

точка банк

НЕ ТАК
СТРАШЕН ИИ,
КАК ЕГО
МАЛЮЮТ

Вячеслав Касимов, октябрь 2026 г.

Сначала закроем старое, потом испугаемся нового

Заголовки обещают апокалипсис: исчезновение рабочих мест, эксплойты за секунды, дипфейки и AGI.

В операционке

Причины большинства инцидентов по-прежнему фишинг, кража учётных данных и мисконфигурации.

Без всякого ИИ

Средний срок обнаружения вторжения — **194 дня**.

80% уязвимостей — неустановленные патчи и слабые пароли.

ИИ меняет скорость и масштаб, но не отменяет базовую гигиену.

Что действительно изменилось

точка банк

Масштабирование и эволюция: ИИ удешевляет знакомые техники и делает их убедительнее и эффективнее.

Фишинг

Персонализация и грамотная подача. Рост конверсии на **30–40%**.

Вредоносный код

Больше полиморфности и обход сигнатур антивирусов.

Deepfake

Клонирование голоса и видео масштабирует таргетинг.

Разведка

OSINT и реверс-инжиниринг ускоряются **в десятки раз**.

0 day

Сокращение времени поиска и создания эксплойтов **с месяцев до дней**.

Вопреки заголовкам

точка банк

Масштабирование и эволюция: ИИ удешевляет знакомые техники и делает их убедительнее и эффективнее.

Массовые атаки ИИ-агентов



Были единичные атаки, проведённые ИИ-агентами с участием атакующего эксперта.

0 day без эксперта



Не зафиксировано ни одного, даже по запросу.

Полная замена SOC-аналитиков



Есть попытки создания SIEM полностью на ИИ.

Восстание LLM



Не произошло пока, однако есть много троянов под видом MCP серверов.

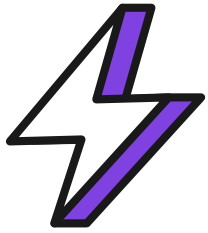
Обход контролей одним промтом



Ни EDR, ни DLP не сбойнули, но поиграться с промтами всегда можно.

Шумная рутина

точка банк



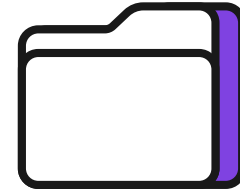
Рост атак низкого качества

К апокалипсису привести
не должно.



Инциденты в чат-ботах

Prompt Injection становится вполне
себе атакой.



Утечки данных

Через использование сторонних
LLM.

5 новых рисков, требующих управления

Теневой ИИ

Конфиденциальные данные уходят в публичные LLM.

Prompt Injection

Манипуляции чат-ботами и RAG-системами.

Data Poisoning

Отравление данных при доступе к ML-пайплайну.

Фишинг

Усиленный дипфейками ВЕС компенсируется аналитикой и верификацией.

AI supply chain

Компрометация модели, API, датасетов на стороне поставщика.

А ещё галлюцинации, использование LLM как MCP и прочие прелести OWASP GenAI LLM Top 10.

Все пять рисков укладываются в знакомые практики управления доступом, данными, поставщиками и процессами.

Надо взять и запретить?

точка банк

Запрет тоже создаёт риск для бизнеса.

Паралич решений: полный запрет рождает теневое использование ИИ и нулевую видимость.

Перекося бюджета: деньги и ФОТ уходят на «анти-ИИ-щиты» вместо полезной «рутины».

Усталость от угроз: команда выгорает от ложных тревог.

Главное: бизнесу сейчас действительно нужен ИИ.

Зрелая позиция:

не отрицать новый риск,
а встроить его в действующую систему
управления.

Полезнее в защите,
чем в нападении

точка банк

Область	ИИ-ассистент	Человек
Оптимизация SOC	Сортировка и обогащение	Приоритизация
Анализ логов	Массовая корреляция	Поиск аномалий
Threat Intelligence	Парсинг и саммари	Интерпретация
Работа с WAF и SIEM	Формирование правил на основе собранной статистики	Валидация правил и их применение
Application Security, Red team, Bug bounty	Предварительный анализ	Принятие решения по уязвимости и критичности
DLP	Валидация срабатываний по регулярным выражениям	Принятие решения

Решения остаются за экспертами.

Прагматичный план: от гигиены к зрелости

точка банк

0–3 месяца

Гигиена

Инвентаризация ИИ, политика допустимого использования, классификация данных, обучение верификации дипфейков и ИИ-фишинга.

3–9 месяцев

Контроль

Корпоративный ИИ-шлюз: DLP, RBAC и аудит промптов, Rate limiting, Threat Model для ML-компонентов, тесты Prompt Injection.

9–18 месяцев · Зрелость

Интеграция в корпоративный риск-менеджмент; AI BOM, регулярный Red Teaming, метрики детектирования и реагирования.

Это эволюция риск-менеджмента, знакомая по облакам, IoT и DevOps.

Наши страхи

точка банк



Вайб-кодинг



Prompt Injection



Бурный рост AI

Страх №1: вайб-кодинг

точка банк



Команда AI x SDLC сделала собственный прокси для вайб-кодеров: централизовали Claude и Codex.

Есть стимул использования: Точка Банк покупает подписки.

Контроль: проверяем было ли ревью от вайб-кодера + полный цикл SSDLC + фильтруем конфиденциальные данные.

Страх №2: Prompt Injection

точка банк



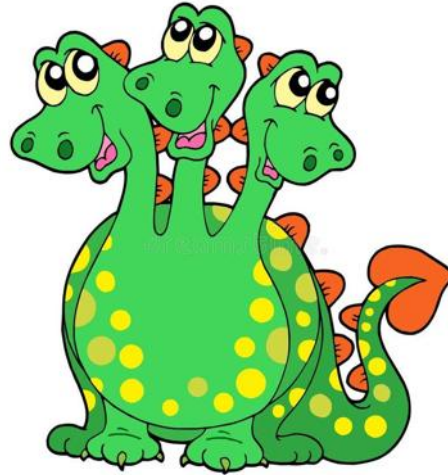
Команда ML сделала свой AI Firewall: ещё одна модель, которая смотрит на входы и выходы в основную LLM.

Сама неотвратимость: доступ только через firewall.

Контроль: посматриваем, кто шалит + фильтруем чувствительную информацию.

Страх №3: бурный рост AI

точка банк



Отдельный круг по AI: обсуждаем прежде, чем делать.

Сформулировали риски: агрегировали OWASP GenAI LLM
Top 10 и методические рекомендации Банка России.

Контроль: уходим в реализацию библиотек, фреймворков,
митигирующих риски, для встраивания в сервисы.

точка банк

СПАСИБО
ЗА
ВНИМАНИЕ!

