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

Внедрение технологий искусственного интеллекта (ИИ) в корпоративные информационные системы (КИС), ввод в эксплуатацию специализированных ИИ-сервисов для конечных пользователей значительно усложнили решение задач кибербезопасности. Эти нововведения, несмотря на свою инновационность и потенциал, повышения эффективности, привнесли новые векторы атак, требующие глубокого анализа и комплексного подхода к их нейтрализации. Исследование В. В. Подтопельного значительно продвигает методологию аудита информационной безопасности систем ИИ. Автор предлагает научно обоснованные решения и инструменты определения возможных действий злоумышленника, что позволяет повысить степень эффективности средств безопасности информации.

В автореферате диссертации В. В. Подтопельного «Модели и методика определения последовательностей атакующих воздействий на системы искусственного интеллекта при аудите информационной безопасности» представлены оригинальные результаты в области защиты систем ИИ. Объектом исследования являются атаки на вычислительные модели и архитектуры ИИ в контексте корпоративных информационных систем.

Автор исследует процесс построения и анализа атакующих последовательностей, выявление которых способствует повышению качества аудита защищенности систем ИИ. Особое внимание в исследовании уделено созданию разных типов моделей формирования атакующих последовательностей с учетом режима их эксплуатации, позволяющих обнаруживать неявные пути реализации атак на системы ИИ, что дает возможность повысить степень подробности аудита защищённости корпоративных информационных систем.

В диссертационном исследовании решаются следующие задачи: изучаются методы и инструменты для оценки безопасности ИИ-систем (1);

определяется пригодность данных для моделирования атак на ИИ и выявить ключевые метрики (2); создаются модели для выявления атакующих последовательностей и анализа атак на ИИ (3); разрабатываются методика и алгоритм для моделирования атак, их анализа и прогнозирования (4); создается программное решение для автоматизации моделирования атак и оценки защитных механизмов (5).

В автореферате приведены результаты, полученные в ходе выполнения диссертационного исследования: модели на основе марковских процессов принятия решений (МППР), методика их применения при аудите, алгоритм определения последовательности атакующих воздействий на системы ИИ при проведении аудита информационной безопасности ИИ-систем. Результаты исследования опубликованы в 17 научных работах, из которых 5 работ в журналах ВАК, также автором получено 7 свидетельств о государственной регистрации программ.

Научная новизна результатов исследования заключается в следующем (все результаты, выносимые на защиту, являются новыми):

1. Разработка моделей на основе марковских процессов принятия решений предоставляет возможность определения последовательностей возможных действий злоумышленников при атаке для последующего создания детализированных сценариев атак, которые могут повысить эффективность аудита систем ИИ. Разработанные модели также позволяют идентифицировать ключевые действия злоумышленников, наиболее значимые для успешного осуществления атак, что существенно повышает качество управления информационной безопасностью. Таким образом, интеграция разработанных моделей МППР в процессы аудита ИБ систем ИИ способствует более глубокому пониманию потенциальных угроз и разработке стратегий их нейтрализации.

2. Применение комплексного подхода к анализу событий, идентифицированных как потенциально вредоносные для систем ИИ, позволяет выявить ключевые этапы потенциальных атак. В рамках данного исследования учитываются специфические особенности инфраструктуры систем ИИ, а также методы формализованного описания действий злоумышленников. Такой подход способствует более глубокому пониманию механизмов потенциальных угроз и разработке эффективных стратегий защиты.

3. Разработаны модели и методика определения атакующих последовательностей для формирования сценариев атак на системы и

подсистем ИИ, которые затем используются в процессе аудита информационной безопасности.

Обоснованность и достоверность научных выводов, представленных в диссертации, подтверждаются тщательным анализом актуальных исследований в этой области. При этом результаты, полученные с помощью компьютерной реализации, согласуются с теоретическими положениями.

Однако стоит отметить, что в автореферате недостаточно подробно рассмотрены особенности MITRE ATLAS, важные для применения в моделях на основе марковских процессов принятия решений.

Указанное замечание не снижают общую оценку диссертационного исследования. В заключение, автореферат позволяет сделать вывод, что диссертация **Владислава Владимировича Подтопельного** представляет собой качественную научно-квалификационную работу, сочетающую актуальность темы, новизну подхода и практическую значимость. Работа удовлетворяет требованиям ВАК и соответствуют паспорту научной специальности 2.3.6 - Методы и системы защиты информации, информационная безопасность. Диссертационная работа соответствует всем требованиям п. 9 - 14 «Положения о присуждении ученых степеней», утвержденным постановлением Правительства Российской Федерации от 24.09.2013 г. № 842. Автор заслуживает присуждения ученой степени кандидата технических наук, что подтверждает его высокий уровень компетентности и фундаментальное знание в области информационной безопасности.

Доцент, к.т.н., доцент Института компьютерных
наук и кибербезопасности
Санкт-Петербургского
политехнического
университета Петра Великого [redacted] Супрун Александр Федорович

Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого»
Почтовый адрес: ул. Политехническая, д. 29 литера Б, вн. тер. г. муниципальный округ Академическое, г. Санкт-Петербург, 195251
Телефон: (812) 297-21-63
Эл. почта: suprun_af@spbstu.ru

Подпись Супрун А.Ф.
УДОСТОВЕРЯЮ
Ведущий специалист
11 2025 г.