МИНИСТЕРСТВО ВНУТРЕННИХ ДЕЛ
РЕСПУБЛИКИ КАЗАХСТАН

АЛМАТИНСКАЯ АКАДЕМИЯ
ИМЕНИ МАКАНА ЕСБУЛАТОВА

ИСПОЛЬЗОВАНИЕ ВОЗМОЖНОСТЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ВЫЯВЛЕНИЯ ИНТЕРНЕТ-МОШЕНИЧЕСТВ

Методические рекомендации

Алматы
2024

Обсуждено и одобрено на заседании научно-методическом совете Алматинской академии МВД Республики Казахстан им. М. Есбулатова (протокол №7 от «19» сентября 2024 года)

Рецензенты:

Смайлов Н.К. – Доктор PhD, профессор кафедры «ЭТиКТ» КазНИТУ им. К. Сатпаева.

Коржумбаева Т.М. – начальник кафедры «Административно-правовых дисциплин» Алматинской академии МВД Республики Казахстан им. М. Есбулатова, к.ю.н., полковник полиции.

Кадырова Р.Т., Дюсетаев Р.С., Ендыбайұлы Е.: Использование возможностей искусственного интеллекта для выявления интернет-мошеничеств: Методические рекомендации. – Алматы: ООНИиРИР Алматинской академии МВД Республики Казахстан им. М. Есбулатова, 2024. – 58 с.

Методические рекомендации посвящены вопросам видов интернет-мошенничеств и возможных методов их выявления с помощью ИИ. Особое внимание уделяется таким технологиям, как машинное обучение, обработка естественного языка (NLP), а также анализ больших данных. Эти подходы позволяют значительно повысить эффективность выявления мошеннических действий в реальном времени. Работа рассчитана на широкий круг исследователей, как для ученых, так и для практических работников, а также для курсантов, слушателей, магистрантов и докторантов высших учебных заведений правоохранительных органов.

Введение

В современном мире, где цифровые технологии играют все более важную роль, онлайн мошенничество становится одной из самых серьезных проблем, с которым потребители сталкиваются ежедневно, в том числе в сети Интернет. Онлайн мошенничество затрагивает не только отдельных пользователей, но и организации, правительства и оказывает значительное влияние на мировую экономику. Те, кто стал жертвой мошенничества, сообщают об эмоциональном смятении, стыде и потере уверенности. Интернет-мошенничество является одним из наиболее заметных проявлений цифрового вреда для потребителей, причиняя потребителям финансовый, эмоциональный и психологический вред. Однако, с развитием искусственного интеллекта (ИИ), появляется возможность борьбы с этой проблемой более эффективными методами.

Как отмечается в [1], даже такие абсолютно понятные пользовательские устройства, как мобильные телефоны уже содержат чипы для искусственного интеллекта (Pixel 6 от Google, iPhone). ИИ меняет то, как компьютеры программируются и как они используются. Благодаря машинному обучению программисты больше не пишут правила. Вместо этого они создают нейронную сеть, которая сама извлекает эти правила в процессе обучения.

Компания Микрософт [2] предложила следующее разделение темы ИИ и кибербезопасность: – Повышение кибербезопасности с помощью ИИ (использование ИИ в кибербезопасности) – Кибератаки с использованием ИИ (использование ИИ для усиления кибератак) – Кибербезопасность систем ИИ (атаки на системы ИИ) – Использование ИИ в злонамеренных информационных операциях (фейки с использованием ИИ)

1 Актуальность и значение проблемы интернет-мошенничеств

С увеличением числа пользователей и объемов онлайн-операций возросло и количество интернет-мошенничеств, которые становятся все более сложными и разнообразными. Эти преступления оказывают серьезное негативное воздействие как на индивидуальных пользователей, так и на организации и экономику в целом.

Актуальность проблемы интернет-мошенничеств обусловлена несколькими ключевыми факторами:

1. *Рост объемов онлайн-транзакций.* В связи с развитием электронной коммерции, онлайн-банкинга и различных цифровых сервисов число интернет-транзакций растет в геометрической прогрессии. Мошенники используют эту тенденцию, разрабатывая все более изощренные схемы для кражи личных данных и финансовых средств.

2. *Усложнение методов атак.* Современные преступники активно используют сложные технологии, такие как социальная инженерия, фишинг и поддельные веб-сайты. Это затрудняет обнаружение мошенничества традиционными методами, так как атаки становятся более целенаправленными и персонализированными.

3. *Ущерб для бизнеса.* Компании сталкиваются с финансовыми потерями из-за мошеннических операций, утраты репутации и доверия клиентов. В частности, банки, платформы электронной коммерции и финансовые компании являются наиболее уязвимыми. Непредотвращенные атаки могут приводить к значительным убыткам и даже к банкротству компаний.

4. *Личные данные и конфиденциальность.* Интернет-мошенничества нередко связаны с кражей персональных данных, что ставит под угрозу конфиденциальность пользователей и повышает риски для их безопасности. Утечка данных может приводить к дальнейшим киберпреступлениям, таким как идентификационные кражи и мошенничество с кредитами.

5. *Недостаточная осведомленность пользователей.* Многие люди не обладают достаточными знаниями и навыками по обеспечению своей кибербезопасности, что делает их легкой мишенью для интернет-мошенников. Даже с применением сложных технологий защиты, человеческий фактор часто остается слабым звеном.

Значение борьбы с интернет-мошенничествами заключается не только в защите финансовых и личных данных, но и в обеспечении устойчивости экономических систем и укреплении доверия к цифровым технологиям. Использование возможностей искусственного интеллекта (ИИ) открывает новые горизонты для выявления и предотвращения мошенничеств, делая эту проблему решаемой на новом уровне технологического прогресса.

1.1 Роль искусственного интеллекта (ИИ) в борьбе с киберпреступностью

Искусственный интеллект внедряется в сферу кибербезопасности с целью не только обнаружения и предотвращения угроз, но и эффективного реагирования на них. Одним из важных аспектов является автоматизация кибербезопасности с использованием искусственного интеллекта. Алгоритмы машинного обучения позволяют системам автоматически анализировать большие объемы данных, выявлять аномалии и предсказывать возможные кибератаки. Это повышает оперативность реакции и позволяет предотвращать угрозы до того, как они нанесут ущерб. Технологии машинного обучения также способствуют созданию интеллектуальных систем, способных адаптироваться к новым видам угроз. Системы, оснащенные искусственным интеллектом, могут обучаться на основе новых данных и улучшать свою эффективность с течением времени. Это особенно важно в контексте появления всё бо лее сложных и хитрых кибератак. Кроме того, искусственный интеллект применяется для создания интеллектуальных систем анализа угроз и решения сложных киберзадач [3, с. 135-

147]. Такие системы могут обрабатывать данные в реальном времени, выявлять нестандартное поведение пользователей, анализировать сетевой трафик и обнаруживать скрытые угрозы.

Однако, вместе с усилением киберзащиты, искусственный интеллект также становится инструментом для создания более сложных кибератак. Атакующие могут использовать технологии машинного обучения для создания интеллектуальных и трудно выявляемых угроз. Это подчеркивает необходимость постоянного совершенствования методов кибербезопасности и адаптации к новым вызовам. Одним из значительных аспектов обсуждения являются этические вопросы, связанные с использованием искусственного интеллекта в кибербезопасности. Вопросы конфиденциальности, ответственности за принятие решений, а также создание стандартов и регулирование в этой области становятся все более актуальными.

В современном мире, где технологии играют огромную роль в нашей повседневной жизни, кибербезопасность становится все более значимым аспектом. Угрозы виртуального пространства постоянно эволюционируют, и потому необходимо развивать инновационные методы и инструменты для борьбы с киберпреступностью. В этом контексте, искусственный интеллект (ИИ) играет все более важную роль в обеспечении кибербезопасности. Искусственный интеллект обладает потенциалом преобразовать современные методы обнаружения и реагирования на киберугрозы. Автоматическое обнаружение инцидентов, анализ больших объемов данных, прогнозирование потенциальных уязвимостей – все это является областью применения ИИ в кибербезопасности. Это позволяет идентифицировать и отражать атаки быстрее, чем это возможно для человека. Повышение кибербезопасности имеет решающее значение в современном цифровом мире. Вот несколько предложений по улучшению кибербезопасности:

Обучение и осведомленность сотрудников:

– Проводите регулярное обучение сотрудников кибербезопасности, чтобы рассказать им о потенциальных угрозах,

фишинге и передовых методах обработки конфиденциальной информации;

– Развивайте культуру заботы о кибербезопасности, чтобы поощрять сотрудников оперативно сообщать о подозрительных действиях. Надежные меры аутентификации:

– Обеспечьте использование надежных и уникальных паролей для всех учетных записей;

– Внедрите многофакторную аутентификацию (MFA), чтобы добавить дополнительный уровень безопасности. Сетевая безопасность:

– Используйте брандмауэры для мониторинга и контроля входящего и исходящего сетевого трафика.

– Используйте системы обнаружения и предотвращения вторжений для выявления потенциальных угроз и реагирования на них.

Искусственный интеллект также способен усилить процессы аутентификации и авторизации. Технологии распознавания лиц и голоса, например, могут предоставить более надежные способы идентификации пользователей [5, с. 81-86]. Благодаря ИИ, можно создать системы, которые могут автоматически обнаруживать необычную активность в сети и применять меры безопасности для предотвращения возможных атак. Кроме того, ИИ может быть использован для разработки сильных систем противодействия фишингу и вредоносному программному обеспечению [4, с. 151-154]. Алгоритмы машинного обучения могут анализировать тысячи электронных писем, обнаруживая подозрительные и небезопасные ссылки, а также прикрепленные файлы. Такие системы способны улучшить процессы фильтрации на самом раннем этапе, что помогает предотвратить многие кибератаки. Искусственный интеллект также может использоваться для разработки системы мониторинга, которая анализирует активность в сети и идентифицирует любые подозрительные действия. Это позволяет предупреждать потенциальные угрозы и принимать меры для их нейтрализации. Однако, несмотря на все преимущества, ИИ также несет некоторые риски. Настолько,

сколько ИИ может быть использован для защиты, он может быть использован и злоумышленниками для создания более сложных и усовершенствованных атак. Соответственно, необходимы такие же сильные системы искусственного интеллекта для борьбы с этими угрозами.

1.2 Классификация интернет-мошенничеств

Интернет-мошенничества представляют собой широкий спектр незаконных действий, которые можно классифицировать по различным критериям, таким как методы осуществления, цели, инструменты и платформы. Эта классификация помогает лучше понять способы борьбы с мошенничеством и выбрать эффективные стратегии защиты.

1. По типу атаки

1.1. Фишинг

Мошенничество, направленное на получение конфиденциальных данных пользователей, таких как пароли, номера кредитных карт и другая персональная информация. Основные виды фишинга:

– *Классический фишинг* (через поддельные сайты или электронные письма).

– *Вишинг* (через голосовые звонки).

– *Смишинг* (через текстовые сообщения).

1.2. Мошенничество с банковскими картами и платежами

Атаки на платежные системы с целью кражи данных карт и проведения несанкционированных транзакций. Включает:

– *Скимминг* – установка устройств на банкоматах для кражи данных карт.

– *Кража данных платежных систем* через поддельные онлайн-платежи.

1.3 Мошенничество с поддельными интернет-магазинами

Создание фальшивых онлайн-магазинов, где товары либо не существуют, либо не будут доставлены после оплаты. Часто используются сайты с поддельными отзывами и низкими ценами.

1.4. Социальная инженерия

Манипуляции с пользователями для того, чтобы обманом заставить их передать конфиденциальную информацию. Методы включают:

– Мошенничество, когда злоумышленник притворяется сотрудником банка, компании или службы поддержки.

1.5. Вредоносное ПО и программы-вымогатели

Вредоносные программы, такие как вирусы и трояны, которые могут блокировать доступ к данным пользователя и требовать выкуп (например, программы-вымогатели).

1.6. Взлом аккаунтов

Целью злоумышленников является взлом учетных записей в социальных сетях, почте, интернет-банкинге для получения доступа к персональной информации или финансовым ресурсам.

1.7. Мошенничество с использованием ботов

Использование автоматизированных программ (ботов) для реализации мошеннических схем, таких как массовые фальшивые транзакции, накрутка кликов, создание поддельных аккаунтов и другие схемы.

1.8. Инвестиционные мошенничества

Предложения для пользователей «выгодных» инвестиционных проектов, которые на деле оказываются фальшивыми. Часто используются схемы с криптовалютами, пирамиды или мошеннические проекты с ценными бумагами.

2. По платформе реализации

2.1. Мошенничество в электронной коммерции

Связано с покупками и продажами товаров в интернете, где мошенники используют поддельные магазины или некорректную информацию о товарах.

2.2. Мошенничество в социальных сетях

Использование соцсетей для распространения фальшивой информации, создания поддельных аккаунтов, проведения розыгрышей, сбора личной информации и дальнейших атак.

2.3. Мошенничество на онлайн-аукционах

Поддельные или недействительные лоты на онлайн-аукционах, целью которых является получение денег от пользователей без передачи товара.

2.4. Мошенничество через электронную почту

Фишинговые письма, содержащие ложные сообщения с просьбами предоставить личные данные или совершить оплату под видом известной компании.

3. По цели мошенничества

3.1. Кража персональных данных

Основной целью атак является кража личных данных пользователей (логины, пароли, номера документов) для их дальнейшей продажи или использования в других схемах мошенничества.

3.2. Финансовое мошенничество

Направлено на кражу денежных средств через несанкционированные транзакции, поддельные магазины или инвестиционные схемы.

3.3. Мошенничество с репутацией

Атаки, направленные на порчу репутации компаний или частных лиц с целью извлечения выгоды, включая негативные отзывы, дискредитацию брендов, создание фальшивых аккаунтов.

4. По используемым технологиям

4.1. Традиционные мошенничества

Основаны на поддельных предложениях, лотереях, фальшивых скидках и продажах.

4.2. Мошенничество с использованием искусственного интеллекта (ИИ)

Применение ИИ для создания более правдоподобных поддельных сайтов, фальшивых сообщений или автоматических схем мошенничества с минимальным участием человека.

4.3. Кибератаки с использованием уязвимостей ПО

Атаки, использующие уязвимости программного обеспечения или операционных систем для установки вредоносного ПО и получения доступа к данным.

5. По уровню взаимодействия с пользователем

5.1. Прямое мошенничество

Осуществляется непосредственно с пользователем через взаимодействие (фишинг, мошенничество в социальных сетях, звонки).

5.2. Непрямое мошенничество

Атаки происходят без прямого контакта с пользователем (например, взлом учетной записи или установка вредоносного ПО на устройство).

2 Повышение кибербезопасности с помощью ИИ

Можно сказать, что это наиболее продвинутая на сегодняшний день область. Ценность, которую привносит здесь машинное обучение, состоит в определении атак, поиске шаблонов и закономерностей, соответствующих вторжениям, быстром анализе и приоритизации угроз, анализе накопленной информации для адаптации методов обнаружения вторжения.

Первый ответ на вопрос, зачем здесь ИИ, согласно [2], заключается в слове «автоматизация». Автор приводит американские данные Бюро статистики труда США о том, что возможности трудоустройства в сфере кибербезопасности вырастут на 33% с 2020 по 2030 год, что более чем в шесть раз превышает средний показатель по стране [7]. Вряд ли картина в других странах отличается от приведенной. При этом, согласно исследованию рынка труда в части кибербезопасности ISC, опубликованному в октябре 2021 года, во всем мире не хватает 2.72 миллиона специалистов по кибербезопасности [8]. Соответственно, альтернативы автоматизации решения задач кибербезопасности просто нет.

Задачи кибербезопасности состоят из предотвращения атак, обнаружения атак, проведения расследований, классификации и анализе угроз, а также обучения и моделирования систем кибербезопасности.

Предотвращение атак (профилактика) – это усилия по снижению количества уязвимостей содержащихся в программном обеспечение. Типичные примеры есть, например, в обзоре [9], который описывает системы машинного обучения, выполняющие поиск вредоносных приложений на Android. Собираются характеристики приложений (рис. 1), на датасетах по приложениям обучаются классификаторы.

Analysis Type	Feature Extraction Method	Features Extracted
Static	Manifest analysis	Package name, Permissions, Intents, Activities, Services, Providers
	Code analysis	API calls, Information flow, Taint tracking, Opcodes, Native code, Cleartext analysis
Dynamic	Network traffic analysis	URLs, IPs, Network Protocols, Certificates, Non-encrypted data
	Code instrumentation	Java classes, intents, network traffic
	System calls analysis	System calls
	System resources analysis	CPU, Memory, and Battery usage, Process reports, Network usage
	User interaction analysis	Buttons, Icons, Actions/Events

Рис.1. Характеристики приложений

Есть даже статистика по используемым методам классификации, где лидирует Random Forest.

Как отмечено в [2], в 2021 году Институт AV-Test [10] обнаружил более 125 миллионов новых вредоносных программ. Способность методов машинного обучения обобщать прошлые шаблоны для обнаружения новых вариантов вредоносных программ и является ключом к построению масштабируемой системы защиты.

Можно отметить, что поиск в Google Scholar работ по запросу «ML for malware detection» показывает более 20 000 статей [11].

Глубокое обучение также активно используется в этой области [12]. В этой работе описывается система, созданная

по государственному гранту Китая для ключевых технологий. Интересный сравнительный анализ моделей глубокого обучения для определения вредоносных приложений есть в работе [13]. Все такие работы имеют практическое применение, например, Microsoft 365 Defender [14] также использует глубокое обучение.

Отметим, что под словом «программы» не следует понимать здесь только код. Например, в работе [15] описывается модель глубокого обучения для определения фишинговых URL. И это только один пример из множества подобных работ. В целом, фишинговые атаки довольно разнообразны и использование машинного обучения для их обнаружения описывалось еще в 2008 году [16]. Обзор современных подходов, использующих машинное обучение в борьбе с фишингом, есть, например, в свежей работе [17].

Обнаружение атак включает выявление подозрительного поведения и оповещение о нем непосредственно по мере его возникновения. Цель состоит в том, чтобы быстро реагировать на атаки, включая определение масштаба атаки, закрытие входов для атакующих и устранение уязвимостей (бэкдоров и т.п.), которые мог эксплуатировать злоумышленник.

Очевидно, что поиск, в общем случае, неизвестных шаблонов атак потенциально может приводить к большому числу ложных срабатываний (false positives) [18]. В литературе отмечается, что основная проблема при обнаружении подозрительной активности как раз и заключается в том, чтобы найти правильный баланс между обеспечением достаточного охвата за счет поиска точных предупреждений системы безопасности и количеством ложных срабатываний.

Можно выделить следующие направления, касающиеся использования машинного обучения для предупреждений об атаках [2]:

(1) расстановка приоритетов для предупреждений о потенциальных атаках [19],

(2) выявление многочисленных попыток взлома с течением времени, которые являются частью более крупных и длительных кампаний по взлому [19],

(3) обнаружение следов действий вредоносных программ, как внутри компьютера, так и в сети [120]

(4) идентификация потока вредоносного программного обеспечения, внедряемого через конкретную организацию. Это так-называемые Living off the Land (LotL) атаки – кибератаки, в которых атакующий использует легальное программное обеспечение в организации для выполнения атакующих действий [21].

(5) определение автоматизированных подходов к смягчению последствий атак, когда требуется быстрое реагирование, чтобы предотвратить распространение атаки. Например, автоматизированная система может отключать сетевое подключение и блокировать устройство, если обнаруживается последовательность предупреждений, которая, как известно, связана с действиями программы-вымогателя [22].

2.1 Кибератаки с использованием ИИ

В связи с атаками используется термин наступательный ИИ. Рисунок 2 из работы [26] суммирует направления атак с использованием систем машинного обучения на матрице угроз MITRE.

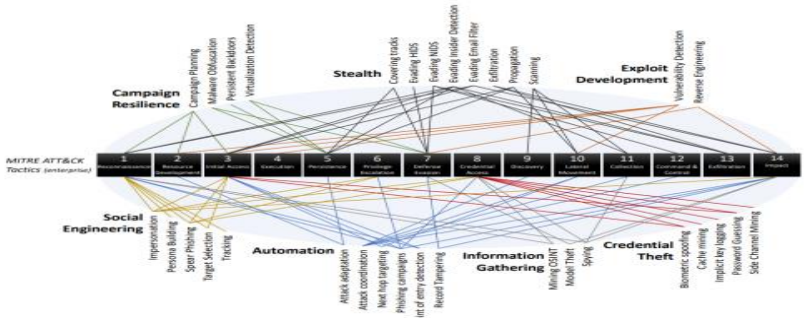

Рис. 2. Машинное обучение в кибератаках [26].

Авторы последнего обзора выделили следующие области атак с использованием ИИ.

1. Прогнозирование – сделать некоторый прогноз на основе ранее наблюдаемых данных. Пример атаки с использованием машинного обучения - идентификация нажатий клавиш на смартфоне на основе движения (вибрации) [28, 29]. Другие приведенные примеры касались предсказания чувствительных данных для пользователей социальных сетей [30] (поиск слабого звена для атаки), поиска уязвимостей программного обеспечения [31, 32, 33].

2. Генерация – создание контента с использованием ИИ. Примеры такой генерации для наступательных целей – фальсификация медиа-данных [34], подбор паролей [35], модификацию трафика [36]. Последнее (в англоязычной литературе – traffic-space attacks) представляет собой, фактически, состязательную атаку на систему машинного обучения, которая используется для анализа трафика (определения вторжений). Цель атаки – скрыть реальное вторжение. Дипфейки – еще один пример наступательного ИИ в этой категории. Дипфейк – это правдоподобный медиафайл. Создаются они с использованием глубокого обучения. Технология может быть использована для того, чтобы выдавать себя за жертву, имитируя ее голос или лицо при совершении фишинговой атаки [37].

3. Анализ – это задача анализа или извлечения полезной информации из данных или модели. Исследование атакуемой

модели ML, с целью определения реальных факторов, влияющих, например, на классификацию. Имеется в виду использование объясняющих подходов (LIME, SHAPLEY и др.). Понимание работы атакуемой модели необходимо для создания эффективных атак или сокрытия вторжений. Если атакуемая модель недоступна, то такие эксперименты могут проводиться на ее теневой копии.

4. Поиск – это задача поиска информации или объектов для атаки по заданным критериям. Приведенные примеры – поиск (идентификация) человека по изображениям на нескольких взломанных камерах [39, 40], поиск возможных инсайдеров по семантическому анализу публикаций в социальных сетях [38], аннотирование (реферирование, суммаризация) документов при сборе данных из открытых источников (OSINT – открытая разведка) [39] (последнее есть пример автоматизации).

5. Принятие решения – это задачи разработки стратегического плана или координации операции (атаки). Примеры в ИИ – использование роевого интеллекта для управления автономной сетью ботов [40] и планирование оптимальных атак на сети [41].

В презентации [42] отмечается, что автоматизировать атаки можно и без машинного обучения, но обучение с подкреплением (reinforcement learning) имеет все шансы стать основным инструментом в осуществлении атак.

Микрософт в отчете [43] ожидает, что использование ИИ в кибератаках начнется с опытных участников, но быстро распространится на более широкую экосистему за счет повышения уровня сотрудничества и коммерциализации используемых инструментов. В частности, инструменты атакующих включают общие базовые тактики обхода защиты, как описано в атласе MITRE [44]. Одна из наиболее успешно используемых систем автоматизации в наступательном ИИ – это боты в социальных сетях [45]. Другой пример автоматизации наступательных действий представлен в работе [46] – автоматизированный тест на проникновение (penetration test), использующий обучение с подкреплением (рис. 3).

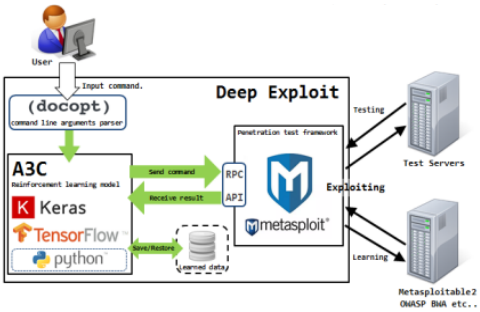

Рис. 3. Deep Exploit [45]

Машинное обучение используется для атак на биометрические системы аутентификации: подделка голоса и т.п. [47, 48].

Выше мы говорили об определении фишинговых атак с помощью машинного обучения. Но машинное обучение используется и при генерации фишинговых атак [49, 50]. Цель – обойти системы защиты, создать более привлекательный контент и побудить пользователей кликнуть злонамеренную ссылку, установить в системе программное обеспечение и т.д. Примеры наступательных действий включают также подбор паролей [50], запутывание исходного кода программ [51], маскировку трафика [52], управление сетью ботов [53]. Наступательному ИИ посвящен отдельный воркшоп (отчет – доступен), организованный компанией Микрософт [54]. Атаки с использованием ИИ рассматриваются также в довольно подробном отчете National Security Commission on Artificial Intelligence (NSCAI) [55].

2.2 Атаки на системы ИИ

Это достаточно новая область для компьютерной безопасности. Атаки могут быть направлены на сами системы ИИ (фактически – на системы машинного обучения). Любая внедренная система машинного обучения есть, в конечном итоге, программа. Но проблема состоит в том, что для таких приложений традиционные методы анализа безопасности не-

применимы. Проблемы с безопасностью именно таких приложений не могут быть решены традиционными методами. Конечно, скомпрометированная среда исполнения программы будет приводить к проблемам. Но это не главная беда.

Системы машинного обучения зависят от данных. На основе представленных тренировочных данных система вырабатывает некие обобщения, которые затем используются при обработке реальных (тестовых) данных. Так вот модификации данных на разных этапах конвейера машинного обучения и приводят к тому, что такие системы могут либо вовсе не работать, либо наоборот, выдавать нужные атакующему результаты. При этом специально модифицированные данные будут, вообще говоря, точно такими же, как и «чистые» данные. В общем случае, их нельзя будет различить. Более того, поскольку обучение всегда производится на некотором тренировочном наборе данных, генеральная совокупность остается, в общем случае, неизвестной. И «изменение» данных на этапе эксплуатации может случиться (и чаще всего случается) безо всяких зловредных действий. Просто потому, что так устроены сами данные. Атаками в данном случае называют именно специальное изменение данных или специальную подстановку данных, на которых система работает неверно (вообще не работает). В общем виде – это проблема устойчивости систем машинного обучения. Этой проблеме сейчас уделяется много внимания, поскольку это основное, что препятствует использованию систем ИИ в критических приложениях (авионика, ядерная безопасность и т.п.) [56, 57].

Другое название атак на системы машинного обучения – состязательные примеры [58]. Таким образом, враждебные воздействия на системы могут осуществляться в форме традиционных уязвимостей, а также с помощью новой категории: состязательных примеров.

Как примеры традиционных уязвимостей можно указать, например, отчет об уязвимостях в программном пакете Tensorflow [59], что, естественно, означает наличие уязвимо-

стей в использующих его системах ИИ. Атаки на программную инфраструктуру ИИ исследовались в работах [60, 61, 62]. В работе [63] исследователи из Нью-Йоркского университета обнаружили, что большинство сред ИИ не проверяют целостность загруженных моделей ИИ, в отличие от общепринятой практики с традиционным программным обеспечением, где криптографическая проверка исполняемых файлов/библиотек является стандартной практикой уже более десяти лет. Публичные датасеты могут содежать ошибки в разметке [64], что, естественно, влияет на работу обученных с их помощью систем [65].

Состязательные примеры принято классифицировать по точке приложения враждебных усилий (этапу конвейера машинного обучения) и знаниям атакующего о системе (белый ящик, черный ящик). Одна из возможных классификаций приведена на рис. 4.

Атака	Этап	Затрагиваемые параметры
Adversarial attack	применение	входные данные
Backdoor attack	тренировка	параметры сети
Data poisoning	тренировка, использование	входные данные
IP stealing	использование	отклик системы
Neural-level trojan	тренировка	отклик системы
Hardware trojan	аппаратное проектирование	отклик системы
Side-channel attack	использование	отклик системы

Рис. 4. Классификация атак на системы ИИ

Также атаки бывают целевые (например, атакующий хочет добиться определенного результата от классификатора) и нецелевые (просто воспрепятствовать правильной работе классификатора).

Модификацию входных данных (по факту, самый распространенный тип атаки) еще называют атаками уклонения. Кража (IP stealing) включает в себя получение сведений о модели (а это нужно для организации атак) [66] и так называемые инверсные атаки, которые направлены на восстановление лежащих в основе частных данных, использованных для обучения целевой системы [67]. Микрософт [68] отмечает, что количество таких атак растет. В первую очередь, это касается,

конечно, критических применений. В работе [69] описываются усилия США и Китая по противодействию системам ИИ друг друга. В целом, в силу отсутствия полной защиты, такие атаки приходится воспринимать как некоторый универсальный риск, связанный с использованием систем машинного обучения. При этом необходимо учитывать как возможность осуществления атаки, так и практическую осуществимость таких атак.

Очевидно, что модифицировать входные данные можно практически всегда. Например, так называемые физические атаки (изменение формы представления), являются одними из наиболее легко осуществимых и опасных для систем распознавания. Простой пример физической атаки – камуфляж (защитная раскраска) [70]. Для организации атак уклонением используют как простые модификации данных (например, атака Salt & Pepper – добавление черных и белых точек к изображению [71]), так и специальные решения с использование машинного обучения, например, порождающих моделей [72].

Отравления данных можно, очевидно, избежать, если использовать собственные проверенные наборы данных, избегать использования данных из неизвестных источников или проверять все такие данные.

Кража данных и модели технически связана с анализом множества откликов атакуемой системы на специальным образом подготовленные входные данные. Если это не решение ML as a service [73], то способа опрашивать систему может просто не быть. Но если нельзя опрашивать саму модель, то можно попробовать создать ее копию (shadow model) и отрабатывать атаки на ней. Отсюда следует вывод о том, что в отличие от классического программного обеспечения, где сами алгоритмы чаще всего открыты, для систем машинного обучения детали реализации моделей в критических областях должны скрываться, поскольку такие знания позволят построить теневую модель (копию модели) для отработки атак.

В целом, атаки уклонением (то есть модификация входных данных) есть главная практическая проблема. На сегодняшний день, атаки в этой области опережают защиту. И это есть основное препятствие для внедрения систем машинного обучения в критические приложения. В отдельных случаях (в зависимости от данных и размера модели) можно говорить о формальных доказательствах устойчивости систем машинного обучения [74]. В других случая подходы к формальному доказательству будут сталкиваться с трендом на увеличение параметров современных сетей (что можно доказывать для сети с миллиардами параметров?). В большинстве случаев “защита” состоит из включения модифицированных данных в тренировочные наборы и учета таким образом возможных модификаций данных, за счет точности системы. Вопрос о том, что это не все возможные модификации, как правило, игнорируется.

Как было уже указано выше, основное направление работ здесь – это создание устойчивых систем (моделей) машинного обучения [76]. Большой обзор такого рода проектов, как академических, так и промышленных есть в работе [56]. С практической точки зрения, для разработки систем машинного обучения для критических применений необходимы так называемые доверенные среды разработки, которые гарантируют отсутствие компрометации инструментальных средств и представляют инструменты для повышения доверия к результатам работы систем [77].

Из британской национальной программы искусственного интеллекта: «Злоумышленники будут стремиться скомпрометировать наши системы искусственного интеллекта, снизить их производительность и подорвать доверие пользователей и общественности, используя множество цифровых и физических средств» [76, 77]. Атаки, снижающие производительность систем машинного обучения, уже существуют [78].

Необходимо отметить, что проблемы с защитой систем ИИ полностью осознаются, как в промышленном, так и в индустриальным сообществе. Есть широко известный каталог

MITRE, поддерживаемый Микрософт и другими организациями, в котором собирается информация по атакам на системы ИИ. В частности, в нем есть так называемая матрица угроз Adversarial ML для каталогизации угроз для систем ИИ [79]. Для инженеров и политиков Microsoft в сотрудничестве с Центром Беркмана Кляйна в Гарвардском университете выпустила таксономию режимов сбоев машинного обучения [80]. DARPA предлагает бесплатные ресурсы для оценки безопасности систем машинного обучения [81]. Микрософт предлагает свой продукт с открытым кодом Counterfit, как инструмент для оценки безопасности систем ИИ [82]. Министерство обороны США включило безопасность систем ИИ в свой список основных принципов построения ИИ [83]. Американский институт стандартов NIST работает над схемой оценки рисков ИИ, направленной на решение множества аспектов систем ИИ, включая надежность и безопасность [84].

2.3 ИИ в операциях со злонамеренной информацией

Достижения в области машинного обучения и компьютерной графики расширили возможности государственных и негосударственных субъектов по производству и распространению высококачественного аудиовизуального контента, называемого синтетическими медиа и дипфейками. Технологии искусственного интеллекта для создания дипфейков теперь могут создавать контент, неотличимый от реальных людей, сцен и событий. Такой контент может реально угрожать национальной безопасности.

Расширение возможностей генеративных методов искусственного интеллекта для синтеза различных сигналов, включая высококачественные аудиовизуальные изображения, имеет значение для кибербезопасности. При персонализации использование ИИ для создания дипфейков может повысить эффективность операций социальной инженерии (программа выдает себя за некоторое реальное лицо) и убедить, например,

конечных пользователей предоставить злоумышленникам доступ к системам и информации [85].

В более широком масштабе, генерирующая мощь методов искусственного интеллекта и синтетических сред имеет важные последствия для обороны и национальной безопасности. Эти методы могут использоваться противниками для создания правдоподобных заявлений мировых лидеров и командующих, для фабрикации убедительных операций под ложным флагом и создания фальшивых новостей [2, 86].

Исследование университета Georgia Tech показывает, что распространение синтетических медиа имело еще один тревожный эффект: злонамеренные субъекты назвали реальные события «фальшивыми», воспользовавшись новыми формами отрицания, которые приходят с потерей доверия в эпоху дипфейков. Видео и фото-доказательства, например, изображения зверств, называют фейком. Распространение синтетических СМИ, известное как «дивиденд лжеца», побуждает людей называть настоящие СМИ «фальшивыми» и создает правдоподобное отрицание их действий [87].

В презентации Микрософт [2] отмечается, что можно ожидать, что синтетические медиа и области их применения будут со временем становиться все более изощренными, включая убедительное чередование дипфейков с реально происходящими событиями в мире и синтез дипфейков в реальном времени. Генерации в реальном времени можно использовать для создания убедительных интерактивных самозванцев (например, появляющихся на телеконференциях и управляемых человеком-контроллером), которые, кажется, имеют естественную позу головы, выражения лица и высказывания. Отметим что, нам, возможно, придется столкнуться с проблемой искусственно созданных людей, которые могут автономно участвовать в убедительных разговорах в реальном времени по аудио и визуальным каналам. Естественно, что в таких условиях определение дипфейков становится весьма актуальной задачей.

Пример – программа DARPA Semantic Forensics (SemaFor) [88]. Программа SemaFor направлена на разработку инновационных семантических технологий для анализа медиа. Эти технологии включают в себя алгоритмы семантического обнаружения, которые определяют, были ли созданы мультимодальные медиаактивы или ими манипулировали. Алгоритмы атрибуции сделают вывод, исходит ли мультимодальное медиа от конкретной организации или отдельного лица. Алгоритмы характеризации будут рассуждать о том, были ли мультимодальные медиа созданы или ими манипулировали в злонамеренных целях. Эти технологии SemaFor помогут выявлять, сдерживать и понимать кампании противника по дезинформации.

Другая программа – DARPA MediaForensics (MediaFor) [89]. Презентация определяет Media Forensic как научное исследование в области сбора, анализа, интерпретации, и представление аудио-, видео- и графических доказательств, полученных в ходе ход расследования и судебного разбирательства. Поставленная цель - разработать технологии автоматизированной оценки целостности изображения или видео (рис. 5).

Микрософт в презентации [2] считает многообещающим подход к противодействию угрозе синтетических носителей на основе технологии происхождения цифрового контента. Происхождение цифрового контента использует криптографию и технологии баз данных для подтверждения источника и истории изменений (происхождения) любых цифровых носителей. Это связано с тем, что в долгосрочной перспективе ни люди, ни методы ИИ не смогут надежно отличить факты от выдумок, созданных ИИ, и, соответственно, мы должны срочно подготовиться к ожидаемой траектории все более реалистичных и убедительных дипфейков. В части создания технологий сертификации аудио-визуального контента появились межотраслевые партнерства Project Origin, Content Authenticity Initiative (CAI) и Coalition to Content Provenance and Authenticity (C2PA) [93, 94, 95, 96].

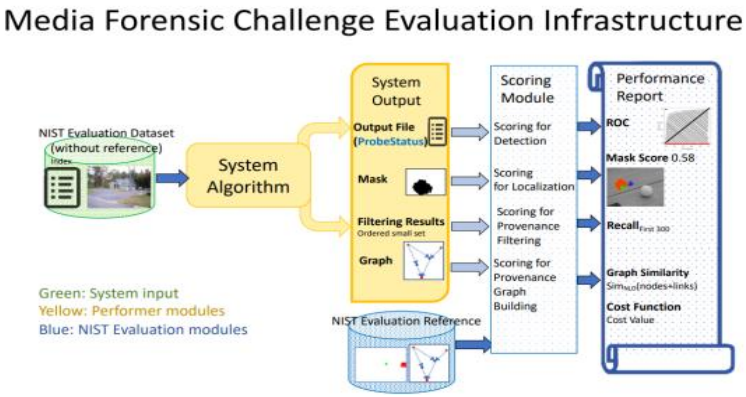

Рис.5. Оценка целостности контента [90]

В январе 2022 года C2PA выпустила спецификацию стандарта, который обеспечивает совместимость систем происхождения цифрового контента [90, 91]. Это позволяет выпускать коммерческие инструменты производства контента в соответствии со стандартом C2PA, которые будут позволять авторам и вещателям уведомлять зрителей об исходном источнике и истории редактирования фото- и аудиовизуальных материалов. В заключительном отчете NSCAI [55] рекомендуется использовать технологии происхождения цифрового контента, чтобы смягчить растущую проблему синтетических медиа. В Конгрессе США двухпартийный Закон о целевой группе по дипфейкам предлагает создать Национальную целевую группу по дипфейкам и цифровому происхождению [92]. Технологии блокчейн предлагается также использовать для подтверждения авторства медиа данных [97].

3. Технологии искусственного интеллекта для выявления мошенничеств

С ростом объема интернет-транзакций и развитием цифровых сервисов использование ИИ для выявления и предотвращения интернет-мошенничеств стало важным элементом современных систем кибербезопасности. Искусственный интеллект позволяет обнаруживать сложные и динамически изменяющиеся мошеннические схемы, реагировать на них в режиме реального времени и минимизировать финансовые потери. Основные технологии ИИ, применяемые для выявления мошенничеств, включают:

1. Машинное обучение

Машинное обучение (ML) является одним из ключевых методов ИИ, который широко применяется для выявления аномалий и прогнозирования мошеннической активности. Системы машинного обучения обучаются на исторических данных для выявления паттернов, характерных для мошенничества.

Примеры технологий машинного обучения:

– *Классификация и регрессия.* Используются для разделения транзакций на мошеннические и законные, на основе признаков, таких как время транзакции, геолокация, сумма и т.д.

– *Обучение с учителем.* Системы учатся на размеченных данных, где указано, какие транзакции или действия были мошенническими, а какие – нет. Это позволяет улучшить точность обнаружения схожих случаев в будущем.

– *Обучение без учителя.* Используется для поиска аномалий в данных, не имея заранее известной информации о том, что является мошенничеством.

– *Глубокое обучение (нейронные сети).* Более сложные модели, которые могут выявлять скрытые взаимосвязи в большом количестве данных, особенно эффективны в ситуациях, где традиционные методы анализа оказываются неэффективными.

2. Нейронные сети и глубокое обучение

Нейронные сети применяются для более сложных задач, таких как анализ поведения пользователей или распознавание образов. Глубокое обучение (deep learning) используется для обработки сложных данных, таких как последовательности транзакций, взаимодействия в соцсетях или анализ поведения на веб-сайтах.

Примеры использования нейронных сетей:

– *Распознавание аномалий.* Нейронные сети могут автоматически обнаруживать необычные паттерны в поведении пользователей, такие как неожиданные переводы больших сумм, нестандартные регионы проведения операций или другие подозрительные действия.

– *Модели на основе временных рядов.* Эти модели анализируют последовательности действий и времени между ними, что позволяет обнаружить мошенничество в динамике (например, когда один аккаунт пытается выполнять множество транзакций за короткий промежуток времени).

3. Обработка естественного языка (NLP)

Технологии обработки естественного языка (Natural Language Processing, NLP) используются для анализа текстов и речевых сообщений с целью выявления фишинговых атак, мошеннических предложений или обманных рекламных кампаний.

Примеры использования NLP:

– *Анализ фишинговых писем и сообщений.* NLP может анализировать содержание электронных писем, текстовых сообщений и социальных сетей, выявляя фишинговые схемы или ложную информацию. Это помогает заблокировать мошеннические письма до их прочтения пользователем.

– *Обнаружение ложных предложений.* NLP может быть использовано для анализа рекламных объявлений, отзывов или других текстов, связанных с мошенническими схемами, и автоматического выявления ложной информации.

4. Анализ больших данных (Big Data)

ИИ также применим в анализе больших данных (Big Data), позволяя обрабатывать огромные объемы информации в реальном времени и выявлять закономерности, связанные с мошенничеством. Большие данные включают информацию о транзакциях, данных пользователей, активности на веб-сайтах, геолокациях и многом другом.

Примеры использования анализа больших данных:

– *Выявление аномалий в реальном времени.* Системы ИИ могут отслеживать транзакции и поведение пользователей в реальном времени, выявляя аномалии, такие как подозрительные изменения в действиях пользователей, необычные IP-адреса или нехарактерные суммы платежей.

– *Прогнозирование мошеннической активности.* На основе исторических данных и текущих паттернов поведения ИИ может прогнозировать вероятные сценарии мошенничества, что помогает принимать превентивные меры до совершения преступления.

5. *Алгоритмы кластеризации и выявления аномалий*

Методы кластеризации и обнаружения аномалий позволяют ИИ – системам выявлять отклонения от нормы в поведении пользователей, транзакциях или сетевом трафике.

Примеры использования алгоритмов:

– *Кластеризация транзакций.* Системы могут группировать схожие транзакции или аккаунты по паттернам поведения и выделять аномальные группы, которые могут быть связаны с мошенничеством.

– *Обнаружение мошенничества в платежных системах.* Анализ данных транзакций и выявление нетипичных операций позволяет системам автоматически блокировать подозрительные платежи.

6. Рекомендательные системы

Рекомендательные системы, основанные на ИИ, помогают предлагать пользователям персонализированные про-

дукты и услуги. Однако они также используются для выявления мошенничества, анализируя поведение пользователей и выявляя отклонения от их привычной активности.

Примеры использования рекомендательных систем:

– *Индивидуальные профили пользователей.* Системы анализируют поведение пользователей в интернете, строя профили активности. Когда поведение отклоняется от нормы, система может заподозрить мошенничество и инициировать проверку.

– *Предотвращение мошенничества в e-commerce.* Анализ привычек покупок помогает выявить подозрительные изменения в поведении, например, если аккаунт внезапно совершает покупки на суммы, нехарактерные для данного пользователя.

3.1 Видео с дипфейками – растущая угроза

Дипфейковые видео сложно обнаружить в режиме реального времени – приложения для обнаружения требуют загрузки видео для анализа, а затем ожидания результатов в течение нескольких часов. Обман, вызванный дипфейками, может нанести вред и привести к негативным последствиям, таким как снижение доверия к СМИ.

Например, в Казнете появилось видео, на котором рассказывается, как мошенники могут оформлять на казахстанцев кредит при помощи нейросети.

Они снимают ваше изображение по видеозвонку, обрабатывают при помощи искуственного интеллекта и оформляют на вас займ в микрокредитных организациях

Другой случай: недавно лицо Нурлана Сабурова использовали в рекламе онлайн-казино. На некоторых кадрах видно, что то, как движутся губы комика, не совпадает с озвучкой.

Рис.6. Обрабатка изображение при помощи искуственного интеллекта

В похожую ситуацию попадал известный вайнер Токтарбек Сергазы, его изображение использовали, чтобы выманить деньги у доверчивых казахстанцев:

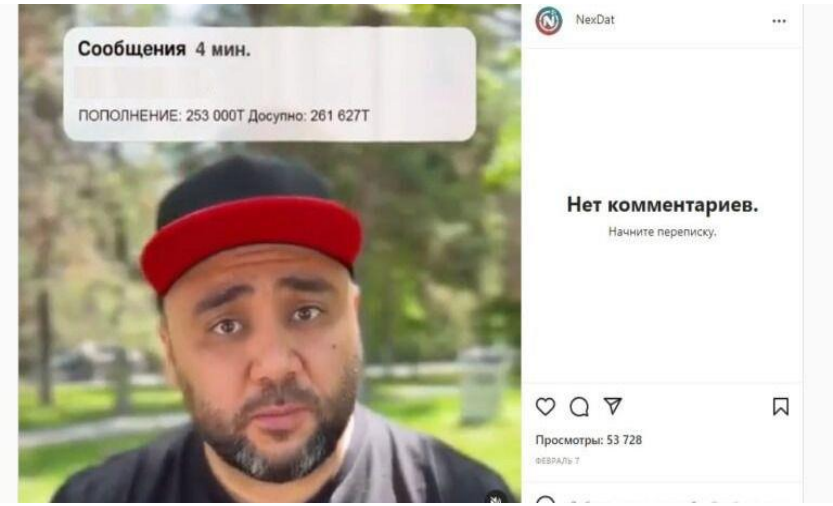

Рис.7. Обрабатка изображение при помощи искуственного интеллекта

3.2 Инструменты и методы обнаружения дипфейков

В эпоху цифровых технологий дипфейки стали серьезной угрозой аутентичности онлайн-контента. Эти сложные видеоролики, созданные искусственным интеллектом, могут убедительно имитировать реальных людей, из-за чего становится все труднее отличить правду от вымысла. Однако по мере развития технологий, лежащих в основе дипфейков, развиваются и инструменты и методы, предназначенные для их обнаружения.

Пять лучших инструментов и методов обнаружения дипфейков:

3.2.1 Sentinel

Рис. 8. Страж

Sentinel – это ведущая платформа защиты на основе искусственного интеллекта, которая помогает демократическим правительствам, оборонным ведомствам и предприятиям противостоять угрозе дипфейков. Технология Sentinel используется ведущими организациями Европы. Система работает, позволяя пользователям загружать цифровые мультимедиа через свой веб-сайт или API, которые затем автоматически анализируются на предмет подделки ИИ. Система определяет,

является ли медиа дипфейком или нет, и обеспечивает визуализацию манипуляции.

Технология обнаружения дипфейков Sentinel предназначена для защиты целостности цифровых носителей. Он использует передовые алгоритмы искусственного интеллекта для анализа загруженного мультимедиа и определения того, были ли им манипулированы. Система предоставляет подробный отчет о своих выводах, включая визуализацию областей носителя, которые были изменены. Это позволяет пользователям точно видеть, где и как манипулировали медиа [98].

Ключевые особенности Sentinel: https://thesentinel.ai/

– Обнаружение дипфейков на основе ИИ.

– Используется ведущими организациями в Европе.

– Позволяет пользователям загружать цифровые медиа для анализа.

– Обеспечивает визуализацию манипуляции.

Использование Sentinel через веб сайт или API:

Рис. 9. Загружаем цифровые медиа через веб сайт или API

Рис. 10. Система автоматически анализирует ИИ подделку

Рис. 12. Определяет, является ли это дипфейком или нет

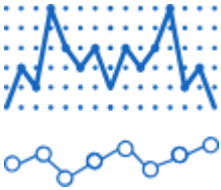

Рис.13. Показана визуализация манипуляция

3.2.2 Attestiv

Attestiv представила решение для обнаружения deepfake, разработанное для частных лиц, влиятельных лиц и предприятий. Эта платформа, доступная для раннего доступа, позволяет пользователям анализировать видео или социальные ссылки на видео на предмет наличия deepfake-контента. Решение Attestiv особенно актуально, учитывая растущую угрозу deepfake для рыночных оценок, результатов выборов и кибербезопасности.

Платформа использует собственный анализ ИИ для оценки и комплексного анализа поддельных элементов, точно определяя, где они находятся в каждом видео. Эта технология особенно ценна для секторов, требующих высокого уровня целостности, безопасности и соответствия, таких как банковское дело, страхование, недвижимость, СМИ и здравоохранение.

Основные характеристики платформы обнаружения дипфейков Attestiv: https://attestiv.com/deepfake-video-detection-software/

– Бесплатная базовая версия с доступными премиум- и корпоративными опциями.

– Анализирует как загруженные видео, так и ссылки в социальных сетях.

– Предоставляет оценку и подробную разбивку поддельных элементов.

– Использует запатентованную фирменную технологию искусственного интеллекта и машинного обучения.

– Рассматривает контент генеративного ИИ, замену лиц, изменения синхронизации губ и другие правки.

– Применяет уникальные «отпечатки пальцев» к видео для будущих проверок подлинности.

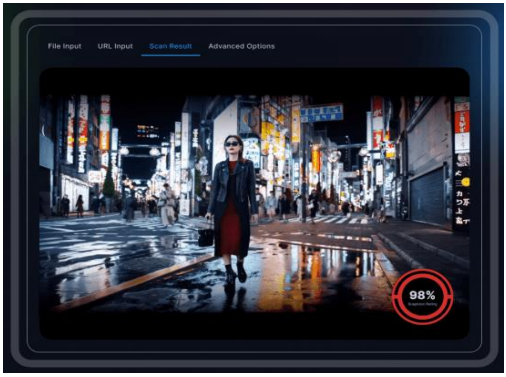

Рис.14. Программное обеспечение Attestiv
для обнаружения поддельных видео

Облачное программное обеспечение Attestiv для обнаружения deepfake-видео использует запатентованную фирменную технологию искусственного интеллекта и машинного обучения (ML) для обнаружения свидетельств фальсификации или синтетических элементов в медиафайлах. Это включает в себя deepfake, а также редактирование и другие изменения.

Процесс включает три основных этапа:

1. Загрузите видео через веб-приложение Attestiv Video или API

2. Анализ и обнаружение несанкционированного доступа

3. Проверка и отчетность

Видео захватываются и анализируются через приложение Attestiv Video или с использованием API. Процесс запускает криминалистическое сканирование, которое генерирует общий рейтинг подозрительности по шкале от 1 до 100, позволяя пользователю оценить подлинность видео.

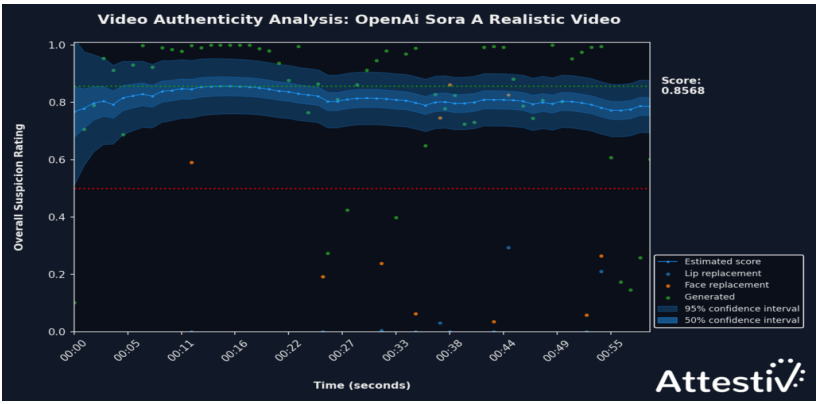

Рис.15. Программное обеспечение Attestiv
для обнаружения поддельных видео

Каждая из обученных моделей ИИ анализирует различные аспекты видео. Различные этапы процесса включают возможность проверки:

– **Контент генеративного ИИ:** Контент, созданный с использованием технологии генеративного ИИ.

– **Замена лица:** Контент, в котором лица субъектов были изменены.

– **Синхронизация губ или их замена:** контент, в котором речь и движения губ субъектов были изменены.

– **Изменения и правки** : Контент, измененный по сравнению с его первоначальной формой подозрительным образом [98].

3.2.3 Детектор дипфейков Intel в реальном времени

Рис. 16. Программное обеспечение FakeCatcher.

Intel представила детектор дипфейков в реальном времени, известный как FakeCatcher. Эта технология может обнаруживать поддельные видео с точностью 96%, возвращая результаты за миллисекунды. Детектор, разработанный в сотрудничестве с Умуром Чифтчи из Университета штата Нью-Йорк в Бингемтоне, использует аппаратное и программное обеспечение Intel, работает на сервере и взаимодействует через веб-платформу.

FakeCatcher ищет подлинные подсказки в реальных видео, оценивая то, что делает нас людьми – тончайший «кровяной поток» в пикселях видео. Когда наши сердца перекачивают кровь, наши вены меняют цвет. Эти сигналы кровотока собираются со всего лица, и алгоритмы переводят эти сигналы в пространственно-временные карты. Затем, используя глубокое обучение, он может мгновенно определить, является ли видео настоящим или фальшивым.

Рис.17. Программное обеспечение FakeCatcher

Основные характеристики детектора дипфейков в реальном времени от Intel:

– Разработано в сотрудничестве с Университетом штата Нью-Йорк в Бингемтоне.

– Может обнаруживать поддельные видео с точностью 96%

– Возвращает результаты в миллисекундах

– Использует тонкий «кровяной поток» в пикселях видео для обнаружения дипфейков [98].

3.2.4 WeVerify

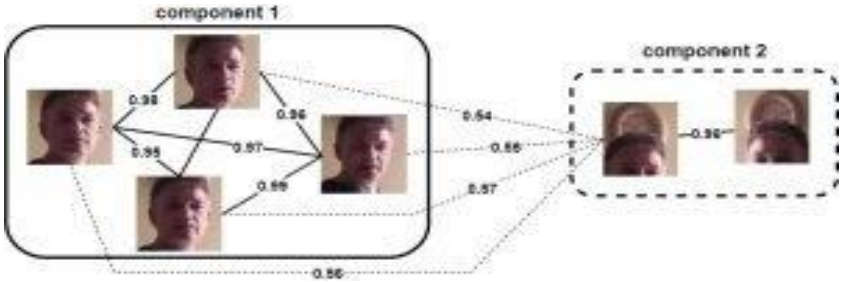

Рис.18. Программное обеспечение WeVerify

WeVerify – это проект, направленный на разработку интеллектуальных методов и инструментов для проверки контента и анализа дезинформации с участием человека. Проект направлен на анализ и контекстуализацию социальных сетей и веб-контента в более широкой онлайн-экосистеме для выявления сфабрикованного контента. Это достигается за счет

кросс-модальной проверки контента, анализа социальных сетей, микроцелевого разоблачения и общедоступной базы данных известных подделок на основе блокчейна.

Ключевые особенности WeVerify:

– Разрабатывает интеллектуальные методы и инструменты для проверки контента и анализа дезинформации с участием человека.

– Анализирует и контекстуализирует социальные сети и веб-контент

– Выявляет сфабрикованный контент посредством кросс-модальной проверки контента, анализа социальных сетей и развенчания микроцелей.

– Использует общедоступную базу данных известных подделок на основе блокчейна. [98]

3.2.5 Инструмент проверки подлинности видео от Microsoft

Рис.18. Программное обеспечение Microsoft Video Authenticator Tool

Инструмент Microsoft Video Authenticator Tool – это мощный инструмент, который может анализировать неподвижное фото или видео, чтобы предоставить оценку достоверности, указывающую, были ли манипуляции с мультимедиа. Он обнаруживает границу смешивания дипфейковых и тонких элементов в градациях серого, которые не видны чело-

веческому глазу. Он также предоставляет этот показатель достоверности в режиме реального времени, что позволяет немедленно обнаруживать дипфейки.

Video Authenticator Tool использует передовые алгоритмы искусственного интеллекта для анализа мультимедиа и обнаружения признаков манипуляций. Он ищет тонкие изменения в элементах медиа в оттенках серого, которые часто являются явным признаком дипфейка. Инструмент обеспечивает оценку достоверности в реальном времени, позволяя пользователям быстро определить, является ли носитель подлинным или нет.

Ключевые особенности инструмента Microsoft Video Authenticator:

– Анализирует неподвижные фотографии или видео.

– Обеспечивает оценку достоверности в реальном времени.

– Обнаруживает незначительные изменения оттенков серого.

– Позволяет мгновенно обнаруживать дипфейки. [98]

3.2.6 Обнаружение дипфейков с использованием несоответствий фонемы и виземы

Рис.19. Обнаружение дипфейков с использованием несоответствий фонемы и виземы

Этот инновационный метод, разработанный исследователями из Стэнфордского университета и Калифорнийского университета, использует тот факт, что висемы, которые обозначают динамику формы рта, иногда отличаются или несовместимы с произносимой фонемой. Это несоответствие является распространенным недостатком дипфейков, поскольку ИИ часто изо всех сил пытается идеально сопоставить движения рта с произнесенными словами.

Техника несоответствия фонемы и виземы использует передовые алгоритмы искусственного интеллекта для анализа видео и обнаружения этих несоответствий. Он сравнивает движения рта (виземы) с произносимыми словами (фонемами) и ищет любые несоответствия. Если обнаружено несоответствие, это явный признак того, что видео является дипфейком.

Ключевые особенности обнаружения дипфейков с использованием несоответствий фонемы и виземы:

– Разработано исследователями из Стэнфордского и Калифорнийского университетов.

– Использует несоответствия между виземами и фонемами в дипфейках.

– Использует передовые алгоритмы искусственного интеллекта для обнаружения несоответствий.

– Обеспечивает четкое указание на дипфейк при обнаружении несоответствия. [98]

3.3 Основные методы и подходы для распознавания речи с целью защиты от интернет-мошенничества

Для борьбы с интернет-мошенничеством, связанным с распознаванием речи, используются различные методы и инструменты на основе искусственного интеллекта (ИИ). Эти технологии направлены на предотвращение мошеннических действий, таких как подделка голосов, мошенничество через телефонные звонки и фальсификация аудио. Ниже приведены

основные методы и подходы, которые используются для распознавания речи с целью защиты от интернет-мошенничества.

1. Биометрия голоса

• *Speaker Verification* (Верификация диктора) – это процесс идентификации личности на основе её голосовых характеристик. Использование биометрии голоса позволяет защитить пользователей от подделки личности при помощи голосов, сгенерированных нейросетями.

– Методы: Свёрточные нейронные сети (CNN), рекуррентные нейронные сети (RNN), LSTM и трансформеры.

– Инструменты: Nuance, Pindrop, VoiceVault.

2. Anti-Spoofing (Антиспуфинг) технологии

– *Антиспуфинг* – это технологии, разработанные для обнаружения мошеннических попыток подделки голоса или использования синтетических голосов для обмана систем распознавания речи.

Методы: Комбинация свёрточных нейронных сетей (CNN) и рекуррентных нейронных сетей (RNN), анализ акустических особенностей голоса (тембр, частота), спектральный анализ.

Примеры использования: Мошенники могут пытаться воспроизводить голос владельца банковского аккаунта, чтобы обмануть голосовую биометрическую систему. Антиспуфинг-системы выявляют, является ли голос подлинным или сгенерированным.

3. Нейросетевые модели для распознавания мошеннических вызовов

– Использование глубоких нейронных сетей (DNN) и сверточных нейронных сетей (CNN) для выявления мошеннических звонков на основе паттернов речи, лексики и других признаков.

– Примеры: Эти модели могут анализировать тон, темп и другие признаки, чтобы выявить аномальные или подозрительные звонки, исходящие от потенциальных мошенников.

4. Системы на основе поведенческих паттернов

– Анализ поведенческих характеристик речи может помочь выявить мошенничество. Например, мошенники часто используют определенные ключевые слова, фразы или агрессивный тон, чтобы манипулировать жертвами.

Методы: Использование моделей машинного обучения для анализа лексики, синтаксиса, интонации и эмоциональной окраски речи.

Примеры: Системы могут анализировать телефонные разговоры для определения мошеннических схем, таких как фишинговые звонки или социальная инженерия.

5. Технологии обработки естественного языка (NLP)

– Использование обработки естественного языка (NLP) для анализа содержания разговоров или аудиозаписей может помочь выявить подозрительную активность или мошенничество. NLP может выявлять подозрительные запросы, манипулятивные вопросы или необычные структуры речи, характерные для мошеннических звонков.

Методы: Языковые модели, такие как BERT, GPT, используются для анализа смысла и контекста фраз.

Примеры: Системы могут анализировать звонки для выявления попыток фишинга, предложений с обманом или манипуляций.

6. AI на основе анализа аномалий

– Анализ аномалий в голосовых данных может помочь выявить мошенничество. Системы ИИ анализируют нормальные паттерны взаимодействий и затем выявляют отклонения, которые могут свидетельствовать о мошенничестве.

Методы: Использование алгоритмов кластеризации и методов аномального анализа (например, Isolation Forest, One-Class SVM).

Примеры: Распознавание нестандартного поведения звонящих, таких как частые звонки с разными акцентами или странные голосовые паттерны, может указывать на мошенничество.

7. Технологии для выявления deepfake-голосов

– Распознавание синтетической речи и deepfake-голосов играет важную роль в защите от мошенников, использующих сгенерированные голоса для подделки личности.

Методы: Использование моделей нейросетей для анализа мелочей, таких как артефакты генерации голоса, отсутствие естественных интонационных колебаний или временных несоответствий.

Примеры: Anti-Deepfake системы могут защитить компании от мошенничества, где синтетический голос используется для подделки идентификации.

8. Системы для анализа эмоциональной окраски речи

– Определение эмоционального состояния звонящего на основе его голоса. Мошенники могут использовать агрессивные или чрезмерно дружелюбные тона, чтобы манипулировать жертвами.

Методы: Нейросети для анализа интонации, тембра и скорости речи.

Примеры: Эмоциональные изменения могут помочь системам выявить стресс или настойчивость, которые могут свидетельствовать о мошеннических намерениях.

9. Технологии анализа звуковой среды

– Мошенники часто используют фоновые шумы или изменяют окружающую среду, чтобы подделать звонки или запутать систему. AI-системы могут анализировать шумы, эхо и другие аспекты окружающей среды, чтобы выявить несоответствия.

Методы: Спектральный анализ, FFT (Fast Fourier Transform) для выявления аномалий в фоне разговора.

Примеры: Системы могут распознавать изменения в звуковой среде, характерные для подделки голосов.

10. End-to-End системы защиты

– Некоторые компании предлагают end-to-end решения, которые интегрируют различные методы ИИ для защиты от голосовых мошенничеств, включая анализ речи, поведенческих паттернов, аномалий и антиспуфинг.

Примеры: Pindrop, VoiceIt, Nuance предлагают решения, которые сочетают биометрию голоса и антиспуфинг для обнаружения мошенников.

11. Инструменты облачных сервисов

– Google Cloud Speech-to-Text, Amazon Transcribe, и Microsoft Azure Speech предлагают встроенные решения для анализа речи, которые могут быть адаптированы для обнаружения мошенничества. Они включают технологии для анализа контекста, эмоциональной окраски и аномалий.

Методы: Комбинация NLP, анализа голоса и методов машинного обучения для выявления подозрительных разговоров.

Эти методы помогают защититься от интернет-мошенничества, связанного с распознаванием и подделкой речи, а также повышают уровень безопасности в таких областях, как банковские услуги, голосовая аутентификация и взаимодействие с клиентами.

3.4 Инструменты ИИ для распознавание генератор голоса

Существуют несколько инструментов на основе ИИ, которые помогают распознавать синтетическую речь, созданную генераторами голоса (такими как deepfake или TTS-системы). Эти инструменты используют различные методы анализа, включая акустические и языковые признаки, для выявления аномалий и артефактов, присущих синтетической речи.

Инструменты ИИ для распознавание генератор голоса:

– **Resemblyzer** – это открытая библиотека Python, которая анализирует и представляет голос в виде эмбеддингов. Эти эмбеддинги можно использовать для сравнения голосов, выявления синтетической речи или подделки. Библиотека может обнаруживать различия в характеристиках голоса, которые могут указывать на то, что речь была сгенерирована.

– **DeFake** – это ИИ-инструмент, специально разработанный для выявления фальшивой речи, созданной deepfake технологиями. Он использует глубокие нейронные сети для анализа аудиоданных и может определить, была ли речь создана искусственно.

– **DeepSonar** – это метод обнаружения поддельных голосов, сгенерированных deepfake технологиями, основанный на анализе временных характеристик. DeepSonar использует рекуррентные нейронные сети (RNN) и отслеживает признаки, характерные для синтетических голосов.

– **Syras** – это специализированный инструмент, разработанный для анализа и распознавания синтетической речи. Он использует алгоритмы машинного обучения для анализа звуковых сигналов и выявления несоответствий, характерных для генераторов голоса.

– **OpenAI's Jukebox** – создан для генерации музыки и аудио, его алгоритмы и технологии также могут использоваться для анализа синтетической речи. В том числе, модель может быть использована для понимания тонких отличий между реальной и синтетической речью.

– **VoiceGAN Detectors** – анализируют синтетическую речь, созданную с помощью генеративно-состязательных сетей (GAN). Эти детекторы пытаются обнаружить артефакты или искажения, которые могут присутствовать в голосах, созданных генераторами.

3.5 Инструменты ИИ для распознавание фальшивых документов

Искусственный интеллект (ИИ) активно используется для распознавания поддельных документов, и его применение в этой области становится все более эффективным благодаря следующим подходам:

1. Оптическое распознавание символов (OCR) с ИИ:

– Текстовый анализ: Технологии OCR на основе ИИ используются для распознавания и анализа текста в документах. ИИ может сравнивать текст с оригинальными базами данных или шаблонами, выявляя несоответствия или ошибки в шрифтах, форматировании и размещении текста, характерных для подделок.

– Семантический анализ: После извлечения текста ИИ может анализировать содержимое с точки зрения контекста.

Например, ошибки в логике, даты, или несоответствие информации могут быть выявлены как потенциальные признаки подделки.

2. Анализ изображений и водяных знаков:

– Распознавание шаблонов: ИИ способен анализировать изображение документа, включая мелкие детали, такие как микропечать, водяные знаки, голограммы и защитные элементы. С помощью машинного обучения ИИ может быть обучен на оригинальных документах и выявлять любые отклонения в изображениях.

– Цифровые следы редактирования: ИИ может выявить следы фотошопа или иных программ для редактирования, используя методы анализа изменений в текстуре, цветах, уровнях освещения и других характеристиках изображения.

3. Проверка метаданных:

– ИИ может анализировать метаданные цифровых документов, такие как информация о создании, изменении и копировании файлов. Несоответствия в этих данных могут свидетельствовать о фальсификации.

4. Анализ почерка и подписей:

– Сравнение с образцами: ИИ может быть обучен распознавать характерные черты почерка или электронной подписи, сравнивая их с подлинными образцами. Аномалии в движении пера или структуре подписи могут указывать на подделку.

– Динамика подписи: С помощью нейросетевых моделей можно анализировать не только статическую подпись, но и динамику ее создания (скорость, давление и т.д.), если данные были собраны, например, с помощью цифрового пера.

5. Идентификация структурных аномалий:

– Структурный анализ: ИИ может анализировать структуру документа на предмет логических и визуальных аномалий. Например, различные части документа, созданные разными шрифтами или форматированием, могут быть индикатором того, что документ был изменен.

6. Сравнение с базами данных:

– ИИ может проверять данные в документе с базами данных для выявления подделок. Например, при проверке паспортов ИИ может сравнивать данные с официальными реестрами.

7. Анализ цифровых подписей и сертификатов:

– В случае цифровых документов ИИ может проверять подлинность цифровых подписей и сертификатов. Используя криптографические методы, ИИ может выявить, была ли цифровая подпись изменена или недействительна.

Преимущества использования ИИ для распознавания поддельных документов:

– Скорость и автоматизация: ИИ может проверять документы быстрее и с меньшей вероятностью ошибок, чем человек.

– Обучаемость: Системы на базе ИИ могут обучаться на новых примерах поддельных документов, становясь все более точными с течением времени.

– Масштабируемость: ИИ может анализировать большое количество документов одновременно, что важно для банков, государственных учреждений и организаций, работающих с большим объемом данных.

Использование ИИ для распознавания фальшивых документов значительно повышает уровень безопасности и снижает риск человеческих ошибок.

Существует несколько программ на базе искусственного интеллекта (ИИ), которые применяются для распознавания поддельных документов. Эти решения используют комбинацию технологий машинного обучения, компьютерного зрения и анализа данных для выявления подделок в документах. Вот несколько популярных программ и сервисов:

1. ABBYY FineReader – это одна из ведущих программ для оптического распознавания символов (OCR), которая использует ИИ для распознавания текста и проверки подлинности документов.

– Функции:

Распознавание текста с высокой точностью.

Верификация документов с помощью встроенных механизмов анализа.

Анализ визуальных элементов, таких как водяные знаки, печати и подписи.

2. DocuSign Identify – сервис для проверки цифровых и бумажных документов, специализирующийся на подтверждении личности и валидации подписей.

– Функции:

Проверка цифровых подписей и сертификатов.

Верификация документов с помощью ИИ.

Сравнение данных с базами данных для выявления несоответствий.

3. Onfido – платформа для верификации личности и проверки подлинности документов с использованием искусственного интеллекта.

– Функции:

Автоматическое распознавание лиц и верификация документов (паспорта, водительские удостоверения и др.).

Использует ИИ для анализа изображения документа, выявления изменений и обнаружения фальшивых элементов.

Проверка метаданных и фотографий на соответствие.

4. TrustID – решение для проверки подлинности документов и верификации личности.

– Функции:

Автоматическое считывание данных с документов и проверка на наличие поддельных элементов.

Анализ изображений и штрих-кодов для выявления подделок.

Поддержка различных видов удостоверений (паспорта, визы, водительские права).

5. Fraud.net – платформа для обнаружения мошенничества с использованием ИИ, предназначенная для различных отраслей, включая финансы и электронную коммерцию.

– Функции:

Анализ транзакций и проверка подлинности документов с использованием ИИ и машинного обучения.

Обнаружение аномалий в документах и данных, что позволяет выявлять фальшивые документы.

Интеграция с другими системами для расширенного анализа.

6. HyperVerge – платформа для идентификации и верификации документов с использованием ИИ и глубокого обучения.

– Функции:

Распознавание документов, таких как паспорта, водительские удостоверения и национальные удостоверения личности.

Автоматическое выявление подделок через анализ изображения, текстуры и визуальных элементов.

Поддержка биометрической верификации (сравнение лица с фотографией в документе).

7. IDAnalyzer – онлайн-инструмент для проверки подлинности документов с использованием искусственного интеллекта.

– Функции:

Поддержка документов из более чем 190 стран.

Анализ документов для выявления поддельных элементов, изменений и подделок.

Проверка соответствия фотографий и данных в документах.

8. Sypht – платформа для анализа документов с использованием искусственного интеллекта.

– Функции:

Автоматическое извлечение данных из документов, таких как счета, контракты, удостоверения личности.

Использование ИИ для проверки подлинности данных и выявления несоответствий.

Обнаружение изменений в документах и выявление подделок.

9. Jumio Identity Verification – система для проверки личности и документов с помощью ИИ, которая используется банками и финансовыми учреждениями.

– Функции:

Автоматическое распознавание и проверка документов.

Верификация личности путем сравнения изображения лица с фотографией в документе.

Выявление подделок и мошенничества через анализ изображения и данных.

10. FacePhi – решение для биометрической идентификации и верификации документов с применением ИИ.

– Функции:

Использование ИИ для распознавания лиц и проверки документов.

Верификация подлинности документов через анализ визуальных элементов и текста.

Интеграция с банками и финансовыми сервисами для проверки документов в реальном времени.

Эти программы и платформы помогают организациям и пользователям автоматизировать процесс проверки подлинности документов и выявлять потенциальные фальсификации с высокой степенью точности.

Заключение

Использование искусственного интеллекта (ИИ) в борьбе с интернет-мошенничеством открывает новые горизонты для повышения уровня кибербезопасности. Методы машинного обучения и обработки больших данных позволяют более эффективно выявлять и предотвращать мошеннические схемы, анализируя поведение пользователей, а также идентифицируя аномалии в транзакциях и взаимодействиях.

Одним из ключевых преимуществ ИИ является его способность адаптироваться к постоянно изменяющимся схемам мошенничества. В отличие от традиционных систем защиты, основанных на фиксированных правилах, ИИ может самостоятельно обновлять свои модели, обучаясь на новых данных и обнаруживая новые угрозы.

Однако, несмотря на все преимущества, применение ИИ требует комплексного подхода. Технология должна использоваться в сочетании с другими методами киберзащиты и постоянно улучшаться. Не менее важно учитывать этические вопросы, такие как защита персональных данных и прозрачность алгоритмов.

Использование ИИ в выявлении интернет-мошенничеств представляет собой мощный инструмент для повышения уровня безопасности в цифровом пространстве. Однако для достижения наилучших результатов необходимо интегрировать ИИ в комплексную стратегию киберзащиты, которая учитывает как технические, так и этические аспекты.

Список использованных источников

1. ИИ переопределил компьютеры https://www.technologyreview.com/2021/10/22/1037179/aireinventing-computers/

2. Applications for artificial intelligence in Department of Defense cyber missions https://blogs.microsoft.com/on-the-issues/2022/05/03/artificialintelligence-department-of-defense-cyber-missions/

3. Намиот Д.Е., Ильюшин Е.А., Чижов И.В. (2022). Искусственный интеллект и кибербезопасность. International Journal of Open Information Technologies, 10(9), 135-147.

4. Афанасьева Д.В. (2020). Применение искусственного интеллекта в обеспечении безопасности данных. Известия Тульского государственного университета. Технические науки, 2, 151-154.

5. Власенко А.В., Киселёв П.С., Склярова Е.А. (2021). Искусственный интеллект и проблемы кибербезопасности. Технология Deepfake. Молодой ученый, (21), 81-86.

6. Magisterskaja programma Iskusstvennyj intellekt v kiberbezopasnosti https://cs.msu.ru/node/3732

7. Information Security Analysts https://www.bls.gov/ooh/computer-andinformation-technology/information-security-analysts.htm

8. Cybersecurity Workforce Study https://www.isc2.org/News-andEvents/Press-Room/Posts/2021/10/26/ISC2-Cybersecurity-Workforce-StudySheds-New-Light-on-Global-Talent-Demand

9. Kouliaridis, Vasileios, and Georgios Kambourakis. «A comprehensive survey on machine learning techniques for android malware detection» Information 12.5 (2021): 185.

10. AV-Test Institute https://www.av-test.org/en/statistics/malware/

11. ML for malware detection https://scholar.google.com/scholar?q=ml+for+malware+detection&hl=en&as _sdt=0,5

12. Yuan, Zhenlong, et al. «Droid-sec: deep learning in android malware detection» Proceedings of the 2014 ACM conference on SIGCOMM. 2014.

13. Vinayakumar, R., et al. «Robust intelligent malware detection using deep learning» IEEE Access 7 (2019): 46717-46738.

14. Using fuzzy hashing and deep learning to counter malware detection evasion techniques https://www.microsoft.com/security/blog/2021/07/27/combing-through-thefuzz-using-fuzzy-hashing-and-deep-learning-to-counter-malware-detectionevasion-techniques/

15. Tajaddodianfar, Farid, Jack W. Stokes, and Arun Gururajan. «Texception: a character/word-level deep learning model for phishing URL detection» ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.

16. Basnet, Ram, Srinivas Mukkamala, and Andrew H. Sung. «Detection of phishing attacks: A machine learning approach» Soft computing applications in industry. Springer, Berlin, Heidelberg, 2008. 373-383.

17. Divakaran, Dinil Mon, and Adam Oest. «Phishing Detection Leveraging Machine Learning and Deep Learning: A Review» arXiv preprint arXiv:2205.07411 (2022).

18. Shenfield, Alex, David Day, and Aladdin Ayesh. «Intelligent intrusion detection systems using artificial neural networks» Ict Express 4.2 (2018): 95-99.

19. Mishra, Preeti, et al. «A detailed investigation and analysis of using machine learning techniques for intrusion detection» IEEE Communications Surveys & Tutorials 21.1 (2018): 686-728.

20. Alsaheel, Abdulellah, et al. «{ATLAS}: A sequence-based learning approach for attack investigation» 30th USENIX Security Symposium (USENIX Security 21). 2021.

21. Ongun, Talha, et al. «Living-Off-The-Land Command Detection Using Active Learning» 24th International Symposium on Research in Attacks, Intrusions and Defenses. 2021.

22. Kok, S., et al. «Ransomware, threat and detection techniques: A review» Int. J. Comput. Sci. Netw. Secur 19.2 (2019): 136.

23. Wu, Yirui, Dabao Wei, and Jun Feng. «Network attacks detection methods based on deep learning techniques: a survey» Security and Communication Networks 2020 (2020).

24. Xin, Yang, et al. «Machine learning and deep learning methods for cybersecurity» IEEE Access 6 (2018): 35365-35381.

25. Noor, Umara, et al. «A machine learning framework for investigating data breaches based on semantic analysis of adversary's attack patterns in threat intelligence repositories» Future Generation Computer Systems 95 (2019): 467-487.

26. Mirsky, Yisroel, et al. «The threat of offensive ai to organizations» arXiv preprint arXiv:2106.15764 (2021)

27. Искусственный интеллект и кибербезопасность https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-i-kiberbezopasnost

28. Abdul Rehman Javed, Mirza Omer Beg, Muhammad Asim, Thar Baker, and Ali Hilal Al-Bayatti. 2020. AlphaLogger: Detecting motion-based side-channel attack using smartphone keystrokes. Journal of Ambient Intelligence and Humanized Computing (2020), 1-14

29. Philip Marquardt, Arunabh Verma, Henry Carter, and Patrick Traynor. 2011. (sp) iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers. In Proceedings of the 18th ACM conference on Computer and communications security. 551-562.

30. Y. Abid, Abdessamad Imine, and Michaël Rusinowitch. 2018. Sensitive Attribute Prediction for Social Networks Users. In EDBT/ICDT Workshop

31. Jian Jiang, Xiangzhan Yu, Yan Sun, and Haohua Zeng. 2019. A Survey of the Software Vulnerability Discovery Using Machine Learning Techniques. In International Conference on Artificial Intelligence and Security. Springer, 308-317.

32. Guanjun Lin, Sheng Wen, Qing-Long Han, Jun Zhang, and Yang Xiang. 2020. Software Vulnerability Detection Using Deep Neural Networks: A Survey. Proc. IEEE 108, 10 (2020), 1825-1848.

33. Serguei A. Mokhov, Joey Paquet, and Mourad Debbabi. 2014. The Use of NLP Techniques in Static Code Analysis to Detect Weaknesses and Vulnerabilities. In Advances in Artificial Intelligence, Marina Sokolova and Peter van Beek (Eds.). Springer International Publishing, Cham, 326-332

34. Yisroel Mirsky, Tom Mahler, Ilan Shelef, and Yuval Elovici. 2019. CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning. In 28th USENIX Security Symposium (USENIX Security 19). USENIX Association, Santa Clara, CA, 461-47

35. Vernit Garg and Laxmi Ahuja. 2019. Password Guessing Using Deep Learning. In 2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC). IEEE, 38-40.

36. Dongqi Han, Zhiliang Wang, Ying Zhong, Wenqi Chen, Jiahai Yang, Shuqiang Lu, Xingang Shi, and Xia Yin. 2020. Practical traffic-space adversarial attacks on learning-based nidss. arXiv preprint arXiv:2005.07519 (2020).

37. Yisroel Mirsky and Wenke Lee. 2021. The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR) 54, 1 (2021), 1-41.

38. Gavai, Gaurang, et al. «Detecting insider threat from enterprise social and online activity data» Proceedings of the 7th ACM CCS international workshop on managing insider security threats. 2015.

39. Zhou, Qingyu, et al. «Neural document summarization by jointly learning to score and select sentences» arXiv preprint arXiv:1807.02305 (2018).

40. Aniello Castiglione, Roberto De Prisco, Alfredo De Santis, Ugo Fiore, and Francesco Palmieri. 2014. A botnet-based command and control approach relying on swarm intelligence. Journal of Network and Computer Applications 38 (2014), 22-33.

Содержание

Верстка:
Туренова Б.Ю.

Отдел организации научно-исследовательской и редакционно-издательской работы Алматинской академии МВД Республики Казахстан имени М. Есбулатова 050060 Алматы, ул. Утепова, 29

Подписано в печать 12 ноября 2024 г.
Формат 60х84 1/16 Бум. тип. №1. Печать на ризографе. Уч.-изд. п.л. 2.
Тираж 50 экз.