Сопин Кирилл Юрьевич,
Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С.М. Штеменко

Баданин Константин Александрович,
Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С.М. Штеменко

Акишин Андрей Владимирович,
Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С.М. Штеменко

Додух Владимир Сергеевич,
Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С.М. Штеменко

Семёнов Алексей Дмитриевич
Краснодарское высшее военное орденов Жукова и Октябрьской Революции Краснознаменное училище имени генерала армии С.М. Штеменко

УГРОЗЫ КИБЕРБЕЗОПАСНОСТИ, СВЯЗАННЫЕ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ПОСТРОЕНИИ СИСТЕМ ЗАЩИТЫ ИНФОРМАЦИИ

Аннотация: однако, по мере того как технологии ИИ усложнялись и при этом становились доступнее для широких масс, усложнялся и ландшафт киберугроз, как для каждого пользователя в отдельности, так и для целых организаций. Далее мы рассмотрим наиболее распространенные киберугрозы для организаций, связанные с искусственным интеллектом.

Цель статьи: – рассмотреть современные угрозы от современного искусственного интеллекта.

Ключевые слова: искусственный интеллект, нейронные сети, дипфейк, социальная инженерия ChatGBT.

Введение

Беспрецедентно быстрое развитие технологий ИИ существенно изменило многие стороны жизни современного человека, привнеся в нее удобство и открыв практически безграничные возможности для работы, обучения и творчества.

Искусственный интеллект (далее – ИИ) – совокупность технологических решений, позволяющих машинам выполнять задачи, для которых обычно требуется человеческий интеллект. В частности, обучение, рассуждение и поиск оптимального решения проблем.

Нейронные сети – математическая модель, воплощенная программно или технически. Воспроизводит принцип работы нейронных связей человека. Понятие не тождественно искусственному интеллекту и является его составляющей частью. Возможность обучаться – важнейшее преимущество нейросетей.

Основная часть

Угроза вредоносного ПО на основе ИИ

Сегодня технологии искусственного интеллекта используются для разработки сложных вредоносных программ, которые в свою очередь могут изучать особенности работы информационной системы организации и ловко обходить традиционные меры безопасности. Также киберпреступники могут использовать ИИ для автоматизации процессов поиска и выявления ИБ-уязвимостей и целевых атак.

Тревожная тенденция заключается в возможности ИИ генерировать огромное количество вариантов вредоносного ПО и перегружать системы безопасности – и ИБ-специалистов, в частности. При этом, искусственный интеллект как инструмент не требует глубоких знаний в области информационных технологий, что значительно снижает порог входа для киберпреступников – и это тоже будет способствовать увеличению числа угроз.

Вредоносное ПО, созданное с помощью искусственного интеллекта, обладает рядом особенностей:

1. Способность имитировать известные семейства вредоносных программ – и с очень высокой точностью. Подражание может ввести в заблуждение службы безопасности, затрудняя выявление угрозы.

2. Полиморфизм. Вредоносное ПО на основе ИИ может автоматически изменять свой код при каждой репликации или заражении. Непрерывные «мутации» затрудняют распознавание и блокировку вредоносного ПО традиционными методами обнаружения на основе сигнатур.

3. Адаптация в реальном времени. Вредоносное ПО на основе ИИ способно в реальном времени адаптировать свое поведение под условия среды.

Еще одно важное обстоятельство, которое стоит учитывать при организации системы информационной безопасности, это относительная новизна атак с использованием ИИ. Согласно последним опросам, у большинства организаций все еще не разработаны планы реагирования на инциденты, связанные с кибербезопасностью и затрагивающие системы ИИ.

Дипфейки на основе ИИ

Как показывает практика, даже самая устойчивая и надежная система информационной безопасности будет неэффективной, если сотрудники организации не осведомлены о правилах безопасного поведения в сети.

С развитием технологий искусственного интеллекта широкое распространение получили дипфейки – и их количество ежедневно растет.

Дипфейк (англ. deepfake от deep learning «глубинное обучение» + fake «подделка») – методика синтеза изображения или голоса, основанная на искусственном интеллекте.

Используя эту ИИ-технологию, киберпреступники могут создавать убедительные видео, изображения или записи голоса ключевых лиц.

Угроза стала настолько очевидной, что в сентябре 2024 года в Госдуму РФ внесли законопроект об уголовной ответственности до шести лет лишения свободы за совершение преступлений с использованием дипфейков.

Согласно данным Human Constanta, в период с 2019 по 2023 год число зафиксированных дипфейковых видео увеличилось на 550%. При этом важно понимать: масштабы атак с использованием дипфейков могут быть значительно шире указанной цифры, ведь далеко не каждый факт атаки реально зафиксировать.

Что важно знать? Технология дипфейков может представлять угрозу для отдельных пользователей, для организаций и даже для национальной безопасности целых государств – особенно в условиях текущей геополитической ситуации.

Цели атак с использованием дипфейков могут быть разными: распространение дезинформации, шантаж, разжигание межнациональной розни, нагнетание паники среди

общественных масс, создание компрометирующих изображений и видео, хищение конфиденциальных данных, выдача себя за высокопоставленное лицо с целью получения финансовой выгоды и проч.

Обнаружение дипфейков сегодня – сложная задача. Алгоритмы ИИ постоянно улучшают способность генерировать высокореалистичный контент, что подчеркивает необходимость обучения сотрудников организаций способам распознавания подделок, а также внедрения программных средств для борьбы с поддельным контентом.

Социальная инженерия с использованием ИИ

Методы социальной инженерии уже давно используются преступниками, а с широким распространением ИИ-технологий кибермошенничество вышло на новый, беспрецедентно высокий уровень.

Социальная инженерия – в контексте информационной безопасности – психологическое манипулирование людьми с целью совершения ими определенных действий или разглашения конфиденциальной информации.

Алгоритмы ИИ могут собирать и обрабатывать огромные объемы данных пользователей из социальных сетей, новостных ресурсов, персональных веб-сайтов и использовать полученную информацию для создания целевых фишинговых атак, то есть направленных на конкретного человека.

Процент успешности атак может быть очень высоким, поскольку ИИ может автоматизировать процесс создания убедительных сообщений, подкрепленных личной информацией о пользователях. При этом, чат-боты или голосовые помощники на базе ИИ могут выдавать себя за доверенных лиц или организации, что затрудняет жертвам выявление мошеннических действий.

Следует учитывать и текущую ситуацию в России и в мире, связанную с утечками высокочувствительных персональных данных. Так, по оценкам «Сбера», в открытый доступ попали персональные данные 90% взрослых россиян.

Важно! Новые технологии опережают законодательство. Многих понятий, о которых мы говорили выше, нормативно еще не существует, поэтому наложение ответственности за правонарушения с их использованием может быть проблематично.

Учитывая обстоятельства, сегодня значительно проще и финансово выгоднее развернуть устойчивую и надежную систему информационной безопасности, чем столкнуться с нарушением, найти виновного, в судебном порядке доказать вину и взыскать компенсацию.

Ограничения и вызовы при использовании ИИ в кибербезопасности

ИИ-инструменты могут содержать собственные уязвимости

Как правило, жизненный цикл инструментов ИИ состоит из 4 этапов: проектирование, разработка, развертывание, обслуживание. На каждом из этих этапов в них могут проникнуть критические уязвимости, которые впоследствии могут быть использованы киберкриминальным группами для обхода механизмов защиты и внедрения вредоносного ПО в информационную систему организации.

Так, например, слабые или неправильно реализованные механизмы аутентификации и авторизации уже на этапах проектирования и разработки могут привести к утечкам информации и несанкционированным изменениям внутри системы.

На этапе разработки злоумышленники могут использовать ошибки кодирования для выполнения произвольного кода, манипулирования моделями ИИ или получения несанкционированного доступа к системным ресурсам.

ИИ-инструменты сложно внедрять и настраивать

Внедрение и настройка системы защиты данных с использованием технологий ИИ требует редких навыков на стыке кибербезопасности и программирования. Помимо этого,

применении ИИ-моделей подразумевает длительный процесс обучения с учетом необходимости подстройки под бизнес-процессы, уникальные для каждой сферы деятельности.

В качестве примера трудностей, с которыми можно столкнуться при внедрении ИИ-инструментов, можно привести настройку правил безопасности для программных комплексов, разработанных для автоматического мониторинга событий безопасности и анализа на предмет нарушений и подозрительной активности. Большое количество выявленных инцидентов указывает, что антивирусные политики настроены недостаточно хорошо. А большое количество ложных сработок говорит о том, что некорректно работают правила, определяющие подозрительную активность. И то, и другое – плохо.

Проблемы безопасности конфиденциальности данных

Исходя из информации, изложенной выше, атаки злоумышленников могут быть направлены не только на информационную систему, но и на сами ИИ-инструменты. При этом процесс обучения ИИ-моделей подразумевает сбор, хранение и обработку больших объемов данных разной степени чувствительности, в том числе информацию об ИТ-активах, бизнес-процессах, архитектуре информационной системы, данные для авторизации пользователей и проч.

Все это увеличивает риски утечек, утраты конфиденциальности и злоупотребления данными.

Примеры утечки информации по вине ИИ

Сотрудники Samsung стали применять ChatGPT в своей работе в середине марта. Однако, по словам источника в корпорации, специалисты допустили ряд ошибок. В частности, в одном из случаев инженер ввел в строку ChatGPT исходный код, связанный с полупроводниковым оборудованием. В другом случае еще один работник также решил поделиться секретным исходным кодом с чат-ботом – сотрудник Samsung хотел упростить его проверку, но в конце концов корпоративные данные были потеряны.

Третья утечка произошла в тот момент, когда специалист компании при помощи ChatGPT решил создать протокол встречи. Во всех трех случаях скрытая информация от Samsung стала частью базы знаний искусственного интеллекта. Корпорация уже предупредила своих сотрудников, что с чат-ботами нельзя делиться конфиденциальной информацией, а также любыми данными, которые могут навредить Samsung. Чтобы предотвратить повторение подобных инцидентов, в организации также приняли ряд мер.

24 марта 2023 года компания OpenAI объявила, что 20 марта ChatGPT был временно отключен из-за обнаружения бага, который позволял некоторым пользователям видеть личные данные других пользователей, включая последние четыре цифры номера кредитной карты и срок ее действия. Утечка коснулась части подписчиков платного плана “ChatGPT Plus” (примерно 1.2% от общего числа членов).

Кроме того, было объявлено о существовании другого бага, который приводил к отображению чата других пользователей в истории чата.

В ответ на это 31 марта 2023 года итальянский орган по защите данных выдал OpenAI предписание о введении временных ограничений на обработку данных итальянских пользователей, поскольку ChatGPT собирал и хранил личные данные без законного основания. В результате OpenAI заблокировала доступ к ChatGPT из Италии. Блокировка была снята 28 апреля того же года после того, как OpenAI улучшила обработку личных данных.

Выводы

Как видно, ИИ может решать довольно большой набор задач кибербезопасности – и в перспективе этот список будет только увеличиваться.

При организации системы информационной безопасности важно закладывать экспоненциальный рост возможностей искусственного интеллекта и машинного обучения, а вместе с ним – рост числа угроз.

Список литературы:

1. Применение нейронных сетей для обнаружения сетевых атак / В. В. Гладков, А. Г. Тараскина, С. И. Кузнецов [и др.] // Региональная информатика и информационная безопасность, Санкт-Петербург, 01–03 ноября 2017 года. Том Выпуск 4. – Санкт-Петербург: Санкт-Петербургское Общество информатики, вычислительной техники, систем связи и управления, 2017. – С. 73-74. – EDN UUJNKC.

2. Алашеев, В. В. Взгляды США на разработку доктрины информационного воздействия в киберпространстве / В. В. Алашеев, А. В. Акишин, М. Н. Чеснаков // Проблемы технического обеспечения войск в современных условиях: Труды III Межвузовской научно-практической конференции, Санкт-Петербург, 16 февраля 2018 года. Том 1. – Санкт-Петербург: ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ КАЗЕННОЕ ВОЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ "ВОЕННАЯ АКАДЕМИЯ СВЯЗИ ИМЕНИ МАРШАЛА СОВЕТСКОГО СОЮЗА С. М. БУДЕННОГО" МИНИСТЕРСТВА ОБОРОНЫ РОССИЙСКОЙ ФЕДЕРАЦИИ, 2018. – С. 67-71. – EDN XMGZPF.

3. Применение нейронных сетей при построении средств защиты информации / С. П. Ушаков, А. В. Акишин, Н. Н. Енин [и др.] // Проблемы научно-практической деятельности. Поиск и выбор перспективных решений: сборник статей VI международной научной конференции, Вологда, 25 февраля 2025 года. – Санкт-Петербург: Общество с ограниченной ответственностью «Международный институт перспективных исследований имени Ломоносова», 2025. – С. 32-39. – EDN ERREXB.

4. Применение искусственного интеллекта для обнаружения атак на веб-приложения / А. В. Акишин, Н. Н. Енин, К. Ю. Сопин [и др.] // Научные инструменты и механизмы перспективного инновационного развития общества: сборник статей XX международной научной конференции, Санкт-Петербург, 08 февраля 2025 года. – Санкт-Петербург: Общество с ограниченной ответственностью «Международный институт перспективных исследований имени Ломоносова», 2025. – С. 17-25. – EDN WMPULQ.