УДК 338.1

ВЫЗОВЫ И РИСКИ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Фадеева Д. А.
студентка кафедры экономики и финансов
АНОВО Московский международный университет,
Москва, Россия,
Филина Ф. В.
к.э.н., доцент
АНОВО Московский международный университет,
Москва, Россия

Аннотация. Дано разграничение понятий «вызовы» и «риски» использования искусственного интеллекта (ИИ). Приведены примеры негативных эффектов применения ИИ в сфере кибербезопасности, включая киберпреступления, фальсификацию данных и технологическое стимулирование угроз. Описаны риски, обусловленные человеческим фактором: алгоритмическая предвзятость, вклад ИИ в поддержание дискриминационных практик, физическая угроза. Проанализированы последствия внедрения ИИ в транспортную сферу и на рынке труда. Сделан вывод о необходимости нормативного и этического регулирования для безопасного использования ИИ.

Ключевые слова: искусственный интеллект (ИИ), вызовы и риски использования ИИ, кибербезопасность, киберпреступления, технологическое стимулирование угроз, фальсификация данных, алгоритмическая предвзятость, дискриминационные практики.

CHALLENGES AND RISKS OF USING ARTIFICIAL INTELLIGENCE

Fadeeva D.A.
Student of the Department of Economics and Finance
ANOVO Moscow International University,
Moscow, Russia
Filina F. V.
Candidate of Economics, Associate Professor
ANOVO Moscow International University,
Moscow, Russia

Annotation. The distinction between the concepts of "challenges" and "risks" associated with the use of artificial intelligence (AI) is presented. Examples of negative effects of AI implementation in the field of cybersecurity are given, including cybercrime, data falsification, and technological stimulation of threats. Risks related to the human factor—such as data breaches, algorithmic bias, and discriminatory practices—are described. The consequences of AI deployment in the transport sector and labor market are analyzed. A conclusion is drawn on the necessity of regulatory and ethical frameworks to ensure the safe and responsible use of AI.

Keywords: artificial intelligence (AI), challenges and risks of using AI, cybersecurity, cybercrime, technological stimulation of threats, data falsification, algorithmic bias, discriminatory practices.

В условиях стремительного развития и всестороннего внедрения инновационных технологий, базирующихся на искусственном интеллекте (ИИ), всё чаще поднимается вопрос не только о преимуществах их применения, но и о связанных с их использованием потенциальными угрозами. Весь спектр негативных эффектов, сопровождающих глобальную цифровизацию, можно разделить на две условные группы в зависимости от природы их возникновения:

Во-первых, это *вызовы*, то есть угрозы, связанные с делинквентным поведением отдельных лиц. Они выражаются в их осознанных действиях, нацеленных на причинение вреда конкретному индивиду, социальной группе либо обществу в целом. К ним можно отнести киберпреступления и их технологическое стимулирование, использование ИИ для фальсификации данных и т. п.

Во-вторых, это *риски*, под которыми подразумеваются угрозы, вызванные, прежде всего, «человеческим» фактором: ненамеренными упущениями, невнимательностью, ошибками, обусловленными недостаточной осведомленностью или слабой вовлеченностью в процессы. Наиболее значимыми, на наш взгляд, рисками применения ИИ являются утечка данных, алгоритмическая предвзятость, поддержание дискриминационных практик с помощью ИИ, физическая угроза и т. д.

Рассмотрим в начале первую группу негативных эффектов применения ИИ, а *именно вызовы*, возникающие при использовании ИИ.

Наиболее значимой областью применения технологий ИИ является выполнение работу по обработке и анализу информации, что способствует интенсификации производственных процессов. Помимо аналитического значения, ИИ также приобрел ценность одновременно и как «хранилище» данных, и как их «защитник». С одной стороны, его участие в хранении и защите информации способствует оптимизации процессов оценивания общего состояния используемых моделей хранения данных. С другой, эти преимущества приводят к тому, что компании начинают доверять хранилищам данных на основе ИИ всю информацию, представляющую для них важность и коммерческую тайну. Но одновременно с этим, такие базы данных все чаще становиться жертвами киберинцидентов.

Так, компания RED Security, осуществляющая свою деятельность в сфере кибербезопасности, провела анализ кибератак, которым российские компании

подвергались в 2024 году. В результате было установлено, что за год совокупное число инцидентов информационной безопасности в компаниях выросло в 2,5 раза по сравнению с 2023 годом и почти достигло 130 тысяч [1]. Согласно статистическим данным за 2022–2023 годы, в России относительная доля утечек информации ограниченного доступа, причиной которых послужили кибератаки, составила более 81% в оба выбранных периода. Рост количества киберпреступлений от года к году составляет 20–25% [4].

Весной 2022 повысилась активность киберпреступников, занимающихся DDoS-атаками и целевыми APT-атаками (Advanced Persistent Threat (APT), что подчеркивает наличие «развитой устойчивой угрозы», представленной таргетированными и целевыми кибератаками против российских ресурсов и значимых объектов.

Бурное развитие технологий на основе ИИ приводит к так называемому технологическому стимулированию угроз. Перспектива автоматизации алгоритмов работы ИИ наталкивает киберпреступников на мысль о том, что можно автоматизировать не только анализ данных для обычных бизнес-процессов, но и создание, а затем и массовое распространение вредоносных программ. Кроме этого, способность имитировать поведение человека, которой обладают генеративные модели, лежащие фундаментом для многих систем искусственного интеллекта - ChatGPT, YandexGPT и др. активизирует фишинг и т. д. В итоге, кибератаки с использованием ИИ управляются програмно, моментально адаптируются к изменениям, подстраиваясь под динамично развивающиеся условия, из-за чего им становится все сложнее противостоять. Официальная статистика МВД России за январь — ноябрь 2023 года показала, что число зарегистрированных преступлений, связанных с задействованием информационных технологий, по сравнению с 2022 годом возросло на 30,8 %. Кроме того, на приблизительно 70% от всей незаконной деятельности, совершенной с применением информационных технологий, приходятся именно хищения [7]. 40% преступлений были совершены с использованием

информационно-телекоммуникационных технологий. Таких деяний зарегистрировано примерно на 13 % больше, чем в 2023 году [1].

Актуальной также является угроза, вызванная участием механизмов искусственного интеллекта в намеренной фальсификации данных и их дальнейшем влиянии на общество или репутацию конкретного лица. Злоумышленники широко используют эти возможности для манипуляций общественным мнением, дестабилизации общественного порядка и т. д.

Что касается второй группы угроз применения продуктов ИИ - рисков, то среди них отметим следующие.

Наравне с перечисленными проблемами стоит еще и вопрос о неграмотной эксплуатации искусственного интеллекта, так как его технологии ставят под угрозу сохранность конфиденциальных данных не только косвенно, через стимулирование и даже развитие киберпреступлений, но и напрямую, в связи с особенностями функционирования. Механизм ИИ базируется на «машинном обучении», которое «позволяет программисту не писать программы, учитывающие все варианты развития событий, а заложить в программу возможность самостоятельного нахождения решений с помощью использования имеющихся … данных» [8], а соответственно, влекущим за собой роботизацию проходящих под его контролем процессов. В качестве источников данных, на основе которых проводится самообучение искусственного интеллекта, выступают различные научные материалы, информация, загруженная в нейросеть пользователями самостоятельно. Проблема заключается в отсутствии законодательства, которое бы регулировало порядок сбора и обработки информации для обучения алгоритмов ИИ. Так, в 2023 году южнокорейская компания Samsung столкнулась с утечкой частных данных ввиду неправильной эксплуатации чата-GPT сотрудниками - все предоставленные ими сведения стали частью его базы данных [11]. С похожей трудностью в 2025 году столкнулась китайская компания DeepSeek, специализирующаяся на разработке

ИИ. Ее база данных была обнаружена в свободном доступе сотрудниками американской организации Wiz, деятельность которой осуществляется в области кибербезопасности. В сети появилось более миллиона единиц конфиденциальной информации, в числе которых запросы пользователей и цифровые программные ключи [13]. Частота утечек данных в первом полугодии 2024 года в России также возросла на приблизительно 10% относительно аналогичного периода 2023 года, был украден почти миллиард единиц персональных данных, а именно - 986 млн записей.

Затрагивая вопрос о возможности причинения ИИ вреда здоровью человека, отдельно стоит отметить группу рисков, обусловленных поступательным ростом присутствия автономных систем в транспортной инфраструктуре. Например, в 2018 году автомобиль, принадлежавший американской компании Uber, насмерть сбил женщину в городе Темпе штата Аризона [12]. Есть и другие печальные примеры.

В России активная разработка беспилотных транспортных средств обеспечивается мерами государственной поддержки, направленными на внедрение инновационных технологий, от которого ожидается укрепление позиций национального транспортного сектора на глобальном рынке за счёт повышения эффективности перевозок и сокращения эксплуатационных затрат. В 2022 году Минэкономразвития РФ запустило эксперимент в области цифровых инноваций по использованию высокоавтоматизированных транспортных средств (ВАТС) на автомобильной дороге М-11 «Нева». По окончании его действия в ноябре 2024 года стали известны промежуточные итоги: всего было зафиксировано 36 дорожно-транспортных происшествий (ДТП), 26 из которых произошли, когда беспилотные автомобильные устройства находились в автоматизированном режиме управления, т.е. вмешательство водителя диктовалось необходимостью, 10 - когда испытатель контролировал машину «вручную». Также были отмечены два случая, где вину за ДТП возложили на

ВАТС [9]. Безусловно, эти работы будут продолжаться и технологии дорабатываться в направлении достижения большей безопасности для жизни и здоровья человека.

Стоит понимать, что описанная ранее категория рисков не является изолированной, напротив, она тесно связана с более масштабными системными изменениями, обусловленными ростом влияния искусственного интеллекта в различных сферах. Так несмотря на то, что на первоначальном этапе характер ее последствий интерпретируют как технический или инфраструктурный, в течение времени становится очевидна их постепенная трансформация в социально-экономическую группу, поскольку вопросы развития человеческого капитала, серьезной трансформации современного рынка труда с тенденцией исчезновения ряда традиционных профессий, и, наряду с этим, потребностью в специалистах новых направлений, необходимость адаптации образовательной и социальной политики становятся все более актуальной. Ряд статистических исследований говорит, что возможное сокращение рабочих мест потенциально будет составлять от 8 до 47%, а темп сокращения профессий - от 1 до 3 профессий ежегодно [6]. По оценкам Goldman Sachs, примерно 300 млн рабочих мест по всему миру вероятно будут затронуты ИИ. То есть, в перспективе около 18% всех рабочих мест будут автоматизированы.

Эту тенденцию можно проследить как на российском, так и на зарубежных рынках труда. Например, ИИ создает серьезную конкуренцию российским фрилансерам, на которых приходится приблизительно 20% от всей рабочей силы страны [5]. По результатам анализа профессиональной активности практически двух тысяч фрилансеров, работающих в шести ключевых направлениях - веб-разработка, программирование, маркетинг, дизайн, обработка текстов и перевод - сфера работы с текстами оказалась наиболее уязвимой к влиянию генеративного ИИ. Так, у переводчиков среднемесячное количество заказов сократилось с 3,25 до 2,32, а у копирайтеров и редакторов — с 2,26 до 1,81 [2]. В

сегменте дизайна также наблюдается снижение: с 2,26 до 1,97 заказов в месяц. При этом в апреле 2023 г. и вовсе был зафиксирован самый большой спад в этой области, сопровождаемый снижением среднего числа заказов относительно предыдущего месяца на 7%. Подобная динамика вызвана запуском отечественной модели генерации изображений «Кандинский 2.1», что отражает общемировую тенденцию: такие ИИ-системы, как Midjourney и DALL·E 2, также оказывают влияние на занятость дизайнеров.

Еще одна группа рисков использования ИИ обусловлена его алгоритмической предвзятостью, которая способствует усилению дискриминационных практик. Так, специализированные алгоритмы осуществляют непрерывный анализ видеопотока, транслирующегося из общественных мест, идентифицируя проявления агрессии, и позволяют, тем самым, оперативно реагировать на возникающие опасные ситуации. В ходе исследований данных технологий было выявлено, что алгоритмы распознавания лиц демонстрировали меньшую точность при идентификации женщин и представителей расовых и этнических меньшинств по сравнению с мужчинами европеоидной внешности. При обработке изображений белых мужчин ошибка в определении пола составляла лишь 0,8%, тогда как для темнокожих женщин этот показатель достигал 34,7% [3].

То есть, можно сделать вывод о наличии серьезных этических рисков, связанных с высокой вероятностью алгоритмической ошибки, нарушением принципов приватности, а также угрозой избирательного контроля и дискриминации отдельных групп населения. Это обстоятельство вызывает серьезные опасения относительно способности искусственного интеллекта усиливать уже имеющиеся социальные предубеждения.

Дополнительные риски связаны с использованием ИИ-систем, обрабатывающих персональные данные без должного уровня прозрачности, согласия со стороны пользователя и адекватных мер защиты, что представляет

угрозу для права на частную жизнь. Более того, высокая степень технологической сложности и непрозрачность алгоритмических решений затрудняют процесс мониторинга и подотчетности, ставя под вопрос справедливость и легитимность принимаемых на их основе действий.

Проведенный анализ вызовов и рисков, связанных с применением технологий искусственного интеллекта, демонстрирует необходимость выработки комплексного нормативного регулирования в данной сфере. Очевидны высокая степень уязвимости цифровой инфраструктуры и наличие существенных социально-этических угроз, обусловленных алгоритмическими решениями. В этой связи приоритетом государственной и международной политики должно стать формирование эффективной системы правовых, организационных и технологических механизмов, обеспечивающих безопасное, прозрачное и ответственное применение ИИ. Такая системная работа позволит минимизировать угрозы, сохранить баланс между инновационным развитием и защитой общественных интересов, а также создать устойчивые условия для интеграции искусственного интеллекта в ключевые сферы общественной жизни.

Библиографический список

1. В России в первом полугодии утекло почти 1 млрд персональных данных. — Текст: электронный // Infowatch URL: https://www.infowatch.ru/company/presscenter/news (дата обращения: 20.03.2025).

2. Желько, Т. Искусственный интеллект отбирает заказы у фрилансеров / Т. Желько. — Текст: электронный // Ведомости. – 01.10.2024: [сайт]. — URL: https://www.vedomosti.ru (дата обращения: 14.04.2025).

3. Искусственный интеллект может быть расистом и сексистом – 13.02.2018.— Текст: электронный // Хайтек. – 13.02.2018: [сайт]. — URL: https://hightech.fm (дата обращения: 14.02.2025).

4. Исследование латентности утечек информации 2022–2023 годах — Текст: электронный // Infowatch URL: https://www.infowatch.ru (дата обращения: 19.03.25).

5. Как много компаний работают с фрилансерами в разных налоговых статусах. — Текст: электронный // Solar Staff. – URL: https://my.solarstaff.com (дата обращения: 14.04.2025).

6. Влияние искусственного интеллекта на рынок труда Российской Федерации / И. Д. Колмакова, М. Е. Бурлаков, Е. М. Колмакова, Н, А. С. Бутаков С. — Текст: непосредственный // Вестник Челябинского государственного университета. — 2023. — № 11 (481). — С. 44–52.

7. Краткая характеристика состояния преступности в Российской Федерации. — Текст: электронный // Министерство внутренних дел Российской Федерации: [сайт]. — URL: https://xn--b1aew.xn--p1ai (дата обращения: 20.04.2025).

8. Макаров, Д. А. Принципы машинного обучения. Теория и практика современной науки. – 2018. – № 6 (36). / Д. А. Макаров, А. Д. Шибанова — Текст: электронный // cyberleninka.ru: [сайт]. — URL: https://cyberleninka.ru/article/ (дата обращения: 23.03.2025).

9. Правительством запущен новый ЕПР для беспилотного грузового автотранспорта на трассе М-11. — Текст: электронный // Министерство экономического развития Российской Федерации [сайт]. — URL: https://www.economy.gov.ru (дата обращения: 25.03.2025).

10. Юрченко, В. Влияние искусственного интеллекта на рынки труда: подготовка к будущему / В. Юрченко. — Текст: электронный // Вестник науки. // cyberleninka.ru: [сайт]. — URL: https://cyberleninka.ru/article/ (дата обращения: 16.04.2025).

11. ChatGPT allegedly leaks confidential information at Samsung [Электронный ресурс] // Interesting Engineering. – URL:

https://interestingengineering.com/culture/chatgpt-alleged-leak-confidential-information-samsung?group=test_b (дата обращения: 23.03.2025).

12.Uber self-driving car fatal crash in Tempe, Arizona [Электронный ресурс] // The Verge. – 19.03.2018. – URL: https://www.theverge.com/2018/3/19/17139518/uber-self-driving-car-fatal-crash-tempe-arizona (дата обращения: 25.03.2025).

13.Wiz research uncovers exposed Deepseek database leak [Электронный ресурс] // Wiz Blog. – URL: https://www.wiz.io/blog/wiz-research-uncovers-exposed-deepseek-database-leak (дата обращения: 23.03.2025).

Оригинальность 76%