УДК 004.8

Яковчиц В. В.
студент 3-го курса,
Филиппова Т. В.
ст. преп.
Брестский государственный технический университет, г. Брест, Беларусь

ВЛИЯНИЕ ГЕНЕРАТИВНОГО ИИ НА КИБЕРБЕЗОПАСНОСТЬ

В настоящий момент сложно представить существование информационных технологий без развития искусственного интеллекта (ИИ). Ведь ИИ сегодня помогает автоматизировать трудоёмкие процессы, адаптировать процессы принятия решений, ускорять процессы обучение. Однако такое обширное внедрение нейросетей бросает вызов кибербезопасности. С чем сегодня пытаются бороться многие эксперты и компании, предотвращая атаки со стороны вредоносного ПО. Для изучения данной темы следует изучить следующие ключевые понятия.

Генеративный искусственный интеллект — это тип системы искусственного интеллекта (ИИ), способной синтезировать текст, изображения или комбинированный медиаконтент в ответ на подсказки. Генеративный ИИ использует генеративные модели, такие как большие языковые модели.

Кибербезопасность — это защита компьютеров, сетей, программных приложений, критически важных систем и данных от потенциальных цифровых угроз.

Генеративный искусственный интеллект становится революционной силой в сфере кибербезопасности, кардинально меняя подходы к прогнозированию, выявлению и нейтрализации угроз. В основе этой технологии лежат алгоритмы машинного обучения, включая генеративно-состязательные сети (GAN), которые способны моделировать кибератаки и генерировать защитные механизмы.

Благодаря возможности создавать синтетические данные, практически неотличимые от реальных, системы безопасности получают способность динамически адаптироваться к новым типам угроз. Обучение таких ИИ-моделей позволяет им глубже анализировать особенности данных, выявляя даже малозаметные аномалии и шаблоны злонамеренной активности, которые часто остаются незамеченными при использовании классических методов. Это открывает путь к опережающему противодействию киберрискам [1].

Кибербезопасность является одним из наиболее важных вариантов использования генеративного ИИ. Использование генеративного ИИ в сфере кибербезопасности оказывает на неё как положительный эффект, так и отрицательный. С одной стороны это инструмент, который позволяет предотвращать и смягчать риск киберпреступности, с другой стороны это так же серьёзный инструмент в руках киберпреступников, который позволяет им совершать данные преступления.

Генеративный ИИ, как ChatGPT и производные от него в настоящее время научились имитировать человеческий стиль общения на уровне, который сложно отличить от реального. Это позволяет злоумышленникам создавать различного рода фишинговые атаки, которые обходят традиционные системы защиты. ИИ анализирует открытые данные жертвы (соцсети, корпоративные рассылки) и генерирует письма с упоминанием личных деталей. Например:

«*Добрый день, [Имя]! Напоминаем, что ваш заказ №5487 на сайте на таком-то готов к выдаче. Для подтверждения перейдите по ссылке...».*

Ссылка ведет на поддельную страницу, крадущую логины и пароли.

Как пример в 2023 году киберпреступники использовали ChatGPT для генерации писем, маскирующихся под уведомления от службы доставки DHL, а данные электронные письма в свою очередь содержали ссылки на поддельные страницы для кражи данных [2].

ИИ не только упрощает создание вредоносного контента, но и позволяет полностью автоматизировать данный процесс. Злоумышленники используют искусственный интеллект для автоматизации всех этапов от разведки вплоть до эксплуатации уязвимостей, что в свою

очередь делает атаки быстрее, масштабнее и сложнее для обнаружения. Для того чтобы автоматизировать атаки ИИ-боты сканируют интернет в поисках уязвимостей различных систем, открытых портов и утечек данных, затем модели вроде WormGPT (неэтичный аналог ChatGPT) пишут код для эксплуатации уязвимостей, например в веб приложения или API (программный интерфейс). Следующий шаг заключается в том, что ИИ анализирует реакцию защитных систем, например срабатывание WAF (Файрвол веб-приложений) и мгновенно меняет вектор атак. В последующем Нейросеть координирует действия ботнетов (Сеть компьютеров, зараженная вредоносным ПО), распределяя нагрузки на цели и имитируя поведение реальных пользователей.

Например, в 2023 году ботнет Mantis стал одним из самых опасных примеров использования генеративного ИИ для проведения DDoS-атак. В отличие от традиционных атак, которые просто «заливают» цель трафиком, Mantis применял ИИ для анализа инфраструктуры жертвы и точечного удара по слабым местам. Mantis — это ботнет, состоящий из тысяч взломанных устройств, а его уникальность — в интеграции генеративного ИИ, который управлял атаками в реальном времени. По данным Cloudflare, пиковая мощность атак достигала 26 млн запросов в секунду (RPS) — это в 5 раз выше, чем у обычных ботнетов. В основном атаковал банки, госучреждения и игровые сервисы [3].

Помимо угроз генеративный искусственный интеллект также может приносить и пользу. Способность ИИ обрабатывать большие объёмы данных в реальном времени, находить паттерны и автоматизировать ответные реакции сокращает время от обнаружения угрозы до её нейтрализации с недель до нескольких секунд.

Современные программы и сервисы генерируют терабайты сетевого трафика, логов и данных. Генеративный ИИ способен анализировать и структурировать данную информацию, выявляя аномалии, которые не способен заметить человек. ИИ способен обнаруживать подозрительные IP-адреса, которые сканируют порты или пытаются подключиться к закрытым сервисам или выстраивать алгоритмы, отслеживающие отклонения от baseline (типичных норм поведения). То есть если сотрудник полезет скачивать корпоративные данные в три часа ночи, то нейросеть сразу идентифицирует это действие как отклонение от нормы поведения и сразу же предпримет меры по безопасности. В настоящие время разрабатывается большое количество инструментов для борьбы с вредоносными ИИ, такие как Splunk AI, которая способна анализировать данные в режиме реального времени, строя графы связей между событиями [4].

Стоит отметить, что одной из основных проблем генеративного искусственного интеллекта является нехватка данных о редких или сложных атаках для обучения. Однако и эту проблему удалось решить с помощью ИИ. Нейросеть решает эту задачу, создавая реалистичные синтетические данные, которые имитируют атаки, поведение злоумышленников и сетевую активность. Это позволяет тренировать алгоритмы без риска для реальной инфраструктуры. Одним из преимуществ данного метода является то, что синтетические данные позволяют безопасно проверять устойчивость систем к новым типам аттак. Но при этом возникает вопрос: каким образом генеративному ИИ удаётся создавать синтетические данные. На самом деле всё очень просто. Две нейросети, где одна выступает в роли генератора, а другая в роли дискриминатора соревнуются, создавая данные неотличимые от реальных, т.е. генератор создает поддельный сетевой трафик, а дискриминатор пытается его отличить от настоящего. Благодаря данной модели обучения компании вроде CrowdStrike и Microsoft Defender улучшили точность обнаружения APT (постоянная серьёзная угроза). Или инструмент вроде DeepExploit который может создавать варианты эксплойтов для тренировки антивирусов и систем анализа поведения. Так стоит отметить компанию McAfee, которая использует синтетические эксплойты, чтобы улучшить обнаружение полиморфных вирусов.

Таким образом генеративный искусственный интеллект — это зеркало, отражающее две стороны цифровой эпохи. С одной стороны, он даёт защитникам беспрецедентные инструменты: предугадывает угрозы, автоматизирует рутину и создаёт «цифровые полигоны»

для тренировки систем. С другой — превращает киберпреступность в высокотехнологичный конвейер, где даже новичок может запустить сложную атаку. Этот парадокс ставит человечество перед выбором: либо мы научимся укрощать технологии, либо рискуем погрузиться в бесконечную гонку с алгоритмами, которые учатся быстрее нас.

Успех в этой борьбе зависит не только от алгоритмов, но от этики, образования и глобальной кооперации. Генеративный ИИ не заменит специалистов, но станет их союзником, если сохранится баланс между автономией и человеческим участием.

Список источников

1. Влияние генеративного ИИ на кибербезопасность [Электронный ресурс] // Securelist : [сайт]. [2025]. URL: https://securelist.ru/story-of-the-year-2023-ai-impact-on-cybersecurity/108558/ (дата обращения: 18.03.2025).

2. What Is Generative AI in Cybersecurity? [Electronic resource] // Palo Alto Networks : [сайт]. URL: https://www.paloaltonetworks.com/cyberpedia/generative-ai-in-cybersecurity (date of treatment: 18.03.2025).

3. Намиот Д. Е., Ильюшин Е. А. О киберрисках генеративного Искусственного Интеллекта [Электоронный ресурс] // International Journal of Open Information Technologies. 2024. Vol. 12. No. 10. URL: https://cyberleninka.ru/article/n/o-kiberriskah-generativnogo-iskusstvennogo-intellekta/viewer (дата обращения: 20.03.2025).

4. Машинное обучение и генеративный ИИ [Электоронный ресурс] // AM Медиа : [сайт]. [2025]. URL: https://live.anti-malware.ru/am-live/primenenie-ii-v-ib-2/ (дата обращения: 20.03.2025).