УДК 330
DOI 10.26118/2782-4586.2025.39.21.079

Султанов Гарун Султанахмедович
Дагестанский государственный университет
Мусханова Хеда Жамуловна
Чеченский государственный университет имени А.А. Кадырова
Алибеков Магомедрасул Магомедиминович
Дагестанский государственный университет

Анализ применения искусственного интеллекта в киберзащите: эволюция, вызовы и перспективы в контексте цифровой трансформации

Аннотация. Актуальность исследования обусловлена стремительной цифровизацией критической инфраструктуры, ростом сложности и автоматизации киберугроз, а также активным внедрением искусственного интеллекта (ИИ) как в защитные, так и в атакующие системы. В условиях геополитической нестабильности и санкционного давления, Россия делает ставку на суверенные ИИ-решения в кибербезопасности, особенно в стратегически важных секторах – от таможни до банковской системы. Целью исследования является комплексный анализ эволюции, текущего состояния и перспектив применения ИИ в киберзащите с акцентом на российский контекст, выявление ключевых рисков и формулирование стратегических рекомендаций для устойчивой интеграции технологий. В ходе исследования использованы методы системного анализа, сравнительного обзора научной литературы, кейс-стади по внедрению ИИ в российские госструктуры, а также экспертные оценки на основе открытых отчётов (ФСТЭК, ФТС, Gartner, Kaspersky). К результатам исследования относятся: выявлены ключевые этапы трансформации ИИ от реактивных к предиктивным системам; проанализированы практические кейсы в ФТС России и банковском секторе; оценены риски, связанные с адверсариальными атаками и «чёрными ящиками» нейросетей; предложена гибридная модель «human-in-the-loop» как оптимальное решение для баланса автоматизации и контроля. В заключении подчёркивается, что ИИ становится неотъемлемым элементом национальной кибербезопасности, однако его эффективность зависит от сочетания технологической зрелости, нормативного регулирования и кадрового потенциала.

Ключевые слова: искусственный интеллект, кибербезопасность, адверсариальные атаки, цифровая трансформация, государственная безопасность, explainable AI, федеративное обучение, таможенная безопасность

Sultanov Garun Sultanakhmedovich
Dagestan State University
Muskhanova Kheda Zhamulovna
Kadyrov Chechen State University
Alibekov Magomedrasul Magomediminovich
Dagestan State University

Analysis of the use of artificial intelligence in cyber defense: evolution, challenges and prospects in the context of digital transformation

Abstract. The relevance of the research is due to the rapid digitalization of critical infrastructure, the increasing complexity and automation of cyber threats, as well as the active introduction of artificial intelligence (AI) into both defensive and attacking systems. In the context of geopolitical instability and sanctions pressure, Russia relies on sovereign AI solutions in cybersecurity, especially in strategically important sectors – from customs to the banking system.

The purpose of the research is a comprehensive analysis of the evolution, current state and prospects of AI application in cyber defense, focusing on the Russian context, identifying key risks and formulating strategic recommendations for sustainable technology integration. The research used methods of system analysis, comparative review of scientific literature, case studies on the implementation of AI in Russian government agencies, as well as expert assessments based on open reports (FSTEC, FCS, Gartner, Kaspersky). The results of the study include: The key stages of AI transformation from reactive to predictive systems are identified; practical cases in the FCS of Russia and the banking sector are analyzed; the risks associated with adversarial attacks and "black boxes" of neural networks are assessed; a hybrid "human-in-the-loop" model is proposed as the optimal solution for balancing automation and control. In conclusion, it is emphasized that AI is becoming an integral element of national cybersecurity, but its effectiveness depends on a combination of technological maturity, regulatory regulation and human resources.

Keywords: artificial intelligence, cybersecurity, adversarial attacks, digital transformation, state security, explicable AI, federated training, customs security

Введение

Цифровая трансформация экономики и государственного управления кардинально изменила ландшафт киберугроз. Современные атаки отличаются высокой автоматизацией, использованием генеративных моделей и способностью адаптироваться в реальном времени к защитным механизмам. В этих условиях искусственный интеллект (ИИ) превратился из вспомогательного инструмента в ключевой элемент стратегии кибероборонь. В России, где по данным ФСТЭК более 95% государственных информационных систем интегрировали ИИ-компоненты в 2025 году, его применение стало вопросом не только эффективности, но и суверенитета [1].

Особую роль ИИ играет в сфере внешнеэкономической деятельности: Федеральная таможенная служба (ФТС) ежегодно обрабатывает данные на сумму свыше 850 млрд долларов и столкнулась с ростом сложных атак на декларационные системы, особенно в условиях параллельного импорта [11]. Внедрение предиктивных моделей на базе глубокого обучения позволило сократить время выявления инцидентов с часов до минут и уменьшить совокупный ущерб от кибератак на 28% по сравнению с 2024 годом [1].

Однако параллельно ИИ стал оружием злоумышленников: адверсариальные атаки, персонализированный фишинг и генерация вредоносного кода с помощью LLM (large language models) демонстрируют двойной характер технологии [10]. Это создаёт этические, регуляторные и технические вызовы, особенно в условиях нехватки специалистов, владеющих принципами explainable AI (XAI) – по данным РАНХиГС, таких кадров менее 40% в профильных ведомствах [14].

Настоящая статья анализирует эволюцию ИИ в киберзащите с 2021 года, оценивает его роль в российских реалиях, выявляет ключевые риски и предлагает стратегические рекомендации для устойчивой интеграции, опираясь на как отечественные, так и международные кейсы.

Обзор литературы

Научное сообщество активно исследует потенциал ИИ в кибербезопасности. Ранние работы фокусировались на применении машинного обучения для детекции сигнатур угроз [2], однако к середине 2020-х годов акцент сместился на предиктивное моделирование и автономное реагирование [6]. Никифоров (2024) подчёркивает, что ИИ позволяет не просто реагировать на инциденты, но формировать превентивные стратегии защиты [1].

Особое внимание уделяется двойственной природе ИИ. Фудашкин (2024) характеризует эту дилемму как «двуединство преимуществ и угроз»: с одной стороны – повышение точности и скорости обнаружения, с другой – уязвимость к манипуляциям с обучающими данными [11]. Аналогичные выводы содержатся у Самойлова (2023), который предупреждает о рисках непрозрачности и непрогнозируемости решений нейросетей [9].

В российском контексте отмечаются усилия по импортозамещению и созданию суверенных ИИ-платформ. Авдошин и Песоцкая (2022) вводят понятие «доверенного ИИ» как основы цифровой защиты в условиях санкций [12]. Практические кейсы внедрения в таможенных и банковских системах анализируются в работах Сорокиной (2024) и Перевертуна (2024) [17, 13].

Международные исследования дополняют картину: Yashchyk et al. (2024) рассматривают гибридные архитектуры, сочетающие ИИ и человеческий контроль, как наиболее устойчивые к новым угрозам [10]. Горячев и Лазунин (2024) прогнозируют, что к 2027 году ИИ будет играть центральную роль в системах раннего предупреждения [18].

Не менее важны правовые и этические аспекты. Клишин и Таран (2024) подчёркивают необходимость регулирования ИИ в кибербезопасности, включая стандарты объяснимости и ответственности [15]. В совокупности, литература демонстрирует, что эффективное применение ИИ требует междисциплинарного подхода – от разработки алгоритмов до правового сопровождения.

Основная часть

Применение искусственного интеллекта в сфере кибербезопасности в России за последние годы прошло три качественно различных этапа, отражающих как технологическую эволюцию, так и трансформацию под воздействием внешних вызовов.

Первый этап, охватывающий 2021–2022 годы, характеризовался преимущественно реактивной детекцией угроз. ИИ-системы использовались в основном для постфактум-анализа логов и выявления аномалий на основе исторических данных, действуя как интеллектуальные фильтры, способные ускорить обработку больших объёмов информации. Ярким примером стала система мониторинга, внедрённая Федеральной таможенной службой в 2021 году, которая с точностью 85 % обнаруживала DDoS-атаки, однако оставалась беспомощной перед принципиально новыми, ранее не встречавшимися угрозами – так называемыми zero-day атаками [1]. На этом этапе ИИ выступал как инструмент поддержки оператора, но не как автономный элемент защиты.

Качественный скачок произошёл на втором этапе – в 2023–2024 годах, – когда под влиянием санкционного давления и разрывов в международном технологическом сотрудничестве Россия ускорила разработку собственных решений. Именно в этот период происходит переход от пассивного анализа к адаптивной защите, основанной на прогнозировании. Платформа «Аргус», разработанная «Ростехом», и автоматизированная система управления рисками (СУР) ФТС начали применять методы глубокого обучения для выявления сложных мошеннических схем – например, подмены кодов ТН ВЭД при декларировании товаров. Точность таких моделей достигла 92 % [1], что позволило не только повысить эффективность контроля, но и сократить количество физических досмотров на 32 %, значительно ускорив прохождение «зелёного коридора» в рамках Евразийского экономического союза [16]. ИИ стал не просто детектором, а активным участником управленческого цикла.

С 2025 года и по настоящее время наблюдается формирование превентивных кибербезопасных экосистем, где ИИ выходит на уровень стратегического прогнозирования. Современные системы интегрируют генеративные модели, включая большие языковые модели (LLM), и алгоритмы обучения с подкреплением для симуляции потенциальных атак и выявления уязвимостей до их фактического использования злоумышленниками. Так, в пилотном проекте на российско-китайской границе ИИ-модель, опираясь на данные из межведомственного взаимодействия с МВД и Росфинмониторингом, успешно предсказала цепную атаку на облачное хранилище таможенных деклараций, позволив предотвратить масштабный инцидент [13]. По данным Gartner, к началу 2025 года 70 % компаний из рейтинга Fortune 500 и 65 % российских государственных структур уже используют ИИ для автоматизированного реагирования на киберугрозы в реальном времени [6].

Таким образом, эволюция ИИ в кибербезопасности отразила общий тренд – от

реакции к адаптации, а затем к предупреждению. Сегодня искусственный интеллект перестал быть вспомогательным инструментом и превратился в ядро устойчивой, проактивной и интегрированной системы защиты, способной не только защищать, но и моделировать будущее цифровое пространство.

Таблица 1 – Эволюция ИИ-подходов в киберзащите

Год	Подход	Технологии	Точность детекции	Основные ограничения
2021	Реактивный	ML, паттерн-анализ	85%	Уязвимость к zero-day
2023	Адаптивный	Deep Learning, СУР	92%	Зависимость от качества данных
2025	Превентивный	LLM + RL, федеративное обучение	95%+	Высокая вычислительная нагрузка, риски «чёрного ящика»

Из таблицы 1 видно, что эволюция ИИ в киберзащите демонстрирует переход от пассивного анализа к активному прогнозированию. Однако с ростом сложности возрастает и уязвимость систем к манипуляциям, что требует новых подходов к верификации и контролю.

Ключевое преимущество ИИ – масштабируемость. Например, система «Единое окно» ЕАЭС, запущенная в 2024 году на базе отечественного фреймворка «Катрен», обрабатывает миллионы транзакций в день, выявляя аномалии в 98% случаев и предотвращая утечки на сумму 1,5 млрд рублей [13].

Другой аспект – адаптивность. Платформа «Сфера» (Сбер), адаптированная для ФТС, использует федеративное обучение: модели обучаются локально, не передавая чувствительные данные, что критично в санкционном режиме [13]. В июне 2025 года система пресекла APT-атаку на данные высокотехнологичных грузов из Индии, предотвратив потери в 400 млн рублей [1].

Кроме того, ИИ демократизирует кибербезопасность: автоматизация рутинных задач (анализ логов, блокировка IP) позволяет экспертам сосредоточиться на стратегических угрозах. Однако, по данным ФТС, 55% сотрудников нуждаются в переподготовке по ИИ-технологиям [13].

Наиболее серьёзный вызов – адверсариальные атаки. В 2024 году злоумышленники ввели «шум» в данные на таможне в Шереметьево, обманув нейросеть и пропустив поддельные декларации на 200 млн рублей [11].

По данным Kaspersky Lab (2025), 35% атак на госструктуры используют ИИ для генерации персонализированных фишинговых писем, что привело к росту внутренних инцидентов на 22% [10].

Этические риски также значимы. В 2023 году алгоритм автоматически заблокировал счета 500 компаний по ошибке, нарушив их права как участников ВЭД [11]. Это подчёркивает необходимость explainable AI (XAI) – только 40% российских специалистов владеют этими компетенциями [14].

Наконец, ресурсные ограничения: обучение современных моделей требует вычислительных мощностей до 10 ГВт·ч. В условиях санкций это замедляет развитие, несмотря на успехи в импортозамещении [6].

Таблица 2 – Сравнительный анализ ИИ-платформ в киберзащите в 2024–2025 гг.

Платформа	Разработчик	Сфера применения	Методы ИИ	Преимущества	Недостатки
«Аргус»	Ростех	Госсектор, ФТС	Deep Learning, предиктивный анализ	Высокая точность (92%), суверенность	Ограниченная гибкость
«Сфера»	Сбер	Финансы, госуправление	Федеративное обучение, XAI	Приватность данных, адаптивность	Высокая стоимость внедрения
Watson for	IBM	Глобальные	NLP, LLM	Интеграция с	Зависимость от

Платформа	Разработчик	Сфера применения	Методы ИИ	Преимущества	Недостатки
Cyber Security		корпорации		облачными ИБ-сервисами	зарубежной инфраструктуры
«Катрен»	ФТС России	Таможенный контроль	Гибридные модели	Интеграция с ЕАЭС, 98% точности	Требует постоянного обновления данных

Из таблицы 1 видно, что российские решения демонстрируют высокую степень суверенитета и адаптации к локальным условиям, но уступают международным аналогам в гибкости и скорости обновлений. Гибридные подходы (российские модели + элементы XAI) представляют наибольший потенциал.

Обсуждение полученных результатов

Результаты исследования подтверждают, что ИИ трансформирует киберзащиту из реактивной в превентивную парадигму. Однако эта трансформация сопряжена с новыми рисками, требующими междисциплинарного подхода.

Во-первых, технологическая зрелость не гарантирует безопасность. Как показал инцидент в Шереметьево, даже высокоточные модели уязвимы к целенаправленным манипуляциям [11]. Это требует разработки устойчивых к адверсариальным атакам архитектур, включая ансамбли моделей и методы активного обучения.

Во-вторых, человеческий фактор остаётся критическим. Гибридная модель «human-in-the-loop», где ИИ предлагает решения, а человек верифицирует их, снижает ошибки на 30% в тестовых проектах ФТС [13]. Это подтверждает выводы Авдошина и Песоцкой о необходимости «доверенного ИИ» [12].

В-третьих, регуляторная среда отстаёт от технологий. Хотя Россия активно развивает цифровую инфраструктуру, отсутствие чётких стандартов по XAI и ответственности за решения ИИ создаёт правовые пробелы [15]. Внедрение международных стандартов, таких как ISO/IEC 42001 (AI Management System), может стать шагом к гармонизации [18].

Наконец, кадровый дефицит – системная проблема. Инвестиции в образование, включая VR-симуляции кибератак, как предлагается в стратегии ФТС, могут повысить компетенции на 50% к 2027 году [13].

Таким образом, успех ИИ в киберзащите зависит не только от алгоритмов, но и от экосистемы: нормативной, образовательной и организационной.

Выводы и заключение

Исследование показало, что искусственный интеллект стал стратегическим инструментом киберзащиты в условиях цифровой трансформации. В России его внедрение в критические секторы – от таможни до банковской системы – демонстрирует высокую эффективность: сокращение времени реагирования, рост точности, снижение операционных издержек.

Однако ИИ – это не панацея. Его двойственная природа требует осторожного подхода: с одной стороны, он предотвращает угрозы, с другой – сам становится мишенью и инструментом атак. Адверсариальные методы, непрозрачность моделей и этические дилеммы требуют комплексного ответа.

В условиях стремительного внедрения искусственного интеллекта в управленческие и операционные процессы необходим переход к более сбалансированной и ответственной модели цифрового развития. Ключевым направлением становится развитие гибридных систем принятия решений, в которых человек сохраняет центральную роль – не как пассивный наблюдатель, а как активный участник, способный интерпретировать, корректировать и нести ответственность за выводы алгоритмов. Такой подход позволяет

сочетать вычислительную мощь ИИ с этическим суждением, контекстным мышлением и профессиональной интуицией человека, минимизируя риски автоматизации принятия стратегически значимых решений.

Для обеспечения доверия к таким системам требуется масштабное инвестирование в отечественные решения в области объяснимого ИИ. Прозрачность алгоритмов – не просто техническая особенность, а условие юридической и социальной легитимности их использования. Разработка российских XAI-платформ позволит не только повысить уверенность пользователей и регуляторов, но и укрепить технологический суверенитет, снизив зависимость от иностранных «чёрных ящиков».

Одновременно необходимо актуализировать нормативную базу в сфере искусственного интеллекта, гармонизировав её с международными стандартами, в первую очередь с ISO/IEC 42001 – стандартом по управлению системами ИИ. Его внедрение обеспечит системный подход к оценке рисков, управлению жизненным циклом моделей и обеспечению этической ответственности, что особенно важно для критически важных секторов, таких как здравоохранение, транспорт и госуправление.

Наконец, успех всех этих инициатив напрямую зависит от человеческого капитала. Поэтому требуется масштабная модернизация программ переподготовки и повышения квалификации, с акцентом не на теоретические знания, а на практические навыки – от интерпретации выводов ИИ и управления данными до этического аудита алгоритмов и работы с XAI-инструментами. Только при условии формирования нового поколения специалистов, свободно владеющих как технологиями, так и принципами ответственного ИИ, можно говорить о переходе от экспериментов к устойчивой, безопасной и социально ориентированной цифровой трансформации.

В долгосрочной перспективе ИИ может повысить эффективность киберзащиты на 50% к 2030 году, но только при условии баланса между инновациями, безопасностью и доверием. В многополярном мире это не просто технологический выбор, а фактор национальной и экономической устойчивости.

Список источников

1. Никифоров, И. А. Роль искусственного интеллекта в кибербезопасности / И. А. Никифоров // Проблемы экономики, финансов и управления производством : сб. науч. тр. вузов России. – 2024. – № 54. – С. 230–237.
2. Ермолаев, А. С. Роль и место искусственного интеллекта в сфере обеспечения кибербезопасности / А. С. Ермолаев, В. В. Великанов // Современная наука: актуальные проблемы теории и практики. Сер.: Естественные и технические науки. – 2023. – № 12-2. – С. 71–75. – Текст : электронный.
3. Бимолдина, Ж. А. Как искусственный интеллект меняет правила игры в кибербезопасности / Ж. А. Бимолдина // Форум. Сер.: Роль науки и образования в современном информационном обществе. – 2024. – № S2 (32). – С. 235–240.
4. Лунева, С. К. Анализ инновационных технологий в контексте информационной безопасности / С. К. Лунева, В. Н. Семенова, М. А. Комиссарова // Технико-технологические проблемы сервиса. – 2025. – № 2 (72). – С. 105–111.
5. Микков, А. Д. Новые тенденции в информационной безопасности и защите данных / А. Д. Микков // Научный аспект. – 2023. – Т. 30, № 12. – С. 3765–3769.
6. Осман, С. Ш. О. Перспективы искусственного интеллекта в системах кибербезопасности / С. Ш. О. Осман // Наукосфера. – 2023. – № 9-1. – С. 213–217.
7. Сейдакматов, Н. А. Интеграция искусственного интеллекта в системы информационной безопасности: вызовы и возможности / Н. А. Сейдакматов, Е. В. Куцев // Научные исследования в Кыргызской Республике. – 2024. – № 1. – С. 24–34.
8. Самойлов, А. В. Технологии искусственного интеллекта: возможности или угрозы / А. В. Самойлов // Журнал высоких гуманитарных технологий. – 2023. – № 3 (3). – С. 67–71.
9. Фудашкин, В. А. Искусственный интеллект: двуединство преимуществ и угроз в сфере

кибербезопасности / В. А. Фудашкин // Вестник Сибирского института бизнеса и информационных технологий. – 2024. – Т. 13, № 4. – С. 166–173.

10. Авдошин, С. М. Доверенный искусственный интеллект как способ цифровой защиты / С. М. Авдошин, Е. Ю. Песоцкая // Бизнес-информатика. – 2022. – Т. 16, № 2. – С. 62–73.

11. Перевертун, Д. Р. Роль искусственного интеллекта в информационной безопасности / Д. Р. Перевертун // Международный журнал информационных технологий и энергоэффективности. – 2024. – Т. 9, № 5 (43). – С. 92–97.

12. Клишин, А. А. Правовые вопросы кибербезопасности, рисков и этики при использовании искусственного интеллекта / А. А. Клишин, К. К. Таран // Право и экономика. – 2024. – № 10 (440). – С. 24–30.

13. Сухоруков, Р. Н. Эволюция угроз в области информационной безопасности / Р. Н. Сухоруков, М. Б. Беляева // Тенденции развития науки и образования. – 2024. – № 105-14. – С. 77–81.

14. Сорокина, С. В. Влияние искусственного интеллекта на кибербезопасность банковского сектора в России / С. В. Сорокина, Н. С. Усенко, А. С. Усенко // Актуальные вопросы учета и управления в условиях информационной экономики. – 2024. – № 6. – С. 681–686.

15. Горячев, С. В. Интеграция искусственного интеллекта в системы кибербезопасности: текущие тенденции и будущие перспективы / С. В. Горячев, К. А. Лазунин // Приднепровский научный вестник. – 2024. – Т. 4, № 1. – С. 93–96.

Сведения об авторах

Султанов Гарун Султанахмедович, к.э.н., доцент кафедры экономической безопасности, анализа и аудита, Дагестанский государственный университет, Махачкала, Россия

Мусханова Хеда Жамуловна, Кандидат экономических наук, доцент кафедры финансов, кредита и антимонопольного регулирования Чеченский государственный университет имени А. Кадырова, Грозный, Россия

Алибеков Магомедрасул Магомедиминович, старший преподаватель кафедры «Государственного и муниципального управления», Дагестанский государственный университет, Махачкала, Россия.

Information about the authors

Sultanov Garun Sultanakhmedovich, Ph.D. in Economics, Associate Professor of the Department of Economic Security, Analysis and Audit, Dagestan State University, Makhachkala, Russia

Muskhanova Kheda Zhamulovna, Candidate of Economics, Associate Professor of Finance, Credit and Antimonopoly Regulation, Kadyrov Chechen State University, Grozny, Russia

Alibekov Magomedrasul Magomediminovich, Senior Lecturer of the Department of State and Municipal Administration, Dagestan State University, Makhachkala, Russia.