ORES
Научно-практический электронный журнал
Оригинальные исследования (ОРИС), *ISSN 2222-0402*
www.ores.su
original-research.ru

УДК 004.89

АНАЛИЗ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ТИПОВ КИБЕРАТАК С ЕГО ПРИМЕНЕНИЕМ. РАЗРАБОТКА МАТРИЦЫ УГРОЗ

Папоротный Никита Владимирович,
соискатель,
Череповецкий государственный университет. г. Череповец,
Вологодской области
nikita.paporotny007@gmail.com

Аннотация

В статье рассматривается применение ИИ в кибератаках.Проведен анализ техник и тактик матриц MITRE ATT&CK. Определены техники в которых возможно применить ИИ. Приведены реальные примеры кибератак с применением ИИ. Перечислены модели ИИ, которые могут использоваться для проведения кибератак. Разработана матрица угроз кибербезопасности с использованием ИИ.

Ключевые слова: вредоносное программное обеспечение, операционная система, дипфейк, MITRE ATT&CK, C&C, кибербезопасность, искусственный интеллект.

ANALYSIS OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES AND TYPES OF CYBER-ATTACKS USING IT. DEVELOPMENT OF A THREAT MATRIX

Paporotny Nikita Vladimirovich,
applicant,
Cherepovets State University. Cherepovets, Vologda Region
nikita.paporotny007@gmail.com

ABSTRACT

The article discusses the use of AI in cyber attacks. The analysis of techniques and tactics of MITRE ATT&CK matrices is carried out. The techniques in which it is possible to apply AI have been identified. Real examples of cyberattacks using AI are given. The AI models that can be used to carry out cyber attacks are listed. A matrix of cybersecurity threats using AI has been developed.

Keywords: malware, operating system, deepfake, MITRE ATT&CK, C&C, cybersecurity, artificial intelligence.

Системный анализ один из самых эффективных инструментов, который могут применять специалисты по информационной безопасности для защиты своих систем [1].

Применение системного анализа позволяет определить факторы, влияющие на производительность и надежность системы, а также выявить угрозы безопасности и возможные пути их решения. Для этого необходимо провести анализ данных, проанализировать существующие угрозы и уязвимости, а также разработать стратегию повышения безопасности системы.

Обратимся к авторитетному ресурсу MITRE ATT&CK [2], который классифицирует и описывает тактики, техники и процедуры, используемые злоумышленниками в ходе кибератак, для анализа возможности применения ИИ.

Тактика «Reconnaissance» тактика «Active Scanning», ИИ может: обучаться на прошлых сканированиях и настраивать параметры сканирование (порты, скорость, тип сканирования) для определение открытых портов, сервисов, используемых операционных системах.

Тактика «Reconnaissance» тактика «Gather Victim Host Information», анализировать веб-сайты, социальные сети и другие источники информации для сбора информации о целевых системах (ОС, ПО, конфигурации).

Тактика «Resource Development» техника «Compromise Infrastructure», ИИ может выбирать наиболее надежные и трудные для обнаружения хостинги и серверы управления (C&C), адаптируясь к блокировкам.

Тактика «Resource Development» техника «Develop Capabilities», ИИ может генерировать новые варианты вредоносного кода, обходя сигнатурные антивирусы, создавать эксплойты на основе анализа уязвимостей.

Тактика «Initial Access» техника «Phishing», ИИ может создавать фишинговые письма, адаптированные к конкретным пользователям, с использованием знаний об их интересах.

Тактика «Execution» техника «User Execution», ИИ может анализировать поведение пользователей, чтобы определить наиболее эффективные способы убедить их запустить вредоносный код.

Тактика «Defense Evasion» техника «Obfuscated Files or Information», ИИ может автоматически обфусцировать код и данные, чтобы усложнить их анализ и обнаружение, адаптируя к конкретным методам защиты.

Тактика «Credential Access» техника «Credential Stuffing», ИИ может анализировать украденные учетные данные (имена пользователей и пароли), чтобы определить вероятные комбинации.

Тактика «Discovery» техника «Network Service Scanning», ИИ может анализировать сетевой трафик и историю сканирования, чтобы оптимизировать процесс сканирования, фокусируясь на наиболее уязвимых сервисах.

Тактика «Lateral Movement» техника «Exploitation of Remote Services», ИИ может автоматически эксплуатировать известные уязвимости в удаленных сервисах, используя готовые эксплойты или генерирую новые.

Тактика «Command and Control» техника «Dynamic Resolution», ИИ может выбрать C&C серверы и изменять их расположение, чтобы избежать блокировок.

Тактика «Exfiltration» техника «Automated Exfiltration», ИИ может автоматически идентифицировать и извлекать конфиденциальную информацию из целевой системы, минимизируя ручную работы злоумышленника.

Тактика «Impact» техника «Data Destruction», ИИ может планировать и оптимизировать процесс уничтожения данных, чтобы нанести максимальный ущерб и избежать возможности восстановления.

Приведем реальные примеры кибератак:

1. DeepLocker (таргентированное ВПО с распознаванием лиц). DeepLocker – экспериментальное ВПО, разработанное компанией IBM X-Force, которое использует ИИ для таргентирование жертв. Оно остается скрытым до тех пор, пока не обнаружит определенное лицо через систему распознавания лиц [3, 4].
2. Фишинговые атаки. Исследователи компании «Barracuda Networks» обнаружили фишинговые атаки, использующие ИИ для создания более убедительных и персонализированных писем. Эти письма сложно обнаружить, поскольку они написаны на естественном языке и имитируют стиль общения конкретных людей [5].
3. Обход CAPTCHA. Злоумышленники используют ИИ для автоматического решения CAPTCHA, обходя эту меру защиты, предназначенную для защиты веб-сайтов от ботов [6].
4. Fuzzing (автоматический поиск уязвимостей). Microsoft использует машинное обучение используется для анализа входных и выходных данных для выявления наиболее вероятных уязвимостей [7, 8].
5. Crefential Stuffing (подстановка учетных данных). В отчете [9] компании «Akamai Technologies» явно не говорится об применении ИИ, но подразумевается, что для анализа больших объемов данных и оптимизации атак часто используются автоматизированные инструменты, в том числи с применением ИИ.
6. Атаки на модели ИИ. В этих статьях обсуждаются способы защиты от атак на ИИ, что подразумевает существование самих атак, в том числи с применением ИИ [10, 11].
7. Генерация дипфейков для социальной инженерии. В 2024 году мошенники использовали технологии DeepFake, чтобы выдать себя за главного финансового директора компании [12].

Перечислим модели ИИ, которые могут применяться для проведения кибератак:

1. Машинное обучение (ML). Общий термин для алгоритмов, которые позволяют учиться на данных без явного программирования. Алгоритм сам строит модель, которую можно использовать для прогнозирования или классификации [13, 14].
2. Обработка естественного языка (NLP). Модель машинного обучения, которая дает возможность компьютерам интерпретировать, манипулировать, генерировать и понимать человеческий язык как текстовый, так и голосовой [15, 16].
3. Компьютерное зрение (CV). Модель ИИ использующая алгоритмы машинного обучения и глубокого обучения для распознавания и интерпретации объектов на изображениях и видео. Компьютерное зрение позволяет компьютерам определять и анализировать предметы на фотографиях и видео точно так же, как это делают люди [17, 18].
4. Обучение с подкрепление (RL). Модель машинного обучения, в которой алгоритм, вместо получения на вход обучающей выборки будет взаимодействовать с некоторой средой , а в роли «разметки» будет выступать «награда»— скалярная величина, которая выдаётся после каждого взаимодействия со средой и показывает, насколько хорошо алгоритм справляется с поставленной ему задачей [19, 20].
5. Генеративные модели (GANs). Тип моделей машинного обучения, способных генерировать новые данные, похожие на те, на которых были обучены [21, 22].

На основании всего вышеизложенного разработана матрица угроз см. Таблица 1.

Таблица 1

Матрица угроз кибербезопасности с использованием ИИ

Модели ИИ / Типы кибератак	Машинное обучение (ML)	Обработка естественного языка (NLP)	Компьютерное зрение (CV)	Обучение с подкреплением (RL)	Генеративные модели (GANs)
	1	2	3	4	5
Фишинг	(4) Пример: Автоматическая генерация персонализированных фишинговых писем на основе анализа данных жертвы	(5) Пример: Генерация убедительных фишинговых писем, имитирующих стиль общения конкретных лиц (компаний)	(3) Пример: Обход CAPTCHA, использование логотипов компаний в фишинге, клонирование веб-сайтов	(2) Пример: Оптимизация рассылки фишинговых писем для максимальных кликов, учет времени и дня недели	(4) Пример: Генерация персонализированных дипфейков для повышения убедительности фишинга
Вредоносное ПО	(5) Пример: Обход антивирусных систем, изменение сигнатур и мутации	(3) Пример: Анализ кода для поиска уязвимостей	(2) Пример: Анализ изображений (скриншотов) для поиска ценной информации	(4) Пример: Адаптация ВПО к защитным механизмам системы безопасности	(5) Пример: Генерация новых, ранее неизвестных типов ВПО (ВПО 0-го дня)
DDoS	(4) Пример: Адаптация атак к изменениям в сетевой инфраструктуре и системе защиты в режиме реального времени			(5) Пример: Оптимизация маршрутов атак для обхода систем обнаружения вторжений (IDS)	(3) Пример: Генерация аномального трафика, сложного для фильтрации
Атаки на пароли	(4) Пример: Интеллектуальный подбор паролей на основе анализа данных о пользовател	(3) Пример: Анализ структуры в известных (слитых) паролях		(2) Пример: Адаптация алгоритмов взлома к системам аутентификации	(4) Пример: Генерация подобных паролей на основе известных

	е и популярных паролей				
Социальная инженерия	(3) Пример: Прогнозирование поведения жертвы для повышения эффективности	(4) Пример: Создание убедительных легенд и сценариев для обмана жертвы	(5) Пример: Использование дипфейков для выдачи себя за доверенных лиц	(2) Пример; Оптимизация методов социальной инженерии на основе обратной связи от жертв	(4) Пример: Генерация персонализированных текстовых и визуальных сообщений
Атаки на цепочки поставок	(4) Пример: Анализ компаний-поставщиков для выявления уязвимых звеньев	(3) Пример: Генерация фейковых писем, дезинформацией	(3) Пример: Анализ визуальных данных для идентификации оборудования и систем, используемых поставщиками	(3) Пример: Поиск и компрометация уязвимых репозиториев, которые используются в качестве зависимостей в других проектах	(2) Пример: Анализ данных о поставщиках для выявления потенциальных скомпрометированных систем
Взлом систем ИИ	(5) Пример: Атака на модели машинного обучения для искажения результатов и получения несанкционированного доступа		(3) Пример: Атака на системы распознавания лиц для обхода аутентификации	(4) Пример: Использование RL поиска для обнаружения уязвимостей в системах ИИ	(5) Генерация вводимых данных для обхода защитных механизмов ИИ

Пояснения к таблице 1:

1. Типы атак (строки): перечень различных типов кибератак.
2. Модели ИИ (столбцы): перечень моделей ИИ, которые могут быть использованы в кибератаке.
3. Уровень информационного риска (шкала от 1 до 5 в скобках): указывает на потенциальный уровень риска (1 – минимальный, 5 –максимальный) для каждой комбинации модели ИИ и типа атаки.
4. Примеры (в ячейке): краткое описание применения ИИ в кибератаках. Искусственный интеллект (ИИ) трансформирует ландшафт кибербезопасности, и не только в положительном ключе. Злоумышленники все чаще используют ИИ для усиления и автоматизации своих атак, делая их более сложными, эффективными и трудными для обнаружения. Это создает серьезные проблемы для организаций, стремящихся защитить себя от киберугроз. Понимание этих и разработка стратегий защиты от них становится

критически важной задачей для всех участников цифрового пространства. Организации должны разрабатывать стратегии кибербезопасности, учитывающие угрозы, связанные с использованием ИИ. Крайне важно следить за новейшими разработками в области ИИ и машинного обучения, чтобы быть в курсе новых угроз и разрабатывать эффективные методы защиты.

Список литературы:

1. Как использовать системный анализ для определения угроз безопасности: кратко для всех: Хабр. [Электронный ресурс]. URL: https://habr.com/ru/companies/otus/articles/727378/#%D1%88%D0%B0%D0%B3%D0%B8 (дата обращения: 08.08.2025)
2. ATT&CK Matrix for Enterprise: MITRE ATT&CK. [Электронный ресурс]. URL: https://attack.mitre.org/matrices/enterprise/ (дата обращения: 08.08.2025).
3. AI's role in new cyber security frontier: IBM Security. [Электронный ресурс]. URL: https://www2.gov.bc.ca/assets/gov/british-columbians-our-governments/services-policies-for-government/information-management-technology/information-security/information-security-awareness/sheik_sahib_-_ibm.pdf (дата обращения: 08.08.2025).
4. DeepLocker - Concealing Targeted Attacks with AI Locksmithing: Black Hat. [Электронный ресурс]. URL: https://www.blackhat.com/us-18/briefings/schedule/index.html#deeplocker---concealing-targeted-attacks-with-ai-locksmithing-11549 (дата обращения: 08.08.2025).
5. 5 Ways cybercriminals are using AI: Phishing: Barracuda. [Электронный ресурс]. URL: https://blog.barracuda.com/2024/03/28/-5-ways-cybercriminals-are-using-ai--phishing (дата обращения: 08.08.2025).
6. Как работает решатель капчи на базе ИИ: от OCR до глубокого обучения: Хабр. [Электронный ресурс]. URL: https://habr.com/ru/articles/913778/ (дата обращения: 08.08.2025).
7. Learn&Fuzz: Machine Learning for Input Fuzzing: Microsoft. [Электронный ресурс]. URL: https://www.microsoft.com/en-us/research/publication/learnfuzz-machine-learning-for-input-fuzzing/ (дата обращения: 08.08.2025).
8. Learn&Fuzz: Machine Learning for Input Fuzzing: Microsoft. [Электронный ресурс]. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2017/01/main-public.pdf (дата обращения: 08.08.2025)
9. Credential Stuffing: Volume 5, Special Media Edition Attacks and Economies: Akamai. [Электронный ресурс]. URL: https://www.akamai.com/site/en/documents/state-of-the-internet/soti-security-credential-stuffing-attacks-and-economies-report-2019.pdf (дата обращения: 08.08.2025).
10. Adversarial attacks and robustness for quantum machine learning: Pennylane. [Электронный ресурс]. URL: https://pennylane.ai/qml/demos/tutorial_adversarial_attacks_QML (дата обращения: 08.08.2025).

11. Making Machine Learning Robust Against Adversarial Inputs: ACM. [Электронный ресурс]. URL: https://dl.acm.org/doi/pdf/10.1145/3134599 (дата обращения: 08.08.2025).
12. AI-Driven Phishing And Deep Fakes: The Future Of Digital Fraud: Forbes. [Электронный ресурс]. URL: https://www.forbes.com/councils/forbestechcouncil/2025/03/10/ai-driven-phishing-and-deep-fakes-the-future-of-digital-fraud/ (дата обращения: 08.08.2025).
13. Машинное обучение: общие принципы и концепции: Хабр. [Электронный ресурс]. URL: https://habr.com/ru/articles/862704/ (дата обращения: 08.08.2025).
14. Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, Xavier Bresson. Graph Neural Networks with Learnable Structural and Positional Representations [Электронный ресурс]. URL: https://openreview.net/pdf?id=wTTjnvGphYj (дата обращения: 08.08.2025).
15. 15 примеров применения Natural Language Processing: Хабр. [Электронный ресурс]. URL: https://habr.com/ru/companies/otus/articles/930130/ (дата обращения: 08.08.2025).
16. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention Is All You Need [Электронный ресурс]. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (дата обращения: 08.08.2025).
17. Шесть шагов для создания более качественных моделей Computer Vision: Хабр. [Электронный ресурс]. URL: https://habr.com/ru/articles/705008/ (дата обращения: 08.08.2025).
18. Gedas Bertasius, Heng Wang, Lorenzo Torresani. Is Space-Time Attention All You Need for Video Understanding? [Электронный ресурс]. URL: https://proceedings.mlr.press/v139/bertasius21a/bertasius21a.pdf (дата обращения: 08.08.2025).
19. 11.1. Обучение с подкреплением: Яндекс Образование. [Электронный ресурс]. URL: https://education.yandex.ru/handbook/ml/article/obuchenie-s-podkrepleniem (дата обращения: 08.08.2025).
20. David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, Demis Hassabis. Mastering the Game of Go without Human Knowledge [Электронный ресурс]. URL: https://discovery.ucl.ac.uk/id/eprint/10045895/1/agz_unformatted_nature.pdf (дата обращения: 08.08.2025).
21. Генеративный ИИ — Будущее или просто Хайп?: Хабр. [Электронный ресурс]. URL: https://habr.com/ru/companies/serverspace/articles/754768/ (дата обращения: 08.08.2025).
22. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Networks. [Электронный

ресурс]. URL: https://dl.acm.org/doi/pdf/10.1145/3422622 (дата обращения: 08.08.2025).

References:

1. How to Use Systems Analysis to Identify Security Threats: Briefly for Everyone: Habr. [Electronic resource]. URL: https://habr.com/ru/companies/otus/articles/727378/#%D1%88%D0%B0%D0%B3%D0%B8 (date accessed: 08.08.2025).
2. ATT&CK Matrix for Enterprise: MITRE ATT&CK. [Electronic resource]. URL: https://attack.mitre.org/matrices/enterprise/ (date accessed: 08.08.2025).
3. AI's role in new cyber security frontier: IBM Security. [Electronic resource]. URL: https://www2.gov.bc.ca/assets/gov/british-columbians-our-governments/services-policies-for-government/information-management-technology/information-security/information-security-awareness/sheik_sahib_-_ibm.pdf (date accessed: 08.08.2025).
4. DeepLocker - Concealing Targeted Attacks with AI Locksmithing: Black Hat. [Electronic resource]. URL: https://www.blackhat.com/us-18/briefings/schedule/index.html#deeplocker---concealing-targeted-attacks-with-ai-locksmithing-11549 (date accessed: 08.08.2025).
5. 5 Ways cybercriminals are using AI: Phishing: Barracuda. [Electronic resource]. URL: https://blog.barracuda.com/2024/03/28/-5-ways-cybercriminals-are-using-ai--phishing (date accessed: 08.08.2025).
6. How an AI Captcha Solver Works: From OCR to Deep Learning: Habr. [Electronic resource]. URL: https://habr.com/ru/articles/913778/ (date accessed: 08.08.2025).
7. Learn&Fuzz: Machine Learning for Input Fuzzing: Microsoft. [Electronic resource]. URL: https://www.microsoft.com/en-us/research/publication/learnfuzz-machine-learning-for-input-fuzzing/ (date accessed: 08.08.2025).
8. Learn&Fuzz: Machine Learning for Input Fuzzing: Microsoft. [Electronic resource]. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2017/01/main-public.pdf (date accessed: 08.08.2025)
9. Credential Stuffing: Volume 5, Special Media Edition Attacks and Economies: Akamai. [Electronic resource]. URL: https://www.akamai.com/site/en/documents/state-of-the-internet/soti-security-credential-stuffing-attacks-and-economies-report-2019.pdf (date accessed: 08.08.2025).
10. Adversarial attacks and robustness for quantum machine learning: Pennylane. [Electronic resource]. URL: https://pennylane.ai/qml/demos/tutorial_adversarial_attacks_QML (date accessed: 08.08.2025).
11. Making Machine Learning Robust Against Adversarial Inputs: ACM. [Electronic resource]. URL: https://dl.acm.org/doi/pdf/10.1145/3134599 (date accessed: 08.08.2025).
12. AI-Driven Phishing And Deep Fakes: The Future Of Digital Fraud: Forbes. [Electronic resource]. URL: https://www.forbes.com/councils/forbestechcouncil/2025/03/10/ai-driven-phishing-and-deep-fakes-the-future-of-digital-fraud/ (date accessed: 08.08.2025).

13. Machine learning: general principles and concepts: Habr. [Electronic resource]. URL: https://habr.com/ru/articles/862704/ (date accessed: 08.08.2025).
14. Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, Xavier Bresson. Graph Neural Networks with Learnable Structural and Positional Representations [Electronic resource]. URL: https://openreview.net/pdf?id=wTTjnvGphYj (date accessed: 08.08.2025).
15. 15 Examples of Natural Language Processing Applications: Habr. [Electronic resource]. URL: https://habr.com/ru/companies/otus/articles/930130/ (date accessed: 08.08.2025).
16. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention Is All You Need [Electronic resource]. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (date accessed: 08.08.2025).
17. Six Steps to Creating Better Computer Vision Models: Habr. [Electronic resource]. URL: https://habr.com/ru/articles/705008/ (date accessed: 08.08.2025).
18. Gedas Bertasius, Heng Wang, Lorenzo Torresani. Is Space-Time Attention All You Need for Video Understanding? [Electronic resource]. URL: https://proceedings.mlr.press/v139/bertasius21a/bertasius21a.pdf (date accessed: 08.08.2025).
19. 11.1. Reinforcement learning: Yandex Education. [Electronic resource]. URL: https://education.yandex.ru/handbook/ml/article/obuchenie-s-podkrepleniem (date accessed: 08.08.2025).
20. David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, Demis Hassabis. Mastering the Game of Go without Human Knowledge [Electronic resource]. URL: https://discovery.ucl.ac.uk/id/eprint/10045895/1/agz_unformatted_nature.pdf (date accessed: 08.08.2025).
21. Generative AI - The Future or Just Hype?: Habr. [Electronic resource]. URL: https://habr.com/ru/companies/serverspace/articles/754768/ (date accessed: 08.08.2025).
22. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Networks. [Electronic resource]. URL: https://dl.acm.org/doi/pdf/10.1145/3422622 (date accessed: 08.08.2025).