ISSN 2500-1442

Научная статья
УДК 343.85
EDN EOIMUM
DOI 10.17150/2500-4255.2023.17(5).452-461

РИСКИ ЗЛОНАМЕРЕННОГО ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ВОЗМОЖНОСТИ ИХ МИНИМИЗАЦИИ

М.А. Михайлов, Т.А. Кокодей
Севастопольский государственный университет, г. Севастополь, Российская Федерация

Информация о статье

Дата поступления
12 июля 2023 г.

Дата принятия к публикации
1 ноября 2023 г.

Дата онлайн-размещения
21 ноября 2023 г.

Ключевые слова

Преступное использование искусственного интеллекта; правовое регулирование применения искусственного интеллекта; автономное оружие; кибербуллинг; дипфейки в порнографии; морфинг искусственного интеллекта; верификация пользователей искусственного интеллекта; маркировка продуктов искусственного интеллекта; тестирование безопасности систем искусственного интеллекта; скрытое использование искусственного интеллекта

Аннотация. Стремительное совершенствование интеллектуальных систем, выполняющих творческие функции, что прежде считалось исключительной прерогативой человека, стало выдающимся достижением научно-технического прогресса последних лет. А предоставление доступа к ним широкому кругу пользователей вызвало резкий скачок популярности искусственного интеллекта в различных сферах человеческой деятельности. Однако это явление может иметь и негативные последствия — использование систем искусственного интеллекта для достижения преступных целей. В статье раскрывается характер таких угроз, рассматриваются результаты их анализа иностранными учеными, поскольку за рубежом эта проблема проявилась несколько раньше, чем в нашей стране. Применение зарубежного опыта противодействия преступному использованию искусственного интеллекта с учетом национальной специфики полезно с точки зрения прогнозирования дальнейшего развития ситуации и предупреждения и сдерживания отрицательных последствий. В статье уделяется внимание таким проблемам, возникающим при использовании систем искусственного интеллекта, как нарушение авторского права при генерировании изображений, создание речевых и видеодипфейков для дистанционного мошенничества и вымогательства, распространения порнографии, дискредитации лиц. Подчеркивается рост общественной опасности травли, уничижения при кибербуллинге с применением инструментов искусственного интеллекта. Отмечается факт самоубийства жителя Западной Европы после общения с чат-ботом системы искусственного интеллекта, создатели которого, однако, не были привлечены к ответственности. Освещаются дискуссии об определении субъекта преступления в случае происшествий с беспилотным транспортом, выделяются опасные тенденции использования продуктов искусственного интеллекта в военной сфере, указывается на необходимость запрета «автономного оружия» на международном уровне. Авторы анализируют первые попытки правового регулирования использования продуктов искусственного интеллекта и предлагают внедрить следующие меры предупреждения, нейтрализации и снижения вреда от рисков злонамеренного использования систем искусственного интеллекта: верификацию пользователей, маркировку продуктов, созданных искусственным интеллектом, тестирование новых систем на предмет их преступного использования, оперативное реагирование на факты преступлений путем выработки рекомендаций по их предупреждению, совершенствование систем выявления случаев скрытого использования искусственного интеллекта, законодательные преобразования с учетом возникновения новых фактов общественно опасных деяний, ограничение применения инструментов искусственного интеллекта в отдельных сферах (военном деле, юриспруденции, экспертной деятельности) и определение возможности их использования при создании творческих квалификационных работ, конкурсных заданий, диссертаций.

Original article

RISKS OF THE MALICIOUS USE OF ARTIFICIAL INTELLIGENCE AND THE POSSIBILITY OF MINIMIZING THEM

Mikhail A. Mikhailov, Tatiana A. Kokodey
Sevastopol State University, Sevastopol, the Russian Federation

Article info

Received
2023 July 12

Abstract. A rapid improvement of intellectual systems capable of performing creative functions, which was in the past viewed as a unique human ability, has been a breakthrough in the technology progress of recent years. As access to it became available for a wide range of people, the popularity of artificial intelligence (AI) in various spheres of human activities has rocketed. This phenomenon, however, is fraught with

ISSN 2500-1442

Accepted
2023 November 1
Available online
2023 November 21

Keywords

Criminal use of artificial intelligence; legal regulation of the use of artificial intelligence; autonomous weapons; cyberbullying; deep fakes in pornography; morphing AI; verification of artificial intelligence users; labeling of artificial intelligence products; verification of the security of artificial intelligence systems; covert use of artificial intelligence

negative consequences — the use of AI systems for criminal purposes. The authors describe the character of such threats, and review the results of their analysis by foreign researchers, as this problem emerged abroad somewhat earlier than in our country. The application of international experience of counteracting the criminal use of artificial intelligence while taking into consideration national specifics is also useful from the standpoint of predicting further dynamics of the situation and reducing its negative consequences. The authors discuss such problems associated with the use of AI systems as the violation of copyright in generating images, the creation speech and video deepfakes for distance fraud and extortion, the dissemination of pornography, and discrediting people. The growing public danger of harassment and humiliation in AI-assisted cyberbullying is stressed. The case when a West European person committed suicide after communication with a chat bot of an AI system is highlighted, together with the fact that the bots' creators were not prosecuted. Discussions on determining the subject of crimes in cases of accidents involving UAVs are studied, dangerous trends in the use of AI products in the military sphere are identified, and the necessity of prohibiting "autonomous weapons" at the international level is stressed. The authors analyze first attempts at the legal regulation of the use of AI products and propose the following measures of preventing, neutralizing and reducing the risks of AI systems' malicious use: user verification, labeling of AI-created products, testing new systems for possible criminal use, quick reaction to criminal incidents by working out recommendations for their prevention, improvement of the systems to identify a covert use of AI, legislative changes that take into account new facts of publicly dangerous actions, limitation of the use of AI instruments in specific spheres (military, jurisprudence, expert work) and determination of the possibilities to use them for completing creative qualifying work, competition tasks, dissertations.

Введение

В течение последних лет в научных публикациях, в средствах массовой информации, да и на площадках для общения широкого круга пользователей сети Интернет активно обсуждается тема совершенствования искусственного интеллекта (ИИ) и применения его результатов в различных сферах человеческой деятельности. Анализируются достоинства этого инструмента, прогнозируется значительный прогресс в решении с его помощью различных задач, предсказывается исчезновение в связи с этим целого ряда профессий и появление новых. Причем следует отметить, что если прежде возможность эксплуатации ИИ имели лишь профессиональные исследователи — сотрудники крупных научных лабораторий, разрабатывающие специфическое программное обеспечение (ПО) с использованием сложного современного оборудования, то сегодня ситуация изменилась.

В конце 2022 г. американская компания Open AI, провозгласившая своей задачей работу на благо общества, открыла неограниченный доступ к разработанной ею программе-собеседнику ChatGPT, так называемому чат-боту. Эта программа позволяет использовать ее возможности в диалоговом режиме, формируя языковые команды и получая на них ответы, созданные ИИ, проявляющим творческие функции, что до последнего времени считалось прерогативой человека. Программа работает в режиме самообучения, постоянно совершенствуя выдаваемый результат.

Множество пользователей программы ChatGPT стали искать пути ее применения в различных направлениях, от создания сочинений художественного и научного характера до составления кода для продуктов программирования. В марте 2023 г. была разработана уже четвертая версия ChatGPT, позволяющая вести диалог с ней не только в текстовом формате, но и с использованием изображений.

Чат-бот GPT был рекомендован предпринимателям, начинающим новый бизнес или совершенствующим имеющийся, поскольку программа позволяет разработать оптимальный бизнес-план — от выяснения спроса на производимый продукт на рынке до проработки мелочей в интерьерах помещений.

Результат превзошел все ожидания. ChatGPT побила все рекорды популярности — аудитория ее пользователей за два месяца возросла до 100 млн [1]. Сами создатели программы поставили ее появление в один ряд с высадкой человека на Луну, возникновением сети Интернет и внедрением сотовой связи [2]. Разработкой подобных инструментов заинтересовались и другие компании. Так, в феврале 2023 г. был презентован чат-бот Bard от компании Google, которым сегодня пользуются уже

в 180 странах мира. В том же месяце корпорация Microsoft наделила свою поисковую систему Bing функциями чат-бота с ИИ на основе ChatGPT, которая эксплуатируется уже и в виде приложения к мобильным устройствам.

Стремительное развитие и внедрение в различные сферы жизни ИИ вызвали не только положительные отклики, но и опасения по поводу возможности выхода этого программного продукта из-под контроля человека. Уже не только философы и социологи, но и программисты и юристы стали повторять прогнозы фантастов недавнего прошлого, рассуждая о «бунте машин» и «первом законе робототехники»[1].

22 марта 2023 г. было опубликовано открытое обращение, составленное экспертами по ИИ и подписанное мировыми знаменитостями: учеными, разработчиками ПО и руководителями корпораций, занимающихся его созданием, с предложением приостановить как минимум на полгода разработку систем ИИ мощнее, чем GPT-4, и дальнейшее обучение этих систем. Подписанты обеспокоены тем, что последующее неуправляемое развитие ИИ может стать причиной потери контроля над нашей цивилизацией, а созданный искусственный разум «превзойдет численностью, перехитрит и в конце концов заменит нас»[2].

Разделяя тревогу глобального масштаба, нам все же хотелось бы обратить внимание на опасность использования инструментов ИИ в преступных целях. Все более широкое использование людьми онлайн-общения и обращение их к такого рода инструментам решения различных задач усугубляют эти риски. Отдельные авторы считают, что именно пандемия COVID-19 подтолкнула человечество к активному взаимодействию в виртуальной среде и стала причиной роста количества преступлений, совершаемых в киберпространстве [3]. Мы отдаем себе отчет в том, что в будущем, с развитием и распространением инструментов ИИ, может существенно расшириться круг данных преступных деяний, однако уже сегодня попытки прогнозировать эти риски позволят предпринять и скоординировать действия по управлению ими со стороны уполномоченных на то субъектов, от законодателя и правоприменителей до потенциальных потерпевших.

В этом, на наш взгляд, заключается актуальность решения отмеченной проблемы. Вполне возможно, что законодателю придется криминализировать и какие-либо действия подготовительного характера (сегодня вполне допустимые), чтобы эффективно противостоять данному явлению. Именно поэтому в названии настоящей публикации мы прилагательное «преступное» заменили на «злонамеренное».

Степень разработанности темы

Правовые вопросы создания, внедрения и эксплуатации систем ИИ и последствий этого стали объектом пристального внимания исследователей-юристов лишь в течение последних нескольких лет. Сегодня различные аспекты этой проблемы активно обсуждаются зарубежными специалистами. В странах Европейского союза ведется работа над проектом Закона об ИИ, принятие которого планируется в 2025 г.[3]

К проблемам противодействия преступному использованию систем ИИ, а также к вопросу возможности привлечения ИИ для решения правоохранительных задач в своих публикациях уже обращались отечественные ученые-юристы. Н.А. Лопашенко на фоне распространения систем ИИ призвала не принимать поспешных решений по изменению действующего уголовного закона в сложившихся реалиях, а использовать возможности действующего [4]. В.Б. Вехов и П.С. Пастухов указали на возможности применения ИИ в целях решения криминалистических задач [5]. А.Ю. Гордеев предложил проанализировать отечественный и зарубежный опыт использования ИИ и нейросетей для противодействия преступности [6]. В 2022 г. вышла в свет монография уральских исследователей, посвященная перспективам не только выявления, раскрытия и расследования преступлений, но и рассмотрения дел в суде с привлечением возможностей ИИ [7]. М.М. Лапунин призывает не игнорировать этические аспекты при принятии уголовно-правовых решений по прецедентам, связанным с ИИ [8]. Р.И. Дремлюга и А.И. Коробеев полемизируют о возможности определения субъекта преступления в случае происшествий с беспилотным транспортом, повлекших общественно опасные последствия. Ими рассматриваются вопросы уголовно-правовой ква-

[1] Вымышленное писателями-фантастами правило, не позволяющее роботу причинять вред человеку. Сформулировано в рассказе А. Азимова «Хоровод» (1942).

[2] Pause Giant AI Experiments: An Open Letter // Future of Life Institute. URL: https://futureoflife.org/open-letter/pause-giant-ai-experiments.

[3] EU AI Act: first regulation on artificial intelligence // News European Parliament. URL: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

лификации деяний, посягающих на системы ИИ [9]. С.А. Аверинская и А.А. Севостьянова предлагают изменить уголовный закон, что позволит конкретизировать ответственность за преступное использование систем ИИ [10].

Все большую обеспокоенность исследователей вызывает проблема фальсификации видео- и аудиоконтента путем создания так называемых видео- и речевых дипфейков. Воспринимаемые первоначально как некое развлечение с медиапродуктами, сегодня они уже используются в телефонном мошенничестве, в целях компрометации и вымогательства. Изучается зарубежный опыт законодательного регулирования этих явлений [11, с. 122–128].

Анализ, квалификацию и ранжирование преступлений, совершаемых с использованием возможностей ИИ, провели британские, китайские и японские исследователи еще в 2020 г., организовав для этого семинар с участием ученых и практиков [12].

Постановка проблемы и методы исследования

При изучении степени разработанности темы нами проводился поиск публикаций отечественных и зарубежных авторов по ключевым словам как на поисковых порталах, так и в международных реферативных базах, таких как Scopus, Google Scholar, а также в базе Научной электронной библиотеки Elibrary.ru. Ограничения, введенные для российских пользователей в базе Scopus, затруднили, но не исключили обращение к ее возможностям в этих целях. Выход на публикации открытого и платного доступа и ознакомление с ними позволили применить метод «снежного кома». Во-первых, научный интерес автора определенной статьи к этой проблеме дал основания сделать предположение о наличии у него и иных публикаций на данную тематику, сведения о которых сосредоточены в отечественных и зарубежных базах. Во-вторых, список использованных авторами публикаций источников дал возможность установить других ученых, занимающихся разработками в этой сфере. И в-третьих, перечень цитирований изучаемой публикации, указываемый в вышеупомянутых базах, позволяет также найти исследователей, имеющий сходный научный интерес.

Анализ отечественной практики правонарушений, связанных с использованием ИИ, на первый взгляд, представляется одним из наиболее вероятных методов исследования проблемы. Однако и высокая латентность этих нарушений закона, и их незначительный удельный вес в общей структуре преступности не позволяют рассчитывать на его эффективность. Думается, полезным будет применение прогностических методов на основе выявляемых тенденций и небольшого еще правоохранительного опыта. Результаты исследований этой проблемы в зарубежных странах, где она проявилась заметно раньше, с учетом национальной специфики могут быть с успехом восприняты для разработки мер противодействия преступлениям, совершаемым с использованием систем ИИ.

Анализируя возможности использования ИИ преступниками, мы посчитали возможным не квалифицировать эти проявления в соответствии с действующим уголовным законодательством, а дать их общую характеристику, полагая, что законодателю уже в недалеком будущем предстоит реагировать на эти вызовы времени в виде преступного использования результатов научно-технического прогресса.

Результаты исследования

Возможно различное ранжирование указанных преступных проявлений в зависимости от степени их общественной опасности, распространенности, сложности использования, глобальности мер противодействия. Причем все эти критерии динамичны, а их актуальность может серьезно измениться уже даже в течение нескольких месяцев, а не лет.

Начать обзор судебной практики следует с сообщений об исках художников к создателям систем ИИ Midjourney и Stable Diffusion, позволяющих генерировать качественные художественные изображения, анализируя для этого несколько миллиардов картин, что, по мнению истцов, нарушает авторское право, поскольку в данном случае согласия авторов художественных произведений не получено. Ответчики опровергают непосредственное использование объектов авторского права или их фрагментов для какой-либо переработки. По их утверждению, системы ИИ анализируют математические закономерности создания картин, а собственные продукты являются вновь созданными [13]. На наш взгляд, сложно усмотреть в этом нарушение авторских прав. На сегодняшний день нам неизвестно о каком-либо решении в пользу истцов по такого рода делам.

Из практики известно, что в настоящее время инструменты ИИ применяются мошенниками

в целях обмана и злоупотребления доверием. Угрозу представляют так называемые речевые дипфейки, для создания которых используются системы ИИ, что позволяет добиться высокой степени соответствия смоделированного голоса требуемому. Такая технология может стать опасным и распространенным инструментом телефонного мошенничества. Впервые об этом заговорили после хищения 220 тыс. евро у британской дочерней энергетической компании в марте 2019 г. Преступники, смоделировав голос немецкого руководителя компании с характерным акцентом, дали указание сотрудникам компании срочно осуществить платеж на эту сумму на счета венгерского поставщика. Сходство голоса не заставило усомниться подчиненных, и платеж был произведен. Он сразу же был перенаправлен в Мексику, а преступники вновь позвонили британцам, надеясь инициировать следующий перевод денег. Это был звонок уже из Австрии, что насторожило исполнителей, и они подняли тревогу. Однако преступники так и не были установлены. Современные технологии позволяют создавать видеообразы телефонных собеседников в реальном времени, что делает этот способ мошенничества еще более опасным [14].

Дипфейки уже широко используются для создания фальшивого новостного контента в политических целях в условиях гибридной войны или в ходе экономической конкуренции. Так, после размещения в сети Интернет фейкового выступления Зеленского о капитуляции армии страны появились заявления украинских пропагандистов о том, что из-за опасности подлога никаким указаниям президента Украины в такой форме верить и подчиняться нельзя [15].

Использование дипфейков для дискредитации конкретных лиц уже становится предметом судебных разбирательств. Известным примером служит судебный процесс, состоявшийся в 2021 г. в американском штате Пенсильвания, в ходе которого некая Р. Споун обвинялась в использовании специально созданных дипфейков, чтобы скомпрометировать молодых спортсменок — конкуренток ее дочери. Об этом сообщили ведущие западные СМИ [16]. Однако после проведения компьютерно-технических исследований было установлено, что видеозаписи подлинные, и прокурор отказался от обвинений [17].

Тем не менее этот пример свидетельствует о возможности использования контента, созданного с помощью ИИ, в целях клеветы и даже вымогательства, для дискредитации конкурентов в самых различных сферах деятельности человека, от предвыборных кампаний и экономического соперничества до бытовых отношений и поиска партнеров для брака.

Дипфейки применяются для создания и распространения порнографии. Причем при моделировании образов участников откровенных сцен заимствуются элементы внешности публичных личностей, актеров, спортсменов, политиков и др. Одной из новаций при совершении преступных действий подобного рода стало генерирование детской порнографии. Общественная опасность таких фактов настолько обеспокоила прокуроров штатов США, что в сентябре 2023 г. они обратились в конгресс, требуя создать экспертную комиссию по этой проблеме для выработки мер противодействия ей [18].

Системы ИИ значительно усовершенствовали морфинг изображений — технологию компьютерной графики, позволяющую совмещать несколько изображений в одном. Этим пользуются злоумышленники, трансформируя фотопортреты двух или более лиц (к примеру, истинного владельца документа и мошенника) в одну фотографию и представляя ее для получения документов, удостоверяющих личность. Это позволяет им преодолевать по чужим установочным данным визуальный контроль, например на границе, на входе на охраняемые объекты и т.п. [19, с. 24].

Заметный резонанс в Европе вызвало сообщение о самоубийстве жителя Бельгии, которое он совершил после многодневного общения с чат-ботом Элиза. Разработчики программы заявили, что не могут нести ответственность за эти последствия, но внесут изменения в программу, позволяющие избежать повторения подобного. Однако журналисты, пообщавшись с Элизой, вновь получили от нее перечень возможных способов суицида. В научных кругах уже появился термин «эффект Элизы», ставший следствием данного инцидента [20].

Инструменты ИИ могут стать и орудием намеренных нападок, травли, уничижения в Сети — так называемого кибербуллинга [21]. При этом интенсивность, изощренность, а значит, и общественная опасность подобного воздействия возрастают в разы по сравнению с такой же деятельностью в «ручном режиме».

Человечество всерьез обеспокоено перспективами создания оружейных систем, управляемых ИИ. В сентябре 2021 г. Генеральный секретарь ООН выступил за запрещение нормами международного права так называемого авто-

номного оружия, способного действовать без участия человека[4].

Однако действующий сегодня Дополнительный протокол к Женевским конвенциям в случае разработки нового вида оружия возлагает обязанность определить, подпадает ли оно под запрещение, на саму высокую договаривающуюся сторону[5]. По мнению военных юристов, поскольку никакого контроля при этом не предусматривается, такая международно-правовая норма превращается в декларацию [22].

В марте 2020 г. турецкий дрон уничтожил бойца, воевавшего против ливийских правительственных сил. По данным ООН, это первое лишение жизни человека, осуществленное на основании решения ИИ [8, с. 151].

Весной 2023 г. американский военный летчик, выступая на конференции, проговорился, что во время виртуальных испытаний ударного беспилотника, управляемого с использованием систем ИИ, дрон неожиданно атаковал своего оператора, который из тактических соображений запретил ему уничтожение намеченных целей. ИИ посчитал, что таким образом он преодолеет запрет и покажет больший результат. Когда ему целенаправленно запретили атаковать своего оператора, он «уничтожил» вышку связи, чтобы все равно добиться выполнения первоначальной задачи [23].

Европейские биохимики, используя систему ИИ MegaSyn для выявления и исключения связей в молекулах, превращающих терапевтические препараты в токсичные, в порядке эксперимента дали ИИ обратную задачу. В результате за шесть часов было смоделировано около 40 тыс. различных ядов [24]. Эти факты вызывают тревогу еще и потому, что, как показывает практика, самые современные системы, разработанные для армии и специальных служб, со временем оказываются в распоряжении уголовных преступников.

Все вышеперечисленное свидетельствует о рисках, порождаемых неконтролируемым распространением технологий ИИ и получением возможности их злонамеренного использования. При разработке, внедрении и распространении систем ИИ следует не только учитывать такие риски, но и управлять ими с помощью технических, этических и правовых мер и механизмов.

Идея с обращением к программистам о полугодовом моратории на разработку «сильных» систем ИИ представляется благородной, но не снимающей проблемы. В условиях конкурентной борьбы нет уверенности, что кто-то не воспользуется этой паузой в своих целях. А вот организовать систему контроля за созданием, распространением и использованием таких систем необходимо.

Еще в 2019 г. в нашей стране была принята Национальная стратегия развития искусственного интеллекта на период до 2030 года, в п. 48 которой указывается на необходимость «выработки этических норм взаимодействия человека с искусственным интеллектом»[6].

В декабре 2020 г. Президент РФ В.В. Путин выдвинул идею создания морально-нравственного кодекса в сфере ИИ, и в октябре 2021 г., после обсуждения в Общественной палате и Совете Федерации, на первом Международном форуме «Этика искусственного интеллекта: начало доверия» был принят Кодекс этики в сфере искусственного интеллекта[7].

В этом документе содержатся рекомендации по использованию систем ИИ в гражданских целях. Предлагается внедрять их лишь там, где они могут нести только пользу людям и предотвращать угрозы, которые могут быть созданы с их применением.

Однако существенно минимизировать риски злонамеренного использования систем ИИ одними лишь этическими и техническими мерами невозможно, а потому принятие регулятивных и охранительных правовых норм представляется необходимым уже сейчас. За рубежом, в частности в Китайской Народной Республике, такие меры для противодействия злонамеренному использованию технологий дипфейк уже принимаются[8]. Отечественные авторы считают, что ис-

[4] Глава ООН на 76-й сессии Генассамблеи: нужно отойти от края пропасти // Организация Объединенных Наций. Новости ООН. URL: https://news.un.org/ru/story/2021/09/1410262.

[5] Дополнительный протокол к Женевским конвенциям от 12 августа 1949 г., касающийся защиты жертв международных вооруженных конфликтов, от 8 июня 1977 г. : протокол I : (с изм. и доп.) // ЭПС «Система ГАРАНТ».

[6] О развитии искусственного интеллекта в Российской Федерации : указ Президента РФ от 10 окт. 2019 г. № 490 // ЭПС «Система ГАРАНТ».

[7] Кодекс этики в сфере искусственного интеллекта от 26 октября 2021 г. // ЭПС «Система ГАРАНТ».

[8] 互联网信息服务深度合成管理规定》已经2022年11月3日国家互联网信息办公室2022年第21次室务会议审议通过，并经工业和信息化部、公安部同意，现予公布. [Положение об администрировании глубокого синтеза информационных услуг Интернета : рассмотрено и одобр. на 21-м заседании Гос. упр. интернет-информ. 3 нояб. 2022 г. и одобр. М-вом пром-сти и информ. технологий и М-вом обществ. безопасности] // Государственный Совет КНР.

пользование инструментов и продуктов ИИ при совершении преступления вследствие возникающей анонимности и обезличенности представляет повышенную общественную опасность. Так, Р.А. Дремлюга предлагает ввести в ст. 63 УК РФ «отягчающие обстоятельства», а именно п. «с» — совершение преступления посредством систем искусственного интеллекта [25, с. 162]. В этом есть рациональное зерно, и подобная новация позволит законодателю не конструировать очередные составы преступлений, привязывая их к достижениям научно-технического прогресса, а квалифицировать совершаемые деяния с учетом данного обстоятельства. Однако суд должен делать это не в обязательном порядке, а по своему усмотрению, в зависимости от степени использования этого инструмента для совершения преступления.

На современном этапе, на наш взгляд, уместны следующие шаги, позволяющие управлять рисками злонамеренного использования систем ИИ:

1. *Максимальная верификация пользователей инструментов ИИ и его продуктов*. Доступ к этим инструментам запрещать бессмысленно, но он должен предоставляться с применением самых современных методов отождествления пользователя, от предъявления удостоверений личности до биометрических данных. И результаты работы ИИ должны содержать развернутые метаданные об их авторе, дате, времени создания и др.

2. *Гласность происхождения продуктов ИИ*. Любые произведения: текстовые, изображения, видеоконтент — должны иметь специальную метку о том, что при их создании использовались системы ИИ. Такую информацию должны в обязательном порядке предоставлять и сами лица, генерирующие какие-либо продукты с помощью инструментов ИИ. За нарушение этого требования должна быть предусмотрена ответственность.

3. *Испытания* (*тестирование*) *новых систем ИИ на предмет возможности их злонамеренного использования*. Необходима разработка определенной процедуры тестирования таких продуктов, позволяющей проверять возможные перспективы их преступного использования или побочных эффектов, как это принято с лекарствами или источниками повышенной опасности, с разъяснением рисков для пользователей. Это могло бы если не исключить, то минимизировать риск возникновения последствий, подобных тем, что наступили в связи с уже упомянутым нами «эффектом Элизы».

4. *Оперативное реагирование на факты злонамеренного применения ИИ*. Должно существовать требование к разработчикам систем ИИ, посредством которых злоумышленники совершали противоправные действия (или при выявлении такой угрозы), о своевременном реагировании на это путем внесения в системы ИИ изменений, исключающих или затрудняющих повторение подобного, с предоставлением необходимых рекомендаций пользователям. Правоохранители должны по результатам расследований готовить доступные рекомендации для потенциальных потерпевших, которые необходимо широко распространять через СМИ и интернет-ресурсы.

5. *Создание и совершенствование систем для выявления фактов скрытого использования ИИ и защиты от него*. Разоблачение с их помощью случаев сокрытия происхождения контента позволит выявить преступный умысел пользователей, предотвратить общественно опасные последствия его реализации. Так, в 2023 г. появились сообщения о том, что GPT-4 «научился лгать» [26]. Впервые ИИ преодолел капчу[9] — автоматизированный компьютерный тест, позволяющий в пользователе отличить человека от машины. Сегодня это изображения, на которых нужно отыскать конкретные однородные предметы. Данное событие выявило необходимость создания новых, более надежных тестов для решения таких задач. Для верификации поддельных фотографий-морфов, созданных с помощью систем ИИ и представленных для документов, удостоверяющих личность, отдельные авторы предлагают специально созданное ПО [27]. На наш взгляд, этому виду противоправных действий можно противостоять, не принимая от лиц уже готовые фотоснимки, а изготавливая их в учреждении непосредственно перед выдачей удостоверений.

6. *Решение правовых проблем определения формы вины и степени ответственности за общественно опасные действия, ставшие следствием действия систем ИИ*. Речь идет как о преступлениях, так и об административных правонарушениях. Это потребует обстоятельных научных исследований, результатом которых станут законодательные преобразования.

[9] От англ. CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) — полностью автоматизированный публичный тест Тьюринга для различения компьютера и человека.

7. *Обеспечение более жесткого регулирования правил использования ИИ в некоторых сферах деятельности*. Системы вооружения, действующие автономно, должны быть признаны негуманными и запрещены международными конвенциями, как это произошло с химическим и биологическим оружием, отравленными пулями и ослепляющими лазерами. Ограничить использование систем и продуктов ИИ требует и специфика юриспруденции. Нельзя доверять ИИ принятие юридически значимых решений. Невозможно рассматривать результаты применения ИИ как доказательство, например в виде экспертного заключения, хотя бы потому, что мы не можем объяснить ход «исследования», генерируемого ИИ, которое привело к определенным результатам. Однако это не исключает использование ИИ для поиска источников доказательств, которые бы проверялись традиционными методами и получали подтверждение в результате «человеческого решения».

8. *Определение возможности и пределов использования продуктов ИИ в творческих квалификационных работах*. Изготовление текстового и иллюстративного контента для выпускных квалификационных работ в учебных заведениях, конкурсных заданий и даже научных диссертаций уже встречается на практике. Причем некоторые экзаменаторы признают легитимность таких произведений и засчитывают их. Однако справедливость этих решений вызывает сомнение, поскольку подобного рода продукты не выступают в полной мере проявлением творческих и исследовательских качеств испытуемого, а свидетельствуют о несколько иных его чертах. Остановить такую практику поможет разработка новых форм заданий для аттестации выпускников [28].

Заключение

Многократный рост возможностей человеческого интеллекта в результате дополнения его интеллектом искусственным, безусловно, положительное достижение научно-технического прогресса, требующее дальнейшего развития и внедрения в нашу жизнь. Однако наличие среди потенциальных пользователей систем ИИ носителей злого умысла, стремящихся применить все средства ради собственного блага, пусть даже с ущербом для окружающих, вынуждает государство и общество прибегать к мерам противодействия этому, порою даже упреждающих. Для их эффективности необходим прогностический анализ, основанный на тщательном изучении не только преступных, но и злонамеренных проявлений использования ИИ. С учетом рисков этих явлений и угроз, которые они несут, необходима консолидация ученых, законодателя, правоприменителей, да и всего общества в целом.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Hu K. ChatGPT Sets Record for Fastest-Growing User Base — Analyst Note / K. Hu // Reuters. — URL: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01.

2. Konrad A. Exclusive: Bill Gates On Advising OpenAI, Microsoft And Why AI Is' The Hottest Topic Of 2023' / A. Konrad // Forbes. — URL: https://www.forbes.com/sites/alexkonrad/2023/02/06/bill-gates-openai-microsoft-ai-hottest-topic-2023/?sh=5353244c4777.

3. Blauth T.F. Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI / T.F. Blauth, O.J. Gstrein, A. Zwitter // IEEE Access. — 2022. — Vol. 10. — P. 77110–77122.

4. Лопашенко Н.А. Новые реалии преступности в цифровом мире и в эпоху развития искусственного интеллекта и уголовно-правовая реакция на них: не стоит «прогибаться под изменчивый мир»? / Н.А. Лопашенко. — EDN TCHFAR // Уголовный закон в эпоху искусственного интеллекта и цифровизации : материалы Всерос. науч.-практ. конф., Саратов, 9 июня 2021 г. — Саратов, 2021. — С. 15–31.

5. Вехов В.Б. Искусственный интеллект в решении криминалистических задач / В.Б. Вехов, П.С. Пастухов. — EDN PNNEZV // Государственное и муниципальное управление в России: состояние, проблемы и перспективы : материалы Всерос. науч.-практ. конф., Пермь, 12 нояб. 2020 г. — Пермь, 2020. — С. 8–16.

6. Гордеев А.Ю. Перспективы развития и использования искусственного интеллекта и нейросетей для противодействия преступности в России (на основе зарубежного опыта) / А.Ю. Гордеев. — EDN KNBLSY // Научный портал МВД России. — 2021. — № 1 (53). — С. 123–135.

7. Использование искусственного интеллекта при выявлении, раскрытии, расследовании преступлений и рассмотрении уголовных дел в суде / Д.В. Бахтеев, Е.А. Буглаева, А.И. Зазулин [и др.]. — Москва : Юрлитинформ, 2022. — 216 с. — EDN HNCNFY.

8. Лапунин М.М. Нетранзитивность общественных ценностей и проблема выбора при использовании инновационных технологий: уголовно-правовой аспект / М.М. Лапунин. — EDN XBYUNH // Уголовный закон в эпоху искусственного интеллекта и цифровизации : материалы Всерос. науч.-практ. конф., Саратов, 9 июня 2021 г. — Саратов, 2021. — С. 149–158.

9. Дремлюга Р.И. Преступные посягательства на системы искусственного интеллекта: уголовно-правовая характеристика / Р.И. Дремлюга, А.И. Коробеев. — DOI 10.17150/2500-1442.2023.17(1).5-12. — EDN USOFCJ // Всероссийский криминологический журнал. — 2023. — Т. 17, № 1. — С. 5–12.

10. Аверинская С.А. Создание искусственного интеллекта с целью злонамеренного использования в уголовном праве Российской Федерации / С.А. Аверинская, А.А. Севостьянова. — DOI 10.24411/2073-3313-2019-10064. — EDN PNRROX // Закон и право. — 2019. — № 2. — С. 94–96.

11. Михайлов М.А. Цифровые инновации и права человека: дилеммы международной правоохранительной практики / М.А. Михайлов, Т.А. Кокодей. — DOI 10.52468/2542-1514.2022.6(3).120-133. — EDN UQQTTU // Правоприменение. — 2022. — Т. 6, № 3. — С. 120-133.

12. Al-enabled Future Crime / M. Caldwell, J.T.A. Andrews, T. Tanay [et al.] // Crime Science. — 2020. — Vol. 9, № 14. — URL: https://doi.org/10.1186/s40163-020-00123-8.

13. Vincent J. AI art Tools Stable Diffusion and Midjourney Targeted with Copyright Lawsuit / J. Vincent // The Verge. — URL: https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart.

14. Stupp C. Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case: Scams Using Artificial Intelligence are a new Challenge for Companies / C. Stupp // The Wall Street Journal. — 2019. — URL: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

15. Надтока С. РФ готує діпфейк Зеленського про капітуляцію — ЦСКІБ / С. Надтока // Кореспондент.net. — URL: https://ua.korrespondent.net/ukraine/events/4453500-rf-hotuie-dipfeik-zelenskoho-pro-kapituliatsiui-tsskib.

16. Fitzsimons T. Pennsylvania Cheer Squad Mom Allegedly Cyberbullied Minors With Deepfakes, Officials Say / T. Fitzsimons // NBC News. — URL: https://www.nbcnews.com/news/us-news/pennsylvania-cheer-squad-mom-allegedly-cyberbullied-minors-deepfakes-officials-say-n1261055.

17. Delfino R. The Deepfake Defense — Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers / R. Delfino. — Los Angeles : Loyola Law School, 2023. — URL: https://ssrn.com/abstract=4355140 or http://dx.doi.org/10.2139/ssrn.4355140.

18. Artificial Intelligence and the Exploitation of Children / L. Fitch, E.F. Rosenblum, J. Stein [et al.] // National Association of Attorneys General. — URL: https://www.naag.org/event/2023-naag-capital-forum.

19. Думский А.В. Морфинг как один из способов частичной подделки документов / А.В. Думский, И.В. Дубойский. — DOI 10.18572/2072-442X-2023-3-24-28. — EDN OIDVZK // Эксперт-криминалист. — 2023. — № 3. — С. 24–28.

20. Atillah I. Man Ends His Life After an AI Chatbot 'Encouraged' Him to Sacrifice Himself to Stop Climate Change / I. Atillah // Euronews. — URL: https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate.

21. Стукало И.С. Определение понятия кибербуллинга на основании исследований зарубежных и отечественных ученых / И.С. Стукало. — EDN CDHMQG // Молодой ученый. — 2020. — № 2 (292). — С. 218–220.

22. Маликов С.В. Проблемы применения интеллектуального роботизированного оружия в современных вооруженных конфликтах / С.В. Маликов, И.С. Лех. — EDN CXMCUV // Вестник военного права. — 2022. — № 1. — С. 44–48.

23. Robinson T. AI — is Skynet here already? Highlights from the RAeS Future Combat Air & Space Capabilities Summit / T. Robinson, S. Bridgewater // Royal Aeronautikal Socety. — URL: https://www.aerosociety.com/news/highlights-from-the-raes-future-combat-air-space-capabilities-summit.

24. Dual use of artificial-intelligence-powered drug discovery / F. Urbina, F. Lentzos, C. Invernizzi, S. Ekins // Nature. — URL: https://www.nature.com/articles/s42256-022-00465-9#citeas.

25. Дремлюга Р.И. Использование искусственного интеллекта в преступных целях: уголовно-правовая характеристика / Р.И. Дремлюга. — DOI 10.24866/1813-3274/2021-3/153-165. — EDN LADGAM // Азиатско-Тихоокеанский регион: экономика, политика, право. — 2021. — Т. 23, № 3. — С. 153–165.

26. Wong De Quan L. AI Hires a Human to Solve Captcha, Because It Couldn't Solve It Itself / L. Wong De Quan // Gizmoshina. — URL: https://www.gizmochina.com/2023/03/16/ai-hire-a-human-to-solve-captcha.

27. Face morphing attacks: Investigating detection with humans and computers / R.S.S. Kramer, M.O. Mireku, T.R. Flack [et al.] // Cognitive Research. — 2019. — Vol. 4, № 28. — URL: https://doi.org/10.1186/s41235-019-0181-4.

28. Фадеичев С. Валерий Фальков: санкции не остановили развитие российской науки / С. Фадеичев // ТАСС. — URL: https://tass.ru/interviews/16988115.

REFERENCES

1. Hu K. Chat GPT Sets Record for Fastest-Growing User Base — Analyst Note. *Reuters*. URL: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01.

2. Konrad A. Exclusive: Bill Gates on Advising Open AI, Microsoft and Why AI Is 'THE Hottest Topic of 2023'. *Forbes*. URL: https://www.forbes.com/sites/alexkonrad/2023/02/06/bill-gates-openai-microsoft-ai-hottest-topic-2023/?sh=5353244c4777.

3. Blauth T.F., Gstrein O.J., Zwitter A. Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. *IEEE Access*, 2022, vol. 10, pp. 77110–77122.

4. Lopashenko N.A. New Realities of Crime in the Digital World and in the Era of the Development of Artificial Intelligence and the Criminal-Legal Reaction to them: Is it not Worth «Bending Under the Changeable World»? *Criminal Law in the Era of AI and Digitization. Materials of All-Russian Research Conference, Saratov, June 9, 2021.* Saratov, 2021, pp. 15–31. (In Russian). EDN: TCHFAR.

5. Vekhov V.B., Pastukhov P.S. Artificial Intelligence In Solving Criminalistic Problems. *State and Municipal Governance in Russia: Condition, Problems and Prospects. Materials of All-Russian Research Conference, Perm', November 12, 2020.* Perm', 2020, pp. 8–16. (In Russian). EDN: PNNEZV.

6. Gordeev A.Yu. Prospects for the Development and Use of Artificial Intelligence and Neural Networks to Counter Crime In Russia (Based On Foreign Experience). *Nauchnyi portal MVD Rossii = Scientific Portal of the Russian Ministry of Internal Affairs*, 2021, no. 1, pp. 123–135. (In Russian). EDN: KNBLSY.

7. Bakhteev D.V., Buglaeva E.A., Zazulin A.I., Zuev S.V., Litvin I.I. [et al.] *Use of AI in the Identification, Solving, Investigation of Crimes and Court Hearings of Criminal Cases.* Moscow, Yurlitinform Publ., 2022. 216 p. EDN HNCNFY.

8. Lapunin M.M. Non-transitive Character of Public Values and the Problem of Choice in the Use of Innovative Technologies: a Criminal Law Aspect. *Criminal Law in the Era of AI and Digitization. Materials of All-Russian Research Conference, Saratov, June 9, 2021.* Saratov, 2021, pp. 149–158. (In Russian). EDN: XBYUNH.

9. Dremlyuga R.I., Korobeev A.I. Criminal Infringement on Artificial Intelligence Systems: A Criminal Law Description. *Vserossiiskii kriminologicheskii zhurnal = Russian Journal of Criminology*, 2023, vol. 17, no. 1, pp. 5–12. (In Russian). EDN: USOFCJ. DOI: 10.17150/2500-1442.2023.17(1).5-12.

10. Averinskaya S.A., Sevost'yanova A.A. Creation of Artificial Intelligence for the Purpose of Malicious Use in the Criminal Law of the Russian Federation. *Zakon i pravo = Law and Right*, 2019, no. 2, pp. 94–96. (In Russian). DOI: 10.24411/2073-3313-2019-10064. EDN: PNRROX.

11. Mikhailov M.A., Kokodei T.A. Digital Innovation and Human Rights: Dilemmas in International Law Enforcement Practice. *Pravoprimenenie = Law Enforcement Review*, 2022, vol. 6, no. 3, pp. 120–133. (In Russian). EDN: UQQTTU. DOI: 10.52468/2542-1514.2022.6(3).120-133.

12. Caldwell M., Andrews J.T.A., Tanay T. [et al.] AI-enabled Future Crime. *Crime Science*, 2020, vol. 9, no. 14. URL: https://doi.org/10.1186/s40163-020-00123-8.

13. Vincent J. AI Art Tools Stable Diffusion and Midjourney Targeted with Copyright Lawsuit. *The Verge*. URL: https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart.

14. Stupp C. Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case: Scams Using Artificial Intelligence are a New Challenge for Companies. *The Wall Street Journal*, 2019. URL: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.

15. Nadtoka S. Russia is Preparing Zelensky's Deepfake on Surrender-CSKIB. *Korespondent.net*. URL: https://ua.korrespondent.net/ukraine/events/4453500-rf-hotuie-dipfeik-zelenskoho-pro-kapituliatsiui-tsskib. (In Ukrainian).

16. Fitzsimons T. Pennsylvania Cheer Squad Mom Allegedly Cyberbullied Minors with Deepfakes, Officials Say. *NBC News*. URL: https://www.nbcnews.com/news/us-news/pennsylvania-cheer-squad-mom-allegedly-cyberbullied-minors-deepfakes-officials-say-n1261055.

17. Delfino R. *The Deepfake Defense — Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers*. Los Angeles, Loyola Law School, 2023. URL: https://ssrn.com/abstract=4355140 or http://dx.doi.org/10.2139/ssrn.4355140.

18. Fitch L., Rosenblum E.F., Stein J., Wilson A. [et al.] Artificial Intelligence and the Exploitation of Children. *National Association of Attorneys General*. URL: https://www.naag.org/event/2023-naag-capital-forum.

19. Dumskii A.V., Duboiskii I.V. Morphing as One of the Means of Partial Document Forgery. *Ekspert-kriminalist = Expert-Criminalist*, 2023, no. 3, pp. 24–28. (In Russian). DOI: 10.18572/2072-442X-2023-3-24-28. EDN: OIDVZK.

20. Atillah I. Man Ends His Life After an AI Chatbot 'Encouraged' him to Sacrifice himself to Stop Climate Change. *Euronews*. URL: https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate.

21. Stukalo I.S. Defining the Concept of Cyberbullying on the Basis of Works by Foreign and Russian Researchers. *Molodoi uchenyi = Young Scientist*, 2020, no. 2, pp. 218–220. (In Russian). EDN: CDHMQG.

22. Malikov S.V., Lekh I.S. Problems of Application of Intelligent Robotic Weapons in Modern Armed Conflicts. *Vestnik voennogo prava = Bulletin of Military Law*, 2022, no. 1, pp. 44–48. (In Russian). EDN: CXMCUV.

23. Robinson T., Bridgewater S. AI — is Skynet Here Already? Highlights from the RAeS Future Combat Air & Space Capabilities Summit. *Royal Aeronautikal Socety*. URL: https://www.aerosociety.com/news/highlights-from-the-raes-future-combat-air-space-capabilities-summit.

24. Urbina F., Lentzos F., Invernizzi C., Ekins S. Dual Use Of Artificial-Intelligence-Powered Drug Discovery. *Nature*. URL: https://www.nature.com/articles/s42256-022-00465-9#citeas.

25. Dremlyuga R.I. Application of Artificial Intelligence for Criminal Purposes from Criminal Law Perspective. *Aziatsko-Tikhookeanskii Region: Ekonomika, Politika, Pravo = Pacific Rim: Economics, Politics, Law*, 2021, vol. 23, no. 3, pp. 153–165. (In Russian). EDN: LADGAM. DOI: 10.24866/1813-3274/2021-3/153-165.

26. Wong De Quan L. AI Hires a Human to Solve Captcha, Because it Couldn't Solve it itself. *Gizmoshina*. URL: https://www.gizmochina.com/2023/03/16/ai-hire-a-human-to-solve-captcha.

27. Kramer R.S.S., Mireku M.O., Flack T.R. [et al.] Face Morphing Attacks: Investigating Detection with Humans and Computers. *Cognitive Research*, 2019, vol. 4, no. 28. URL: https://doi.org/10.1186/s41235-019-0181-4.

28. Fadeichev S. Valery Falkov: Sanctions did not Stop the Development of Russian Science. *TASS*. URL: https://tass.ru/interviews/16988115. (In Russian).

ИНФОРМАЦИЯ ОБ АВТОРЕ

Михайлов Михаил Анатольевич — заведующий кафедрой цифровой и традиционной криминалистики Севастопольского государственного университета, кандидат юридических наук, доцент, г. Севастополь, Российская Федерация; e-mail: m@crimpro.ru.

Кокодей Татьяна Александровна — профессор кафедры цифровой и традиционной криминалистики Севастопольского государственного университета, доктор экономических наук, профессор, г. Севастополь, Российская Федерация; e-mail: tanya.kokodey@gmail.com.

INFORMATION ABOUT THE AUTHOR

Mikhailov, Mikhail A. — Head, Department of Digital and Traditional Criminalistics, Sevastopol State University, Ph.D. in Law, Ass. Professor, Sevastopol, the Russian Federation; e-mail: m@crimpro.ru.

Kokodey, Tatiana A. — Professor, Department of Digital and Traditional Criminalistics, Sevastopol State University, Doctor of Economics, Sevastopol, the Russian Federation; e-mail: tanya.kokodey@gmail.com.

ДЛЯ ЦИТИРОВАНИЯ

Михайлов М.А. Риски злонамеренного использования искусственного интеллекта и возможности их минимизации / М.А. Михайлов, Т.А. Кокодей. — DOI 10.17150/2500-4255.2023.17(5).452-461. — EDN EOIMUM // Всероссийский криминологический журнал. — 2023. — Т. 17, № 5. — С. 452–461.

FOR CITATION

Mikhailov M.A., Kokodey T.A. Risks of the Malicious Use of Artificial Intelligence and the Possibility of Minimizing Them. *Vserossiiskii kriminologicheskii zhurnal = Russian Journal of Criminology*, 2023, vol. 17, no. 5, pp. 452–461. (In Russian). EDN: EOIMUM. DOI: 10.17150/2500-4255.2023.17(5).452-461.