

УДК 004.056.53

Уязвимости информационной безопасности в российских системах искусственного интеллекта: анализ через призму модели угроз кибербезопасности искусственного интеллекта

Ворона Ангелина Алексеевна^{1,2}

angelinavorona@yandex.ru

SPIN-код: 8622-6746

Шаброва Анна Сергеевна¹

shabrova.anna.2410@list.ru

Цирлов Валентин Леонидович¹

v.tsirlov@npo-echelon.ru

¹ МГТУ им. Н.Э. Баумана, Москва, Россия² ООО «Технологии Доверия — Консультирование», Москва, Россия

Аннотация. В статье представлен обзор уязвимостей ИБ в российских системах ИИ. Анализ проведен на основе комплексной модели угроз кибербезопасности ИИ от СберБанка, охватывающей весь жизненный цикл системы. Систематизированы ключевые риски: уязвимости open-source компонентов (скрытые закладки, biased-данные) и несоответствие регуляторным требованиям. Сформулированы практические предложения по защите, включая внедрение систем фильтрации промптов и отечественных криптографических стандартов. Результаты подтверждают эффективность комплексных моделей угроз для проактивного выявления уязвимостей.

Ключевые слова: искусственный интеллект, информационная безопасность, модель угроз, регуляторные требования, аудит информационной безопасности, жизненный цикл ИИ

Введение. Кибербезопасность ИИ, активно внедряемого в критически важные сферы России, является вопросом национальной безопасности. Отечественные разработки уязвимы из-за использования импортного open-source кода и несоответствия строгим стандартам. В статье проводится комплексный анализ уязвимостей на основе модели угроз СберБанка, целью которого является систематизация рисков и формулирование практических рекомендаций по защите.

Методы и материалы; результаты. Для комплексного анализа уязвимостей российских систем ИИ была применена «Модель угроз для кибербезопасности AI» СберБанка, охватывающая все этапы их жизненного цикла. Методология исследования включала системный и сравнительный анализ, что позволило рассматривать ИИ-систему как целостный комплекс и выявлять значимые риски [1, 2]. В результате были систематизированы угрозы безопасности, классифицированные по этапам жизненного цикла и объектам воздействия. На основе проведенного анализа методом синтеза сформулированы практические предложения по усилению защиты. Данные рекомендации направлены на нивелирование наиболее актуальных для российского контекста рисков.

Анализ на основе модели СберБанка выявил специфические векторы атак российских систем ИИ, выходящие за рамки традиционных угроз. Систематизированы риски на всех этапах жизненного цикла, ключевым из которых является зависимость от импортного open-source кода. Особое внимание уделено угрозам информационно-психологического воздействия и каскадному эффекту уязвимостей. Результаты исследования демонстрируют необходимость адаптации методик защиты к российским реалиям. Полученные выводы имеют практическую ценность для формирования национальных стратегий и нормативных требований к безопасности ИИ [3–8].

Заключение. Исследование подтверждает критическую важность кибербезопасности российских систем ИИ для национальной безопасности, требуя специализированных подходов, охватывающих весь их жизненный цикл. Ключевыми рисками признаны зависимость от импортного open-source кода и необходимость адаптации международных стандартов. Практические рекомендации включают разработку национальных стандартов, усиление контроля цепочек поставок и развитие систем мониторинга. Дальнейшие исследования должны быть сосредоточены на разработке конкретных технических решений для устранения выявленных угроз. Реализация комплексного подхода обеспечит безопасное развитие ИИ в РФ.

Список источников

- [1] *Модель угроз для кибербезопасности AI*. URL: https://www.sberbank.ru/common/img/uploaded/kibrary/themap/model_ugroz_kiberbezopasnosti_ai.pdf (дата обращения 08.09.2025).
- [2] Костогрызов А.И., Нистратов А.А. Анализ угроз злоумышленной модификации модели машинного обучения для систем с искусственным интеллектом. *Вопросы кибербезопасности*, 2023, № 5 (57), с. 9–24. <https://doi.org/10.21681/2311-3456-2023-5-9-24>
- [3] Василенко В.В., Климов С.М., Швед В.В. Модель угроз информационно-психологических воздействий и алгоритм оценки информационной обстановки. *Известия Российской академии ракетных и артиллерийских наук*, 2019, № 4 (109), с. 3–7. EDN: JVOWBJ.
- [4] Костогрызов А.И. Прогнозирование рисков по данным мониторинга для систем искусственного интеллекта. *Безопасные информационные технологии. Десятая Междунар. науч.-техн. конф.: сб. тр.* Москва, Изд-во МГТУ им Н.Э. Баумана, 2019, с. 220–229. EDN: NZNGRH.
- [5] Феденева К.Д., Панов М.А. Потенциальные угрозы искусственного интеллекта и его критическая роль в жизни современного человека. *Молодой ученый*, 2024, № 40 (539), с. 34–38. URL: <https://moluch.ru/archive/539/118022/> (дата обращения 10.09.2025).
- [6] Антонов А.С. Безопасность искусственного интеллекта. Риски и угрозы. *BIS Journal*, 2023, № 3 (50), с. 24–28. URL: <https://ib-bank.ru/bisjournal/post/2101> (дата обращения 10.09.2025).
- [7] Босенко О.А. *Формирование защищенных информационных процессов с применением искусственного интеллекта на корпоративном уровне. Официальный сайт компании IBS*. URL: <https://ibs.ru/media/formirovanie-zashchishchennykh-informatsionnykh-protsessov-s-primeneniem-iskusstvennogo-intellekta-na-korporativnom-urovne/> (дата обращения 10.09.2025).
- [8] Соловьев М.М. Искусственный интеллект как новое измерение информационных угроз. *Общество: политика, экономика, право*, 2025, № 3, с. 87–91. <https://doi.org/10.24158/pep.2025.3.9>