

КОНФЕРЕНЦИЯ АБИСС ·

ИИ внедряют быстрее, чем **защищают**: регуляторика безопасности ИИ и что с ней делать уже сегодня



Альбина Аскерова

Руководитель направления по взаимодействию с регуляторами Swordfish Security

13.10.2026

Наш вклад в индустрию

1 Постоянный член ТК 362

Состоим в ПК-2 по разработке серии ГОСТ по безопасной разработке ПО

2 Участники Консорциума исследований безопасности технологий ИИ

Участвуем в разработке методологии тестирования решений ИИ на предмет доверия и безопасности

3 Поддерживаем проект РБПО.РФ

Информационный ресурс Центра компетенций ФСТЭК России и ИСП РАН — публикуем экспертные материалы в блоге портала

4 Участники Рабочей группы № 5 «Информационная безопасность в сфере ИИ» Правительства РФ

Создаём технологические и регуляторные условия в экспертном органе Правительственной подкомиссии по развитию и внедрению технологий ИИ

Технологии опережают **защиту**

- 1** ИИ встраивают и в продукты, и в саму разработку быстрее, чем выстраивают его безопасность
- 2** Требования уже обязательны, но распределены по документам разных регуляторов
- 3** Двойной риск: несоответствие требованиям и избыточные меры «на всякий случай»

23%

в среднем внедрено базовых мер защиты ИИ

у каждой третьей организации — ни одной

Требования к безопасности ИИ уже есть

Уровень	Документ	Суть
Закон	243-ФЗ(26.07.2026)	Риск-ориентированный подход, обязанности разработчиков
ФСТЭК	Приказ № 117, п. 60/61(ред. 08.05.2026 № 137) + БДУ ФСТЭК и методика (дек. 2025 · 12.04.2026)	Защита моделей ИИ, описанные угрозы
Банк России	МР № 3-МР(16.06.2026)	Безопасность ИИ на финансовом рынке
ГОСТ / ТК 362	ГОСТ Р 56939-2024 — основа РБПО + ПНСТ 1046-2026 «ИИ в КИИ» + Проект ГОСТ РБПО ИИ	Требования к безопасной разработке + первый национальный стандарт по ИИ в КИИ

Что изменится в ближайшее время

Кто формирует требования

243-ФЗ



Правительственная подкомиссия по ИИ



7 рабочих групп



РГ № 5 «ИБ в сфере ИИ»

ФСТЭК

Проект ГОСТ

**«Защита информации.
Разработка безопасного ПО,
реализующего технологии ИИ»**

Требования РБПО, закреплённые для ИИ

Приказ № 117 — проект

- + Агенты ИИ
- + Выделенный сегмент и усиленная МФА
- + п. 14 «д» — обязательный внутренний документ «порядок защиты информации при использовании ИИ»

Приказ № 21 — проект

Защита ИИ в ИСПДн также выделена в отдельную меру

Разработка ИИ ≠ классическая РБПО

	Классическая РБПО		Разработка ИИ
Что защищаем	код и зависимости	→	+ данные, модель (веса, параметры), запросы к модели, подключаемые базы знаний
Поведение	предсказуемое	→	вероятностное: на один и тот же запрос — разные ответы
Контрольные точки	модульные тесты, статический анализ, анализ состава ПО	→	+ оценка качества ответов, тестирование атаками на модель, защитные фильтры
Угрозы	уязвимости кода	→	+ вредоносные запросы и обход ограничений модели, отравление данных, кража и подмена весов
Цепочка поставок	перечень компонентов ПО	→	+ перечень компонентов ИИ: внешние модели, наборы данных, подключаемые модули
С чего начинается защита	с написания кода (проектирования?)	→	со сбора данных: сбор → обучение → оценка → эксплуатация
Жизненный цикл	завершается выпуском версии	→	продолжается: поведение модели меняется со временем, нужен постоянный контроль

ИИ как инструмент разработки (безопасной)



Где применяют

Генерация кода

Проверка кода

Автоматическое
документирование

Разбор результатов анализаторов

Моделирование угроз



Риски

Утечка кода и секретов во
внешние модели

Уязвимости в сгенерированном
коде

Ложные выводы об уязвимостях



Контроль

Доверенные и локальные модели

Запрет на передачу
чувствительных данных

Проверка результатов человеком

Журналирование действий

Измеримая зрелость

Открытая методика: учитывает международные модели зрелости, построена на действующих регуляторных требованиях к безопасности ИИ

7 доменов РБПО на ЖЦ

Управление

Данные и конфиденциальность

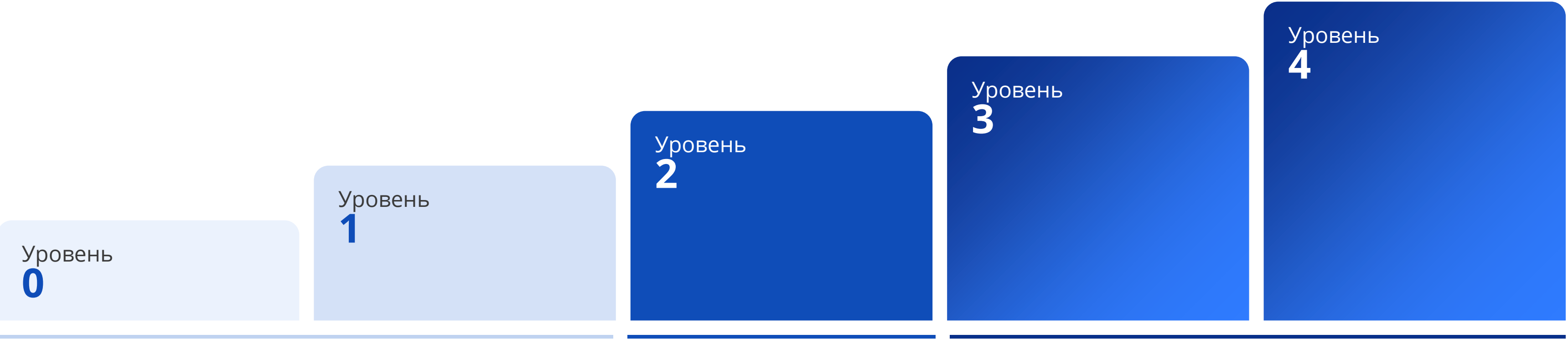
Модели

Цепочка поставок

Инфраструктура

Агенты

Базы знаний



Несоответствие · риск санкций

Готовность к
проверкам регулятора

Что делать уже сегодня

Цель

Довести базовые домены до Уровня 2 — готовности к проверкам

Следующий шаг

Домены на Уровне 2 поднять до 3 за счёт автоматизации

- 1

Инвентаризация ИИ-активов и данных

Реестр моделей и наборов данных, где они используются, выявление неучтённого ИИ

243-ФЗ

152-ФЗ
- 2

Распространить контрольные точки разработки на модель и данные

+ оценка качества ответов, тестирование атаками, защитные фильтры, проверка данных на отравление

Модели

Данные
- 3

Перечень компонентов ИИ и проверка поставщиков

Учёт и проверка внешних моделей, наборов данных, подключаемых модулей

Цепочка поставок
- 4

Контроль данных

Что передаётся в модель, где хранится, кто проверяет результат

152-ФЗ

Данные
- 5

Внутренний регламент по ИИ

Порядок безопасной разработки и использования — заранее, под п. 14 «д»

Приказ № 117

Управление

Типовые ошибки



«Безопасность в конце»

вместо управления рисками на каждом этапе ЖЦ



Код и секреты во внешних моделях

передача без контроля



Забытая цепочка поставок

внешние модели и наборы данных без учёта



Меры «на всякий случай»

без привязки к рискам



Ждать окончательного набора регуляторных требований

вместо повышения зрелости уже сейчас

Безопасная разработка ПО (ПО ИИ) — это **процесс**, а не документ

1

РБПО распространяется на данные и модель; защита начинается со сбора данных и обучения

2

Основа требований есть — ГОСТ Р 56939, проект ГОСТ по РБПО для ИИ, приказ № 117

3

Повышение зрелости и постоянный пересмотр требований с учетом стремительного развития технологии и способов ее применения

Спасибо за внимание

Альбина Аскерова

Руководитель направления по
взаимодействию с регуляторами Swordfish
Security

aaskerova@swordfishsecurity.ru

Мой Habr



SAIMM 2.0

Модель зрелости безопасности ИИ

saimm.swordfish-security.ru



Оценка зрелости онлайн

Уровни зрелости ИИ-процессов

saimm.swordfish-security.ru/matrix.html