DOI 10.37882/2223-2966.2026.02-2.08

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СИСТЕМАХ КИБЕРБЕЗОПАСНОСТИ

THE USE OF ARTIFICIAL INTELLIGENCE IN CYBERSECURITY SYSTEMS

N. Verezubova
O. Yakovleva
N. Sakovich

Summary. The rapid adoption of artificial intelligence (AI) technologies is fundamentally transforming the cybersecurity landscape, creating a paradoxical situation. On the one hand, AI has become an indispensable tool for defenders, allowing them to automate threat detection, analyze behavioral anomalies, and promptly respond to incidents in the face of huge amounts of data. On the other hand, these same technologies are actively used by hackers to create sophisticated phishing attacks, generate polymorphic malware and produce convincing deepfakes, which significantly reduces the threshold for entry into cybercrime and increases the effectiveness of attacks. The relevance of the research is due to the need to understand this dual nature of AI and develop strategies that would maximize the protective potential of the technology while minimizing the risks associated with it. The purpose of the article is to analyze the security problems generated and exacerbated by AI, and to develop a comprehensive adaptive protection model based on them. As an author's development, the concept of an integrated cybersecurity platform based on the principles of machine learning security, hybrid human-machine analytics and proactive data privacy protection is proposed. The practical significance of the work lies in structuring a step-by-step approach to building systems resistant to modern threats, balancing technological efficiency and ethical and legal requirements.

Keywords: artificial intelligence, cybersecurity, adaptive protection, data privacy, data protection.

Верезубова Наталья Афанасьевна
Кандидат экономических наук, доцент, Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина
nverez@mail.ru
Яковлева Ольга Анатольевна
Кандидат с/х наук, доцент, Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина
yakovleffo@yandex.ru
Сакович Наталия Евгениевна
Доктор технических наук, доцент, Брянский государственный аграрный университет
nasa2610@mail.ru

Аннотация. Стремительное внедрение технологий искусственного интеллекта (ИИ) коренным образом трансформирует ландшафт кибербезопасности, создавая парадоксальную ситуацию. С одной стороны, ИИ стал незаменимым инструментом для защитников, позволяя автоматизировать обнаружение угроз, анализировать поведенческие аномалии и оперативно реагировать на инциденты в условиях огромных массивов данных. С другой — эти же технологии активно используются злоумышленниками для создания изощренных фишинговых атак, генерации полиморфного вредоносного ПО и производства убедительных дипфейков, что значительно снижает порог входа в киберпреступность и повышает эффективность атак. Актуальность исследования обусловлена необходимостью осмысления этой двойственной природы ИИ и выработки стратегий, которые позволили бы максимизировать защитный потенциал технологии при одновременной минимизации связанных с ней рисков. Целью статьи является анализ проблем безопасности, порождаемых и обостряемых ИИ, и разработка на их основе комплексной модели адаптивной защиты. В качестве авторской разработки предлагается концепция интегрированной платформы кибербезопасности, основанной на принципах безопасности машинного обучения, гибридной человеко-машинной аналитики и проактивной защиты конфиденциальности данных. Практическая значимость работы заключается в структурировании поэтапного подхода к построению устойчивых к современным угрозам систем, балансирующих между технологической эффективностью и этико-правовыми требованиями.

Ключевые слова: искусственный интеллект, кибербезопасность, адаптивная защита, конфиденциальность данных, защита данных.

Введение

Цифровая эпоха привела к экспоненциальному росту поверхности атаки предприятий, превратив кибербезопасность в задачу по обработке сотен миллиардов сигналов, которая уже не может быть решена исключительно человеческими силами. В этот переломный момент искусственный интеллект выступает как катализатор перемен, кардинально меняя баланс сил в цифровом пространстве. Однако его роль далеко не однозначна. Если для специалистов по защите ИИ — это мощный инструмент, способный к предиктивной аналитике и автоматизации рутинных операций, то для злоумышленников он стал универсальным инструмен-

том для масштабирования и усложнения атак [2, 6]. Генеративные модели, такие как большие языковые модели (LLM), стирают границу между реальным и сгенерированным контентом, что несет фундаментальные риски для целостности информации и может провоцировать серьезные социальные и дипломатические кризисы [1, 3, 8]. Таким образом, современный этап характеризуется переходом от дискуссий о потенциальных угрозах к необходимости управления реальными, материализовавшимися рисками. Актуальность данного исследования продиктована необходимостью преодоления фрагментарного подхода к безопасности ИИ. Проблемы часто рассматриваются изолированно: либо как технические уязвимости моделей, либо как вопросы нормативного регулирования. Однако реальный вызов заключается в их системной взаимосвязи. Атака на обучающие данные может привести к утечке коммерческой тайны, а усовершенствованный фишинг с помощью ИИ обнуляет инвестиции в обучение сотрудников, если не интегрирован в общую стратегию защиты [3, 6]. Поэтому актуальным становится поиск целостных моделей, объединяющих технологические, процессные и человеческие аспекты. Цель статьи — провести комплексный анализ дуалистической природы применения ИИ в кибербезопасности, выявить системные проблемы на стыке технологий, данных и человеческого фактора, а также разработать и обосновать модель построения адаптивной системы защиты.

Материалы и методы исследования

Методологическую основу исследования составил многоуровневый анализ, сочетающий изучение технологических трендов, нормативных инициатив и практических кейсов. Для выявления глобальных тенденций и прогнозов на 2025 год были проанализированы отраслевые обзоры и экспертные мнения, опубликованные ведущими компаниями в сфере кибербезопасности (такими как ESET) и ИТ-сектора. Особое внимание было уделено материалам профессиональных дискуссий, в частности, итогам круглого стола «Современные вызовы в безопасности AI», прошедшего в рамках Kazan Digital Week 2025, где экспертами обсуждались конкретные уязвимости и методологии защиты. Вторым, практический блок метода, включал кейс-анализ конкретных технологий и инцидентов. На основе синтеза полученных данных методом системного проектирования была разработана авторская концепция интегрированной платформы, вбирающая в себя лучшие практики и направленная на устранение выявленных системных противоречий.

Результаты и обсуждения

Проведенный анализ позволил систематизировать несколько взаимосвязанных групп проблем, формирующих современный комплекс угроз в области использования ИИ в кибербезопасности.

Первая группа проблем — технологическая дуальность и усиление атак. Наиболее очевидной проблемой является использование преступниками тех же инструментов, что и у защитников. Генеративный ИИ радикально повысил эффективность социальной инженерии. С помощью языковых моделей создаются персонализированные фишинговые письма без грамматических ошибок, имитирующие стиль коллег или руководителей. Дипфейк-технологии позволяют генерировать поддельные аудио- и видеосообщения для обмана систем биометрической аутентификации или мошеннических звонков [3]. Более того, ИИ используется для автоматизации создания вредоносного ПО, способного динамически изменять свой код (полиморфизм), чтобы избегать обнаружения сигнатурными методами. Также развиваются атаки на сами ИИ-системы, такие как «отравление данных» (внесение искажений в наборы для обучения) и «джейлбрейк» (обход этических и safety-ограничений модели для получения запрещенной информации) [6, 7, 11].

Другая группа проблем — уязвимости данных и конфиденциальности. ИИ-системы, особенно в корпоративном контексте, становятся новыми критически важными активами и одновременно — новыми векторами атаки. Как отмечают эксперты, модель, обученная на уникальных корпоративных данных, сама по себе несет угрозу утечки конфиденциальной информации, коммерческой тайны или персональных данных. Риск возникает на всех этапах: при сборе и подготовке данных, в процессе обучения модели (когда злоумышленник может попытаться извлечь сведения) и в ходе эксплуатации, через специально сформулированные промпты. Проблема усугубляется практикой сотрудников, которые могут неосознанно загружать конфиденциальные документы в публичные ИИ-сервисы для анализа, что ведет к непреднамеренному раскрытию информации [5, 10].

Также необходимо рассмотреть операционные и кадровые вызовы. Внедрение ИИ в SOC (Security Operations Center) не отменяет, а трансформирует роль человека. Возникает проблема эффективного человеко-машинного взаимодействия. С одной стороны, ИИ великолепно справляется с обработкой больших объемов данных и первичным анализом, освобождая аналитиков от рутины. С другой — сохраняется риск избыточной зависимости от «черного ящика» алгоритмов, ложного чувства безопасности и, как следствие, деградации экспертных навыков у специалистов [8, 9]. Необходимы новые компетенции: способность «допрашивать» модель, понимать принципы ее работы, интерпретировать и проверять ее выводы. Как подчеркивают в «Лаборатории Касперского», построением комплексных систем защиты по-прежнему должны заниматься профессионалы, способные корректно интегрировать ИИ-инструменты в существующие процессы [4].

По результатам исследования была предложена целостная концепция построения системы безопасности, основанная на трех фундаментальных принципах: безопасность жизненного цикла ИИ (MLSecOps), гибридный интеллект и проактивная защита данных.

Ядром концепции является внедрение практик MLSecOps — методологии, которая распространяет принципы безопасной разработки (DevSecOps) на полный жизненный цикл машинного обучения. Это подразумевает непрерывный мониторинг и безопасность на всех этапах: от управления версиями и целостности обучающих данных, через безопасное проведение экспериментов с моделями, до контроля качества и устойчивости развернутых моделей к враждебным атакам [6]. Предлагаемая концепция предполагает создание технологического контура, где каждая модель перед внедрением проходит тестирование не только на точность, но и на устойчивость к попыткам манипуляции ее выводами или извлечения данных.

Второй элемент — архитектура гибридного интеллекта для SOC. В рамках предложенной концепции ИИ выступает не как замена аналитика, а как его «второй пилот» или сильный ассистент [3, 5]. Система строится по принципу взаимного контроля: ИИ обрабатывает потоки телеметрии, выявляет аномалии, кластеризует инциденты и предлагает гипотезы и варианты реагирования. Человек-аналитик выполняет функцию валидатора, стратега и принимает финальные решения на основе контекста, недоступного машине (например, знаний о политических мотивах атаки или внутренних бизнес-процессах). Такое разделение труда позволяет, как в случае с Kaspersky MDR, значительно снизить нагрузку на специалистов, фильтруя ложные срабатывания, и одновременно сохранить человеческий контроль над критическими решениями [4].

Третья составляющая — инфраструктура проактивной защиты данных для ИИ. Предлагаемая концепция предлагает встроить защиту конфиденциальности в сам процесс работы с данными для ИИ. Это включает [6]:

— строгую сегментацию и управление доступом к корпоративным данным, используемым для обучения или запрашиваемых моделями в режиме RAG (Retrieval-Augmented Generation). LLM-агент не должен иметь неограниченного доступа ко всей корпоративной базе знаний; его права должны соответствовать принципу минимальных привилегий;
— применение методов дифференциальной приватности и федеративного обучения там, где это возможно, чтобы минимизировать риски утечки исходных данных;
— обязательный аудит всех запросов к внешним и внутренним ИИ-сервисам для предотвращения утечек через промпты сотрудников. Организация должна иметь четкую политику, определяющую, какие данные можно обрабатывать в публичных облачных ИИ, а какие — только в изолированных, локальных контурах.

Реализация предлагаемой концепции носит итеративный характер. Она начинается с аудита существующих ИИ-активов и потоков данных, продолжается через пилотные проекты по внедрению MLSecOps-практик для наиболее критичных моделей и построения пилотного гибридного центра мониторинга и реагирования на инциденты информационной безопасности., и завершается интеграцией этих компонентов в единую управляемую платформу с общим интерфейсом и скоординированными процессами реагирования.

Выводы

Таким образом, проведенное исследование подтверждает, что искусственный интеллект стал не просто новым инструментом в арсенале кибербезопасности, а фактором, формирующим принципиально новую, более сложную и динамичную среду цифровых угроз. Проблемы носят системный характер: технологическое усиление атак со стороны злоумышленников неразрывно связано с уязвимостями в жизненном цикле корпоративных ИИ-систем, а операционные выгоды от автоматизации SOC могут быть нивелированы рисками утраты экспертного контроля и новых утечек данных. Успешное противостояние современным угрозам требует отказа от точечного внедрения ИИ-инструментов в пользу выстраивания целостных стратегий.

В качестве концептуального ответа на эти вызовы предложена Концепция интегрированной платформы адаптивной защиты, представляющая собой авторскую разработку. Ее отличие от традиционных подходов — в синтезе трех ранее разрозненных направлений: операционной безопасности машинного обучения (MLSecOps), архитектуры эффективного взаимодействия человека и ИИ в процессах защиты (гибридный интеллект) и встроенных механизмов охраны конфиденциальности данных на всех этапах работы с ИИ. Модель акцентирует, что безопасность должна быть заложена в сам процесс создания и эксплуатации ИИ-активов, а не являться надстройкой над ними. Таким образом, будущее кибербезопасности лежит в построении адаптивных, «умных» систем управления безопасностью, где технологии искусственного интеллекта выполняют роль мощного усилителя человеческих компетенций, а не их замены. Практическая реализация подходов, подобных предложенному, позволит организациям трансформировать вызовы эпохи ИИ в возможности для построения более устойчивой и проактивной цифровой защиты.

ЛИТЕРАТУРА

1. Аветисян, А. Обеспечение кибербезопасности в эпоху ИИ: проблемы, методы и институционализация / А. Аветисян, А. Белеванцев // Пути к миру и безопасности. — 2025. — № 1(68). — С. 40–52. — DOI 10.20542/2307-1494-2025-1-40-52.
2. Белорусов, М.М. Основные проблемы, которые можно решать с помощью ИИ в сфере кибербезопасности / М.М. Белорусов, Е.А. Дмитриева, А.Е. Мартынова // Социосфера. — 2025. — № 1. — С. 240–246.
3. Будущее технологий или угроза: что ожидать от искусственного интеллекта в 2025 году // ESET. — URL: https://www.eset.com/ge-ru/about/newsroom/press-releases/malware/budushcheye-tekhnologiy-ili-ugroza-chto-ozhidat-ot-iskusstvennogo-intellekta-v-2025-godu/ (дата обращения: 09.12.2025).
4. «В эпоху искусственного интеллекта заниматься кибербезопасностью должны профессионалы». Интервью с представителем «Лаборатории Касперского» // Infocity.tech. — URL: https://infocity.tech/2025/11/v-epohu-iskusstvennogo-intellekta-zanimatsya-kiberbezopasnostyu-dolzhny-professionaly/ (дата обращения: 09.12.2025).
5. Журина, А. Искусственный интеллект в кибербезопасности: польза, риски и реальная эффективность // Anti-Malware.ru. URL: https://www.anti-malware.ru/analytics/Technology_Analysis/Artificial-Intelligence-in-Information-Security (дата обращения: 09.12.2025).
6. Заметки по результатам участия в круглом столе «Современные вызовы в безопасности AI и пути их решения» на KDW — 2025 // Хабр. — URL: https://habr.com/ru/articles/953374/ (дата обращения: 09.12.2025).
7. ИИ в кибербезопасности: друг или враг // DDOS-guard. — URL: https://ddos-guard.ru/blog/ii-v-kiberbezopasnosti (дата обращения: 09.12.2025).
8. Марков, Я.А. Преимущества и риски применения ИИ в кибербезопасности / Я.А. Марков // Интеллектуальные ресурсы — региональному развитию. — 2025. — № 1. — С. 119–124.
9. Намиот, Д.Е. О кибербезопасности ИИ-агентов / Д.Е. Намиот, Е.А. Ильюшин // International Journal of Open Information Technologies. — 2025. — Т. 13, № 9. — С. 13–24.
10. Татаринов, К.А. Влияние искусственного интеллекта на кибербезопасность / К.А. Татаринов // Вопросы российского и международного права. — 2025. — Т. 15, № 5-1. — С. 197–205.
11. Gadelshina, V.D. The future of cybersecurity: how AI is changing approaches to data protection in the era of digital transformation / V.D. Gadelshina. — 2025. — №1. — pp. 451–453.