Использование методов машинного обучения для повышения эффективности систем противодействия многоэтапных кибератак

О.Б. Лебедев, Д.Д. Левченко, Р.И. Черкасов

МИРЭА – Российский технологический университет, Москва

Аннотация: Статья посвящена анализу влияния технологий искусственного интеллекта (ИИ) и машинного обучения на развитие, трансформацию киберугроз и создание высокоэффективных систем киберзащиты. Рассматриваются ключевые направления эволюции ИИ, включая ориентированные на данные, модели, приложения и человека подходы, и их роль в формировании как защитных, так и наступательных возможностей. Показано, что злоумышленники активно используют ИИ для автоматизации разведки, персонализации атак, обхода систем обнаружения и проведения сложных многоэтапных кибератак. Анализируются основные типы воздействий на системы машинного обучения: манипуляция данными, состязательные примеры, атаки на модели и их инфраструктуру. Представлены современные методы защиты, повышающие робастность моделей, защищённость данных и устойчивость ИИ-систем. Выдвигается идея о необходимости интеграции интеллектуальных подходов на всех уровнях архитектуры киберзащиты и разработки доверенных, интерпретируемых и устойчивых моделей машинного обучения для противодействия новым классам угроз.

Ключевые слова: искусственный интеллект, кибербезопасность, кибератака, машинное обучение, инновация защищенность, информация, защищенность.

Введение

Стремительное развитие технологий искусственного интеллекта, в целом, и машинного обучения (МО), в частности, привело к их широкому внедрению их в системы кибербезопасности. Эти технологические решения показали высокий уровень эффективности при обработке больших объемов разнородных данных и решении ряда фундаментальных задач защиты информации, включая обнаружение вторжений, управление уязвимостями, оценку уровня защищённости, мониторинг событий безопасности и повышение точности классификации сетевых аномалий [1,2].

Одновременно с этим наблюдается эволюция методов проведения кибератак. Злоумышленники активно адаптируют технологии ИИ для автоматизации разведки, генерации вредоносного кода, обхода механизмов обнаружения и реализации целевых атак. Появление так называемого

наступательного ИИ существенно усложняет применение традиционных средств защиты, основанных на сигнатурах и статических правилах, поскольку такие атаки характеризуются высокой скоростью, масштабируемостью, а входящие в их состав программные пакеты способны к самостоятельной адаптации [1].

В результате возрастает потребность в создании интеллектуальных систем противодействия сложным и многоэтапным кибератакам, способных обучаться, обобщать зависимости и выявлять скрытые закономерности в сетевой активности. Потенциал методов машинного обучения в адаптивной обработке данных открывает возможности для формирования нового поколения систем киберзащиты, способных противостоять масштабируемым, индивидуализированным и человекоподобным атакам [1,2].

1. Классификация и анализ современных инноваций в области искусственного интеллекта

Современное развитие технологий ИИ характеризуется их переходом от локальных алгоритмических решений к комплексным интеллектуальным платформам, ориентированным на полную интеграцию данных, моделей, приложений и аспектов взаимодействия человека с системой. Инновации в данной области могут быть систематизированы в рамках четырёх основных категорий [2,3]:

- ориентированный на данные ИИ;
- модельно-ориентированный ИИ;
- ориентированный на приложения ИИ;
- человеко-ориентированный ИИ.

Ключевой особенностью ИИ, ориентированного на данные, является смещение исследовательского акцента с совершенствования архитектуры моделей на повышение качества данных, обеспечивающих их обучение.

Данное направление ИИ, включает такие инновационные инструменты, как синтетические данные, графы знаний, средства интеллектуальной разметки и автоматизированной аннотации [3,4].

Синтетические данные, генерируемые методами статистического моделирования, генеративными состязательными сетями или имитационными моделями сложных процессов и обеспечивают [5]:

– исключение использования конфиденциальной информации, заменяемой синтетическими аналогами;

– снижение стоимости и трудозатрат при формировании обучающих выборок;

– повышение масштабируемости ML-моделей за счёт увеличения объёма данных;

– улучшение качества обучения при условии предотвращения переобучения.

В условиях роста требований к приватности и безопасности данных синтетические выборки становятся одним из ключевых инструментов разработки современных интеллектуальных систем.

Модельно-ориентированный ИИ объединяет инновации, направленные на развитие свойств моделей: робастности, интерпретируемости, адаптивности и способности к обобщению [3-5].

К числу перспективных технологий относятся:

– причинно-следственный ИИ, основанный на моделировании причинно-следственных зависимостей;

– композитный ИИ, интегрирующий различные типы моделей;

– генеративный ИИ;

– глубокие нейронные сети;

– базовые модели, способные решать широкий спектр задач после минимальной адаптации.

Причинно-следственный ИИ отличают следующие преимущества:

– сокращение объёма требуемых данных благодаря включению предметных знаний;

– расширение автономности и качества принятия решений;

– повышение объяснимости моделей, за счёт явного представления причинно-следственных структур;

– устойчивость к изменению внешних условий и сценариев функционирования.

Эти качества особенно значимы для задач анализа поведения нарушителя, прогнозирования развития атаки и построения устойчивых моделей для динамических систем киберзащиты.

В ориентированном на приложения ИИ сосредоточены практические реализации ИИ, встроенные в корпоративные, облачные, инфраструктурные и автономные системы [4,5]. Основные его области:

– инженерия и операционные процессы ИИ;

– аналитика принятия решений;

– облачные сервисы ИИ;

– обработка естественного языка;

– компьютерное зрение;

– автономные транспортные платформы;

– пограничный ИИ.

Пограничный ИИ обеспечивает выполнение вычислений непосредственно на устройствах периферии, что позволяет: минимизировать задержки при обработке данных; уменьшить трафик между устройствами и облачной инфраструктурой; повышать устойчивость в условиях отсутствия сетевой доступности; повышать точность и скорость аналитики при мониторинге сложных процессов, включая задачи информационной безопасности [6,7]. Фокус данной категории направлен на обеспечение

доверия, безопасности, прозрачности и этичности применения ИИ. К ключевым аспектам которой относятся: управление рисками ИИ; цифровая этика; ответственный ИИ; обеспечение прозрачности и устранение технологических предвзятостей [8-10].

Согласно рекомендациям Европейской комиссии по разработке доверенного ИИ, такие системы должны быть [11,12]:

– легитимными, соответствующими действующему законодательству;

– этическими, учитывающими социальные нормы и ценности;

– робастными, обеспечивающими технологическую устойчивость и безопасность.

В России вопросы стандартизации систем искусственного интеллекта регулируются ТК 164 при Росстандарте. На данный момент уже введен целый ряд стандартов, определяющих требования к доверенным ИИ-системам, их классификации и жизненному циклу (ГОСТ Р 59276–2020, ГОСТ Р 59277–2020 и др.).

Совокупность рассмотренных инновационных направлений демонстрирует, что современный ИИ развивается одновременно по нескольким взаимодополняющим траекториям, формируя основу для создания высокоадаптивных интеллектуальных систем. Однако стремительное развитие ИИ приводит не только к расширению возможностей защитных технологий, но и к эволюции средств наступательного воздействия. Появление генеративных моделей, автономных алгоритмов принятия решений и инструментария для обработки больших данных значительно трансформирует характер современных киберугроз. В этих условиях особую актуальность приобретает анализ тенденций развития кибератак [10].

2. Эволюция кибератак в условиях развития искусственного интеллекта

Развитие технологий (ИИ) и машинного обучения, оказывает существенное влияние не только на средства защиты информации, но и на характер современных киберугроз. Интеллектуальные методы активно внедряются как в системы кибербезопасности, так и в арсенал киберпреступников, что формирует качественно новую конфигурацию «гонки вооружений» в цифровой среде [2].

Одной из ключевых областей применения ИИ является компьютерное зрение, ориентированное на воспроизведение отдельных компонентов системы человеческого зрения и обеспечивающее автоматический анализ изображений и видеопотоков. Данный функционал открывает широкие возможности для решения задач идентификации объектов, мониторинга инфраструктуры, видеоаналитики и ситуационного анализа [2]. Масштаб внедрения технологий компьютерного зрения постоянно растёт: повышение вычислительной мощности и совершенствование моделей приводят к росту точности распознавания и созданию более сложных прикладных систем [6].

Согласно Указу Президента Российской Федерации «О развитии искусственного интеллекта в Российской Федерации» от 10.10.2019 г., одной из приоритетных задач является активное внедрение технологий ИИ в государственном и корпоративном секторах, развитие партнёрства с научными организациями, а также формирование стандартов открытых данных для решения актуальных задач цифровой трансформации. Вместе с тем широкое использование систем ИИ требует учёта взаимовлияния ИИ и кибербезопасности, поскольку компьютерная преступность эволюционирует с опорой на те же самые технологические тренды [10].

Кибератаки в настоящее время рассматриваются как один из наиболее значимых глобальных рисков [5, 8]. Злоумышленники адаптируют методы ИИ, прежде всего машинного обучения, для автоматизации разведки, отбора целей, персонализации атак и обхода средств обнаружения. Технологические

разработки предоставляют киберпреступникам новые возможности благодаря относительной анонимности, отсутствию географических границ, менее жёстким правовым ограничениям и высокой доступности инфраструктурных сервисов [11].

С точки зрения жизненного цикла, большинство кибератак включает следующие базовые этапы:

– поиск потенциальных жертв;

– компрометация целевого ресурса или пользователя;

– распространение в инфраструктуре;

– координированное управление атакой и её последствиями [9, 10].

В таблице 1 представлена обобщённая классификация основных типов кибератак с указанием их назначения и характерных этапов реализации.

Таблица № 1

Основные типы кибератак

№ п/п	Тип атаки	Этапы реализации
1	2	3
1	Повышение уровня автоматизации и скорости компрометации средств нападения.	Ранее уязвимости эксплуатировались после завершения широкомасштабного сканирования. Теперь средства атаки используют уязвимости как часть сканирования, что увеличивает скорость распространения. Инструменты распределенных атак способны более эффективно запускать DDoS-атаки, сканировать потенциальные жертвы и компрометировать уязвимые системы. Функции координации используют легкодоступные общедоступные протоколы связи.

1	2	3
2	Использование уязвимостей глобальной политики безопасности в Интернете.	Один злоумышленник может относительно легко использовать большое количество распределенных систем для проведения разрушительных атак. Инфраструктурные атаки (DoS, черви, DNS и атаки на маршрутизаторы). Повышение сложности инструментов атаки (сложнее обнаружить атаки. Анти-форензика. Анализ часто включает лабораторные испытания и обратный инжиниринг. Динамическое поведение (на основе случайного выбора, предопределенных путей принятия решений или прямого управления злоумышленниками).
3	Увеличение скорости обнаружения уязвимостей.	Злоумышленники могут обнаружить новые уязвимости до того, как поставщики смогут их исправить. Время на исправление становится все меньше.
4	По степени информированности атакующего:	Стратегия «белого ящика»: атакующему известны как входные данные, так и алгоритм, функции, параметры (полное или почти полное знание). Стратегия «серого ящика»: могут быть известны входные данные или алгоритм (частичное знание). Стратегия «черного ящика»: атакующему неизвестно ничего (доступ максимум ко входу и / или выходу обученной модели).
5	По специфике проведения:	Атаки с нарушением политики безопасности системы ИИ. Атаки, влияющие на компоненты системы. Атаки без нарушения политики безопасности системы ИИ – Adversarial attack (влияющие на поступающие данные до попадания в систему).

1	2	3
6	По времени проведения атаки:	В процессе обучения модели (например, отравляющие). В процессе тестирования модели (например, искажающие). В процессе использования (например, исследовательские, состязательные).
7	По способу воздействия:	Казуальные – оказывающие непосредственное влияние на систему ИИ или ее поведение. Исследовательские – не оказывающие влияния на систему ИИ.
8	По специфике результата:	Таргетированные – приведение ответа к определенному виду/значению. Массовые – приведение ответа к любому виду/значению, кроме верного.

К числу наиболее распространённых типов атак относятся: внедрение вредоносного программного обеспечения, фишинговые и целевые фишинговые кампании, программы-вымогатели, атаки отказа в обслуживании, SQL-инъекции, эксплойты нулевого дня, парольные атаки, межсайтовый скриптинг, атаки типа «человек посередине», а также атаки на устройства Интернета вещей. Однако, развитие ИИ позволяет автоматизировать множество этапов защиты: от интеллектуального подбора социальной инженерии до автоматического сканирования уязвимостей и адаптивного управления ботнетами [12].

Помимо использования ИИ в качестве инструмента реализации атак, всё чаще можно встретить инциденты, когда угрозы направлены непосредственно на системы ИИ. Такие атаки призваны исказить результаты работы моделей, снизить их надёжности или извлечение конфиденциальной информации. Основные классы атак на системы ИИ представлены на рисунках 1-5.

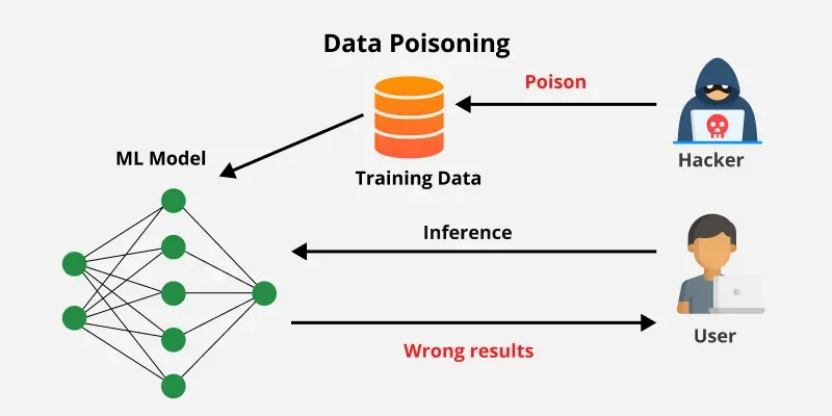

Рис. 1. – Манипуляция данными [1]

На рисунке 1 изображено схематичное представление процесса манипуляции данными: источник легитимных данных, канал их подмены или искажения, формирование обучающей выборки с вредоносными записями и влияние на обученную модель (смещение границ классификации, внедрение скрытых триггеров).

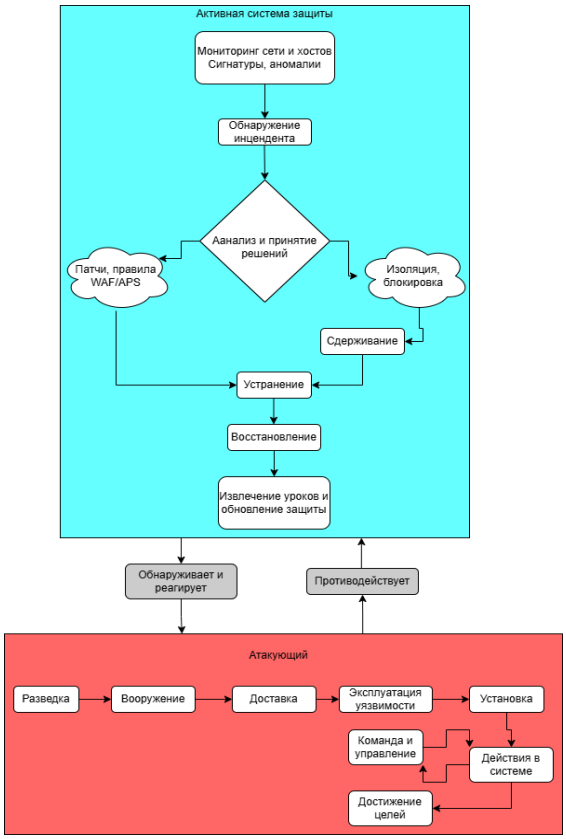

Рис. 2. – Состязательные атаки

Схема на рисунке 2 формализует состязательную кибератаку как систему двух параллельных и взаимозависимых процессов: контура атаки, реализующего последовательность стадий от разведки до достижения цели в рамках парадигмы Cyber Kill Chain, и контура защиты, опирающегося на жизненный цикл реагирования на инциденты [2, 5, 8].

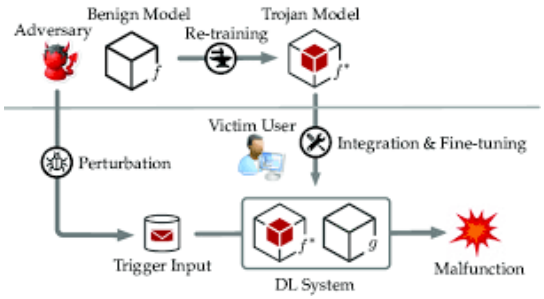

Рис. 3. – Атаки на модели [5]

На рисунке 3 изображена структурная схема переноса обучения при атаке на модель ИИ: базовая (предварительно обученная) модель, этап дообучения на специфической выборке и возможные векторы атаки (компрометация данных дообучения, внедрение бэкдоров, использование уязвимых публичных моделей) [2, 10].

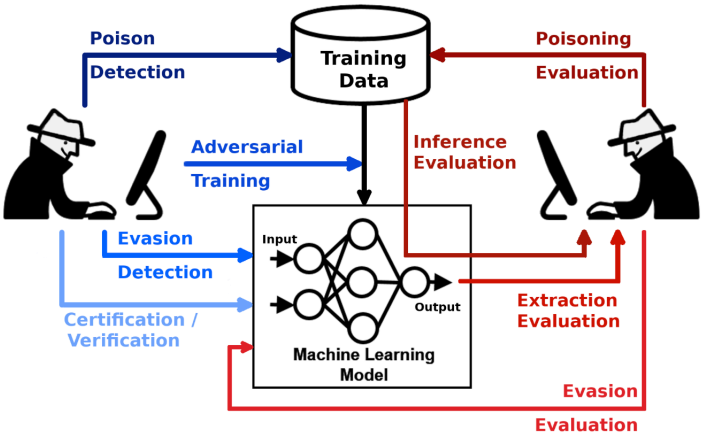

Рис. 4. – Атаки на инфраструктуру и окружение ИИ

Рисунок 4 представляет собой архитектурную схема развертывания ИИ-системы (модель, API, контейнеры, облачная платформа, хранилища моделей) с указанием точек атаки: компрометация API, подмена модели, атаки на контейнеры/оркестраторы, доступ к хранилищам [2, 10].

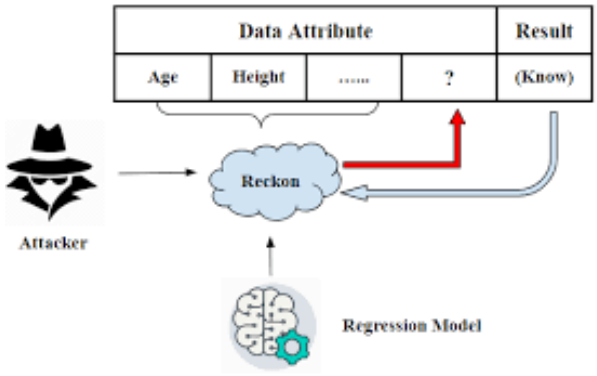

Рис. 5. – Извлечение конфиденциальных данных

Рисунок 5 иллюстрация атак вида «model inversion», «membership inference» и «model extraction»: внешний клиент, обращающийся к модели по API, получает ответ и использует его для восстановления фрагментов обучающих данных или копирования модели [2, 10].

Для минимизации рисков, связанных с атаками на системы искусственного интеллекта, применяются специализированные методы защиты, дополняющие классические средства информационной безопасности. Обобщение основных подходов приведено в таблице 2.

Таблица № 2

Основные типы кибератак

№ п/п	Метод	Объяснение действия
1	2	3
1	Состязательное обучение	Добавление состязательных примеров в обучающую выборку. Недостаток: время на обучение модели

		возрастает многократно.
1	2	3
2	Гауссовское зашумление данных	Добавление состязательных примеров с гауссовским шумом в обучающую выборку.
3	Сглаживание меток	Сглаживание меток / классов. Регуляризация модели в задаче классификации, что делает ее более устойчивой к шумам.
4	Использование ансамблей моделей	Совокупность моделей сложнее обойти.
5	Сжатие признаков	Обнаружение состязательных примеров путем сравнения с результатами контрольной модели, обученной на ограниченном наборе признаков.

Ключевые направления противодействия включают [10-12]:

– обеспечение доверия к данным: верификация источников, контроль целостности и качества, выявление аномалий в обучающих выборках;

– повышение робастности моделей: использование состязательного обучения, специальных архитектур и методов сглаживания;

– защиту инфраструктуры ИИ: применение DevSecOps-подходов, шифрование, журналирование и аудит, сегментация окружения;

– защиту конфиденциальности: применение дифференциальной приватности, ограничение и мониторинг доступа к API моделей, рандомизация ответов и оценка риска утечек.

Защита информационных систем от кибератак носит характер асимметричного противоборства, при котором атакующая сторона зачастую обладает преимуществом за счёт инициативы, возможности гибко выбирать векторы атаки и комбинировать методы воздействия [3, 6]. В течение последних десятилетий подходы к обеспечению безопасности эволюционировали от статической модели, ориентированной на создание «безотказных» систем, к динамическим и адаптивным концепциям,

исходящим из неизбежности сбоев и успешных атак. В рамках такого апостериорного подхода ключевое значение приобретают устойчивость к инцидентам, способность к быстрому восстановлению и снижению последствий нарушений [7, 8].

Дополнительный импульс развитию динамических средств защиты связан с распространением сервис-ориентированных архитектур, облачных вычислений, виртуализации и семантических технологий, позволивших гибко управлять ресурсами и конфигурацией систем. В совокупности с внедрением методов машинного обучения это создаёт предпосылки для построения интеллектуальных систем противодействия кибератакам, включая сложные многоэтапные сценарии.

3. Применение искусственного интеллекта для повышения эффективности киберзащиты

Развитие технологий искусственного интеллекта оказывает существенное влияние на подходы к обеспечению кибербезопасности, позволяя решать широкий спектр задач, связанных с предотвращением атак, обнаружением вторжений, выявлением закономерностей в поведении нарушителей и автоматизацией процессов реагирования. Интеллектуальные методы дают возможность обрабатывать разнородные и масштабные данные, анализировать скрытые зависимости и принимать решения в условиях неполной или быстро изменяющейся информации, что особенно важно при противодействии угрозам, отличающимся высокой динамичностью и сложностью [1, 2].

Применение ИИ в системах защиты связано прежде всего с повышением качества предупреждений о потенциальных атаках. Алгоритмы машинного обучения способны автоматически расставлять приоритеты среди множества событий безопасности, выделяя критические отклонения,

указывающие на подготовку, или начало атаки. С их помощью определяется принадлежность отдельных действий к более крупным многоэтапным угрозам, что позволяет своевременно выявлять стратегию злоумышленника. Такие методы эффективны также при обнаружении следов вредоносной активности как в операционных системах, так и в сетевой инфраструктуре, а также при идентификации атак, основанных на использовании легитимных программных механизмов (так называемых Living off the Land-атак), что представляет особую сложность для традиционных средств обнаружения. Интеллектуальные инструменты обеспечивают и быстрое реагирование на инциденты: изоляцию заражённых узлов, блокировку трафика или автоматическую остановку процессов, что критически важно при атаках программ-вымогателей [5, 7, 11].

Однако, как отмечалось ранее, искусственный интеллект используется не только защитной стороной. Киберпреступники также активно адаптируют эти технологии, что ведёт к появлению новых форм и механизмов атак. Одним из наиболее заметных направлений здесь, является использование ИИ для прогнозирования – от восстановления нажатий клавиш по данным вибрации смартфона до поиска уязвимостей программного обеспечения и предсказания конфиденциальной информации пользователей. Другой важный аспект связан с генерацией контента: алгоритмы глубинного обучения позволяют подделывать изображения, голос и видеозаписи, автоматически подбирать пароли, формировать фишинговые сообщения и модифицировать сетевой трафик для обхода систем обнаружения [2, 7].

Всё более существенным фактором становятся синтетические медиа и дипфейки, которые используются для проведения операций социальной инженерии, выдачи себя за доверенных лиц, фабрикации новостей и создания видимости событий, которые в действительности не происходили. Распространение таких технологий приводит к эффекту «дивиденда лжеца»,

когда граница между истинной и ложной информацией становится менее различимой [10, 12].

Значительное внимание необходимо уделять и атакам, направленным непосредственно на системы машинного обучения. Подобные угрозы, называемые состязательными атаками, предполагают внесение целенаправленных искажений в данные или окружение модели, что приводит к ошибкам классификации или неправильным решениям. Они могут осуществляться как на этапе подготовки обучающих данных (манипуляция обучающих выборок), так и на этапе эксплуатации модели, когда внешний злоумышленник формирует специальные входные данные с минимальными изменениями, незаметными для человека, но критичными для алгоритма. В зависимости от степени осведомлённости злоумышленника о структуре модели выделяют атаки «белого ящика» и «чёрного ящика». Кроме того, всё большее распространение получают способы извлечения информации о модели, включая восстановление её параметров, архитектуры или даже данных, использованных в обучении [2, 4, 6].

Использование ИИ расширяет и возможности проведения разведывательных и поисковых операций. Системы анализируют изображения с большого числа взломанных камер, выявляют паттерны поведения в социальных сетях, выполняют автоматическое аннотирование и суммаризацию текстов из открытых источников, что значительно облегчает сбор данных для последующей атаки. В сочетании с методами оптимизации и моделями коллективного поведения возможно создание распределённых управляемых ботнетов и автоматизированных систем планирования действий.

Особое значение приобретает прогнозная аналитика, обеспечивающая автоматизацию принятия решений при анализе угроз. Методы глубокого байесовского прогнозирования, всплескового анализа, генеративных моделей

с временными ограничениями и нейронных сетей на основе временных графов дают возможность предсказывать развитие атак, оценивать риски и формировать рекомендации по выбору оптимальных контрмер. Полученные в ходе аналитических процедур знания используются для автоматической реализации механизмов защиты, таких как сегментация сетей, моделирование угроз, восстановление инфраструктуры после атак и автоматическая установка обновлений и исправлений [9-12].

Заключение

Проведённое исследование показало, что развитие технологий ИИ оказывает принципиальное влияние на формирование современной экосистемы кибербезопасности. Интеллектуальные методы перестали быть вспомогательным инструментом и превратились в ключевой элемент как защитных механизмов, так и наступательных средств, что приводит к необходимости комплексного анализа роли ИИ в противоборстве сторон. В работе выявлены основные направления использования ИИ злоумышленниками, включая построение адаптивных и труднообнаружимых атак, генерацию синтетического медиаконтента, автоматизацию сбора информации и использование методов предсказательного анализа. Одновременно показано, что системы машинного обучения сами становятся объектом атак, требующих разработки новых принципов их защиты и повышения устойчивости к целенаправленным воздействиям.

Рассмотрение данных аспектов в сопоставлении с современными защитными методами позволило определить ключевые векторы применения ИИ в кибербезопасности: от интеллектуального обнаружения аномалий и корреляции событий до автоматизации реагирования и адаптации моделей к динамически изменяющейся обстановке. Проанализированные исследования в области интеллектуального мониторинга и обнаружения вторжений

демонстрируют возрастающую способность ИИ выявлять сложные, многоэтапные атаки, отличающиеся минимальной сигнатурной выраженностью и высокой степенью маскировки.

Полученные результаты указывают на необходимость переосмысления традиционных средств обеспечения кибербезопасности. В условиях, когда злонамеренные субъекты обладают доступом к тем же интеллектуальным технологиям, что и защитники, возрастает потребность в создании устойчивых, робастных моделей машинного обучения, способных противостоять манипуляциям с данными, состязательным воздействиям и попыткам извлечения информации. Важной задачей становится формирование подходов, обеспечивающих проверяемость, интерпретируемость и доверенность решений, принимаемых системами ИИ, что особенно актуально для критически важных областей.

Таким образом, можно сказать, что ИИ одновременно расширяет возможности киберзащиты и увеличивает «площадь» атаки, существенно усложняя характер информационных угроз. Эффективное развитие систем безопасности в этих условиях возможно только при интеграции ИИ в архитектуру защиты на всех уровнях, при одновременном учёте рисков, связанных с его применением.

Литература

1. Kotenko I., Saenko I., Lauta O., Kribel K., Vasiliev N. Attacks on artificial intelligence systems: classification, the threat model and the approach to protection // Proceedings of the Sixth International Scientific Conference «Intelligent Information Technologies for Industry 2022». Lecture Notes in Networks and Systems, vol. 566. Springer, Cham. 2023. – pp. 293-302.

2. Котенко И. В., Саенко И. Б., Лаута О. С., Васильев Н. А., Садовников В.Е. Подход к обнаружению атак на системы машинного

обучения с использованием генеративно-состязательной сети // Двадцать первая Национальная конференция по искусственному интеллекту с международным участием, 2023. – С. 366-376.

3. Xu H., Ma Y., Liu H.C. et al. Adversarial Attacks and Defenses in Images, Graphs and Text: A Review // International Journal of Automation and Computing, 2020, vol. 17. – pp. 151-178.

4. Ren K., Zheng T., Qin Zh., Liu X. Adversarial Attacks and Defenses in Deep Learning // Engineering, 2020, vol. 6. – pp. 346-360.

5. Rosenberg I., Shabtai A., Elovici Y., Rokach L. Adversarial Machine Learning Attacks and Defense Methods in the Cyber Security Domain // ACM Computing Surveys, 2021, vol. 54, №5. – 36 p.

6. Zhou Sh., Liu Ch., Ye D., Zhu T., Zhou W., Yu Ph.S. Adversarial Attacks and Defenses in Deep Learning: From a Perspective of Cybersecurity // ACM Computing Surveys, 2022, vol. 55, №8. – 39 p.

7. Li Y., Cheng M., Hsieh Ch. -J., Lee Th. C. M. A Review of Adversarial Attack and Defense for Classification Methods // The American Statistician, 2022, vol. 76, №4. – pp. 329-345.

8. Федорченко Е.В., Федорченко А.В., Новикова Е.С., Саенко И.Б. Оценивание защищенности информационных систем на основе графовой модели эксплойтов // Вопросы кибербезопасности, 2023. №3 (55). – С. 23-36.

9. Лебедев В.Б., Лебедев О.Б. Синтез разделяющих функций при распознавании образов с использованием композитной архитектуры многоагентной системы // Материалы X Международной научно-технической конференции «Технологии разработки информационных систем». – Таганрог: Южный федеральный университет, 2020. – С. 194-200.

10. Котенко И.В., Левшун Д.А. Методы интеллектуального анализа системных событий для обнаружения многошаговых кибератак:

использование методов машинного обучения // Искусственный интеллект и принятие решений, 2023, № 3. – С. 3-16.

11. Zhang Y., Yang Q. A survey on multi-task learning // IEEE transactions on knowledge and data engineering, 2022, № 34. – pp. 5586–5609.

12. Girdhar M., Hong J., Moore J. Cybersecurity of Autonomous Vehicles: A Systematic Literature Review of Adversarial Attacks and Defense Models // IEEE Open Journal of Vehicular Technology, 2023, vol. 4. – pp. 417–437.

References

1. Kotenko I., Saenko I., Lauta O., Kribel K., Vasiliev N. Proceedings of the Sixth International Scientific Conference «Intelligent Information Technologies for Industry 2022». Lecture Notes in Networks and Systems, vol. 566. Springer, Cham. 2023. pp. 293- 302.

2. Kotenko I. V., Saenko I. B., Lauta O. S., Vasiliev N. A., Sadovnikov V. E. Dvadtsat pervaya Natsionalnaya konferentsiya po iskusstvennomu intellektu s mezhdunarodnym uchastiem. 2023. pp. 366-376.

3. Xu H., Ma Y., Liu H.C. et al. A Review International Journal of Automation and Computing, 2020, vol. 17. pp. 151- 178.

4. Ren K., Zheng T., Qin Zh., Liu X. Engineering, 2020, vol. 6. pp. 346-360.

5. Rosenberg I., Shabtai A., Elovici Y., Rokach L. ACM Computing Surveys, 2021, vol. 54, №5. 36 p.

6. Zhou Sh., Liu Ch., Ye D., Zhu T., Zhou W., Yu Ph.S. ACM Computing Surveys, 2022, vol. 55, №8. 39 p.

7. Li Y., Cheng M., Hsieh Ch. J., Lee Th. C. M. The American Statistician, 2022, vol. 76, №4. pp. 329 - 345.

8. Fedorchenko E. V., Fedorchenko A. V., Novikova E. S., Saenko I. B. Voprosy kiberbezopasnosti. 2023. № 3 (55). pp. 23- 36.

9. Lebedev V. B., Lebedev O. B. Materialy X Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii "Tekhnologii razrabotki informatsionnykh sistem". Taganrog: Yuzhnyi federalnyi universitet, 2020. pp. 194 - 200.

10. Kotenko I. V., Levshun D. A. Iskusstvennyi intellekt i prinyatie reshenii. 2023. № 3. pp. 3 - 16.

11. Zhang Y., Yang Q. IEEE transactions on knowledge and data engineering, 2022, № 34. pp. 5586 - 5609.

12. Girdhar M., Hong J., Moore J. IEEE Open Journal of Vehicular Technology, 2023, vol. 4. pp. 417- 437.

Дата поступления: 12.12.2025

Дата публикации: 7.02.2026