Экз. № 1

УТВЕРЖДАЮ
Заместитель начальника
Военного учебно-научного центра
Военно-воздушных сил
«Военно-воздушная академия имени
профессора Н.Е. Жуковского и
Ю.А. Гагарина» (г. Воронеж)
по учебной и научной работе
кандидат военных наук, доцент

[redacted] В.Г. Казаков

« 11 » ноября 2025 г.

ОТЗЫВ

на автореферат диссертации Подтопельного Владислава Владимировича на тему: «Модели и методика определения последовательностей атакующих воздействий на системы искусственного интеллекта при аудите информационной безопасности», представленной на соискание ученой степени кандидата технических наук по специальности 2.3.6 – «Методы и системы защиты информации, информационная безопасность»

Актуальность темы.

Внедрение механизмов искусственного интеллекта в корпоративные информационные системы выдвигает на первый план задачу формирования целостной стратегии обеспечения информационной безопасности (ИБ). Это связано с тем, что технологии искусственного интеллекта (ИИ) могут повысить эффективность и автоматизацию процессов, но одновременно создают новые риски и уязвимости.

Традиционные подходы к аудиту и мониторингу событий ИБ оказываются недостаточно результативными перед лицом новых угроз безопасности информации, что обусловливает необходимость создания новых подходов к оценке и управлению рисками.

В сфере ИБ ключевым приоритетом становится не просто выявление существующих угроз, но и прогнозирование возможных сценариев их реализации. В диссертационной работе рассматриваются модели и методики построения и анализа содержания компьютерных атак на системы ИИ с учетом современной специфики формирования сценариев атак и рекомендаций в области ИБ при аудите информационных систем.

В существующих сегодня публикациях по этой тематике содержатся лишь качественные суждения авторов об угрозах безопасности информации, обрабатываемой с использованием технологий ИИ. Ряд публикаций лишь поднимает проблемные вопросы обеспечения ИБ технологий ИИ, применяемых в различных сферах или описывают варианты применения технологий ИИ в задачах обеспечения ИБ.

В этой связи тема диссертации В.В. Подтопельного, посвященной разработке моделей и методики определения последовательностей атакующих воздействий на системы ИИ для повышения качества процессов аудита ИБ, является актуальной.

Судя по автореферату, автор решил **научную задачу**, суть которой состояла в разработке на основе методов марковских процессов принятия решений моделей и методики построения и анализа атак на системы ИИ с учетом современной специфики формирования сценариев атак и рекомендаций в области ИБ при аудите информационных систем. Для решения данной научной задачи в работе поставлены и решены следующие частные задачи:

1. Произведен анализ существующих подходов поиска и оценки событий безопасности в информационных системах с элементами ИИ.
2. Определены возможности использования доступных наборов данных для разработки моделей формирования последовательностей атакующих воздействий на системы ИИ, определены параметры используемых при моделировании данных.
3. Разработаны модели построения и анализа атак на ПИИ при определении мер противодействия атакам.
4. Разработаны методика и алгоритм, позволяющие моделировать атакующие воздействия, используемые в процессе аудита ИБ ПИИ, изучать их динамику, использовать для составления сценариев атак.
5. Разработана архитектура программного решения, которое позволяет автоматизировать процессы моделирования.

В ходе исследований автором получен целый ряд новых научных результатов. Их содержание и **научная новизна** состоит в следующем.

1. Разработаны математические модели построения сценария атаки как последовательности событий безопасности с учетом специфики систем ИИ,

позволяющие с использованием методологии марковских процессов принятия решений количественно описать действия, необходимые для реализации угроз системам ИИ. Модели **отличаются**, введением функции вознаграждения за каждое действие нарушителя на основе оценок CVSS, эксплуатируемой уязвимости, что позволяет оценивать атаки как последовательности атакующих воздействий и серии сменяющихся состояний с учетом значений вознаграждения.

2. Разработана методика определения последовательностей атакующих воздействий на системы ИИ в процессе построения сценариев атак, используемых при проведении аудита ИБ, **отличающаяся от известных** включением этапа «Определения актуального уровня детализации»и последующим выбором уровня моделирования сценария атаки, что позволяет адаптивно изменять меру сложности моделирования и изменять скорость расчетов.

3. В ходе экспериментальных оценок разработанных моделей получены **новые** закономерности изменения количества действий в зависимости от количества узлов (компонентов), присутствующих в инфраструктуре системы ИИ.

Теоретическая значимость работы состоит в том, что в ней:

- впервые предложено для определения последовательностей атакующих воздействий на системы ИИ и обоснования формируемых сценариев атак, используемых при проведении аудита ИБ, использовать аппарат марковских процессов принятия решений, что позволяет рассматривать атаки как последовательности атакующих воздействий и серии сменяющихся состояний, фиксирующих приближение нарушителя к цели с учетом значений функции ценности (полезности) каждого действия.

- в интересах описания сценария атаки при наличии множества учитываемых параметров и компонентов предложено рассматривать три уровня абстракции состояний: позволяющих реализовать множества действий, связанных с общей логикой функционирования систем ИИ в период атакующих воздействий; связанных с инфраструктурой и логикой функционирования ИИ в период атакующих воздействий; позволяющих реализовать множества действий, являющихся этапами компьютерной атаки, описанными в тактиках MITRE, что позволило связать этапы атаки с состояниями моделируемой инфраструктуры.

Практическая значимость работы состоит в том, что результаты, полученные в диссертации, внедрены в научно-исследовательскую работу, образовательный процесс и практику деятельности в ФГБОУ ВО «Калининградский государственный технический университет» (г. Калининград), ФГАОУ ВО «Балтийский федеральный университет им. И. Канта» (г.Калининград), ООО «Центр защиты информации»

(г. Калининград), ООО «Радиоэлектронные системы» (г. Екатеринбург). Автором также получено 7 свидетельств о государственной регистрации программ для ЭВМ.

Содержание автореферата достаточно полно отражает логику проведения исследования, обладает внутренним единством.

Результаты диссертации прошли достаточную апробацию: обсуждены на 8 научных конференциях, опубликованы в 17 научных работах, включая 5 статей в журналах по перечню ВАК России.

Основными недостатками работы, на наш взгляд, являются следующие:

1. Из текста автореферата не очевидно достижение цели работы, так как не приведен критерий выбора моделей (максимум полноты) и количественные значения повышения степени полноты описания атак на системы искусственного интеллекта при аудите их информационной безопасности. Утверждения о «превосходстве» предложенных моделей по критерию полноты требуют более строгой аргументации. Кроме того, судя по содержанию таблицы 2, должна была проводиться многокритериальная оценка с привлечением дополнительных показателей «Точность», «Время выполнения» и «Адаптивность».

2. Недостаточная конкретизация научной новизны. Пункты научной новизны (стр. 6-7) сформулированы как декларации о намерениях или возможностях предлагаемого аппарата («появится возможность...», «позволяет выявлять...»), а не как конкретные, уже достигнутые научные результаты. Требуется более четкое формулирование, в чем именно заключается новизна разработанных моделей, методики и алгоритма по сравнению с известными аналогами.

3. В автореферате в явном виде не представлена формализованная постановка задачи исследования и структурно-логическая схема исследования; что затрудняет понимание роли и места разработанных моделей и методик при аудите информационной безопасности.

Несмотря на отмеченные недостатки, работа, несомненно, заслуживает внимания специалистов в области моделей, методов и средств обеспечения аудита и мониторинга состояния объекта, находящегося под воздействием угроз нарушения его информационной безопасности, и расследования инцидентов информационной безопасности в системах ИИ.

Вывод:

Содержание автореферата позволяют сделать вывод о том, что диссертационная работа на тему: «Модели и методика определения последовательностей атакующих воздействий на системы искусственного интеллекта при аудите информационной безопасности»представляет собой законченную научно-квалификационную работу, в которой содержится

решение научной задачи, имеющее существенное значение для повышения защищенности технологий искусственного интеллекта в Российской Федерации. Диссертационная работа соответствует всем требованиям п. 9-14 «Положения о присуждении ученых степеней», утвержденным постановлением Правительства Российской Федерации от 24.09.2013 г. № 842, а ее автор Подтопельный Владислав Владимирович заслуживает присуждения ученой степени кандидата технических наук по научной специальности 2.3.6– «Методы и системы защиты информации, информационная безопасность».

Отзыв рассмотрен на заседании 55 кафедры автоматизированных систем управления (и информационной безопасности) ВУНЦ ВВС «ВВА» (г. Воронеж), протокол № 5 от « 10 » ноября 2025 г.

Начальник 55 кафедры автоматизированных
систем управления (и информационной
безопасности) ВУНЦ ВВС «ВВА»
кандидат технических наук [redacted] Кущев Сергей Сергеевич

« 11 » ноября 2025 г.

Профессор 55 кафедры автоматизированных
систем управления (и информационной
безопасности) ВУНЦ ВВС «ВВА»
доктор технических наук, профессор [redacted] Будников Сергей Алексеевич

« 11 » ноября 2025 г.

Федеральноегосударственноеказенноевоенноеобразовательное учреждение высшего образования «Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е.Жуковского и Ю.А.Гагарина» (г. Воронеж) МинистерстваобороныРоссийскойФедерации,
394064, Воронежская обл., г. Воронеж, ул. Старых Большевиков, д. 54 а
номер контактного тлф.
адрес электронной почты: vva@mil.ru