УДК 004.056

Каримова А.Р., Эгит Д.О., Кагарманов Д.Э.
Уфимский университет науки и технологий, Уфа

Научный руководитель:
Шагапов И.А.
Уфимский университет науки и технологий, Уфа

ОСОБЕННОСТИ РАССЛЕДОВАНИЯ ИНЦИДЕНТОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СИСТЕМАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. Расследование инцидентов безопасности информации в системах искусственного интеллекта представляет собой сложную задачу, требующую глубокого понимания как технологий искусственного интеллекта (ИИ), так и методов информационной безопасности. Успешное расследование может быть достигнуто с помощью разработки четких политик, обучения персонала и внедрения специализированных инструментов.

Постоянный мониторинг новых угроз и сотрудничество между экспертами в области ИБ и ИИ также играют важную роль в обеспечении защиты.

Ключевые слова: искусственный интеллект, инциденты безопасности, расследование, политики и процедуры, обучение персонала, методы анализа больших данных, Explainable AI, состязательные примеры, мониторинг и защита, сотрудничество специалистов.

Современные системы искусственного интеллекта (ИИ), активно внедряемые в различных сферах деятельности, становятся все более привлекательными целями для кибератак. Это связано с тем, что такие системы обрабатывают большие объемы конфиденциальной информации, обеспечивают работу критически важных инфраструктурных объектов и оказывают значительное влияние на бизнес-процессы компаний. Именно поэтому обеспечение защиты систем ИИ становится важнейшей задачей, особенно в контексте расследований инцидентов информационной безопасности.

В работе [4] приведены аспекты инцидентов ИИ-систем.

1. Автоматизация процессов принятия решений

Одна из ключевых особенностей систем искусственного интеллекта – автоматическое принятие решений. Если злоумышленник получит доступ к таким решениям, последствия могут оказаться непредсказуемыми. Например, атаки на алгоритмы распознавания лиц могут привести к ошибочным идентификациям людей, ложному допуску посторонних сотрудников на охраняемые объекты и другим негативным последствиям.

2. Уязвимость машинного обучения

Алгоритмы машинного обучения являются основой большинства современных ИИ-решений. Однако эти алгоритмы часто уязвимы перед атаками типа adversarial attacks, когда специально подготовленные данные приводят к некорректной работе модели. Такие атаки позволяют обойти механизмы идентификации и аутентификации, делая возможным несанкционированный доступ к ресурсам компании.

3. Атаки на инфраструктуру хранения данных

Важная особенность ИИ заключается в зависимости от больших объемов данных, используемых для тренировки моделей. Эти данные также представляют собой потенциальную цель для хакеров. Инцидент, связанный с утечкой данных, способен поставить под угрозу эффективность работы всей системы искусственного интеллекта и нарушить доверие пользователей.

Этапы расследования инцидентов в ИИ-системах.

В работе [2] предложен процесс расследования инцидентов информационной безопасности в системах искусственного интеллекта:

Анализ инцидента

На данном этапе происходит первичный сбор данных о произошедшем инциденте. Важно выявить следующие моменты: что именно произошло?

Какие данные были затронуты? Каковы возможные причины происшествия? Кто мог стать инициатором атаки? Эти вопросы помогают сформировать предварительную картину произошедшего и определить дальнейшие шаги.

Оценка последствий

После анализа инцидента проводится оценка возможных последствий нарушения безопасности. Необходимо оценить масштаб ущерба, включая финансовые потери, репутационные риски и угрозы для бизнеса организации. Важность данного этапа связана с необходимостью принимать оперативные меры по минимизации потерь.

По итогам расследования разрабатываются рекомендации по повышению уровня защищенности системы искусственного интеллекта. В работе [3] они подробно описываются, сюда входят мероприятия по улучшению архитектуры инфраструктуры, обновлению программного обеспечения, проведению регулярных проверок безопасности и обучению персонала правилам безопасной работы с системой. В работе [1] указан ряд специфических проблем, который возникает при расследовании инцидентов информационной безопасности в системах искусственного интеллекта:

Сложность выявления вредоносных действий из-за высокой автоматизации процесса обработки данных.

Недостаточная прозрачность механизмов принятия решений алгоритмами машинного обучения.

Возможность влияния третьих сторон через обучение нейронных сетей на основе искаженных данных.

Кроме того, существуют трудности правового характера, поскольку законодательство многих стран пока не предусматривает четких норм регулирования вопросов, связанных с искусственным интеллектом.

Расследование инцидентов информационной безопасности в системах искусственного интеллекта требует особого подхода, учитывающего особенности технологии и сложности ее функционирования. Грамотное расследование позволяет минимизировать ущерб от нападения и предотвратить подобные инциденты в будущем. Поэтому развитие компетенций специалистов по обеспечению информационной безопасности является приоритетной задачей современного общества.

Список использованных источников:

1. Коваленко А.И. Угрозы информационной безопасности в системах ИИ. Издательский дом РАЭК, 2020. URL: https://www.raeonline.ru/publications/ugrozy-info-bezopasnosti.

2. Борисов О.В. Методология расследования киберинцидентов. М.: Юрайт, 2019. URL: https://www.jurait.ru/book/metodologiya-rassledovaniya-kibertincidentov.

3. Романов, С. В. Защита информации в системах искусственного интеллекта. Издательство «БХВ-Петербург», 2022. URL: https://www.bhvpeterburg.ru/book/zashchita-informacii-v-sistemah-iskusstvennogo-intellekta.

4. Березкин, А. Г. Управление инцидентами безопасности в ИТ. Издательство «Гардарика», 2021. URL: https://www.gardarik.com/books/upravlenie-incidentami-bezopasnosti.

Karimova A.R., Egit D.O., Kagarmanov D.E.
Ufa University of Science and Technology, Ufa

Scientific supervisor:
Shagapov I.A.
Ufa University of Science and Technology, Ufa

FEATURES OF INVESTIGATING INFORMATION SECURITY INCIDENTS IN ARTIFICIAL INTELLIGENCE SYSTEMS

Abstract. Investigating information security incidents in artificial intelligence systems is a complex task that requires a deep understanding of both AI technologies and information security methods. Successful investigations can be achieved through the development of clear policies, staff training, and the implementation of specialized tools. Continuous monitoring of new threats and collaboration among experts in the fields of information security and AI also play a crucial role in ensuring protection.

Keywords: artificial intelligence, security incidents, investigation, policies and procedures, staff training, big data analysis methods, Explainable AI, adversarial examples, monitoring and protection, collaboration of specialists.