

АНАЛИЗ РИСКОВ УТЕЧКИ КОНФИДЕНЦИАЛЬНЫХ ДАННЫХ ПРИ ИСПОЛЬЗОВАНИИ ИИ-АССИСТЕНТОВ В ПРОЦЕССЕ РАЗРАБОТКИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

В.А. Черный¹

¹Самарский университет, г. Самара
e-mail: stavalrecrut@yandex.ru

Аннотация

В работе рассмотрены риски утечки конфиденциальных данных при использовании ИИ-ассистентов в разработке программного обеспечения. Выделены основные сценарии передачи чувствительной информации внешним сервисам, предложена классификация рисков и приведен демонстрационный пример проверки сценария передачи журнала событий в облачный ИИ-сервис. Сформулированы меры снижения угроз.

Ключевые слова: информационная безопасность, конфиденциальные данные, утечка данных, большие языковые модели, ИИ-ассистенты, безопасная разработка, secure SDLC.

Введение

В последние годы инструменты искусственного интеллекта, основанные на больших языковых моделях, активно применяются в разработке программного обеспечения. Они используются для генерации кода, рефакторинга, анализа ошибок и документирования, ускоряя типовые операции, но одновременно формируя новые риски в области информационной безопасности [1, 2].

Особую значимость приобретает риск утечки конфиденциальных данных. При взаимодействии с ИИ-ассистентом разработчик может передавать в запросе фрагменты исходного кода, конфигурации, журналы событий, сведения о внутренней архитектуре системы и учетные данные. Целью работы является анализ основных сценариев такой утечки и определение мер по снижению соответствующих рисков [2, 4, 5].

1. Использование ИИ-ассистентов в современной разработке программного обеспечения

ИИ-ассистенты встраиваются в редакторы кода, системы управления версиями и средства проверки качества. Они применяются для генерации и пояснения кода, составления unit-тестов, анализа логов и интерпретации конфигураций. Следовательно, такие инструменты получают доступ к значимому контексту разработки и должны рассматриваться как элемент технологического контура, влияющий на конфиденциальность и управляемость данных [1, 3].

Подходы NIST подчеркивают необходимость идентификации источников риска и внедрения механизмов контроля на всех этапах жизненного цикла ПО [1-3].

2. Основные сценарии утечки конфиденциальных данных

Первый типовой сценарий связан с прямой передачей во внешний облачный сервис фрагментов внутреннего исходного кода, конфигураций или бизнес-логики. Если в них присутствуют токены, пароли, адреса внутренних сервисов и идентификаторы клиентов, возникает риск раскрытия чувствительных сведений за пределами контролируемой среды [4, 5].

Второй сценарий связан с передачей журналов событий и дампов ошибок, которые нередко содержат персональные данные, параметры соединений и сведения о сетевой инфраструктуре. Третий сценарий обусловлен интеграцией ИИ-моделей с внешними инструментами и плагинами: при недостаточном разграничении доступа либо успешной prompt injection-атаке возможно раскрытие чувствительной информации через ответы модели [4-7].

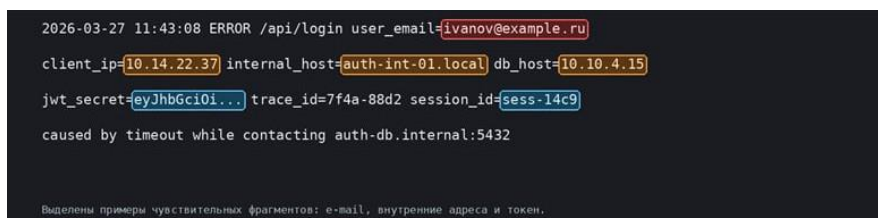
3. Классификация рисков и возможные последствия

Риски утечки конфиденциальных данных целесообразно классифицировать по источнику возникновения, характеру обрабатываемых данных и механизму реализации. По источнику выделяются человеческий фактор, архитектурные особенности сервиса, ошибки настройки доступа и отсутствие локальных регламентов. По характеру данных различаются утечки исходного кода, учетных, персональных и служебных технических данных. Последствия таких инцидентов носят технический, экономический и правовой характер и могут выражаться в раскрытии структуры системы, утрате интеллектуальной собственности и нарушении требований к защите персональных данных и конфиденциальной информации [2, 4-7].

4. Практический пример проверки сценария передачи журнала событий

Для демонстрационной проверки был смоделирован тестовый стенд авторизации веб-приложения. Использовался синтетический журнал, отражающий ошибку 500 при обращении к ресурсу /api/login. В журнал были включены чувствительные поля: e-mail пользователя, внутренние IP-адреса и имена узлов, фрагмент JWT-секрета, а также идентификаторы сессии и трассировки.

В ИИ передавался запрос: «Проанализируй причины ошибки авторизации и предложи действия по исправлению». На рисунке 1 приведен фрагмент тестового журнала до обезличивания.



```
2026-03-27 11:43:08 ERROR /api/login user_email=ivanov@example.ru
client_ip=10.14.22.37 internal_host=auth-int-01.local db_host=10.10.4.15
jwt_secret=eyJhbGciOiI... trace_id=7f4a-88d2 session_id=sess-14c9
caused by timeout while contacting auth-db.internal:5432
```

Выделены примеры чувствительных фрагментов: e-mail, внутренние адреса и токен.

Рисунок 1 – Фрагмент синтетического журнала событий
в демонстрационном эксперименте

Проверка включала два этапа: ручную классификацию полей журнала и поиск маркеров по регулярным выражениям для e-mail, IPv4-адресов, токеноподобных строк и внутренних DNS-имен. В исходном фрагменте были выявлены 6 чувствительных

элементов. Затем запрос повторялся после маскирования полей значениями [MASKED_EMAIL], [INTERNAL_HOST], [TOKEN] и [SESSION_ID]. Результаты приведены в таблице 1.

Таблица 1 – Результаты демонстрационной проверки

Категория оценки	До маскирования	После маскирования
Идентификаторы пользователя	1 e-mail, 1 session_id	Прямые идентификаторы отсутствуют
Внутренняя инфраструктура	3 внутренних адреса/узла	Адреса и имена узлов скрыты
Аутентификационные данные	1 JWT-фрагмент, 1 служебный маркер	Секреты удалены
Итоговая оценка риска	Высокая	Умеренная

Сравнение показало, что после обезличивания модель сохранила возможность локализовать причину сбоя: в обоих вариантах указывались вероятные причины – ошибка валидации токена, некорректная обработка исключения и недоступность узла auth-db. Следовательно, предварительная минимизация данных заметно снижает риск утечки без существенной потери аналитической ценности.

5. Меры снижения рисков в процессе безопасной разработки

Снижение вероятности утечки требует сочетания организационных и технических мер. На организационном уровне необходимо определить допустимые сценарии использования ИИ-ассистентов, запретить передачу кода закрытых систем без предварительной очистки и обучить сотрудников правилам безопасной работы с такими сервисами [1-3]. На техническом уровне целесообразно применять маскирование токенов, паролей, персональных данных и внутренних идентификаторов, автоматическое обнаружение секретов, фильтрацию входных и выходных данных, аудит обращений и, при повышенных требованиях к конфиденциальности, локально развернутые модели или изолированные корпоративные контуры обработки [2, 3, 5].

Заключение

Использование ИИ-ассистентов дает заметный прикладной эффект, но одновременно формирует новые каналы утечки конфиденциальных данных. Наиболее существенные угрозы связаны с передачей во внешние сервисы исходного кода, журналов событий, конфигураций и другой чувствительной информации. Демонстрационный пример показывает, что предварительная очистка запроса позволяет заметно снизить риск утечки без существенной потери аналитической ценности ИИ-инструмента.

Список использованных источников

1. Artificial Intelligence Risk Management Framework (AI RMF 1.0) / E. Tabassi, K. Burns, M. Hadjimichael [et al.]. – NIST, 2023. – URL: <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf> (дата обращения: 15.03.2026).
2. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile / NIST. – 2024. – URL: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf> (дата обращения: 15.03.2026).
3. Secure Software Development Framework (SSDF) Version 1.1 / M. Souppaya, K. Scarfone, H. Dodson [et al.]. – NIST, 2022. – URL: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218.pdf> (дата обращения: 19.03.2026).
4. OWASP Top 10 for LLM Applications 2025. – URL: <https://owasp.org/www-project-top-10-for-large-language-model-applications/assets/PDF/OWASP-Top-10-for-LLMs-v2025.pdf> (дата обращения: 19.03.2026).
5. AI Privacy Risks & Mitigations - Large Language Models (LLMs). - European Data Protection Board, 2025. – URL: <https://www.edpb.europa.eu/system/files/2025-04/ai-privacy-risks-and-mitigations-in-llms.pdf> (дата обращения: 19.03.2026).
6. Федеральный закон от 27.07.2006 N 152-ФЗ «О персональных данных». – URL: https://www.consultant.ru/document/cons_doc_LAW_61801/ (дата обращения: 23.03.2026).
7. Федеральный закон от 29.07.2004 N 98-ФЗ «О коммерческой тайне». – URL: https://www.consultant.ru/document/cons_doc_LAW_48699/ (дата обращения: 23.03.2026).

ANALYSIS OF CONFIDENTIAL DATA LEAKAGE RISKS WHEN USING AI ASSISTANTS IN THE SOFTWARE DEVELOPMENT PROCESS

V.A. Chernyi¹

¹*Samara University, Samara*
e-mail: stavalrecrut@yandex.ru

Abstract

The paper examines confidential data leakage risks when AI assistants are used in software development. Main leakage scenarios, a risk classification and mitigation measures are described; a brief demonstrational check of log transfer to a cloud AI service is presented.

Keywords: information security, confidential data, data leakage, large language models, AI assistants, secure software development, secure SDLC.