится в стадии становления, развития, демонстрирует появление новых свойств и состояний, вследствие чего открываются новые перспективные направления дальнейшего изучения систем ИИ и информационного общества в целом. Исследование этих вопросов невозможно ограничить рамками технических наук и технологического дискурса. Требуется проведение множества междисциплинарных, философских исследований, рассматривающих процессы информатизации сквозь призму проблем человека, с позиции признания человека высшей ценностью, без чего преимущества, предлагаемые системами ИИ, все больше вызывают недоверие в обществе и ставят под сомнение достижимость широко декларируемых в любом государстве целей улучшения качества жизни человека, прогресса в развитии социальных систем и гармонизации общественных отношений.

СПИСОК ЛИТЕРАТУРЫ

I. Нормативная правовая информация

1. Восьмой закон о внесении изменений в Закон о дорожном движении от 16 июня 2017 года. – URL: http://robopravo.ru/uploads/s/z/6/g/z6gj0wkwhv1o/file/5MZOclyT.pdf
2. Дорожная карта развития «сквозной» цифровой технологии «Нейротехнологии и искусственный интеллект». – URL: https://digital.gov.ru/ru/documents/6658/ (дата обращения 19.11.2022 г.); Дорожная карта развития «сквозной» цифровой технологии «Компоненты робототехники и сенсорика». – URL: https://digital.gov.ru/ru/documents/6666/ (дата обращения 11.12.2022 г.).
3. Законопроект № 922869-7. – URL: https://sozd.duma.gov.ru/bill/922869-7 (дата обращения: 20.04.2020)
4. Заседание Президиума Российской академии наук «О внедрении робототехники в отечественную медицину – проблемы и пути решения». URL: http:// www.ras.ru/news/news_release.aspx?ID=2eaccb0cd887-4d02-bb10-7caae48b083b&print=1 (дата обращения: 23.02.2022)
5. «О проведении эксперимента по опытной эксплуатации на автомобильных дорогах общего пользования высокоавтоматизированных транспортных средств» : Постановление Правительства РФ от 26 ноября 2018 г. № 1415. – URL:http://publication.pravo.gov.ru/File/GetFile/0001201811270008?type=pdf (дата обращения: 01.10.2012).
6. «Об утверждении Правил государственной регистрации медицинских изделий». Постановление Правительства РФ от 27.12.2012 № 1416 // Собрание законодательства РФ. 2013. № 1. – Ст. 14.
7. Проект Федерального закона «О внесении изменений в Гражданский кодекс Российской Федерации в части совершенствования правового регулирования отношений в области робототехники» / Проект Д. С. Гришина (Grishin Robotics) (не вносился в Госдуму). –

URL://https://www.dentons.com/ru/insights/alerts/2017/january/27/dentons-develops-first-robotics-draft-law-in-russia (дата обращения: 20.08.2022).

8. Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»). – URL: https://base.garant.ru/72838946/ (дата обращения: 14.05.2022).
9. «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных». Федеральный закон от 24.04.2020 № 123-ФЗ // СЗ РФ. 2020 № 17. Ст. 2701
10. «Об основах охраны здоровья граждан в Российской Федерации». Федеральный закон от 21.11.2011 № 323-ФЗ // Собрание законодательства РФ. 2011. № 48. Ст. 6724.
11. «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.». Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р // СЗ РФ. № 35, 2020 г. Ст. 5593
12. Acm code of ethics and professional conduct. – URL: https://www.acm.org/binaries/content/assets/about/acmcode-of-ethics-booklet.pdf (дата обращения:15.03.2022)
13. AI Strategy 2019. AI for Everyone: People, Industries, Regions and Governments. – URL: https://www8.cao.go.jp/cstp/english/humancentricai.pdf (дата обращения: 21.06.2021)
14. Asilomar AI Principles. – URL: https://www.artificial-intelligence.blog/news/asilomar-ai-principles (дата обращения: 12.12.2021)

15. Automated and Electric Vehicles Act. – URL: https://www.legislation.gov.uk/ukpga/2018/18/contents/enacted (дата обращения: 03.06.2022).
16. Australia's AI Ethics Principles. – URL: https://www.industry.gov.au/data-andpublications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles (дата обращения:30.01.2022)
17. Artificial Intelligence. An Accountability Framework for Federal Agencies and Other Entities. – URL: https://www.gao.gov/assets/gao-21-519sp.pdf (дата обращения:25.05.2022)
18. Artificial Intelligence at Google: Our Principles. – URL: https://ai.google/principles/ (дата обращения: 30.08.2022)
19. Artificial intelligence principles and ethics. – URL: https://www.digitaldubai.ae/initiatives/ai-principles-ethics (дата обращения: 23.09.2022)
20. Artificial Intelligence Initiative Act. 2019. – URL: https://www.congress.gov/bill/116th-congress/senate-bill/1558/text (дата обращения: 12.09.2022)
21. B. BS8611, Robots and robotic devices, guide to the ethical design and application of robots and robotic systems, British Standards Institute, 2016.
22. Beijing AI Principles. – URL: https://www.baai.ac.cn/news/beijing-ai-principles-en.html. (дата обращения: 12.09.2022)
23. "Bundesdatenschutzgesetz". – URL:https://www.datenschutz-wiki.de/BDSG (дата обращения: 12.05.2022)
24. Bundesregierung, Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz. – URL: https://www.bmbf.de/files/180718Eckpunkte_KI-StrategiefinalLayout.pdf (дата обращения: 13.08.2022.).
25. CAHAI – Ad hoc Committee on Artificial Intelligence. – URL: https://www.coe.int/en/web/artificial-intelligence/cahai (дата обращения: 19.04.2021 г.).

26. Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) // OJ C 252, 2018. – P. 239–257. – URL: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN (дата обращения: 09.04.2022 г.).
27. Cyber security law. – URL: https://www.tradecommissioner.gc.ca/china-chine/cybersecuritycybersecuritechina-chine.aspx?lang=eng (дата обращения:30.04. 2021)
28. Data protection act. – URL: https://www.hhs.gov/hipaa/for-professionals/index.html (дата обращения: 15.03.2022)
29. Elaboration of a Recommendation on the ethics of artificial intelligence. – URL: https://en.unesco.org/artificial-intelligence/ethics#drafttext (дата обращения:27.01.2022)
30. Executive Order on Maintaining American Leadership in Artificial Intelligence. – URL: https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/ (дата обращения: 20.06.2022)
31. Estonia accelerates artificial intelligence development. – URL: https://e-estonia.com/estonia-accelerates-artificial-intelligence/ (дата обращения: 19.10.2022 г.); Kratid Eesti heaks (на эстонском языке). URL: https://www.kratid.ee/ (дата обращения: 19.10.2022 г.).
32. Ethics Commission. Automated and Connected Driving. – URL: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connecteddriving.pdf?__blob=publicationFile (дата обращения: 14.01.2022)
33. Ethical design and use of automated decision systems. – URL: https://viewer.joomag.com/ciosc-standardcan-ciosc-101-2019/0672323001570024164 (дата обращения: 16.05.2022)

34. Ethical Principles for AI in Medicine. The Royal Australian and New Zealand College of Radiologists. – URL: https://www.ranzcr.com/documents/4952-ethical-principles-for-ai-in-medicine/file (дата обращения: 19.09.2022)
35. Eun-jin K. Korean Government to Repeal Regulations in AI Industry December 18, 2019. URL: http://www.businesskorea.co.kr/news/articleView.html?idxno=39324 (дата обращения: 10.03.2022
36. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment (Dec. 3/4, 2018), European Commission for the Efficiency of Justice (CEPEJ)(2018)14 // Adopted at the 31st plenary meeting of the CEPEJ. – Strasbourg, 2018. – URL: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (дата обращения: 19.11.2022г.)
37. Executive Order on Maintaining American Leadership in Artificial Intelligence. – URL: https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/ (дата обращения: 20.06.2022)
38. G20 Ministerial Statement on Trade and Digital Economy. – URL: https://www.mofa.go.jp/files/000486596.pdf (дата обращения: 20.05.2022)
39. General data protection regulation. – URL: https://gdpr-info.eu/ (дата обращения: 17.02.2022)
40. Government offices of Sweden, National Approach to artificial intelligence. – URL: https://www.regeringen.se/4aa638/contentassets/a6488ccebc6f418e9ada18bae40bb71f/nationalapproach-to-artificial-intelligence.pdf (дата обращения: 21.03.2022 г)
41. Government of Canada, “Directive on Automated Decision-Making”, 2019.

42. Government of Canada. Responsible use of artificial intelligence (AI). – URL: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsibleuse-ai.html (дата обращения:21.07.2022)
43. High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. – URL: https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения 19.10.2022 г.).
44. IEEE. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. IEEE Standards Association. – URL: https://standards.ieee.org/industryconnections/ec/autonomous-systems.html (дата обращения:20.04.2022
45. Intel. Artificial Intelligence: The Public Policy Opportunity. – URL: https://blogs.intel.com/policy/files/2017/10/Intel-Artificial-Intelligence-Public-Policy-White-Paper-2017.pdf (дата обращения (16.02.2022)
46. Intelligent Robots Development and Promotion Act, Act № 9014, Mar. 28, 2008. – Act № 13744. – 2016, art. 2. – URL: http://elaw.klri.re.kr/eng_service/lawView.do?hseq=39153&lang=ENG (дата обращения: 10.06.2022 г.)
47. ISO/IEC JTC 1/SC 42 Artificial intelligence. – URL: https://www.iso.org/ru/committee/6794475.html?view=participation (дата обращения: 20.12.2022)
48. Loi pour une Republique numerique. – № 2016-1321. – URL: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id (дата обращения:23.09.2022)
49. Model Artificial Intelligence Governance Framework (Second Edition). Infocomm Media Development Authority Singapore. – URL: https://www.imda.gov.sg/-/media/Imda/Files/Infocomm-Media-Landscape/SG-Digital/Tech-Pillars/Artificial-Intelligence/Primer-for-

second-edition-of-the-Model-Framework.pdf?la=en (дата обращения - 19.10.2022)

50. Montréal Declaration: Responsible AI. – URL: https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_ Artificial_Intelligence_2018.pdf (дата обращения: 16.08.2022).
51. Microsoft. The future computed: Artificial Intelligence and its role in society: chapter 2. – URL://news.microsoft.com/cloudforgood/_media/downloads/the-future-computed-english.pdf (дата обращения:17.09.2022)
52. Mid-to Long-term Master Plan in Preparation for the Intelligent Information Society: Managing the Fourth Industrial Revolution. – URL: https://english.msit.go.kr/cms/english/pl/policies2/__icsFiles/afieldfile/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf (дата обращения: 20.02.2022)
53. Ministry of Finance and Ministry of Industry, Business and Financial Affairs. National Strategy for Artificial Intelligence. – URL: https://eng.em.dk/media/13081/305755-gb-version_4k.pdf (дата обращения: 11.07.2022 г.).
54. National Artificial Intelligence Strategy. – URL: https://www.smartnation.gov.sg/why-Smart-Nation/NationalAIStrategy (дата обращения:12.12.2022)
55. Oceanis. – URL: https://ethicsstandards.org/repository/ (дата обращения:23.07.2022)
56. OECD moves forward on developing guidelines for artificial intelligence (AI). – URL: https://www.oecd.org/going-digital/ai/oecd-moves-forward-on-developing-guidelines-for-artificial-intelligence.htm (дата обращения: 17.08.2022).

57. Pan-Canadian Artificial Intelligence Strategy, Invest in Canada. – URL: https://www.investcanada.ca/why-invest/pancanadian-artificial-intelligence-strategy (дата обращения: 23.06.2022 г.)
58. Personal information protection and electronic documents act. – URL: https://lawslois.justice.gc.ca/ENG/ACTS/P-8.6/index.html (дата обращения: 21.02.2022)
59. Proposition de loi constitutionnelle relative à la Charte de l'intelligence artificielle et des algorithmes. – URL: http://www.assemblee-nationale.fr/dyn/15/textes/l15b2585_proposition-loi (дата обращения: 20.10.2022)
60. Protection from online falsehoods and manipulation act. – URL:https://sso.agc.gov.sg/Acts-Supp/18-2019 (дата обращения: 11.03.2021)
61. Putting principles into practice at Microsoft. – URL: https://www.microsoft.com/en-us/ai/our-approach?activetab=pivot1:primaryr5 (дата обращения:12.09.2022)
62. Research ICT Africa (RIA). – URL: https://researchictafrica.net/people/ (дата обращения: 21.09.2022)
63. SAP. European Prosperity Through Human-Centric Artificial Intelligence // The Intelligent Enterprise. – 2018. – P.28. – +URL://www.sap.com/documents/2018/01/3e67a134-ee7c-0010-82c7-eda71af511fa.html (дата обращения:12.04.2022)
64. SAP's Guiding Principles for Artificial Intelligence. – URL: https://news.sap.com/2018/09/sap-guiding-principles-for-artificial-intelligence/ (дата обращения: 21.07.2022)
65. Select Committee on Artificial Intelligence. AI in the UK: ready, willing and able // House of Lords, UK, Apr. 2018.
66. Social Principles of Human-Centric AI. – URL: https://www.cas.go.jp/jp/seisaku/jinkouchinou/pdf/humancentricai.pdf (дата обращения: 12.09.2022)

67. State Council, Notice of Issuing New Generation Artificial Intelligence Development Plan. – Guo Fa, 2017. – № 35. – URL: http://www.gov.cn/zhengce/content/2017- 07/20/content_5211996.html (на китайском языке) дата обращения: 16.05.2022 г.)
68. Strategic Council for AI technology. – URL: http://www.nedo.go.jp/content/100865202.pdf (дата обращения: 13.10.2022 г.).
69. The Different Faces of AI Ethics Across the World: A Principle-Implementation Gap Analysis Lionel Nganyewou Tidjon and Foutse Khomh, Senior Member, IEEE. – URL: https://arxiv.org/pdf/2206.03225.pdf (дата обращения: 21.09.2022)
70. The Guidelines on Artificial Intelligence and Data Protection // Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data (Convention 108). – URL: https://rm.coe.int/lignesdirhttps:/rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8 (дата обращения: 18.11.2022 г.).
71. The Public Voice. Universal Guidelines for Artificial Intelligence. – URL: //thepublicvoice.org/ai-universal-guidelines (дата обращения:12.01.2022)
72. The Toronto Declaration. – URL: https//www.torontodeclaration.org/declaration-text/english/ (дата обращения:21.01.2022
73. UNI Global Union. Top 10 principles for ethical artificial intelligence. – URL://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf (дата обращения: 12.08.2022)
74. United Nations Activities on Artificial Intelligence (AI). – URL: https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2019-1-PDF-E.pdf (дата обращения: 01.12.2022 г.).
75. Villani C. For a meaningful artificial intelligence towards. A French and European strategy. – URL:

https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf (дата обращения 12.12.2022).

76. Visually probe the behavior of trained machine learning models, with minimal coding. – URL: https://pair-code.github.io/what-if-tool/ (дата обращения: 21.09.2022)

II. Источники (опубликованные):

77. Азимов А. Я, Робот / А. Азимов. – Санкт-Петербург : МП Издатель, 1991. – 800 с.

78. Алексеев А. Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты / А. Ю. Алексеев. – Москва : ИИнтеЛЛ, 2013. – 304 с.

79. Алексеев П. В. Социальная философия / П. В. Алексеев. – Москва : ООО «ТК Велби», 2003. – 256 с.

80. Антимонов Б. С. Гражданская ответственность за вред, причиненный источником повышенной опасности / Б. С. Антимонов. – Москва : Юрид. лит., 1952. – 300 с.

81. Аристотель. Сочинения в 4 томах / Аристотель. – Т. 4. – Москва : Мысль, 1984. – 832 с.

82. Бодрийяр Ж. Система вещей / Пер. с фр. С.Н. Зенкина. – Москва : «Рудоми-но», 1999. – 224 с.

83. Болотова Л. С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях: учебник / Л. С. Болотова. – Москва : Финансы и статистика, 2012. – С.38-39.

84. Бостром Н. Искусственный интеллект. Этапы. Угрозы. Стратегии / пер. с англ. С. Филина. – Москва : Манн, Иванов и Фербер, 2016. – 496 с.

85. Васильев В. В. Трудная проблема сознания / В. В. Васильев. — Москва : Прогресс-Традиция, 2009. – 272 с.

86. Гринбаум А. Машина-доносчица: как избавить искусственный интеллект от зла / А. Гринбаум. – Москва : ТрансЛит, 2017. – 76 с.

87. Де Гарис Х. Искусственный мозг: подход с развитым модулем нейронной сети / Х. Де Гарис // World Scientific. – 2010. – 400 с.

88. Деннет Д. Виды психики. На пути к пониманию сознания / пер. А. Веретенникова. – Москва : Идея-Пресс, 2004.

89. Иванов Д. В. Виртуализация общества / Д. В. Иванов. — Санкт-Петербург, 2002. – С. 24.

90. Интеллектуальные системы управления. Теория и практика: учеб. пособие. Москва : Радиотехника, 2009. – 392 с.

91. Кутырев В. А. Культура и технология: борьба миров / В. А. Кутырев. – Москва : Прогресс-Традиция, 2001. – 240 с.

92. Ленк Х. Размышления о современной технике / Х. Ленк. – Москва : Аспект-Пресс, 1996. – 183 с.

93. Маклюэн М. Понимание медиа: внешнее расширение человека / М. Маклюэн. – Москва : Жуковский : «Канон-пресс-Ц», 2003. – 464 с.

94. Маркузе Г. Одномерный человек. / Г. Маркузе // Исследование идеологии Развитого Индустриального Общества. – Москва : Reefl-book, 1994. – 368 с.

95. Майер-Шёнбергер В., Кукьер К. Большие данные: Революция, которая изменит то, как мы живем, работаем и мыслим / Пер. с англ. – Москва : Издательство «Манн, Иванов и Фербер», 2014. – 240 с.

96. Мелюхин И. С. Информационное общество: истоки, проблемы, тенденции развития / И. С. Мелюхин. – Москва : Издательство Московского университета, 1999. – 206 с.

97. Мински М. Фреймы для представления знаний / Мински М. – Москва : Мир, 1979. – 151 с.

98. Мосечкин И. Н. Искусственный интеллект и уголовная ответственность: проблемы становления нового вида субъекта преступления / И. Н. Мосечкин // Вестник Санкт-Петербургского университета. Право. – № 10 (3). – С. 461-476.

99. Морхат П. М. Искусственный интеллект: правовой взгляд / П. М. Морхат // Институт государственно-конфессиональных отношений и права. – Москва : Буки Веди, 2017. – 257 с.
100. Морхат П. М. Право интеллектуальной собственности и искусственный интеллект / – Москва: ЮНИТИ-ДАНА, 2018. – 121 с.
101. Мэмфорд Л. Техника и природа человека / Д. Мэмфорд // Новая технократическая волна на Западе. Москва : Прогресс-Традиция, 1986. – С. 225-239.
102. Нейсбит Д. Высокая технология, глубокая гуманность: Технологии и наши поиски смысла / Д. Нейсбит при участии Наны Нейсбит и Дугласа Филипса; пер. с англ. А. Н. Анваера // Москва: АСТ : Транзиткнига, 2005. – 381 с.
103. Нильсон Н. Искусственный интеллект: методы поиска решений / Пер. с англ. В. Л. Стефанюка; под редакцией С.В. Фомина. – Москва : Мир, 1973. – 272 с.
104. Ортега-и-Гассет Х. Размышления о технике / Х. Ортега-и-Гассет // Избранные труды пер. с исп.; сост., предисл. и общ. ред. А. М. Руткевича. – Москва : Весь Мир, 1997. – С. 164-232.
105. Пенроуз Р. Новый ум короля: О компьютерах, мышлении и законах физики / Пер. с англ. под. ред. В.О. Малышенко, 3-е изд. – Москва : Издательство ЛКИ, 2008. – С. 328.
106. Рело Ф. Техника и ее связь с задачею культуры / Ф. Рело. – Санкт-Петербург : Типография министерства путей сообщения, 1885. – 27 с.
107. Ридлер А. Германские высшие учебные заведения и запросы двадцатого столетия / А. Ридлер. – Санкт-Петербург : типография Р. Голике, 1900. – 30 с.
108. Розин В. М. Понятие и современные концепции техники / В. М. Розин. – Москва : Институт философии РАН, 2006. – 255 с.

109. Ручкин В. Н., Фулин В. А. Универсальный искусственный интеллект и экспертные системы / В. Н. Ручкин, В. А. Фулин. – Санкт-Петербург : БХВ-Петербург. – 2009. – 240 с.

110. Степин В. С., Горохов В. Г., Розов М. А. Философия науки и техники. / В. С. Степин, В. Г. Горохов, М. А. Розов. – Москва, 1996. – 380 с.

111. Смирнов С. А. Человек перехода / Отв. за вып. П. А. Носова. – Новосибирск, 2006. – 177 с.

112. Серль Дж. Р. Сознание, мозг и программы / Дж. Р. Серль // Аналитическая философия: Становление и развитие: Антология / Общ. ред. и сост. А. Ф. Грязнов. – Москва, 1998. – 528 с.

113. Тополь Э. Будущее медицины: Ваше здоровье в ваших руках / Э. Тополь. – Москва : Альпина нон-фикшн, 2016. – 491 с.

114. Турчин А. В. Футурология. XXI век. Бессмертие или глобальная катастрофа? / А. В. Турчин, М. А. Бахтин. — Москва, 2013. — URL: https://libking.ru/books/nonf-/nonf-publicism/205876-aleksey-turchin-rossiyskaya-akademiya-nauk.html. (дата обращения: 12.09.2022).

115. Тьюринг А. Может ли машина мыслить / А. Тьюринг. – Москва : Едиториал УРСС, Ленанд. – 2016. – 128 с.

116. Философский энциклопедический словарь / Л. Ф. Ильичев, П. Н. Федосеев, С. М. Ковалев, В. Г. Панов. – Москва : Советская энциклопедия, 1983. – С. 622.

117. М. Форд. Роботы наступают: развитие технологий и будущее без работы / пер.с англ. С. Чернина. – Москва : Альпина нон-фикшн, 2016. – 572 с.

118. Фукуяма Ф. Наше постчеловеческое будущее: Последствия биотехнологической революции; пер. с англ. М. Б. Левина / Ф. Фукуяма. – Москва : АСТ : ЛЮКС, 2004. – 349 с.

119. Хабермас Ю. Будущее человеческой природы. Москва : Весь мир, 2002. – 144 с.

120. Хабермас Ю. Техника и наука как «идеология» / Ю. Хабермас. – Москва : Праксис, 2007. – 208 с.

121. Хайдеггер М. Вопрос о технике / М. Хайдеггер // Время и бытие : статьи и выступления : пер. с нем. – Москва, 1993. – 49 с.

122. Юдковски Э. Систематические ошибки в рассуждениях, потенциально влияющие на оценку глобальных рисков. Новые технологии и продолжение эволюции человека? / Э. Юдковски // Трансгуманистический проект будущего. – Москва : URSS, 2008. – с. 182-225.

III. Периодические издания:

123. Аверинская С. А., Севостьянова А. А. Создание искусственного интеллекта с целью злонамеренного использования в уголовном праве Российской Федерации / С. А. Аверинская, А. А. Севостьянова // Закон и право. – 2019. – № 2. – С. 94-96.

124. Апресян Р. Г. Этика и дискуссии об искусственном интеллекте / XI международная конференция «Теоретическая и прикладная этика: Традиции и перспективы - 2019. К грядущему цифровому обществу. Опыт этического прогнозирования (100 лет со дня рождения Д. Белла - 1919-2019)». Санкт-Петербургский Государственный Университет, 21-23 ноября 2019 г. Материалы конференции / Отв. ред. В. Ю. Перов. – Санкт-Петербург : ООО «Сборка», 2019. – С. 169-170.

125. Архипов В. В., Наумов В. Б. Искусственный интеллект и автономные устройства в контексте права: о разработке первого в России закона о робототехнике / В. В. Архипов, В. Б. Наумов // Труды СПИИРАН. – 2017. – № 6. – С. 157-170.

126. Архипов В. В., Наумов В. Б. О некоторых вопросах теоретических оснований развития законодательства о робототехнике: аспекты воли и правосубъектности / В. В. Архипов, В. Б. Наумов // Закон. – 2017. – № 5. – С. 157-170.

127. Асеева И. А. Искусственный интеллект и большие данные: этические проблемы практического использования. (Аналитический обзор) / И.

А. Асеева // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 8 : Науковедение. – 2022. – № 2. – С. 89-98.

128. Бадмаева М. Х. Повседневная жизнь человека в умном городе / М. Х. Бадмаева // Вестник Бурятского государственного университета. Философия. – 2020. – Вып. 4. – С. 34.

129. Баева Л. Социокультурные изменения в условиях развития высоких технологий / Л. Баева // Инноватика и экспертиза : научный журнал. – 2012. – №2 (11). – С. 110-119.

130. Баева Л. В., Храпов С. А. Цифровизация образовательного пространства: эмоциональные риски и эффекты / Л. В. Баева, С. А. Храпов // Вопросы философии. – 2022. – №4. – С. 16-24.

131. Бахтеев Д. В., Тарасова Л. В. Применение искусственного интеллекта в деятельности арбитражных судов РФ: перспективные направления и проблемы / Д. В. Бахтеев, Л. В. Тарасова // Вестник Костромского государственного университета. – 2020. – Т. 26, № 4. – С. 249-254.

132. Готовцев П. М., Ройзенсон Г. В. Характеристика проектов стандартов на этичный искусственный интеллект IEEE / П. М. Готовцев, Г. В. Ройзенсон // 390 Этика и «цифра». – 2020. – URL: https://ethics.cdto.center/ieee (дата обращения: 12.04.2022).

133. Горохов В. Г. Место и роль философии техники и современной философии и её органическая связь с философией науки / В. Г. Горохов // Философия науки. – 2011. – № 1. – С. 181-199.

134. Горохов В. Г. Социальные проблемы нанотехнологии / В. Г. Горохов // Высшее образование в России. –2008. – № 3. – С. 84-98.

135. Гусейнов А. А. Размышления о прикладной этике / Доклад на основе статьи: Размышления о прикладной этике // Ведомости НИИПЭ, Вып. 25: Общепрофессиональная этика. – Тюмень : НИИПЭ, 2004.

136. Евстратов А. Э., Гученков И. Ю. Пределы применения искусственного интеллекта (правовые проблемы) / А. Э. Евстратов, И. Ю. Гученков // Правоприменение. – 2020. – Т. 4(2). – С.13-19.

137. Карпов В. Э, Готовцев П. М., Ройзензон Г. В. К вопросу об этике и системах искусственного интеллекта / В. Э. Карпов, П.М. Готовцев, Г. В. Ройзензон // Философия и общество. – 2018. – №2 (87). – С. 84-105.

138. Козаев Н. Ш. Состояние уголовной политики и вопросы преодоления кризисных явлений в уголовном праве / Н. Ш. Козаев // Юридический вестник ДГУ. – 2016. – №1. – С. 98.

139. Лаптев В. А. Электронные доказательства в арбитражном процессе/ В. А. Лаптев // Российская юстиция 2. – 2017. – С. 56-59.

140. Маслов С. Ю. Обратный метод установления выводимости в классическом исчислении предикатов / С. Ю. Маслов // ДАН СССР. – 1964. – Т. 159, № 1. – С. 17-20.

141. Нагель Т. Каково быть летучей мышью? // Хоф-штадтер Д., Деннет Д. Глаз разума / пер. с англ. М. А. Эскиной. Самара : «Бахрах-М», 2003. – С. 349-360.

142. Наумов В. Б. Право в эпоху цифровой трансформации: в поисках решений / В. Б. Наумов // Российское право: образование, практика, наука. – 2018. – № 6 (108). – С. 4-11.

143. Незнамов А. В. О концепции регулирования технологий искусственного интеллекта и робототехники в России / А. В. Незнамов // Закон. – 2020. – № 1. – С. 171-185.

144. Осипов Г.С. Искусственный интеллект: состояние исследований и взгляд в будущее / Г. С. Осипов // Новости искусственного интеллекта. – 2001. – № 1. – С. 3-13.

145. Понкин И. В. Искусственный интеллект с точки зрения права / И. В. Понкин, А. И. Редькина // Вестник Российского университета дружбы народов. Серия: Юридические науки. – 2018. – Т. 22, №1. – С. 91–109.

146. Попова А. В. Этические принципы взаимодействия с искусственным интеллектом как основа правового регулирования // Правовое государство: теория и практика. – 2020. – № 3 (61). – С. 34-43.

147. Попова А. В. Этические принципы взаимодействия с искусственным интеллектом как основа правового регулирования / А. В. Попова // Правовое государство: теория и практика. – 2020. – № 3 (61). – С. 34-43.
148. Разин А. В. Этика искусственного интеллекта / Разин А. В. // Философия и общество. – 2019. – №1. – С. 57-73.
149. Ракитов А. И. Наш путь к информационному обществу / А. И. Ракитов // Теория и практика общественно-научной информации. – Москва : ИНИОН, 1989. – С. 50-68.
150. Резаев А. В., Трегубова Н. Д. Искусственный интеллект и искусственная социальность: новые явления и проблемы для развития медицинских наук / А. В. Резаев, Н. Д. Трегубова // Эпистемология и философия науки. — Москва, 2019. — Т. 56, №4. – С. 183-199.
151. Ройзензон Г. В. Проблемы формализации понятия этики в искусственном интеллекте / Г. В. Ройзензон // XVI национальная конференция по искусственному интеллекту с международным участием КИИ-2018. – Москва, 2018. – С. 245-252.
152. Синельникова В. Н., Ревинский О. В. Права на результаты искусственного интеллекта / Синельникова В. Н., Ревинский О. В. // Копирайт. – 2017. – № 4. – С. 24-27.
153. Степин В.С. Научное познание и ценности техногенной цивилизации / В.С. Степин // Вопросы философии. – 1989. – № 10. – С. 3-18.
154. Флоренский П. Органопроекция / П. Флоренский // Русский космизм: антология философской мысли. – Москва : Педагогика-пресс, 1993. - С. 149-162.
155. Храпов С. А., Лопатинская Т. Д., Кашкаров А. М. Artificial intelligence as cognitive and sociocultural phenomenon / С. А. Храпов, Т. Д. Лопатинская, А. М. Кашкаров // The Turkish Online Journal of Design, Art and Communication. – 2018. – С. 107
156. Шевченко А. И. К вопросу о создании искусственного интеллекта / А. И. Шевченко // Искусственный интеллект. – 2016. – № 2. – С. 7-15.

157. Юдин Б. Г. Социальные технологии, их производство и потребление / Б. Г. Юдин // Эпистемология и философия науки. – 2012. – Вып. 31. – № 1. – С. 55-64.

158. Ясперс К. Истоки истории и ее цель / К. Ясперс. Смысл и назначение истории. – Москва, 1994. – с. 139.

159. Ястреб Н. А. Индустрия 4.0: киберфизические системы и интернет вещей / Н. А. Ястреб // Человек в технической среде: сборник научных статей / Под ред. доц. Н.А. Ястреб. – Вологда : ВолГУ, 2015. – С. 136 - 141.

160. Abrams J. J. Pragmatism, artificial intelligence, and posthuman bioethics: Shusterman, Rorty, Foucault / J. J. Abrams // Human Studies. – 2020. – Vol. 27(3). – Pp. 241-258.

161. Abdulkareem M., Leiner T., Petersen S.E. Artificial intelligence will transform cardiac imaging – opportunities and challenges / Abdulkareem M., Leiner T., S. E. Petersen // Front Cardiovasc Med. – 2019. – Vol. 6. – 133 p.

162. Angwin J., Larson J. et al. Machine Bias. There is software that is used across the county to predict future criminals. And it is biased against blacks / J. Angwin, J. Larson, S. Mattu, L. Kirchner. – URL: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (дата обращения: 13.08.2022)

163. Ahmad M. A. et al. Fairness, accountability, transparency in AI at scale: Lessons from national programs / M. A. Ahmad, A. Teredesai, C. Eckert // Proceedings of the 2020 Conference on Fairness, Accountability and Transparency. – Barcelona, Spain, 2020. – Vol.27-30. – Pp. 690-699.

164. Araujo T., Helberger N. et al. In AI we trust? Perceptions about automated decision-making by artificial intelligence / T. Araujo, N. Helberger, S. Kruikemeier, C. H. De Vreese // AI & SOCIETY, 2020. – Vol.35. – Pp.611–623.

165. Arendt H. Collective Responsibility / H. Arendt // Responsibility and Judgment. – New York : Schoken books, 2003. – 295 p.

166. Arkin R. C. Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture / Eds. P. Wang, B. Goertzel, S. Franklin // Artificial general intelligence, 2008: Proceedings of the first AGI conference. – Washington DC : ISO Press, 2008. – Pp.51-62.

167. Bahner J. E. et al. Misuse of automated decision aids: Complacency, automation bias and the impact of training experience / J. E. Bahner, A. D. Hüper, D. Manzey // International Journal of Human-Computer Studies. – 2008. – Vol. 66(9). – Pp. 688-699.

168. Balkin J. B. The Path of Robotics Law / J. B. Balkin // California Law Review. – 2015. – Vol. 6. – Pp. 45-60.

169. Bian Y et al. Artificial intelligence–assisted system in postoperative follow-up of orthopedic patients: exploratory quantitative and qualitative study / Y. Bian, Xiang, B. Tong, B. Feng, X. Weng // J. Med. Internet Res. – 2020. – Vol. 22 (5). – 23 p.

170. Bishop S. Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm / S. Bishop // Convergence, 2018. – Vol. 24(1). – Pp. 69-84.

171. Blease C., Kaptchuk T.J et al. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views / C. Blease, T. J. Kaptchuk, M. H. Bernstein, K. D. Mandl, J. D. Halamka, C. M. Desroches // Journal of Med. Internet Research. – 2019. – Vol. 21 (3). – Pp. 1-24.

172. Bloustein E. J. Privacy as an aspect of human dignity: An answer to Dean Prosser / E. J. Bloustein // Philosophical Dimensions of Privacy : An Anthology. – 1984. – Pp. 156-202.

173. Brayne S., Christin A. Technologies of crime prediction: The reception of algorithms in policing and criminal courts / S. Brayne, A. Christin // Social Problems. – 2020. – Vol. 68(3). – Pp. 608-624.

174. Broadhurst R et al. Artificial Intelligence and Crime / R. Broadhurst, P. Brown, D. Maxim, H. Trivedi, J. Wang // Research Paper, Korean Institute of

Criminology and Australian National University Cybercrime Observatory, College of Asia and the Pacific. – Canberra, 2019. – Pp. 1-70.

175. Block N. Troubles with functionalism. Minnesota Studies in the Philosophy of Science / N. Block // Troubles with functionalism, Minnesota Studies in the Philosophy of Science. – 1978. – Pp. 261-325.

176. Brooks R. A. Elephants don´t play chess / R. A. Brooks // Robotics and Autonomous Systems. – 1990. – Vol. 6(1–2). – Pp. 3-15.

177. Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms / J. Burrel // Big Data and Society. – 2016. – Vol. 3(1). – Pp. 1-12.

178. Carolan M. Automated agrifood futures: robotics, labor and the distributive politics of digital agriculture / M. Carolan // J Peasant Stud. – 2020. – Vol. 47. – Pp.184-207.

179. Caruana R, Lou Y. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission / R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, N. Elhadad // Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. – Sydney, NSW, Australia : ACM, 2015. – Vol. 30. – P. 17-21.

180. Case A., Deaton A. Deaths of despair and the future of capitalism / A. Case, A. Deaton // Princeton University Press. – 2020.

181. Castro D., New J. The Promise of Artificial Intelligence / D. Castro, J. New // Center for data innovation. – 2016. – P. 3. – URL: https://www2.datainnovation.org/2016-promise-of-ai.pdf (дата обращения: 23.09. 2022)

182. Chorzempa M. et al. China's social credit system: a mark of progress or a threat to privacy? / M. Chorzempa, P. Triolo, S. Sacks // Policy Briefs PB18–14, Peterson Institute for International Economics. – 2018. – № PB18-14.

183. Clark A., Oswald A. Unhappiness and unemployment / A. Clark, A. Oswald // Economic Journal. – 1994. – Vol. 104(424). – Pp. 648–659.

184. De Garis H. The Artilect War: A bitter controversy concerning whether humanity should build godlike massively intelligent machines / Eds. P. Wang, B. Goertzel, S. Franklin // Artificial general intelligence, 2008: Proceedings of the first AGI conference. – Washington DC : ISO Press, 2008. – Pp. 362- 373.
185. Dyson G. Turing's cathedral: The origins of the Digital Universe / G. Dyson. – New-York : Vintage. 2012. – 464 p.
186. European Parliament Report on artifcial intelligence in education, culture and the audiovisual sector Committee on Culture and Education. 2021. – URL: https://www.europarl.europa.eu/doceo/document/A-9-2021-0127_EN.html (дата обращения: 12.07.2022)
187. Executive Order on Maintaining American Leadership in Artificial Intelligence. – URL: https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/ (дата обращения: 20.06.2022)
188. Floridi L. The end of an era: from self-regulation to hard law for the digital industry / L. Floridi // Philosophy & Technology. – 2021. – Vol. 34, № 4. – Pp. 612-622.
189. Floridi L., Cowls J. A Unified Framework of Five Principles for AI in Society / L. Floridi, J. Cowls // Harvard Data Science Review. – 2022. – Vol. 1 (1). – Pp. 535-545.
190. Floridi L., et al. AI4People - An ethical framework for a good AI society / L. Floridi, J. Cowls, M. Beltrametti et al. // Minds and Machines. – 2018. – Vol. 28(4). – Pp. 689-707.
191. Fodor J. A. In critical condition: Polemical essays on cognitive science and the philosophy of mind / J. A. Fodor. – Cambridge, MA : MIT Press. – 1998. – P. 17.
192. Frankfurt H. G. Freedom of the Will and the Concept of a Person / H. G. Frankfurt // Journal of Philosophy. – 1971. – Vol. 68. – Pp. 5-20.
193. Friedman M. Autonomy, Gender, Politics / M. Friedman. – New York : Oxford University Press, 2003. – 263 p.

194. Gasser U, Lenca M et al. Digital tools against COVID-19: Taxonomy, ethical challenges, and navigation aid / U. Gasser, M. Lenca, J. Scheibner, J. Sleigh, E. Vayena // Healthpolicy. – 2020. – Vol. 2. – Pp. 425-434.

195. Goertzel B. Artificial General Intelligence: Concept, State of the Art, and Future Prospects / B. Goertzel // Journal of Artificial General Intelligence. – 2014. – Vol. 5(1). – Pp. 1- 46.

196. Goertzel B. Should humanity build a global AI nanny to delay the singularity until it's better understood? // Journal of consciousness studies. – 2012. – Vol. 19. – Pp. 96-111.

197. Grinbaum A. et al. Ethics in Robotics Research/ A. Grinbaum, R. Chatila, L. Devillers, J. G. Ganascia, // IEEE Robotics and Automation Magazine. – 2017. – № 24. – Pp. 139-145.

198. Hackman J. R., Oldham G. R. Motivation through the design of work: Test of a theory / J. R. Hackman, G. R. Oldham // Organizational Behavior and Human Performance. – 1976. – Vol. 16(2). – P. 258

199. Haefner K. Mensch und Computer im Jahre 2000 / K. Haefner. – Basel - Boston - Stuttgart : Birkhauser, 1994. – 402 p.

200. Hagendorff T. The Ethics of AI Ethics: An Evaluation of Guidelines / T. Hagendorff // Minds & Machines. – 2020. – Vol. 30. – Pp. 99-120.

201. Haridasani G.A. Are Algorithms Sexist? / G. A. Haridasari. – URL: https://www.nytimes.com/2019/11/15/us/apple-card-goldman-sachs.html (дата обращения: 21.05. 2022)

202. Hassabis D., Revell T. With AI, you might unlock some of the secrets about how life works / D. Hassabis, T. Revell. – New Scientist. – 2021. – Vol. 249(3315). – Pp. 44-49.

203. Hashimoto D.A et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy / D. A. Hasimoto, G. Rosman, E. R. Witkowski // Ann Surg. – 2019. – Vol. 270 (3). – P.21.

204. Holstein K., Wortman J. Improving fairness in machine learning systems: What do industry practitioners need? / K. Holstein, V. Wortman, J. Daumé, M. Dudík, H. Wallach // Proceedings of the 2019 CHI conference on human factors in computing systems. – 2019. – Pp.1-16.

205. Goertzel B. Should humanity build a global AI nanny to delay the singularity until it's better understood? // Journal of consciousness studies. –2012. – Vol. 19. – Pp. 96-111.

206. Grogger J., Ivandic R., Kirchmaier T. (2020). Comparing conventional and machine-learning approaches to risk assessment in domestic abuse cases / J. Grogger, R. Ivandiv, T. Kirchmaier // Journal of Empirical Legal Studies. – 2020. – Vol. 18(1). – Pp. 90-130.

207. Jobin A., Ienca M., Vayena E. Artificial Intelligence: The Global Landscape of Ethics Guidelines / A. Jobin, M. Lenca, E. Vayena // Nature Machine Intelligence. – 2019. – Vol.1. – Pp. 389-399.

208. Kellogg K. C., Valentine M. A. Algorithms at work: The new contested terrain of control / K. Kellogg, M. A. Valentine, A. Christin // Academy of Management Annals. – 2020. – Vol. 14(1). – Pp. 366-410.

209. Kim T. W., Scheller-Wolf A. Technological unemployment, meaning in life, purpose of business, and the future of stakeholders/ T. Kim, A. Scheller-Wolf // Journal of Business Ethics. – 2019. – Vol. 160(2). – Pp. 319-337.

210. Langlois R. N. Cognitive comparative advantage and the organization of work: Lessons from Herbert Simon's vision of the future / R. N. Langlois // Journal of Economic Psychology. – 2020. – Vol. 24(2). – P. 174

211. Legg S., Hutter M. A collection of definitions of intelligence / B. Goertzel, P. Wang (Eds.) // Advances in artificial general intelligence: concept, architectures and algorithms. – Amsterdam : IOS Press, 2007. – Vol. 157. – Pp. 17-24.

212. Rawls J. A Theory of Justice / J. Rawls. – Cambridge, Massachusetts : Belknap Press, 1971. – 562 p.

213. Ryan M., Antoniou J. Research and Practice of AI Ethics: A Case Study Approach Juxtaposing Academic Discourse with Organisational Reality / J. Antoniou, M. Ryan, L. Brooks, T. Jiya, K. Macnish, B. Stahl // Science and Engineering Ethics. – 2021. – Vol. 27**(**2). – P. 16.

214. Shortliffe E. H., Sepúlveda M. J. Clinical decision support in the era of artificial intelligence / E. H. Shortliffe, M. J. Sepulveda // JAMA. –2018. – Vol. 320 (21). – Pp. 2199-2200.

215. Walsh T., Levy N et al. The effective and ethical development of artificial intelligence / T. Walsh, N. Levy, G. Bell, A. Elliott, J. Maclaurin et al. // ACOLA. – 2019. – URL: https://acola.org/wp-content/uploads/2019/07/hs4_artificial-intelligence-report.pdf (дата обращения: 30.01. 2022).

216. Zerilli J., Knott A et al. Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard? / J. Zerilli, A. Knott, J. Maclaurin, C. Gavaghan // Philosophy and Technology, 2019. – Vol. 32 (4). – Pp. 661-683.

IV. Диссертации и авторефераты диссертаций

217. Заладина М. В. Социально-философский анализ духовного отчуждения: автореферат диссертации на соискание ученой степени кандидата философских наук / М. В. Заладина. – Нижний Новгород, 2020. –25 с.

218. Маслов В. М. Высокие технологии и феномен постчеловеческого в современном обществе: автореферат диссертации на соискание ученой степени кандидата философских наук / В. М. Маслов. – Нижний Новгород, 2014. – 38 с.

219. Морхат П. М. Правосубъектность искусственного интеллекта в сфере права интеллектуальной собственности: гражданско-правовые проблемы: диссертация на соискание ученой степени доктора юридических наук / П. М. Морхат. – Москва, 2019. – 420 с. – С. 30–31.

220. Цуркан Е. Г. Социокультурная динамика и интернет-технологии: социально-философский анализ: автореферат диссертации на соискание

ученой степени кандидата философских наук / Е. Г. Цуркан. – Москва, 2021. – 33 с.

221. Чепьюк О. Р. Экономическая бессубъектность как фактор дегуманизации социальных отношений: диссертация на соискание ученой степени доктора философских наук / О. Р. Чепьюк. – Нижний Новгород, 2020. – 371 с.

222. Щитова А. А. Правовое регулирование информационных отношений по использованию систем искусственного интеллекта: диссертация на соискание ученой степени кандидата юридических наук /А. А. Щитова. – Москва, 2021. – 225 с.

V. Интернет-источники:

223. Греф: Сбербанк сможет принимать 80 % решений искусственным интеллектом. – URL: https://ria.ru/economy/20160908/1476449735.html (дата обращения: 19.07.2022).

224. Европарламентарий могут предоставить роботам юридический статус. – URL: https://ria.ru/world/20170114/1485715425.html?inj=1 (дата обращения: 19.08.2022).

225. Ефимова Е. Новое слово в живописи: искусственный интеллект «пишет» картины в уникальном стиле / Е. Ефимова. – URL: http://www.vesti.ru/doc.html?id=2905726 (дата обращения: 16.09.2022).

226. Искусственный интеллект как ключевой фактор цифровизации глобальной экономики. – URL: https://www.iksmedia.ru/news/5385191-Iskusstvennyj-intellekt-II-Artifici.html (дата обращения: 29.09.2021).

227. Конев С. И. Этико-правовые проблемы регулирования искусственного интеллекта и робототехники в отечественном и зарубежном праве / С. И. Конев. – URL: https://cyberleninka.ru/article/n/etiko-pravovye-problemy-regulirovaniya-iskusstvennogo-intellekta-i-robototehniki-v-otechestvennom-i-zarubezhnom-prave (дата обращения: 19.09. 2022).

228. Ксенофонтова А. Бесправный механизм: почему ученые выступили против присвоения роботам статуса «электронной личности» / А. Ксено-

фонтова. – URL: https://russian.rt.com/science/article/504118-roboty-evroparlament-yuridicheskoye-lico (дата обращения: 17.05.2022).
229. Об испытаниях высокоавтоматизированных автотранспортных средств. – URL: https://lovdata.no/dokument/NL/lov/20171215112?q=Lov%20om%20utpr%C3%B8ving%20av%20selvkj%C3%B8rende (дата обращения: 12.09.2022).
230. Серль Дж. Р. Разум мозга - компьютерная программа? / Дж. Р. Серль. – URL: https://psychosearch.ru/teoriya/psikhika/338-searle-john-razum-mozga-kompyuternaya-programma (дата обращения: 12.03.2022).
231. Специальный комитет по искусственному интеллекту (CAHAI). – URL: https://sk.ru/media/documents/CAHAI_AI_Research.pdf (дата обращения: 19. 11. 2022).
232. Чат-бот от Microsoft за сутки научился ругаться и стал расистом. – URL: https://www.interfax.ru/world/500152 (дата обращения: 16.09.2022).
233. Boeing во время аудита обнаружила ошибки в обновленном ПО для самолетов модели 737 MAX. – URL: https://habr.com/ru/news/t/484434/ (дата обращения: 12.09.2022).
234. Work Fusion объединяет интеллектуальную автоматизацию с облачным хранилищем. – URL: https://robroy.ru/intellektualnuyu-avtomatizacziyu-s-oblachnyim-xranilishhem.html (дата обращения: 12.02.2023).
235. ACM FAccT Conference. – URL: https://facctconference.org// (дата обращения: 04.01.2022).
236. Asher-Schapiro A. Amazon AI van cameras spark surveillance concerns / A. Asher-Schapiro // News.Trust.Org. – 2021. – URL: https://news.trust.org/item/20210205132207-c0mz7/ (дата обращения: 12.09.2022).
237. Bossmann Dzh. Top 9 Ethical Issues in Artificial Intelligence. – URL: https://hr-portal.ru/article/9-glavnyh-eticheskih-problem-iskusstvennogo-intellekta (дата обращения: 22.03.2022).

238. Castro D., New J. The Promise of Artificial Intelligence. Center for data innovation. – 2016. – URL: https://www2.datainnovation.org/2016-promise-of-ai.pdf (дата обращения: 23.09. 2022).
239. Dawson D et al. Artificial Intelligence - Australia's Ethics Framework / D. Dawson / D. Dawson, E. Schleiger, J. Horton, J. McLaughlin, C. Robinson et al. // Data 61 CSIRO. – Australia, 2019. – URL: https://www.csiro.au/-/media/D61/Reports/Artificial-Intelligence-ethics-framework.pdf (21.09.2022).
240. Report of COMEST on robotics ethics. – 2017. – URL: https://unesdoc.unesco.org/ark:/48223/pf0000253952 (дата обращения: 31.11.2022).
241. Dignum V. Responsible Artificial Intelligence – from Principles to Practice Virginia Dignum. – URL: file:///C:/Users/oem/Downloads/Responsible_Artificial_Intelligence_--_from_Princi.pdf (дата обращения: 12.09.2021).
242. Ellen M Harper , Sara Parkerson, Powering Big Data for Nursing Through Partnership. – URL: https://pubmed.ncbi.nlm.nih.gov/26340243/ (дата обращения: 14.03.2022).
243. Equifax Launches NeuroDecision Technology. – 2018. – URL: https://investor.equifax.com/news-events/press-releases/detail/203/equifax-launches-neurodecision-technology (дата обращения: 12.09.2022).
244. Government of Canada, "Directive on Automated Decision-Making", 2019. – URL: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 (дата обращения:12.09.2022).
245. GPT 4 – OPEN AI. – URL: https://openai.com/product/gpt-4 (дата обращения:02. 04. 2023)
246. Grossman L. 2045: The year man becomes immortal / L. Grossman // Time. – 2011. – URL: http://content.time.com/time/magazine/article/0,9171,2048299,00.html (дата обращения: 21.03.2022).

247. Iamus, a music-making computer, could be the next Mozart. – URL: https://www.vice.com/en/article/pgg8yy/iamus-a-music-making-computer-could-be-the-next-mozart (дата обращения: 21.01.2023).
248. McCarthy J. What is Artificial Intelligence? // Stanford University. - 2007. – URL: http://www-formal.stanford.edu/jmc/whatisai. (дата обращения: 21.09.2022).
249. Pardes A. AI can run your work meetings now. Wired. – URL: https://www.wired.com/story/ai-can-run-work-meetings-now-headroom-clockwise (14.09.2022).
250. Roger C. Regulatory alternatives for AI. Science Direct (2019). – URL: http://www.rogerclarke.com/EC/AIR-Final.pdf (дата обращения: 12.12. 2022)
251. Search Enterprise AI. Artificial intelligence. – URL: https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence. (дата обращения: 15.05.2022).
252. Shakey the Robot – SRI International. – URL: https://www.sri.com/hoi/shakey-the-robot/ (дата обращения: 12.05. 2022)
253. Singularity Institute for Artificial Intelligence, Seed AI. General Intelligence and Seed AI. — URL: http://singinst.org/ourresearch/publications/GISAI/paradigms/seedAI.html#glossary_crystalline (дата обращения: 12.09.2021).

ПРИЛОЖЕНИЕ

№	Название документа	Организация	Тип
1	California Consumer Privacy Act (CCPA), «Закон Калифорнии о конфиденциальности потребителей», 2018 г.	Законодательное собрание штата Калифорния, США	Закон
2	EU-U.S. and Swiss-U.S. Privacy Shield, («Закон о защите конфиденциальности»), 2016	Министерство торговли США, Европейская комиссия и администрация Швейцарии	Закон
3	Personal information protection and electronic documents act, («Закон о защите личной информации и электронных документов»), 2000	Правительство Канады	Закон
4	Data Protection Act, («Закон о защите данных»), 2018	Палата лордов, Великобритания	Закон
5	General data protection regulation, («Общее положение о защите данных»), 2016	ЕС	Закон
6	Fair Credit Reporting Act, («Закон о достоверной кредитной отчетности»), 2018	Федеральная торговая комиссия США, Бюро финансовой защиты потребителей, США	Закон
7	Personal information protection law, («Закон о защите личной информации»), 2021	ВСНП, КНР	Закон
8	Cyber security law, («Закон о кибер-защите»), 2016	ВСНП, КНР	Закон
9	Protection from online falsehoods and manipulation act, («Закон о защите от лжи и манипуляций в Интернете »)	Парламент, Сингапур	Закон
10	Data protection law, («Закон о защите данных»), 2021	Федеральный национальный совет, ОАЭ	Закон
11	Bundesdatenschutzgesetz, («Федеральный закон о защите данных»), 2009	Парламент, Германия	Закон
12	Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р «Об ут-	Правительство, РФ	Концепция

	верждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.», 2020		
13	Федеральный закон от 24.04.2020 г. № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных"», 2020	Парламент, РФ	Закон
14	«Об экспериментальных правовых режимах в сфере цифровых инноваций в Российской Федерации» (Законопроект № 922869-7), 2020	Правительство РФ	Законопроект
15	Loi pour une Republique numerique. No 2016-1321.,(«Закон Франции о цифровой республике»), 2016	Национальное собрание Франции	Закон
16	Intelligent robots development and distribution promotion act, («Закон о развитии и распространении интеллектуальных роботов»),2016	Парламент Южной Кореи	Закон
17	Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года») // Собрание законодательства РФ. 2019. № 41. Ст. 5700; Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 года», 2020	Правительство РФ	Национальная стратегия РФ

18	State Council, Notice of Issuing New Generation Artificial Intelligence Development Plan, («План развития искусственного интеллекта Нового поколения ») 2017	Государственный совет КНР	Национальная стратегия КНР
19	Bundesregierung, Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz, («Основные положения Федеральной правительственной Стратегии по искусственному интеллекту»), 2018	Федеральное правительство Германии	Национальная стратегия Германии
20	For a meaningful artificial intelligence towards a french and european strategy, («На пути к осмысленному искусственному интеллекту»), 2018	Парламент Франции	Национальная стратегия Франции
21	National Strategy for Artificial Intelligence, («Национальная стратегия по искусственному интеллекту») , 2020	Правительство Дании, Министерство финансов и Министерство промышленности, бизнеса и финансов.	Национальная стратегия Дании
22	Pan-Canadian Artificial Intelligence Strategy, («Панканадская стратегия в области искусственного интеллекта»), 2018	CIFAR (Canadian Institute for Advanced Research) - канадская государственная научно-исследовательская организация	Национальная стратегия Канады
23	AI Strategy 2019 AI for Everyone: People, Industries, Regions and Governments, («Стратегия в области искусственного интеллекта, предназначенная для всех: людям, отраслям, регионам и правительству») 2019	Совет по продвижению комплексной инновационной стратегии	Национальная стратегия Японии
24	National Approach to artificial intelligence, («Национальный подход к искусственному интеллекту»), 2018	Государственные учреждения Швеции	Национальная стратегия Швеции
25	Eesti 2035, («Эстония 2035»), 2021	Государственная канцелярия и	Национальная стратегия Эс-

		Министерство финансов	тонии
26	Australia's AI Ethics Principles, 2019	Правительство Австралии, Департамент промышленности, науки и ресурсов	Руководство
27	Artificial intelligence principles and ethics, 2020	Правительство ОАЭ	Руководство
28	Directive on Automated Decision-Making, («Директива по автоматизированному принятию решений») 2019.	Правительство Канады	Директива
29	Asilomar AI Principles, ("Азиломарские принципы искусственного интеллекта"), 2017	Института будущего жизни – добровольная ассоциация ученых, предпринимателей, экспертов в Бостоне.	Результаты международной научной конференции
30	Montreal Declaration: Responsible AI, ("Монреальская декларация по ответственному ИИ") 2017	Монреальский университет	Декларация
31	Draft Recommendation on the Ethics of Artificial Intelligence, («Первоначальный вариант проекта рекомендации об этических аспектах искусственного интеллекта»), 2020	Специальная группа экспертов по подготовке проекта рекомендации об этических аспектах искусственного интеллекта от ЮНЭСКО, Франция	Рекомендации
32	IEEE Ethically Aligned Design, («Этически согласованный дизайн»), 2019	IEEE - Международная некоммерческая ассоциация специалистов в области техники	Трактат
33	Preliminary study on a possible standard-setting instrument on the	ЮНЭСКО	Программный документ и до-

	ethics of artificial intelligence, Предварительное исследование возможности подготовки нормативного акта по вопросам этики применения искусственного интеллекта, 2019		кумент совещаний
34	AI in the UK: ready, willing and able? («Пять всеобъемлющих принципов для кода ИИ»), 2018	Комитет по искусственному интеллекту Палаты лордов, Великобритания	Доклад
35	Terms of Reference, («Круг полномочий»), 2020	Global Partnership on Artificial Intelligence, Франция и Канада	Рекомендации
36	Civil Law Rules on Robotics (2015/2103(INL)), («Резолюция 2015/2103 (INL) от Парламента ЕС»), 2017	Европейский парламент, Комиссии по гражданско-правовым нормам в отношении робототехники, ЕС	Резолюция
37	Strategy for automated and connected driving («Стратегия автоматизированного и подключенного к сети вождения»), 2019	Федеральное министерство транспорта и цифровой инфраструктуры, Германия	Стратегия
38	Automated vehicle trial guidelines, («Руководство по испытаниям автоматизированных транспортных средств)», 2017	Национальная комиссия по автомобильному транспорту, Австралия	Руководство
39	Résolution visant à interdire l'utilisation, par la Défense belge, de robots tueurs et de drones armés, («Резолюция о запрете применения автономных смертельных систем вооружения, Бельгия)», 2018	Бельгийская палата народных представителей, Бельгия	Резолюция
40	Auromated and connected driving: report («Рекомендация по беспилот-	Федеральное министерство	Доклад, рекомендация

	ным автомобилям от Комиссии по этике при Министерстве транспорта и цифровой инфраструктуры Германии»), 2017	цифровых технологий и транспорта является министерством на уровне кабинета министров, Этическая комиссия, Германия	
41	Report of COMEST on robotics ethics, («Доклад об этике робототехники от Юнеско»),2017	ЮНЭСКО	Доклад, рекомендация
42	中国《人工智能标准化白皮书2018》发布完整版（附下载, («Белая книга по стандартизации ИИ »), Китай, 2018	Национальная общая группа по стандартизации искусственного интеллекта и Консультативная группа экспертов, КНР	Рекомендации
43	新一代人工智能伦理规范》发布, («Этический кодекс искусственного интеллекта нового поколения»), 2021	Национальный профессиональный комитет по управлению искусственным интеллектом нового поколения, Министерство науки и технологий, КНР	Кодекс
44	Модельная конвенция робототехники и искусственного интеллекта, 2018	Исследовательский центр проблем регулирования робототехники и искусственного интеллекта, РФ	Конценция
45	Developing Standards for Artificial Intelligence: Hearing Australia's Voice, («Разработка стандартов для искусственного интеллекта: Слышать голос Австралии»), 2019	Standards Australia, неправительственная, некоммерческая организация по стандартизации	Доклад
46	A Plan for Federal Engagement in	Национальный	Стандарт

	Developing Technical Standards and Related Tools, Проект плана федерального участия в разработке технических стандартов ИИ и связанных с ними инструментов, 2019	институт стандартов и технологий (NIST) – лаборатория физических наук и нерегулирующее агентство Министерства торговли США	
47	European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, («Европейская этическая хартия Совета Европы по использованию ИИ в судебных системах)», 2018	Европейская комиссия по эффективности правосудия (CEPEJ).	Хартия
48	Guidelines on artificial intelligence and data protection, («Руководство по защите данных при использовании ИИ»), Strasbourg, 2019	Консультативный комитет конвенции о защите лиц в отношении автоматической обработки персональных данных, Главное управление по правам человека и верховенству закона, Франция	Конвенция
49	Декларация Комитета Министров о манипулятивных возможностях алгоритмов, (“Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes”),2019	Группа экспертов Совета Европы	Декларация
50	Unboxing Artificial Intelligence: 10 steps to protect Human Rights,(«10 шагов для защиты прав человека при использовании ИИ»), 2019	Комиссар СЕ по правам человека	Рекомендации
51	Recommendation of the Council on Artificial Intelligence, («Принципы ИИ и рекомендации по национальной политике Экспертной группы по искусственному интеллекту»), 2019	Совет по искусственному интеллекту, ОЭСР	Ркомендации

52	Addressing the impacts of Algorithms on Human Rights («Устранение воздействия алгоритмов на права человека»), 2019	Комитет экспертов по правам человека по аспектам автоматизированной обработки данных и различных форм искусственного интеллекта, Комитет министров Совета Европы	Проект рекомендаций
53	Artificial Intelligence. Australia's Ethics Framework: A Discussion Paper Department of Industry Innovation and Science, («Этические рамки искусственного интеллекта в Австралии»), 2019	Организация научных и промышленных исследований Common Well (CSIRO), под руководством Министерства инноваций и науки правительства Австралии	Дискуссионный документ
54	Work in the Age of Artificial Intelligence. Four Perspectives on the Economy, Employment, Skills and Ethics, («Работа в эпоху искусственного интеллекта: четыре взгляда на экономику, занятость, навыки и этику»), 2018	Министерство экономики и занятости, Финляндия	Доклад
55	Tieto's AI Ethics Guidelines, («Этические принципы в области искусственного интеллекта»), 2018	Tieto - скандинавская компания по предоставлению ИТ-услуг для промышленности и сферы обслуживания, Финляндия	Рекомендации
56	How Can Humans Keep the Upper Hand? Report on the Ethical Matters	Национальная комиссия по ин-	Отчет

	Raised by AI Algorithms («Как люди могут одержать верх? Отчет об этических вопросах, поднятых алгоритмами ИИ»), Франция, 2017	форматике и свободам, Французское управление по защите данных (CNIL), Франция	
57	The Ethics of AI Ethics – An Evaluation of Guidelines, («Этика ИИ - оценка руководящих принципов »), 2020	Т. Hagendorff, Университет Тюбинген, Германия	Руководство
58	Understanding artificial intelligence ethics and safety, («Понимание этики и безопасности искусственного интеллекта»), 2019	Д. Лесли, Институт Алана Тьюринга, США	Руководство
59	The globall and scape of AI ethics guidelines, («Руководство по этике ИИ в мире »), 2019	А. Джобин и др., Высшая техническая школа Цюрих, Швейцария	Руководство
60	AI HLEG. “Ethics Guidelines for Trustworthy AI, («Руководство по этике для надежного ИИ»), 2019	Европейская Комиссия	Руководство
61	From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices («От «чего» к вопросу «как»: первоначальный обзор общедоступных инструментов, методов и исследований по этике ИИ для воплощения принципов в практику»), 2020	J. Morleyetal, Оксфордский институт Интернета, Великобритания	Руководство
62	From Principles to Practice: A ninter disciplinary framework to operationalize AI ethics («От принципов к практике: междисциплинарная основа для практической реализации этики ИИ»), 2020	С. Халленслебен, AIEI Group, Германия	Руководство
63	Building Ethics into AI: Lessons Learned from Pioneers in the Trenches («Внедрение этических норм в ИИ: уроки, полученные от первопроходцев»), 2019 г	К. Бакстер, Salesforce Research, США	Руководство
64	Technical and Organizational Best Practices, («Технические и организа-	FBPML (The Foundation for	Руководство

	ционные передовые практики»), 2021	Best Practices in Machine Learning), Нидерланды	
65	Advancing AI ethics beyond compliance From principles to practice («Продвижение этики ИИ за пределы соблюдения требований. От принципов к практике»), 2020	В. Геринг, IBM, США	Руководство
66	Guidelines for AI Procurement, («Руководство по закупкам с использованием ИИ»), 2019	World Economic Forum, Швейцария	Руководство
67	Руководство по ИИ и защите данных (“Guidance on AI an data protection”), 2020	Information Commissioner Office, Управление комиссара по информации, Великобритания	Руководство
68	Guidance on the AI auditing framework («Руководство по системе аудита ИИ»), 2020	Information Commissioner Office, Управление комиссара по информации, Великобритания	Руководство
69	Building Ethics into Privacy Frameworks for Big Data and AI, («Встраивание этики в рамки конфиденциальности для больших данных и искусственного интеллекта»), 2018	IAPP, США	Руководство
70	Using Big Data to Improve the Quality of Care and Outcomes, («Руководящие принципы использования больших данных в сестринском деле — использование больших данных для улучшения качества ухода и результатов»), 2018	HIMSS (Общество медицинских информационных и управленческих систем)	Руководство
71	Responsible use of artificial intelligence (AI), («Ответственное использование искусственного интеллекта»), 2019	Правительство Канады	Руководство

72	Artificial Intelligence - Australia's Ethics Framework, («Искусственный интеллект — Австралийская система этических норм»), 2019	CSIRO, Организация научных и промышленных исследований Содружества, Австралия	Дискуссионное исследование
73	An Accountability Framework for Federal Agencies and Other Entities, («Система подотчетности федеральных агентств и других организаций»), 2021	Счетная палата, США	Рекомендации
74	Universal Guidelines for Artificial Intelligence, («Универсальные принципы для искусственного интеллекта»), 2018	The Public Voice, комиссия по интересам сообщества, созданная при Информационном центре электронной конфиденциальности (EPIC), Бельгия	Рекомендации
75	African Observatory on Responsible Artificial Intelligence, («Африканская обсерватория ответственного искусственного интеллекта») 2022	Research ICT Africa, африканский аналитический центр, работающий над устранением стратегического пробела в развитии устойчивого информационного общества и цифровой экономики, ЮАР	Рекомендации
76	Top 10 principles for ethical artificial intelligence, («10 принципов для этического искусственного интеллекта»)	UNI Global Union, глобальной федерацией профсоюзов в сфере навыков и услуг, Швейцария	Рекомендации
77	The Toronto Declaration. Protecting the right to equality in machine learn-	Amnesty International	Декларация

	ing, («Декларация Торонто. Защита права на равенство в машинном обучении,»), 2018	(глобальное движение, выступающее за защиту прав человека) и Access Now (институт Нью-Йоркского университета, занимающийся исследованием социальных последствий ИИ), Канада	
78	Artificial Intelligence and its role in society, («Искусственный интеллект и его роль в обществе»), 2020	Microsoft, США	Рекомендации
79	SAP's Guiding Principles for Artificial Intelligence, («Руководящие принципы для искусственного интеллекта»), 2018	SAP, Германия	Руководство
80	Ethical Principles for Artificial Intelligence in Medicine, («Этические принципы искусственного интеллекта в медицине»), 2019	The Royal Australian and New Zealand College of Radiologists, Австралия	Рекомендации
81	Automated and Connected Driving, («Автоматизированное и подключенное к сети вождение»), 2017	Министерство транспорта и цифровой инфраструктуры, Германия	Доклад
82	Artificial Intelligence at Google: Our Principles, («Искусственный интеллект у Goggle: Наши принципы»), 2018	GOOGLE	Рекомендации
83	Beijing AI Principles, («Пекинские принципы искусственного интеллекта»), 2019	Пекинская академия искусственного интеллекта (BAAI), КНР	Рекомендации

84	B.BS8611, «Robots and robotic devices, guide to the ethical design and application of robots and robotic systems», 2016	British Standards Institute, Великобритания	Стандарт
85	Ethical design and use of automated decision systems (CAN/CIOSC 101:2019), 2019-настоящее время	CIO Strategy Council, Канада	Стандарт
86	ISO/IEC JTC 1/SC 42 Standards for Artificial Intelligence, 2017-настоящее время («Information technology — artificial intelligence — overview of trustworthiness in artificial intelligence»), 2020	ISO	Стандарт
87	Artificial Intelligence and User Trust (NISTIR 8332), 2021 (B. Stanton, T. Jensen et al., “Trust and artificial intelligence”), 2021	NIST, США	Стандарт
88	Artificial Intelligence and Public Standards, 2020	Committee on Standards in Public Life UK, Великобритания	Стандарт
89	The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019	IEEE SA	Стандарт
90	Code of Ethics and Professional Conduct, (Acm code of ethics and professional conduct),2018	ACM, США	Стандарт