УДК 004.056:351.74

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ПРИ ИНТЕГРАЦИИ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ В ГОСУДАРСТВЕННЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ ОРГАНОВ ВНУТРЕННИХ ДЕЛ

М.А. Елькина[1], В.А. Епифанцева[1]

[1]Краснодарский университет МВД России, г. Краснодар
e-mail: mari.elkina.2004@mail.ru

Аннотация

В работе представлена комплексная методология обеспечения информационной безопасности при внедрении генеративных моделей искусственного интеллекта в информационные системы органов внутренних дел Российской Федерации. Рассмотрены ключевые угрозы, связанные с использованием облачных LLMсервисов, и обоснована необходимость перехода к модели локального развертывания (on-premise) с применением архитектуры Retrieval-Augmented Generation (RAG). Предложена многоуровневая система защиты, включающая механизмы идентификации и аутентификации, фильтрации входных и выходных данных, контроля доступа и непрерывного аудита. Показано, что реализация предложенной методологии позволяет обеспечить соответствие требованиям Федерального закона № 149ФЗ и Приказа ФСТЭК России № 17. Работа имеет прикладной характер и может служить основой для разработки ведомственных нормативных актов по безопасному внедрению ИИ в деятельности ОВД.

Ключевые слова: информационная безопасность, генеративный искусственный интеллект, государственные информационные системы, RAG, защита информации, органы внутренних дел, ФСТЭК, Zero Trust, эшелонированная оборона.

Введение

Современный этап развития правоохранительной деятельности характеризуется активным внедрением цифровых технологий, направленных на повышение эффективности обработки больших объемов данных. В частности, генеративные модели искусственного интеллекта (LLM) демонстрируют значительный потенциал для автоматизации аналитической работы, подготовки процессуальных документов и поддержки принятия решений в оперативной деятельности органов внутренних дел (ОВД). Однако использование публичных облачных сервисов (например, ChatGPT, Gemini) для обработки ведомственной информации сопряжено с высокими рисками нарушения конфиденциальности, целостности и доступности данных.

Согласно Федеральному закону от 27 июля 2006 г. № 149ФЗ «Об информации, информационных технологиях и о защите информации» [1], государственные органы обязаны обеспечивать защиту информации, обрабатываемой в информационных системах. При этом Приказ ФСТЭК России от 11 февраля 2013 г. № 17 [2] устанавливает обязательные требования к защите информации, не составляющей государственную

тайну, в государственных информационных системах (ГИС). Применение облачных LLM сервисов, размещенных на зарубежных платформах, противоречит данным требованиям.

1. Анализ угроз при использовании генеративных моделей в ОВД

Внедрение генеративных моделей искусственного интеллекта в информационную инфраструктуру органов внутренних дел открывает значительные возможности для автоматизации рутинных процессов, анализа оперативной информации и поддержки принятия решений. Однако сопровождающие этот процесс технологические и организационные изменения влекут за собой появление новых векторов киберугроз, требующих системного подхода к оценке рисков и построению защищённой архитектуры. В контексте деятельности ОВД, где обрабатывается значительный объём персональных данных, сведений, составляющих служебную тайну, а также оперативно-розыскной информации, обеспечение информационной безопасности при использовании LLM становится приоритетной задачей [3]. Совокупность угроз может быть классифицирована по трем основным категориям: угрозы конфиденциальности, целостности и доступности информации, в соответствии с фундаментальной моделью информационной безопасности.

Таблица 1 – Классификация угроз при внедрении LLM в ГИС ОВД

Категория	Угроза	Краткое описание
Конфиденциальность	Утечка через облачные сервисы	Передача данных на внешние API – потеря контроля, нарушение ФЗ №149 и Приказа ФСТЭК №17
	Компрометация весов модели	Наличие «бэкдоров» в open-weight моделях, активируемых специфическим триггером
	Галлюцинации LLM	Генерация ложных, но правдоподобных фактов → риск дезинформации сотрудников
Целостность	Prompt Injection	Внедрение в запрос команд, заставляющих модель игнорировать системные инструкции
	AG Poisoning	Подмена документов в векторной базе данных → распространение фальсифицированных сведений
Доступность	DoS-атака	Перегрузка GPU- кластера серией сложных запросов → деградация/остановка сервиса
	Зависимость от провайдера	Отключение облачного сервиса по политическим или техническим причинам

2. Методология построения защищенного контура генеративного ИИ

Для нейтрализации выявленных угроз мною была предложена соответствующая методология создания защищенного ИИ контура, основанная на принципе «эшелонированной обороны» (Defensein Depth). Ключевым требованием является переход от облачных вычислений к on-premise развертыванию, обеспечивающему физическую и логическую изоляцию данных, что позволит применять такие системы

внутри ведомственного сегмента сети. Архитектура системы строится на базе технологии RAG, которая разделяет процессы хранения «знаний» и их обработки. Векторная база данных выступает изолированным хранилищем ведомственных документов, что позволяет исключить необходимость дообучения самой нейросети на конфиденциальных данных [4].

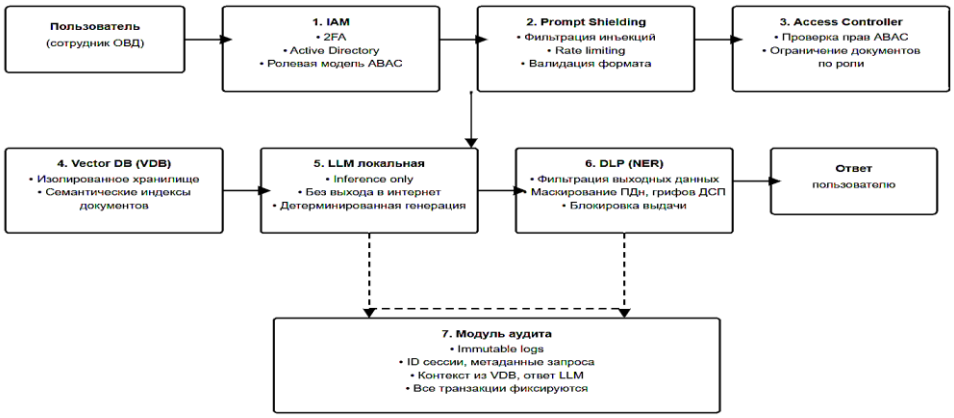


Рисунок 1 – Блок-схема защищённого конвейера обработки (RAG on-premise) [5]

Таким образом, система должна опираться на строго детерминированную цепочку обработки (pipeline), которая бы исключала возможность несанкционированного вывода данных или утечки информации [6].

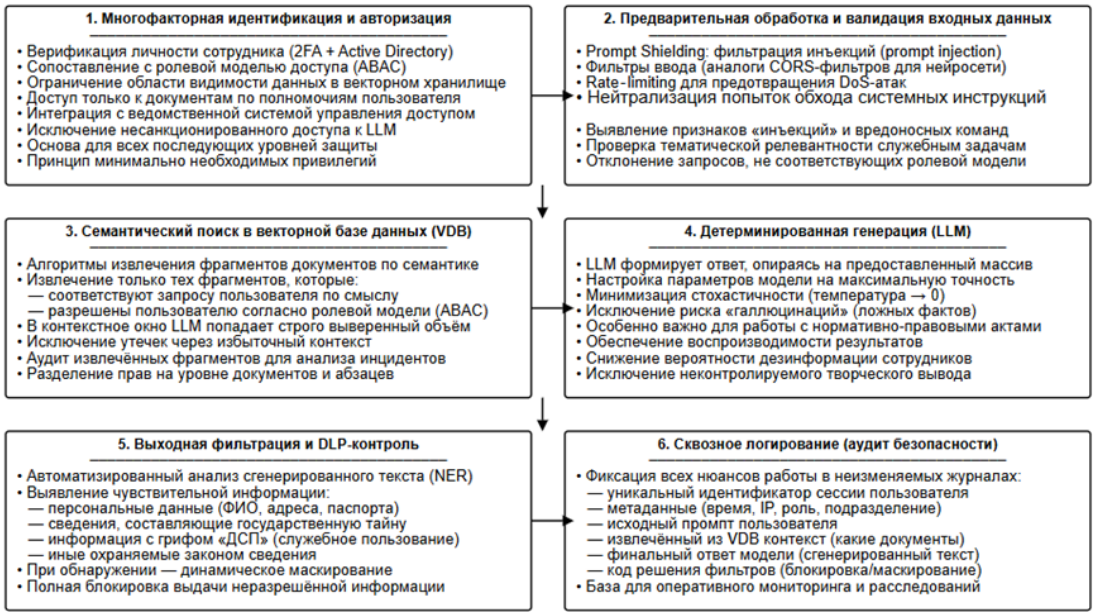


Рисунок 2 – Уровни эшелонированной защиты информации в LLM для ОВД

Заключение

Внедрение генеративных моделей ИИ в деятельность ОВД является перспективным направлением. Предложенная методология построения защищенного контура позволяет минимизировать риски и обеспечить соответствие требованиям действующего законодательства.

Список использованных источников

1. Федеральный закон от 27 июля 2006 г. № 149ФЗ «Об информации, информационных технологиях и о защите информации» // Собрание законодательства РФ. – 2006. – № 31 (ч. 1). – Ст. 3448.

2. Приказ ФСТЭК России от 11 февраля 2013 г. № 17 «Об утверждении Требований о защите информации, не составляющей государственную тайну, содержащейся в государственных информационных системах» // Информационный бюллетень ФСТЭК России. – 2013. – № 2. – С. 12–45.

3. Попов, И.В. Информационная безопасность государственных информационных систем: учебник / И.В. Попов, А.В. Федоров. – М.: Горячая линияТелеком, 2021. – 384 с.

4. Галатенко, В.А. Современные подходы к обеспечению безопасности искусственного интеллекта / В.А. Галатенко, А.В. Кудинов // Труды Института системного программирования РАН. – 2023. – Т. 35, № 2. – С. 45–62.

5. Lewis, P. et al. RetrievalAugmented Generation for KnowledgeIntensive NLP Tasks // arXiv preprint arXiv:2005.11401. – 2020.

6. Методические рекомендации по обеспечению безопасности при использовании технологий искусственного интеллекта в государственных информационных системах / ФСТЭК России. – М., 2024. – 56 с.

ENSURING INFORMATION SECURITY DURING THE INTEGRATION OF GENERATIVE MODELS INTO STATE INFORMATION SYSTEMS OF INTERNAL AFFAIRS BODIES

M.A. Elkina[1], V.A. Epifantseva[1]

[1]Krasnodar University of the Ministry of Internal Affairs of Russia, Krasnodar
e-mail: mari.elkina.2004@mail.ru

Abstract

This paper presents a comprehensive methodology for ensuring information security during the integration of generative artificial intelligence (AI) models into the information systems of the Internal Affairs Bodies of the Russian Federation. Key threats associated with the use of cloudbased LLM services are examined, and the necessity of transitioning to an onpremise deployment model using the RetrievalAugmented Generation (RAG) architecture is substantiated. A multilayered security system is proposed, which includes mechanisms for identification and authentication, input and output data filtering, access control, and continuous auditing. It is demonstrated that the implementation of the proposed methodology ensures compliance with the requirements of Federal Law No. 149FZ and FSTEC of Russia Order No. 17. The work has an applied nature and can serve as a basis for developing departmental regulatory acts for the safe integration of AI into the activities of the Internal Affairs Bodies.

Keywords: information security, generative artificial intelligence, state information systems, RAG, information protection, Internal Affairs Bodies, FSTEC, Zero Trust, defenseindepth.