Научная статья
УДК 004:34
https://doi.org/10.36511/2078-5356-2024-3-77-81

Потенциальные угрозы, связанные с применением искусственного интеллекта в преступных действиях

***Сафронов Дмитрий Владимирович*[1, 2]**

[1]Нижегородская академия МВД России, Нижний Новгород, Россия,
[2]Волго-Вятский филиал Московского технического университета связи и информатики, Нижний Новгород, Россия, dv_safronov@mail

Аннотация

Статья посвящена вопросам оценки угроз, связанных с применением искусственного интеллекта в реализации преступных действий с учетом стремительного развития технических возможностей и областей применения современных нейросетей, поскольку степень готовности к противодействию непосредственно влияет на эффективность такого противодействия и оперативность его применения.

Ключевые слова: искусственный интеллект, нейронные сети, преступление

Для цитирования

Сафронов Д. В. Потенциальные угрозы, связанные с применением искусственного интеллекта в преступных действиях // На страже экономики. 2024. № 3 (30). С. 77–81. https://doi.org/10.36511/2588-0071-2024-3-77-81.

Potential threats associated with the use of artificial intelligence in criminal acts

***Dmitry V. Safronov*[1, 2]**

[1]Nizhny Novgorod Academy of the Ministry of Internal Affairs of Russia, Nizhny Novgorod, Russian Federation
[2]Volga-Vyatky branch of the Moscow Technical University of Communication and Informatics, Nizhny Novgorod, Russian Federation, dv_safronov@mail

Abstract

The article is devoted to the issues of assessing the threats associated with the use of artificial intelligence in the implementation of criminal actions, taking into account the rapid development of technical capabilities and areas of application of modern neural networks, since the degree of readiness for counteraction directly affects the effectiveness of such counteraction and the efficiency of its application.

Keywords: artificial intelligence, neural networks, crime

For citation

Safronov D. V. Potential threats associated with the use of artificial intelligence in criminal acts. *The Economy under Guard,* 2024, no. 3 (30), pp. 77–81. (In Russ.). https://doi.org/10.36511/2588-0071-2024-3-77-81.

Стремительно ворвавшийся в нашу жизнь искусственный интеллект (далее — ИИ) не оставил равнодушным ни одного специалиста, активно занимающегося деятельностью в различных областях общественной жизни: на производстве, в сферах СМИ, сельского хозяйства, здравоохранения, телекоммуникаций, науки и культуры, финансов, транспорта, вооружения, безопасности и так далее.

Перечень нейростей, реализующих функции ИИ, включает в себя более 100 тысяч наименований и с каждым днем только растет, что доказывает не только актуальность вопроса разработки, применения и совершенствования нейросетей в различных областях, но и растущую значимость данного процесса.

Сегодня нейросети способны обрабатывать звук, статические графические изображения, видеоинформацию, текст, что позволяет в режиме реального времени выделять из звуковых фрагментов голосовые дорожки, обрабатывать и заменять их новыми, выделять из графических изображений и видеокадров объекты и производить замену с наложением новых теней, цветовых переходов, аналогичных реальным, преобразовывать текстовую информацию согласно стилевым особенностям, заданным в тексте-образце, а также на основе постоянно пополняемой базы данных генерировать новые данные и решения, вплоть до открытий. Примером является случай успешного испытания жидкостного реактивного двигателя мощностью 5 кН (20000 лошадиных сил), спроектированного ИИ в течение двух недель [1].

Вместе с тем история развития человечества доказала неоднократно, что все новые положительные технические достижения сразу же находят применение для достижения преступных целей нечистых на руку личностей. При этом надо отметить, что уровень технической подготовки преступников для успешной работы с ИИ и соответствия процессу развития постоянно растет.

Немаловажно, что для подготовки и эффективного противодействия в реализации преступных замыслов с применением ИИ актуальным является вопрос рассмотрения потенциальных возможностей ИИ и направлений его преступного применения, который рассмотрим ниже.

Во-первых, несмотря на то, что большинство реально функционирующих нейросетей способны эффективно выполнять достаточно ограниченный набор функций, последовательно и структурировано обработанная несколькими нейросетями аудио- и видеоинформация может стать основой создания так называемых глубоких фейков или дипфейков (*deepfake*) ничего не имеющих общего с действительностью.

Используя дипфейки, преступники способны вводить в заблуждение и шантажировать размещением дезинформации в СМИ и соцсетях граждан (в том числе имеющих высокий социальный статус и/или финансовые накопления) для получения от них финансовых средств или выгодных услуг. Изменение голоса и даже трансляция лица известного потенциальной жертве человека в режиме реального времени, замена лица в видеоряде не являются сложной задачей, а поэтому количество таких преступлений будет расти.

Во-вторых, задачи, связанные с безопасностью хранения и передачи данных (в том числе составляющих тайну различного уровня), устойчивостью и бесперебойностью работы локальных и сетевых программ, лежащих в основе работы крупных организаций, всегда были и останутся в приоритете, поскольку

безопасности никогда не бывает много: потенциальные уязвимости при запуске новых технических решений будут априори, и устранить их полностью невозможно. Это подтверждают недавние нашумевшие случаи преодоления защиты систем безопасности: взлом хакерами *LockBit* Федеральной резервной системы США и кражи 33 терабайт данных с целью получить за них выкуп в сумме $50000 [2]; случай блокирования сервиса доставки грузов СДЭК [3].

Таким образом, следует ожидать применения преступниками ИИ для определения уязвимостей программного обеспечения крупных организаций и состоятельных граждан, а также написания вредоносных кодов с применением ИИ для создания новых угроз.

С увеличением использования ИИ в различных отраслях, включая медицину, финансы и транспорт, растет риск атак на эти системы. Злоумышленники могут пытаться манипулировать полученными данными для обучения моделей ИИ, используемых в этих отраслях, чтобы повлиять на их поведение (изменять медицинские диагнозы или финансовые прогнозы и так далее). Внедрение вредоносных данных в тренировочные наборы данных, используемых при подготовке специалистов на основе систем машинного обучения, приведет к неправильной работе системы ИИ, управляющего системой организации, что потребует значительного времени, сил и средств для поиска и устранения. Зачастую выплата выкупа является более дешевой альтернативой работе квалифицированных специалистов.

С ростом количества подключенных устройств (умные дома, автомобили, устройства для здоровья) увеличивается и количество уязвимостей для атак. Устройства *IoT* производители стараются делать достаточно простыми и дешевыми, с ограниченным ресурсом памяти и вычислительных мощностей, а потому не оснащенными эффективными программами защиты. Такие устройства часто имеют уязвимости, которые могут быть использованы для атак ИИ с целью вымогательства денег у владельцев.

Финансовые ограничения со стороны различных организаций и даже государств сделали популярными финансовые операции, связанные с блокчейном и криптовалютами, что автоматически повышает их привлекательность со стороны преступного мира для получения финансовых выгод. Преступники для выведения средств могут использовать ИИ при определении недостатков в смарт-контрактах, создании фальшивых *ICO* (*Initial Coin Offerings*) для обмана инвесторов или проведении атак на криптовалютные биржи.

Виртуальная и дополненная реальность становятся все более популярными в играх, образовании и других сферах, что открывает новые возможности для атак на базы данных (в том числе персональных) и устройства пользователей. Внедрение вредоносного кода с помощью ИИ в *VR / AR*-приложения позволит проводить кражи виртуального имущества в компьютерных сетевых играх, на покупку которого люди тратят достаточно большие суммы реальных денег.

С увеличением зависимости от автоматизированных систем управления экологическими ресурсами (водоснабжение, энергетика) атаки на эти системы могут иметь катастрофические последствия. Применение преступниками ИИ возможно с целью выкупа или манипуляции системами управления ресурсообеспечения для создания дефицита ресурсов или загрязнения территорий.

Немаловажным является тот факт, что обнаружение преступников, использующих ИИ для своих преступных действий, будет являться достаточно затруднительным, поскольку действия ИИ высокого уровня как программы, размещенной на удаленных серверах общественного пользования, пока детально не отслеживаются. В то же время для противодействия противоправным действиям с применением ИИ эффективным может быть только применение другого ИИ, способного с высокой долей вероятности распознать преступный алгоритм. В связи с этим в будущем следует ожидать «битву технологий» для достижения двух противоположных целей, которые ставят перед собой преступники и правоохранительные органы, контролирующие работу конкретных ИИ.

С учетом скорости развития технологии искусственного интеллекта и появлением новых более точных и высокопроизводительных архитектур (*KAN*) для организации работ по применению ИИ в целях противодействия противоправным действиям необходимо включаться в данную работу немедленно. По прогнозам профессора Университета Луисвилля Романа Ямпольского, общий ИИ, представляющий для человека угрозу и обладающий способностями, существенно превосходящими способности человека, будет функционален уже в 2026 году [4].

Рассмотренные направления основаны на текущих тенденциях в развитии технологий и наблюдаемых уязвимостях. Полностью и однозначно предсказать будущее невозможно, но понимание потенциальных угроз позволяет сделать вывод о необходимости постоянного контроля за функционированием ИИ со стороны многих структур (в том числе со стороны правоохранительных органов) и может помочь в разработке мер защиты и превентивных стратегий.

Список источников

1. ИИ за две недели с нуля спроектировал ракетный двигатель. URL: https://3dnews.ru/1106988/sproektirovanniy-ii-s-nulya-raketniy-dvigatel-zarabotal-s-pervoy-popitki-na-razrabotku-ushlo-dve-nedeli (дата обращения: 15.06.2024). (In Russ.)

2. Хакеры из LockBit заявили о взломе Федеральной резервной системы США. URL: https://aif.ru/society/hakery-iz-lockbit-zayavili-o-vzlome-federalnoy-rezervnoy-sistemy-ssha (дата обращения: 15.06.2024). (In Russ.)

3. СДЭК подверглась атаке хакеров. URL: https://realnoevremya.ru/news/309734-sdek-podverglas-atake-hakerov (дата обращения: 15.06.2024). (In Russ.)

4. Осталось два года. Ученый предупредил о смертельной угрозе человечеству. URL: https://ria.ru/20240603/ugroza-1950074180.html (дата обращения: 15.06.2024). (In Russ.)

References

1. AI designed a rocket engine from scratch in two weeks. URL: https://3dnews.ru/1106988/sproektirovanniy-ii-s-nulya-raketniy-dvigatel-zarabotal-s-pervoy-popitki-na-razrabotku-ushlo-dve-nedeli (accessed 15.06.2024). (In Russ.)

2. Hackers from LockBit claimed to have hacked the US Federal Reserve System. URL: https://aif.ru/society/hakery-iz-lockbit-zayavili-o-vzlome-federalnoy-rezervnoy-sistemy-ssha (accessed 15.06.2024). (In Russ.)

3. SDEK was attacked by hackers. URL: https://realnoevremya.ru/news/309734-sdek-podverglas-atake-hakerov (accessed 15.06.2024). (In Russ.)

4. Two years left. The scientist warned of a deadly threat to humanity. URL: https://ria.ru/20240603/ugroza-1950074180.html (accessed 15.06.2024). (In Russ.)

Информация об авторе | Information about the author

Д. В. Сафронов — кандидат технических наук, доцент, профессор кафедры управления Нижегородской академии МВД России; доцент кафедры инфокоммуникационных и профессиональных дисциплин Волго-Вятского филиала Московского технического университета связи и информатики

D. V. Safronov — Candidate of Sciences (Technical), Docent, Professor of the Department of Management, Nizhny Novgorod Academy of the Ministry of Internal Affairs of Russia; Professor of the Department of Infocommunication and Professional Disciplines Volga-Vyatky branch of the Moscow Technical University of Communication and Informatics

Статья поступила в редакцию 23.06.2024; одобрена после рецензирования 05.07.2024; принята к публикации 24.09.2024.

The article was submitted 23.06.2024; approved after reviewing 05.07.2024; accepted for publication 24.09.2024.