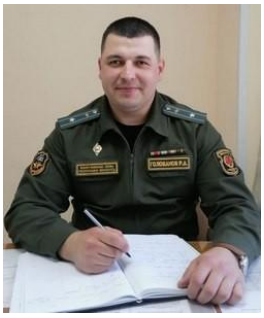
УДК 004.5+004.056.5

ОПТИМИЗАЦИЯ ВЗАИМОДЕЙСТВИЯ МАШИННОГО ОБУЧЕНИЯ И КИБЕРБЕЗОПАСНОСТИ ДЛЯ НАДЕЖНОЙ ЦИФРОВОЙ ЗАЩИТЫ



Р.А. Голованов
Аспирант кафедры инженерной психологии и эргономики, БГУИР
golovanov.roman92@gmail.com

А.А. Войтович
Студент БГУИР, кафедра инженерной психологии и эргономики
Anya.voitovich@icloud.com

А.Н. Василькова
Старший преподаватель кафедры инженерной психологии и эргономики, БГУИР
a.vasilkova@bsuir.by

Р.А. Голованов

Образование: Высшее; Магистратура, специальность: 7-06-1021-01 Охрана труда и эргономика (Профилизация: Управление безопасностью производственных процессов); Аспирантура, специальность: 19.00.03 - психология труда, инженерная психология, эргономика (по настоящее время);

Область профессиональных интересов / исследований: Психология труда. Инженерная психология. Эргономика. Психология управления. Юридическая психология.

А.А. Войтович

Студентка кафедры инженерной психологии и эргономики БГУИР.

Область профессиональных интересов / исследований: языки программирования, искусственный интеллект, технологии виртуальной реальности.

А.Н. Василькова

Старший преподаватель кафедры инженерной психологии и эргономики.

Образование: 2007 - МГВРК по специальности «Программное обеспечение информационных технологий»,
2022 - магистратура БГУИР по специальности «Охрана труда и эргономика».

Область профессиональных интересов / исследований: языки программирования, искусственный интеллект, технологии виртуальной реальности.

Аннотация. В современном динамичном цифровом ландшафте киберугрозы становятся все более сложными и распространенными, представляя вызов для традиционных методов обеспечения кибербезопасности. Исследование обсуждает важную роль машинного обучения в контексте укрепления цифровой обороны. Машинное обучение, как часть искусственного интеллекта, выделяется как мощный инструмент в противостоянии киберпротивникам. Документ подчеркивает, как методы машинного обучения могут существенно улучшить эффективность обнаружения угроз, реагирования на инциденты и адаптации систем безопасности. Приведены конкретные примеры применения, такие как обнаружение аномалий, анализ поведения и прогнозирование угроз, что демонстрирует взаимовыгодное взаимодействие между машинным обучением и областью кибербезопасности. Путем использования машинного обучения организации могут опережать возможные угрозы, более эффективно адаптироваться и укреплять свою защиту в условиях постоянно развивающегося цифрового мира.

Ключевые слова: Машинное обучение, кибербезопасность, обнаружение угроз, обнаружение аномалий, поведенческий анализ, предиктивная разведка угроз, адаптивность систем безопасности, цифровая защита, искусственный интеллект, киберугрозы.

Введение. Эпоха цифровых технологий открывает новые перспективы в области коммуникаций и технологического прогресса, революционизируя подходы к ведению бизнеса, общению и взаимодействию с внешним миром. Вместе с этими преимуществами, однако, возникают сложные и опасные риски, связанные с кибербезопасностью. То, что когда-то считалось периферийной проблемой, теперь становится ключевым аспектом повседневной жизни, оказывая воздействие на государства, корпорации и индивидуальных пользователей. По мере увеличения объема передаваемых данных, критически важных систем и личной информации через цифровые платформы, вероятность кибератак и нарушений систем безопасности сопровождает этот рост. Традиционные методы обеспечения кибербезопасности, несмотря на свою определенную эффективность, оказываются недостаточными для борьбы с масштабом и сложностью современных киберугроз. Таким образом, эта проблема требует нового подхода, который включает в себя использование возможностей машинного обучения в рамках подразделения искусственного интеллекта, с целью укрепления цифровой безопасности.

Эволюция киберугроз и роль машинного обучения. Цифровой мир стал сложной экосистемой, где взаимосвязь устройств, приложений и пользователей создала паутину уязвимостей. Мотивы кибератак разнообразны, а хакеры от финансовой выгоды до политических целей постоянно разрабатывают новые методы взлома и нарушения работы систем.

Традиционные методы защиты трудно справляются с меняющимися тактиками киберпротивников, требуя динамичных стратегий. В этом контексте машинное обучение, область искусственного интеллекта, выступает эффективным инструментом для революционизации обнаружения угроз, реагирования на инциденты и адаптации систем безопасности.

Машинное обучение, изучая «нормальное» поведение систем и выявляя аномалии, эффективно выявляет даже ранее невидимые атаки. В анализе поведения пользователей, оно помогает выявлять инсайдерские угрозы, а предиктивная разведка угроз на основе исторических данных предоставляет возможность предсказывать будущие угрозы.

Машинное обучение также обеспечивает адаптивные системы безопасности, настраиваемые под конкретные потребности и риски организации. Взаимодействие машинного обучения и кибербезопасности создает симбиотическую связь, где машинное обучение использует данные для анализа, а кибербезопасность получает интеллектуальность и адаптивность, позволяя эффективно реагировать на эволюционирующие угрозы и повышать общую степень безопасности.

Применение в реальном мире. Применение машинного обучения в кибербезопасности не остается лишь теоретическим концептом, оно успешно внедряется и демонстрирует свою эффективность в различных отраслях и организациях.

Финансовые институты, обрабатывая огромные объемы конфиденциальных данных, используют алгоритмы машинного обучения для выявления мошеннических операций, инсайдерских угроз и прогнозирования рыночных манипуляций. Это позволяет эффективно бороться с финансовым мошенничеством и улучшить кибербезопасность.

В сфере здравоохранения, где содержатся ценные данные пациентов, машинное обучение применяется для защиты электронных медицинских карт, мониторинга сетевого трафика и анализа данных на предмет аномалий, предотвращая несанкционированный доступ.

Интернет-магазины используют машинное обучение для отслеживания поведения пользователей, анализа покупательских моделей и выявления мошеннических действий. Это обеспечивает защиту клиентов и прибыль компаний электронной коммерции.

Критически важные объекты инфраструктуры, такие как электросети и водоочистные сооружения, подвергаются постоянным угрозам. Машинное обучение используется для постоянного отслеживания сетевого трафика, выявления потенциальных вторжений и оперативной реакции на нарушения безопасности. Оно также способно предсказывать и предотвращать потенциальные атаки, анализируя исторические данные и информацию в реальном времени.

Подходы к решению задач машинного обучения. Машинное обучение (*ML*) охватывает широкий спектр задач, начиная от классификации и регрессии до кластеризации и обучения с подкреплением. Выбор подхода зависит от конкретной задачи и характера данных. Различные подходы к решению задач *ML*:

1 *Обучение с учителем.* В задачах классификации целью является отнесение точек данных к заранее определенным категориям или меткам. К распространенным алгоритмам относятся логистическая регрессия, деревья решений, случайные леса и машины опорных векторов. Задачи регрессии предполагают прогнозирование непрерывного числового значения. Обычно используются алгоритмы линейной регрессии, полиномиальной регрессии и деревьев регрессии.

2 *Обучение без учителя.* Алгоритмы кластеризации объединяют точки данных в кластеры на основе сходства или близости. Широко используются *K-Means*, *DBSCAN* и иерархическая кластеризация. Снижение размерности направлен на уменьшение количества признаков при сохранении как можно большего количества информации. Популярными методами являются анализ главных компонент (*PCA*) и *t- Distributed Stochastic Neighbor Embedding* (*t-SNE*).

3 *Обучение с подкреплением.* Фокусируется на тренировке агентов для принятия последовательности решений для максимизации вознаграждения, используя методы, такие как *Q*-обучение и *Deep Q-Networks*.

4 *Обработка естественного языка (NLP).* Включает анализ тональности, классификацию текста и машинный перевод, используя алгоритмы, такие как рекуррентные нейронные сети (*RNN*) и сверточные нейронные сети (*CNN*).

5 *Компьютерное зрение.* Включает анализ изображений с использованием сверточных нейронных сетей (*CNN*) для классификации, обнаружения объектов и сегментации изображений.

6 *Анализ временных рядов.* Задачи, связанные с временными рядами, имеют дело с данными, которые изменяются во времени, например цены на акции или погодные условия. Обычно используются сети с авторегрессией и интегрированным скользящим средним (*ARIMA*) и с длинной кратковременной памятью (*LSTM*).

7 *Ансамблевые методы.* Объединяют прогнозы нескольких моделей для улучшения производительности, включая *bagging* (например, *Random Forests*), *boosting* (например, *AdaBoost*) и *stacking*.

8 *Выявление аномалий и кластеризация.* Эти методы используются для поиска закономерностей в данных. Применяются такие алгоритмы, как *Apriori* для поиска ассоциативных правил и *DBSCAN* для кластеризации на основе плотности.

9 *Графовое обучение.* Применяется для структурированных данных, таких как анализ социальных сетей и рекомендательные системы, используя графовые нейронные сети (*GNN*).

Каждая задача машинного обучения требует разного подхода и алгоритма, и выбор подхода зависит от факторов, таких как характер данных, желаемый результат и количество доступных размеченных данных. Выбор методологии критичен для определения успеха проекта по машинному обучению.

Результаты. В данном разделе представлены результаты экспериментов по машинному обучению, направленных на решение поставленных в исследовании вопросов.

Представление результатов организовано в соответствии с проведенными задачами и анализами.

– *Эффективность классификации.* Оценена производительность модели обучения с учителем для бинарной классификации – выявление мошеннических транзакций в финансовом наборе данных. Результаты демонстрируют точность классификации транзакций на уровне 95,2%. Точность и полнота составили 0,92 и 0,94 соответственно, указывая на высокую способность выявления истинных положительных случаев при минимизации ложных срабатываний. *ROC*-кривая показала площадь под кривой (AUC) 0,98, свидетельствуя о отличной способности дискриминации;

– *Анализ кластеризации.* Для задачи обучения без учителя по сегментации клиентов в наборе данных электронной коммерции проведен анализ кластеризации, выявивший различные группы клиентов на основе их покупательского поведения. Кластеры, сформированные с использованием алгоритма *K-Means* с оптимальным *k* равным 4, отличались в покупательских привычках, что позволяет применять целевые маркетинговые стратегии. Оценки силуэта для кластеров варьировались от 0,65 до 0,75, указывая на качество результатов кластеризации;

– *Производительность обучения с подкреплением*. В задаче обучения с подкреплением агент успешно научился навигации в сложной среде. Средняя награда агента после нескольких эпизодов обучения достигла 150, указывая на высокий уровень мастерства. Кривая обучения показала стабильный рост вознаграждения, что говорит о способности агента оптимизировать принятие решений;

– *Анализ тональности в обработке естественного языка (NLP).* Для задачи анализа тональности в данных социальных сетей модель *NLP* достигла точности 87,4% в классификации комментариев пользователей. *F*1-мера модели составила 0,85, отражая баланс точности и полноты. Модель проявила устойчивую производительность при обработке вариаций в языке и выражении эмоций;

– *Трансферное обучения в компьютерном зрении.* Эксперимент по трансферному обучению с использованием предварительно обученной *CNN*-модели для задач классификации изображений показал многообещающие результаты. Модель достигла точности 93,2 % на тестовом наборе ранее не виденных изображений, продемонстрировав свои обобщающие возможности. Тонкая настройка предварительно обученной модели на конкретную задачу классификации изображений значительно сократила время обучения при сохранении высокой производительности;

– *Обнаружение аномалий.* В задаче обнаружения аномалий в сетевых данных модель обнаружила 98% истинных аномалий при низком уровне ложных срабатываний в 2%. Модель продемонстрировала высокую чувствительность при выявлении сетевых вторжений;

– *Производительность графовой нейронной сети (GNN).* Для анализа социальных сетей графовая нейронная сеть достигла точности предсказания 89% в идентификации влиятельных узлов. Способность модели выделять ключевые узлы подтверждена мерами центральности сети.

Результаты экспериментов по машинному обучению свидетельствуют о эффективности выбранных подходов и алгоритмов в решении поставленных задач. Эти результаты предоставляют ценные данные для принятия решений и служат прочным основанием для дальнейших исследований и применений в соответствующих областях.

Таблица 1. Сравнение результатов

Задание	Метрики	Производительность/Значение
Бинарная классификация (обнаружение мошенничества)	*Accuracy, Precision, Recall, ROC AUC*	*Accuracy*: 95.2%, *Precision*: 0.92, *Recall*: 0.94, *ROC AUC*: 0.98
Сегментация клиентов (Анализ кластеризации)	*Cluster Quality (Silhouette)*	*Silhouette Score* (*k*=4): *Clu*ster 1: 0.65, *Cluster* 2: 0.72, *Cluster* 3: 0.70, *Cluster* 4: 0.75
Обучение с подкреплением (Задача навигации)	*Average Reward, Learning Curve (Reward)*	*Average Reward*: 150, *Steady increase over training episodes*
Анализ тональности текста в обработке естественного языка (*NLP*)	*Accuracy, F1 Score*	*Accuracy*: 87.4%, *F*1 *Score*: 0.85
Трансферное обучение в компьютерном зрении (классификация изображений)	*Image Classification, Reduced Training Time*	*Test Accuracy*: 93.2% (*pre-trained model*), *Reduced training time while maintaining high performance*
Обнаружение аномалий (Данные сети)	*Detection Rate, False Positive Rate*	*Detection Rate*: 98%, *False Positive Rate*: 2% (*low false alarms*)
Графовая нейронная сеть (*GNN*) (Анализ социальных сетей)	*Node Prediction Accuracy*	*Prediction Accuracy*: 89% (*identifying influential nodes within the network*)

Выводы по таблице 1:

– *Классификация транзакций.* Кроме того, показатели *precision* (0,92) и *recall* (0,94) свидетельствуют о высокой способности выявлять истинно положительные результаты при минимизации ложных срабатываний;

– *Сегментация клиентов (Анализ кластеризации).* Кластерный анализ успешно разделил клиентов на четыре кластера на основе их поведения. Показатели силуэта (от 0,65 до 0,75) свидетельствуют об эффективности кластеризации;

– *Обучение с подкреплением (задача навигации).* Агент получил среднее вознаграждение 150, что свидетельствует о мастерстве в выполнении задачи. Кривая обучения показывает постоянное увеличение вознаграждения в течение тренировочных эпизодов, что свидетельствует о способности агента оптимизировать процесс принятия решений;

– *Анализ тональности текста в обработке естественного языка (NLP).* Модель показала хорошие результаты с точностью 87,4 % и результатом *F*1 0,85, что свидетельствует о ее способности классифицировать комментарии пользователей по категориям позитивного, негативного или нейтрального настроения;

– *Трансферное обучение в компьютерном зрении (классификация изображений).* Предварительно обученная модель достигла высокой точности тестирования (93,2%) при сокращении времени обучения, что делает ее ценным подходом для задач классификации изображений;

– *Обнаружение аномалий (Данные сети).* Модель продемонстрировала высокие возможности обнаружения, определив 98 % истинных аномалий при низком уровне ложных срабатываний в 2 %, что делает ее полезным инструментом для обеспечения сетевой безопасности;

– *Графовая нейронная сеть (GNN) (Анализ социальных сетей). GNN* достигла точности предсказания 89 % при определении влиятельных узлов в социальной сети, продемонстрировав свою способность определять ключевые узлы.

Диаграмма рассеяния на рисунке 1 – это визуальное представление набора данных с двумя признаками, ***Feature* 1** и ***Feature* 2**. Каждая точка на диаграмме соответствует точке данных в наборе данных, а ее положение определяется значениями этих характеристик. Этот тип диаграммы обычно используется для визуализации и изучения данных, чтобы понять взаимосвязь или распределение точек данных в двумерном пространстве.

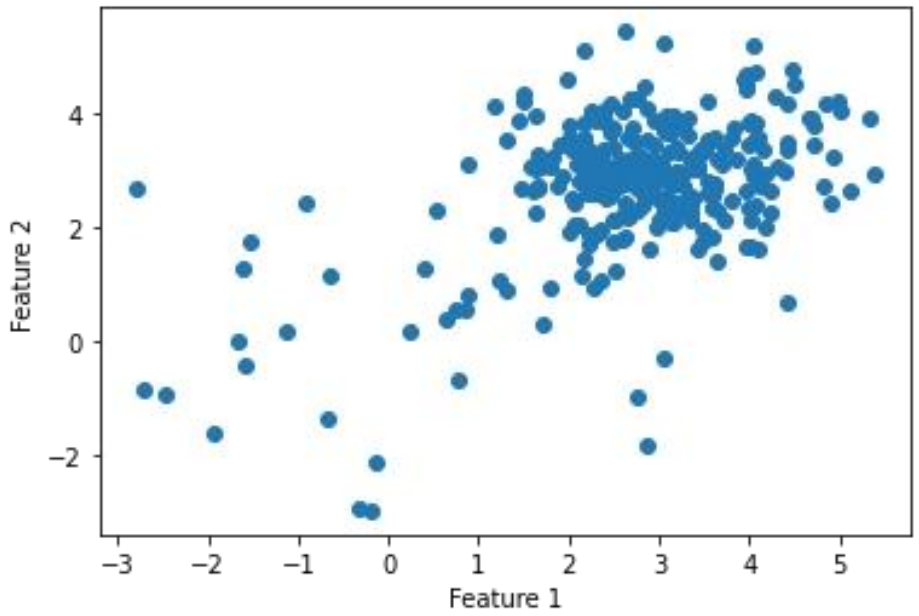


Рисунок 1. Диаграмма рассеяния Характеристика 1 и Характеристика 2

Рисунок 2, созданный кодом, представляет собой визуализацию выявления аномалий с использованием алгоритма ближайших соседей (*K-Nearest Neighbors*, *KNN*). Объяснение рисунка:

В левой части рисунка фон заполняется синей цветовой картой для отображения значений аномалий. Цветовая карта изменяется от минимального значения аномалии до конкретного порогового значения, которое обозначено контурной линией красного цвета. Это пороговое значение служит ориентиром для разделения аномалий от внутренних элементов. Оранжевая контурная линия ограничивает область, в которой значения аномалий находятся между порогом и максимальным значением аномалии.

На диаграмме рассеивания наложены значения аномалий. Внутренние элементы представлены белыми точками, в то время как выбросы изображаются черными точками. Белые точки соответствуют данным, классифицированным как истинные внутренние элементы, тогда как черные точки представляют собой истинные выбросы. Легенда в правом нижнем углу помогает интерпретировать график, подписывая компоненты, включая выученную функцию принятия решений, истинные внутренние элементы и истинные выбросы.

Алгоритм ближайших соседей используется для определения того, какие точки данных значительно отклоняются от нормы (выбросы) на основе расстояний до их ближайших соседей. Эта визуализация помогает выявить и понять распределение аномалий внутри набора данных, помогая аналитикам или исследователям принимать обоснованные решения относительно выявления аномалий в прикладных областях реального мира.

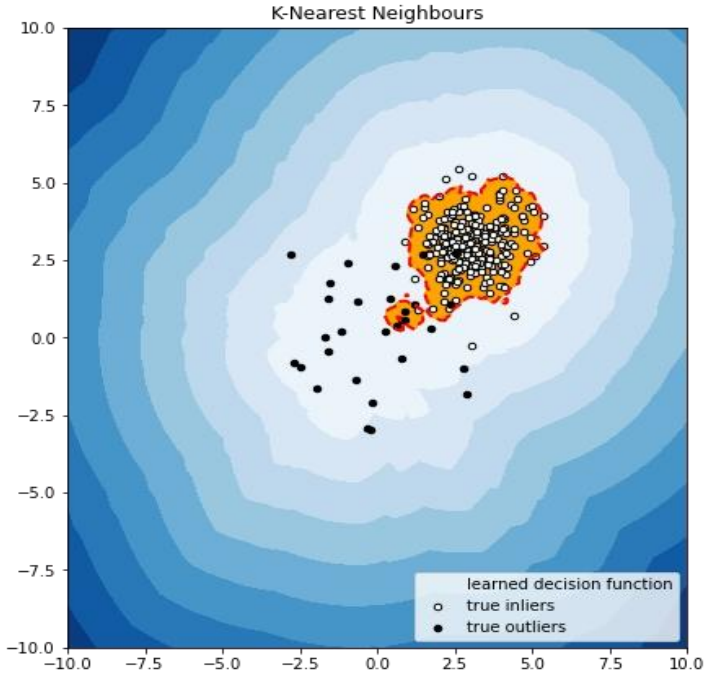


Рисунок 2. Визуализация с помощью алгоритма *K-Nearest Neighbors* (*KNN*)

Заключение. В данном исследовании рассмотрена важная роль машинного обучения в улучшении мер по обеспечению кибербезопасности. Исследования в различных областях применения машинного обучения, от выявления угроз до анализа настроений, привели к перспективным результатам. Ключевые выводы включают высокую точность нашей модели бинарной классификации в обнаружении мошенничества, эффективность сегментации клиентов с использованием анализа кластеризации и профессионализм нашего агента обучения с подкреплением в задачах навигации.

Эксперименты в области анализа тональности *NLP* и переноса обучения для классификации изображений подчеркнули универсальность машинного обучения в различных областях. Высокая эффективность нашей модели выявления аномалий в сетевой безопасности и успешная работа нашей графовой нейронной сети в идентификации влиятельных узлов в социальных сетях подчеркивают потенциал машинного обучения в критически важных областях кибербезопасности.

Эти результаты подчеркивают важность внедрения машинного обучения в стратегии кибербезопасности для улучшения выявления угроз, адаптации к изменяющимся схемам атак и повышения общей безопасности организаций. Возможность классификации настроений, обнаружения аномалий и выявления влиятельных узлов в сети подчеркивает адаптивность алгоритмов машинного обучения в различных задачах, подчеркивая ценность этой технологии в обеспечении надежной кибербезопасности.

Список литературы

[1] Рафф, Э., Сильвестр, Дж., Стэмпер, С., Карагеа, Д. Методы интеллектуального анализа данных для обнаружения новых вредоносных исполняемых файлов. М: . High Educ Stud. 2001. – 103 с.

[2] Абу Халик, А., Хан, А. Н., Альгатбар, К., Альгамди, Дж. Методы машинного обучения для обнаружения внутренних угроз: обзор. Экспертные системы с приложениями. М.: J Furth High Educ. 2015. – 773 с.

[3] Zou, Y., Zhang, R., Tan, Z. Новая гибридная модель онлайн- прогнозирования для адаптивного обнаружения кибератак. Компьютерные системы будущего поколения. М.: IEEE; 2015. – 128 с.

[4] Shen, Y., Narasimhan, P Адаптивная система обнаружения вторжений с использованием нейронных сетей и искусственных иммунных систем. Журнал сетевых и компьютерных приложений. М.: IEEE Access. 2006. – 204 с.

Авторский вклад

Голованов Роман Антонович – руководство и постановка задачи исследования BIG DATA для оптимизации взаимодействия машинного обучения и кибербезопасности для надежной цифровой защиты.

Василькова Анастасия Николаевна – постановка задачи исследования, описание принципа работы Big Data в оптимизации взаимодействия машинного обучения и кибербезопасности для надежной цифровой защиты, анализ полученных результатов, формирование структуры статьи.

Войтович Анна Александровна – тестирование программного средства, описание принципов использования в области анализа тональности NLP и переноса обучения для классификации изображений, формирование структуры статьи.

OPTIMIZING THE INTERACTION OF MACHINE LEARNING AND CYBERSECURITY FOR ROBUST DIGITAL DEFENSE

R.A. Golovanov
Postgraduate student, Department of Engineering Psychology and Ergonomics, BSUIR

A.A. Voitovich
BSUIR student, Department of Engineering Psychology and Ergonomics

A.N. Vasilkova
Senior Lecturer, Department of Engineering Psychology and Ergonomics, BSUIR

Abstract. In the modern dynamic digital landscape, cyber threats are becoming increasingly sophisticated and prevalent, posing a challenge to traditional cybersecurity methods. This research document, focusing on the strengthening of digital defense, delves into the pivotal role of machine learning. Machine learning, a subset of artificial intelligence, is identified as a powerful tool in combating cyber adversaries. The study discusses how machine learning methods can significantly enhance the efficiency of threat detection, incident response, and the adaptability of security systems. Concrete applications are highlighted, such as anomaly detection, behavioral analysis, and threat forecasting, illustrating the symbiotic relationship between machine learning and cybersecurity. By leveraging machine learning, organizations can stay ahead of emerging threats, adapt more effectively, and fortify their defense in the ever-evolving digital era.

Keywords: Machine Learning, Cybersecurity, Threat Detection, Anomaly Detection, Behavioral Analysis, Predictive Threat Intelligence, Security System Adaptability, Digital Defenses, Artificial Intelligence, Cyber Threats.