DOI 10.37882/2223-2966.2026.02.14

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КОМПЬЮТЕРНЫХ АТАКАХ И ПРОТИВОДЕЙСТВИИ ТАКОВЫМ: КОМПЛЕКСНЫЙ ФИШИНГ, DEEPFAKE-МЕНЕДЖЕРЫ И АВТОМАТИЗИРОВАННЫЙ OSINT

APPLICATION OF ARTIFICIAL INTELLIGENCE IN COMPUTER ATTACKS AND THEIR COUNTERACTION: ADVANCED PHISHING, DEEPFAKE MANAGERS, AND AUTOMATED OSINT

A. Kim
I. Krepak

Summary. This scientific article is devoted to the analysis of the transformation of information security threats and information protection mechanisms under the influence of modern artificial intelligence technologies. The paper examines key areas of artificial intelligence application in offensive scenarios—from complex phishing with personalized social engineering attacks to deepfake managers capable of impersonating individuals in real time, as well as automated OSINT that accelerates the collection and correlation of target-related data. Defensive applications of artificial intelligence are also considered, including intelligent anomaly detection, countering deepfake content, behavioral authentication, and automated incident response. Special attention is paid to the asymmetry between attackers and victims, ethical and legal risks, and practical recommendations for actively countering AI-enhanced attacks.

Keywords: artificial intelligence, deepfake technologies, social engineering, automated OSINT, generative model, cyberattack, anomaly detection.

Ким Артур Сергеевич
Финансовый университет при Правительстве Российской Федерации, г. Москва
archikim1441@gmail.com
Крепак Иван Павлович
руководитель группы информационной безопасности, ООО Клиника Будь Здоров;
аспирант, Финансовый университет при Правительстве Российской Федерации, г. Москва
krepak.2311@yandex.ru

Аннотация. Статья посвящена анализу трансформации угроз информационной безопасности и средств защиты информации под влиянием современных технологий искусственного интеллекта. Рассматриваются ключевые направления использования искусственного интеллекта в наступательных сценариях — от комплексного фишинга с персонализированными социально-инженерными атаками до deepfake-менеджеров, способных имитировать физических лиц в режиме реального времени, а также автоматизированного OSINT, ускоряющего сбор и корреляцию данных о целях. Рассматриваются защитные применения искусственного интеллекта: интеллектуальное выявление аномалий, противодействие deepfake-содержимому, поведенческая аутентификация и автоматизация реагирования на инциденты. Особое внимание уделяется асимметрии между злоумышленником и жертвой, этическим и правовым рискам, а также практическим рекомендациям по активному противодействию ИИ-усиленным атакам.

Ключевые слова: искусственный интеллект, deepfake-технологии, социальная инженерия, автоматизированный OSINT, генеративная модель, кибератака, обнаружение аномалий.

Введение

Искусственный интеллект сейчас становится одним из ключевых факторов эволюции угроз информационной безопасности и средств защиты информации. Генеративные модели, машинное обучение и автоматизированный анализ данных позволяют злоумышленникам производить высокоточные компьютерные атаки, в рамках которых применяются комплексный фишинг, deepfake-менеджеры и авто-OSINT, значительно повышая их эффективность и латентность. Одновременно, искусственный интеллект (ИИ) активно внедряется в защитные механизмы — от поведенческого анализа и выявления аномалий до автоматизированного реагирования на инциденты информационной безопасности. В условиях нарастающей асимметрии между злоумышленниками и жертвами, особое значение приобретает понимание принципов работы ИИ-усиленных компьютерных атак и методов противодействия таковым, что и определяет цель данной статьи.

Эволюция киберугроз в эпоху искусственного интеллекта

Искусственный интеллект существенно изменил характер современных угроз информационной безопасности, сместив фокус от массовых атак к точечным и адаптивным сценариям. Генеративные модели и ал-

горитмы машинного обучения позволяют автоматизировать подготовку компьютерных атак, анализировать поведение целей и создавать правдоподобные социально-инженерные сценарии [5].

Доступность ИИ-инструментов привела к росту количества целевых компьютерных атак, в которых фишинг сообщения и коммуникации формируются с учётом контекста, роли и цифрового профиля жертвы. Это снижает эффективность традиционных защитных механизмов, основанных на сигнатурах и шаблонах.

Отдельную роль играет автоматизированный OSINT, ускоряющий сбор и корреляцию данных из открытых источников. Использование ИИ позволяет масштабировать этап разведки без потери качества, что напрямую повышает успешность кибератак. В результате формируется асимметричная среда, где злоумышленники быстрее адаптируются к изменениям, а защитные системы вынуждены переходить к интеллектуальным и поведенческим методам анализа.

Комплексный фишинг и deepfake-менеджеры как ключевые инструменты злоумышленника

Сегодня комплексный фишинг, который так же имеет название «Фишинг 2.0», представляет собой результат эволюции социально-инженерных компьютерных атак, усиленных возможностями искусственного интеллекта. Генеративные модели позволяют создавать персонализированные сообщения, адаптированные под роль, контекст и стиль коммуникации конкретной жертвы, что сильно повышает эффективность компьютерных атак по сравнению с традиционным фишингом [1].

Особое место занимают deepfake-менеджеры. Это атаки, основанные на синтезе голоса и видео доверенных лиц (коллеги, руководители, родственники и друзья). Такие сценарии часто реализуются через мессенджеры и средства видеосвязи. Сопровождаются элементами срочности и давления, характерными для компьютерной атаки «CEO fraud». Жертва при этом взаимодействует с ИИ-системой, способной в реальном времени имитировать поведение реального человека. Ход такой атаки обозначен в соответствующей блок-схеме, обозначенной на Рисунке 1.

Автоматизированный OSINT и роль искусственного интеллекта в подготовке компьютерных атак

Автоматизированный OSINT является важным элементом подготовки современных целевых компьютерных атак. Использование искусственного интеллекта позволяет быстро собирать и анализировать данные из открытых источников, выявляя связи и характеристики целей, недоступные при ручной разведке [4].

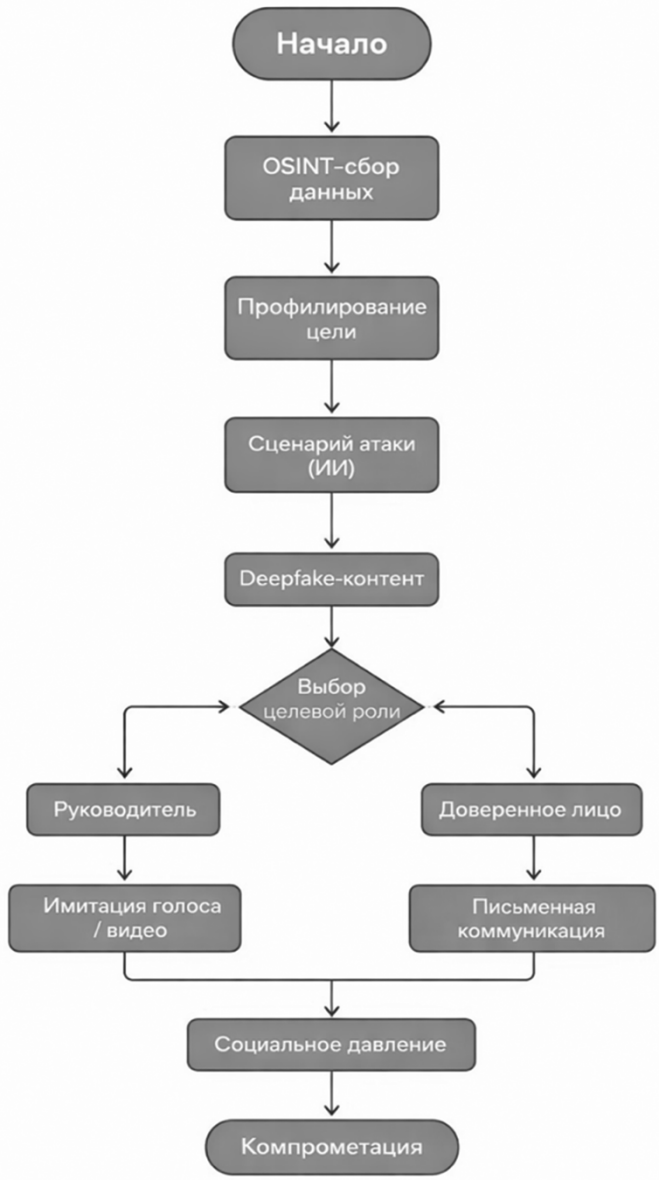

Рис. 1. Блок-схема реализации комплексного фишинга через deepfake-менеджера

Технически авто-OSINT реализуется в виде многоэтапной кибератаки, включающей агрегацию, фильтрацию и корреляцию данных. Алгоритмы машинного обучения формируют профиль цели, повышая точность и эффективность социальной инженерии [4]. ИИ также применяется для приоритизации целей и выбора необходимого сценария компьютерной атаки, включая канал и форму коммуникации. Это позволяет масштабировать целевые кибератаки при минимальных финансовых затратах и снижении частоты ошибок.

Искусственный интеллект в средствах защиты информации

Искусственный интеллект активно применяется для противодействия ИИ-усиленным компьютерным атакам,

таким как комплексный фишинг и deepfake-социальная инженерия. В отличие от традиционных сигнатурных средств защиты, ИИ-ориентированные решения используют преимущественно поведенческий анализ, выявление аномалий и контекстную оценку риска, что позволяет обнаруживать ранее неизвестные и модифицированные правонарушения [3].

Ключевым элементом защиты является многоэтапный анализ входящих коммуникаций: сообщений, звонков и видеосессий. На данном этапе проверяется источник, стиль коммуникации, частотные и лингвистические признаки, а также поведенческие отклонения от нормы. Общая логика такого процесса представлена на Рисунке 2, где показан конвейер обнаружения и реагирования на ИИ-усиленные атаки.

Рис. 2. Блок-схема алгоритма применения искусственного интеллекта в средствах защиты информации

Программное обоснование

На практике защитные механизмы часто строятся по принципу каскадной оценки риска. Сначала извлекаются простые признаки (наличие срочности, упоминание переводов денежных средств, запросы учетных данных), затем, вычисляется итоговый скоринг, на основе которого принимается решение о блокировке или эскалации инцидента в SOC. Даже без сложных нейросетевых моделей такой подход позволяет отсеивать значительную часть фишинговых атак на раннем этапе.

Ниже (Листинг 1) приведен пример Python-скрипта, реализующего базовый детектор фишинг сообщений. Он решает задачу первичной оценки риска и может использоваться как вспомогательный модуль в почтовом и мессенджер фильтре.

```
def phishing_risk_score(message: str) -> int:
keywords = [«срочно», «немедленно», «перевести», «подтвердить», «пароль», «доступ», «финансы», «платеж»]
score = 0
message_lower = message.lower()
for word in keywords:
if word in message_lower:
score += 1
if «@» in message_lower or «http» in message_lower:
score += 1
return score
msg = «Срочно переведите средства и подтверди доступ к аккаунту»
risk = phishing_risk_score(msg)
if risk >= 3:
print («Высокий риск фишинга»)
else:
print («Низкий риск»)
```

Листинг 1. Пример оценки риска фишингового сообщения

Данный код анализирует текст сообщения, определяет количество подозрительных признаков и формирует итоговую оценку риска. Несмотря на простоту, подобный механизм эффективно выявляет сообщения с элементами давления и срочности, характерные для комплексного и компьютерных атак класса «CEO fraud» [1] [5]. В сочетании с более сложными моделями машинного обучения, анти-deepfake проверками и автоматизированным реагированием такие решения формируют многоуровневую систему защиты информации, способную адаптироваться к современным ИИ-усиленным угрозам.

Сравнительный анализ и практические рекомендации

Развитие ИИ приводит к одновременному усилению наступательных и защитных возможностей. Для главных действий злоумышленника приоритетом являются автоматизация, персонализация и масштабируемость, тогда как защитные решения ориентированы на раннее выявление аномалий и снижение влияния человеческого фактора. Сравнение ключевых подходов представлено в Таблице 1.

Таблица 1.
Сравнение ИИ-усиленных компьютерных атак и средств защиты информации

Аспект	Наступательные действия (комплексный фишинг, deepfake и OSINT автоматизация)	Защита (ИИ в информационной безопасности)
Цель	Компрометация данных и несанкционированный доступ к финансовым активам.	Предотвращение и минимизация количества инцидентов информационной безопасности
Основные технологии	Генеративные модели, OSINT автоматизация и deepfake	ML-анализ аномалий и анти-deepfake
Уровень автоматизации	Высокий	Высокий
Масштабируемость	Массовая персонализация	Централизованная защита
Участие человека	Минимальная на этапе атаки	Ключевая на этапе реагирования
Основные признаки	Срочность, давление, имитация авторитета	Поведенческие отклонения и аномалии
Время реакции	Минимальное	Зависит от зрелости SOC

На основе проведенного анализа можно сформулировать ряд практических рекомендаций. Организациям следует внедрять многоуровневые системы защиты, сочетающие ИИ-инструменты с процессами подтверждения действий через независимые каналы связи. Особое внимание необходимо уделять обучению сотрудников распознаванию признаков комплексного фишинга и компьютерных атак с использованием deepfake. Кроме того, важно интегрировать автоматизированные средства анализа и реагирования в процессы SOC, обеспечивая быструю эскалацию инцидентов и снижение потенциального ущерба.

Заключение

Искусственный интеллект существенно трансформирует как методы компьютерных атак, так и средства киберзащиты. Фишинг 2.0, deepfake-менеджеры и автоматизированный OSINT демонстрируют переход к более точным, адаптивным и латентным компьютерным атакам. В то же время ИИ становится ключевым инструментом защиты информационных систем, обеспечивая постоянный поведенческий анализ, выявление аномалий и автоматизацию реагирования на инциденты.

Искусственный интеллект радикально меняет ландшафт кибератак и методов защиты, создавая новую реальность информационной безопасности. Современные угрозы информационной безопасности становятся всё более сложными и скрытыми. Однако, ИИ также играет ключевую роль в защите корпоративных сетей и персональных устройств, предлагая инструменты для непрерывного мониторинга поведения пользователей, выявления подозрительных действий и автоматического устранения угроз. Для эффективного противостояния таким рискам необходима целостная стратегия, включающая современные технические средства, строгие внутренние регламенты и постоянное обучение сотрудников принципам цифровой гигиены.

ЛИТЕРАТУРА

1. Фишинг 2.0: как мессенджеры и дипфейки стали оружием кибермошенников // Anti-Malware [Электронный ресурс]. — 2025. — URL: https://www.anti-malware.ru/analytics/Threats_Analysis/Phishing20-messengers-deepfakes (дата обращения: 18.12.2025).
2. Искусственный интеллект в кибербезопасности: баланс угроз и защитных технологий // Хабр [Электронный ресурс]. — 2025. — URL: https://habr.com/ru/companies/solarsecurity/articles/954744/ (дата обращения: 18.12.2025).
3. Искусственный интеллект в киберзащите // Positive Technologies [Электронный ресурс]. — 2025. — URL: https://ptsecurity.com/research/analytics/iskusstvennyi-intellekt-v-kiberzaschite/ (дата обращения: 19.12.2025).
4. OSINT 2025: Полное Руководство по Инструментам, AI и Автоматизации Разведки // Codeby.net [Электронный ресурс]. — 2025. — URL: https://codeby.net/threads/osint-2025-polnoye-rukovodstvo-po-instrumentam-ai-i-avtomatizatsii-razvedki.88590/ (дата обращения: 19.12.2025).
5. ИИ в информационной безопасности: от генерации фишинга до анализа уязвимостей // Хабр [Электронный ресурс]. — 2025. — URL: https://habr.com/ru/companies/ru_mts/articles/966440/ (дата обращения: 20.12.2025).

Журнал «Современная наука: актуальные проблемы теории и практики»