Жизненный цикл ML моделей в ИБ onprem решениях

Обеспечиваем практическую кибербезопасность

22 года

опыта исследований и разработок

3 + тыс.

сотрудников: инженеров по ИБ, разработчиков, аналитиков и других специалистов

20 +

продуктов для защиты компаний и отраслей

205 + тыс.

акционеров

13 лет

проводим крупнейший в России кибер-фестиваль Positive Hacking Days

10 лет

Выпускаем журнал о кибербезопасности Positive Research

- Создаем продукты и решения
- Проводим аудиты безопасности
- Расследуем инциденты
- Исследуем угрозы

whoami

Данила Ваганов,
ML Team Lead, Positive Technologies

Коля Лыфенко,
Head of ML, Positive Technologies

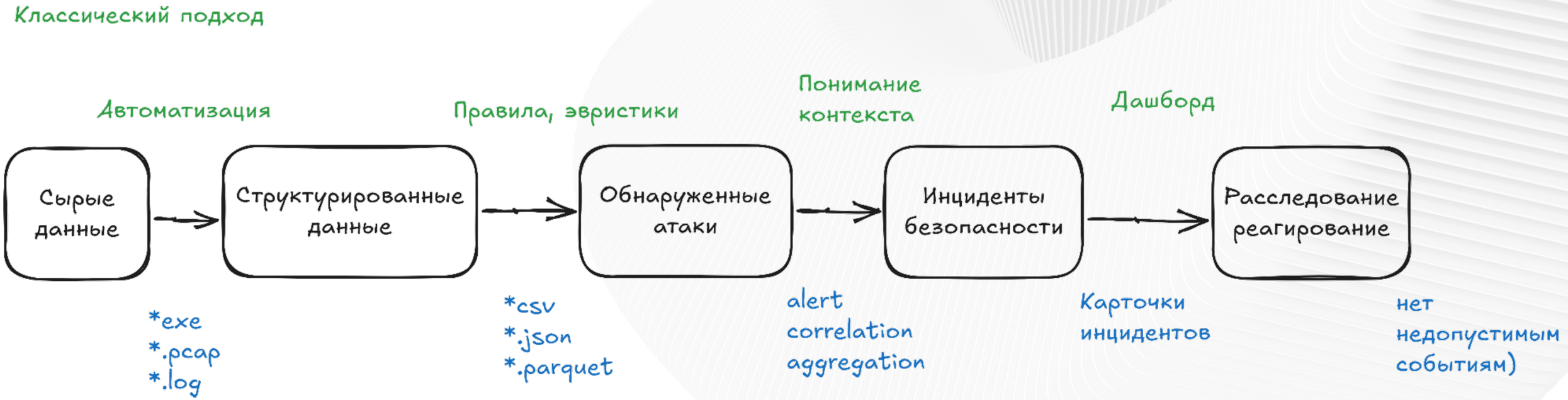
Что расскажем?

- Какая специфика ML в кибербезе?
- Какие задачи решаем?
- С какими проблемами сталкиваемся?
- Как ML решения могут попасть в прод?

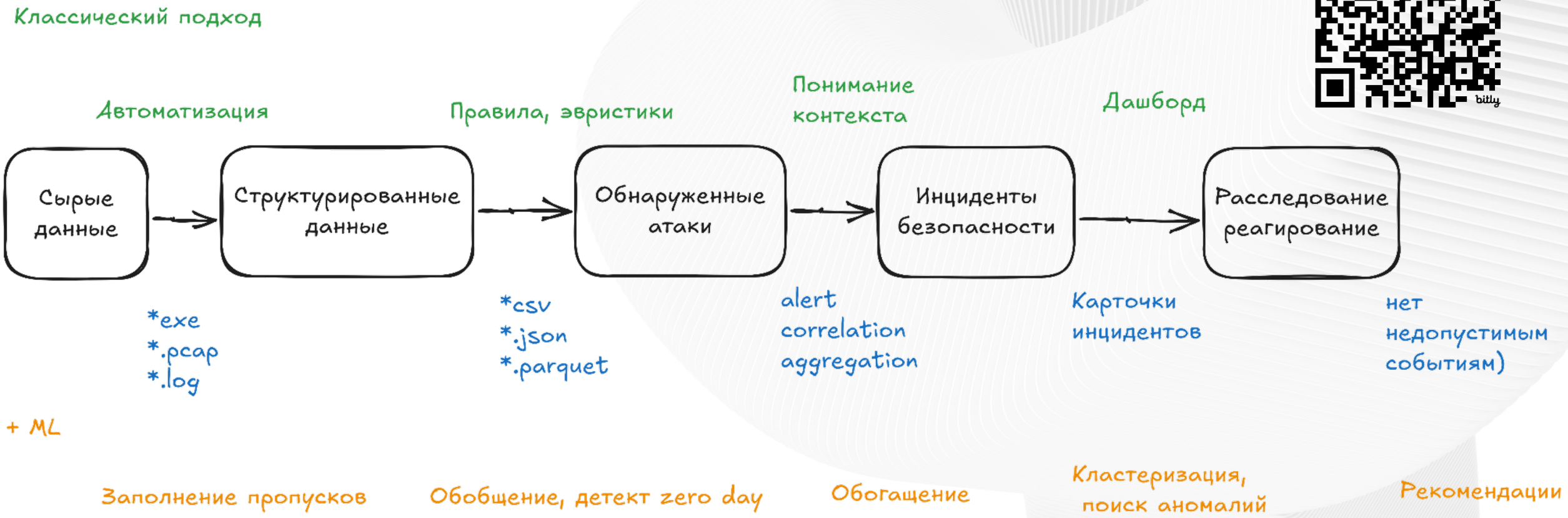
Чуть-чуть про ИБ. Концепт продуктов

Чуть-чуть про ИБ. Концепт продуктов

Чуть-чуть про ИБ. Концепт продуктов



Чуть-чуть про ИБ. Концепт продуктов



Какие задачи решаем?

- Детектирование атак
  - Вредоносное ПО (статика, поведение, сетевое взаимодействие)
  - Детектирование шелов
  - Обфускации скриптов
  - Цепочки процессов
  - Новые протоколы, приложения и малвари в шифрованном трафике
  - И ~~просто~~ хакера в инфраструктуре находим
- Помощь оператору
  - Text to query language
  - Интерпретация сработок
- Помощь экспертам
  - Трендовые уязвимости (какой софт будут ломать завтра?)
  - NER/NEL на OSINT данных
  - Группировка, поиск похожих инцидентов

С какими сложностями сталкиваемся?

- Разметки практически нет
  - Решаем экспертной разметкой, сохраненными дампами атак, LLM
- Наша разметка далека от генеральной совокупности
- Поставляем МЛ решения на стенды клиентам
- Отсутствие доступа к данным клиентов

Rule-based vs ML подходы к данным

Rule Based

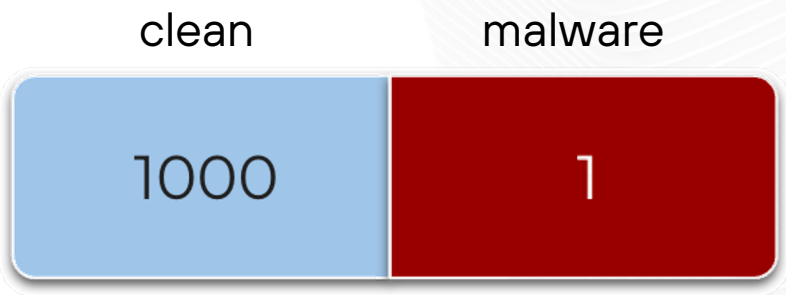
- Рассматриваем только опасные сэмплы
- Каждое семейство рассматривается независимо в рамках правила
- Вердикт по одному правилу не влияет на другие
- Не экстраполируются на новые данные

Наши данные

ML based:

- Обобщаем все в рамках одной модели
- Учим модель отличать чистые сэмплы от вредоносных
- Стараемся приблизить решения модели к генеральной совокупности
- Нужна разметка (supervised)

Real life



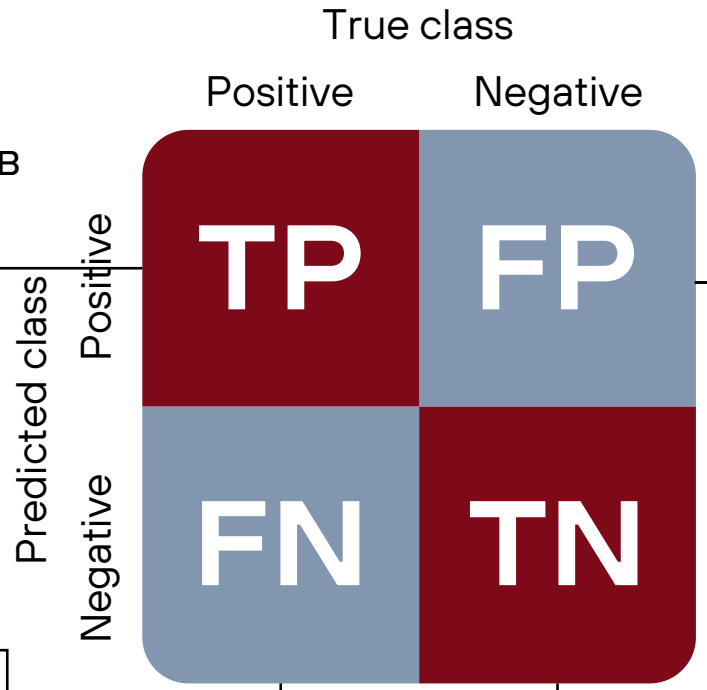
Разметка? Не, не слышал

Разметки практически нет

Решаем экспертной разметкой, сохраненными дампами атак, LLM

Как решаем?

Наша разметка:
- Вердикты правил
- Ручная разметка экспертов

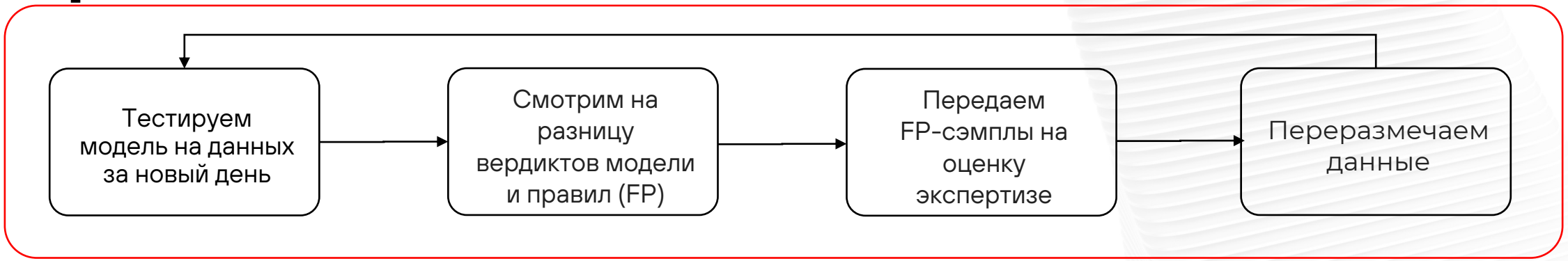


Уникальный детект:
- Value модели

Ложнопозитивная сработка:
- Очень плохо =(

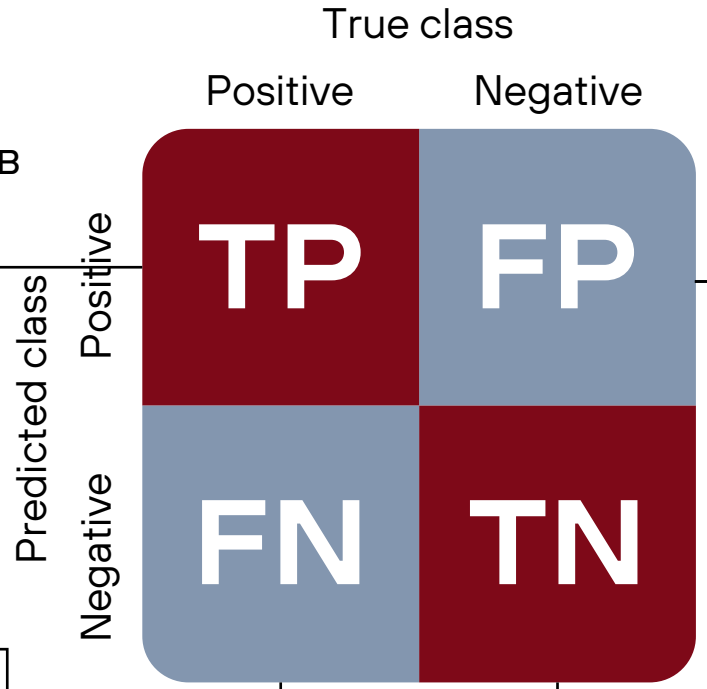
Покрывается правилами и не влияет на опыт клиента

Как решаем?



Наша разметка:

- Вердикты правил
- Ручная разметка экспертов



Уникальный детект:

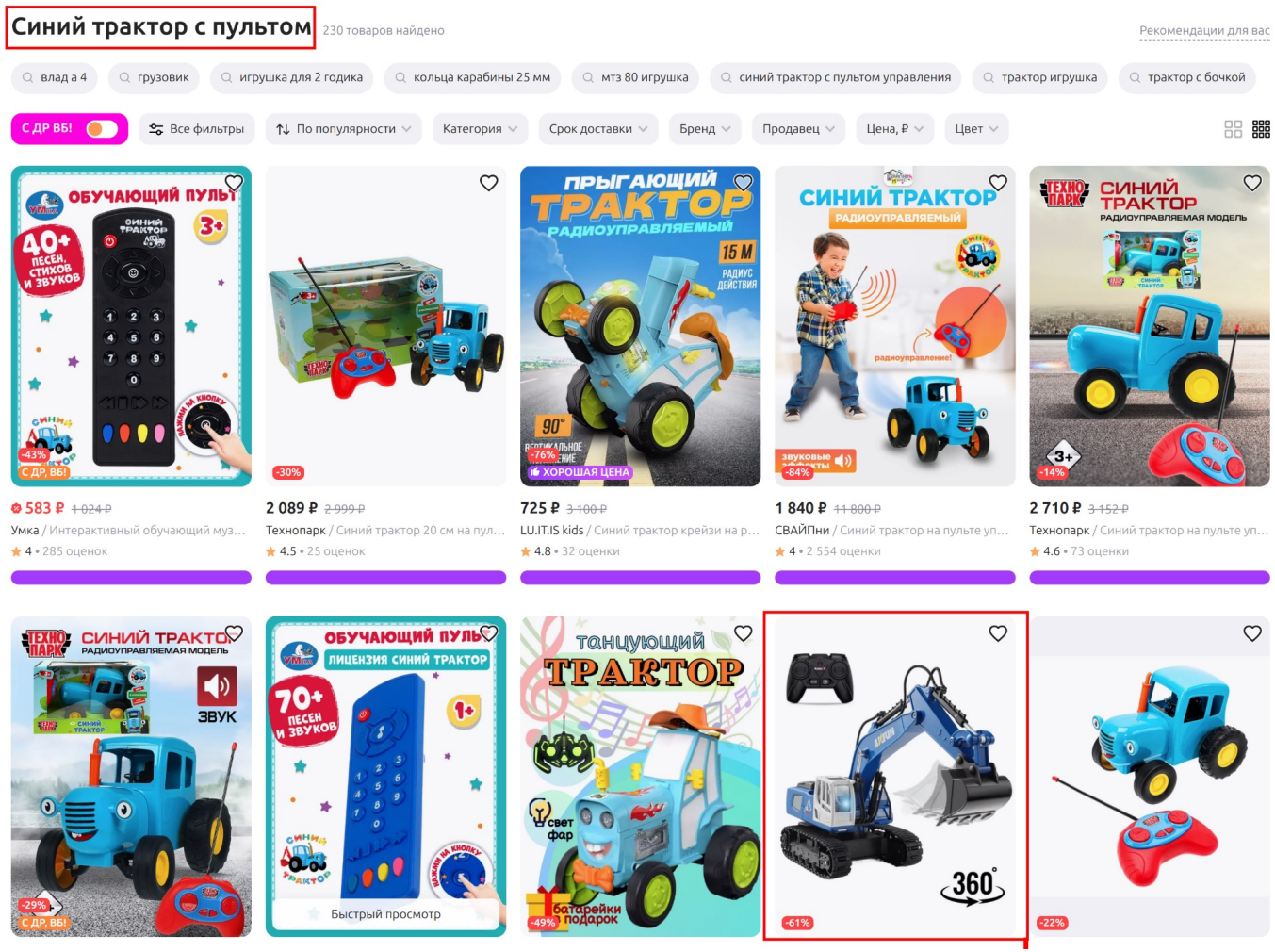
- Value модели

Ложнопозитивная сработка:

- Очень плохо =(

Покрывается правилами и не влияет на опыт клиента

Цена ошибки


Синий трактор с пультом
230 товаров найдено
Рекомендации для вас
влад а 4
грузовик
игрушка для 2 годика
кольца карабины 25 мм
мтз 80 игрушка
синий трактор с пультом управления
трактор игрушка
трактор с бочкой
С ДР ВБ!
Все фильтры
По популярности
Категория
Срок доставки
Бренд
Продавец
Цена, ₽
Цвет
ОБУЧАЮЩИЙ ПУЛЬТ
40+ ПЕСЕН, СТИХОВ И ЗВУКОВ
3+
-43%
С ДР, ВБ!
583 ₽ 1 024 ₽
Умка / Интерактивный обучающий муз…
4 • 285 оценок
-30%
2 089 ₽ 2 999 ₽
Технопарк / Синий трактор 20 см на пул…
4.5 • 25 оценок
ПРЫГАЮЩИЙ ТРАКТОР
РАДИОУПРАВЛЯЕМЫЙ
15 М
РАДИУС ДЕЙСТВИЯ
90°
-76%
ХОРОШАЯ ЦЕНА
725 ₽ 3 100 ₽
LU.IT.IS kids / Синий трактор крейзи на р…
4.8 • 32 оценки
СИНИЙ ТРАКТОР
РАДИОУПРАВЛЯЕМЫЙ
радиоуправление!
-84%
1 840 ₽ 11 800 ₽
СВАЙПни / Синий трактор на пульте уп…
4 • 2 554 оценки
ТЕХНОПАРК
СИНИЙ ТРАКТОР
РАДИОУПРАВЛЯЕМАЯ МОДЕЛЬ
3+
-14%
2 710 ₽ 3 152 ₽
Технопарк / Синий трактор на пульте уп…
4.6 • 73 оценки
ТЕХНОПАРК
СИНИЙ ТРАКТОР
РАДИОУПРАВЛЯЕМАЯ МОДЕЛЬ
ЗВУК
-29%
С ДР, ВБ!
ОБУЧАЮЩИЙ ПУЛЬТ
ЛИЦЕНЗИЯ СИНИЙ ТРАКТОР
70+ ПЕСЕН И ЗВУКОВ
1+
Быстрый просмотр
ТАНЦУЮЩИЙ ТРАКТОР
свет фар
батарейки подарок
-49%
360°
-61%
-22%


feels good man

Цена ошибки (как не надо делать)

```
alert.msg == "[ML-TEST] Looks like ktalk.exe" and dst.dns == "ptsecurity.ktalk.ru"
```

484 сессии · 74 узла · 0 атак · 0 индикаторов компрометации

```
alert.msg == "[ML-TEST] Looks like ktalk.exe" and dst.dns != "ptsecurity.ktalk.ru"
```

11 230 сессий · 785 узлов · 0 атак · 3 индикатора компрометации

Цена ошибки (как допустимо делать)

```
app_service=="Telegram" && !dst.geo.org == "Telegram Messenger Inc" and app_proto=="tls" && flags == "APP_SERVICE_ML"
```

37 сессий · 15 узлов · 0 атак · 0 индикаторов компрометации

```
app_service=="Telegram" && dst.geo.org == "Telegram Messenger Inc" and app_proto=="tls" && flags == "APP_SERVICE_ML"
```

21 001 сессия · 241 узел · 0 атак · 0 индикаторов компрометации

Эффективный анализ трафика, или Как ничего не упустить

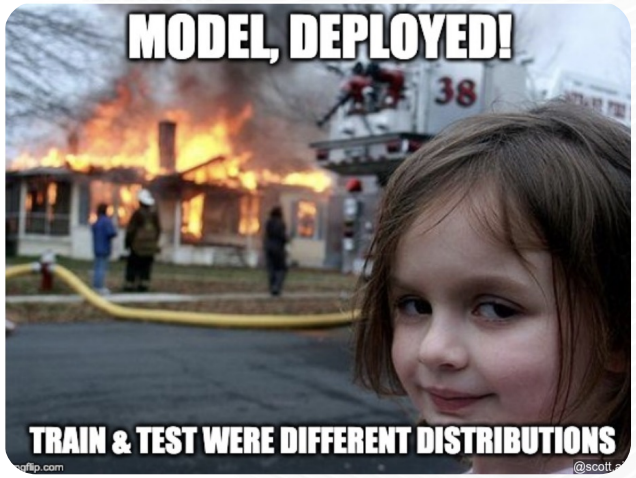
Как ML может попасть в прод?

- **Модель как сервис**
- **Онпрем**
  - Сериализованный артефакт (pickle, onnx, json, joblib, bentoML)
  - Препроцессинг в ядре продукта
  - Сервис в контуре продукта, взаимодействующий с брокером сообщений

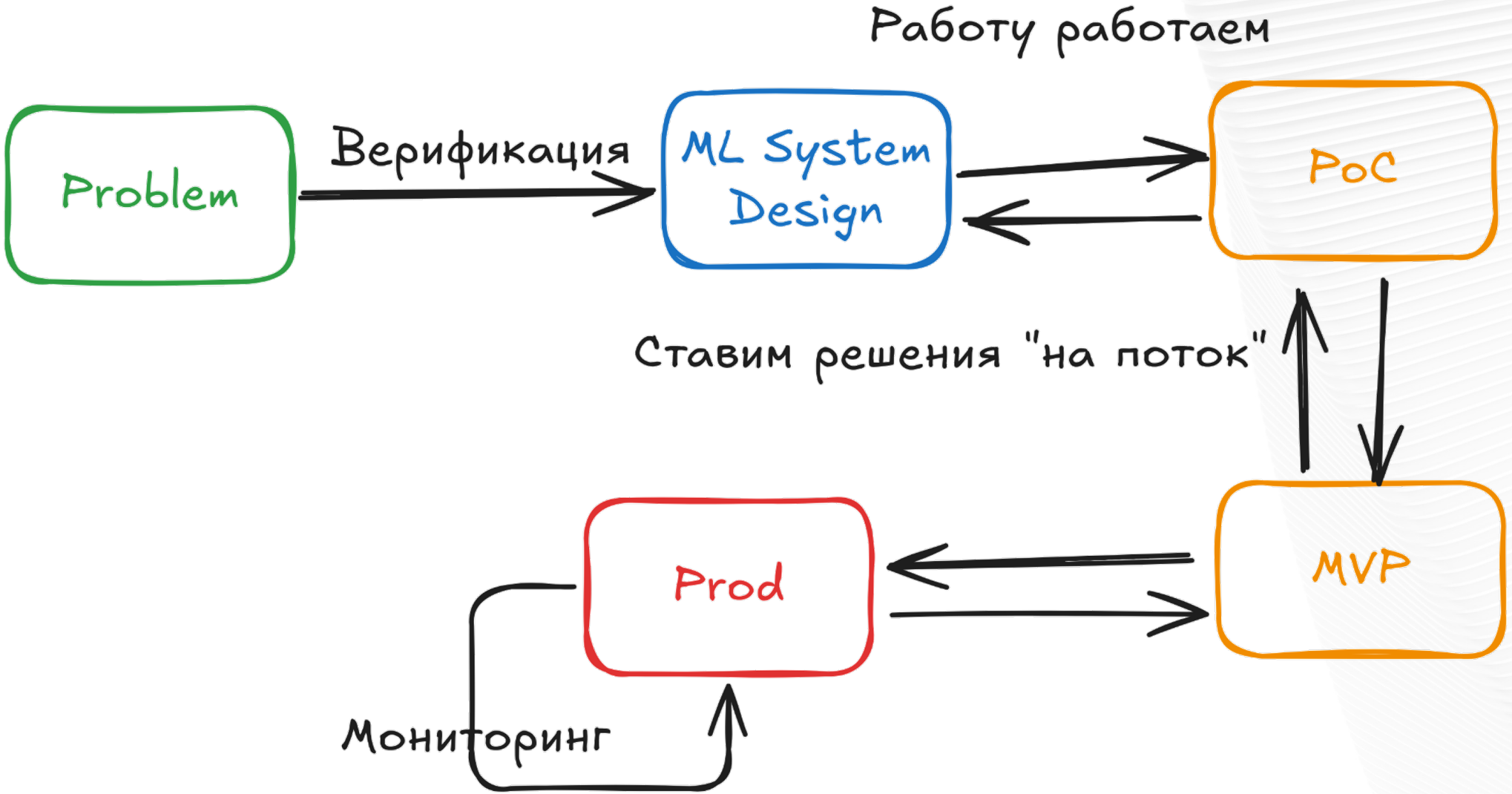
Нормально делай, нормально будет!

- **Технические требования**
  - <0.1 мс на ответ от ML
  - Инференс в onprem на CPU
  - Не хотим увеличивать “размер” железки

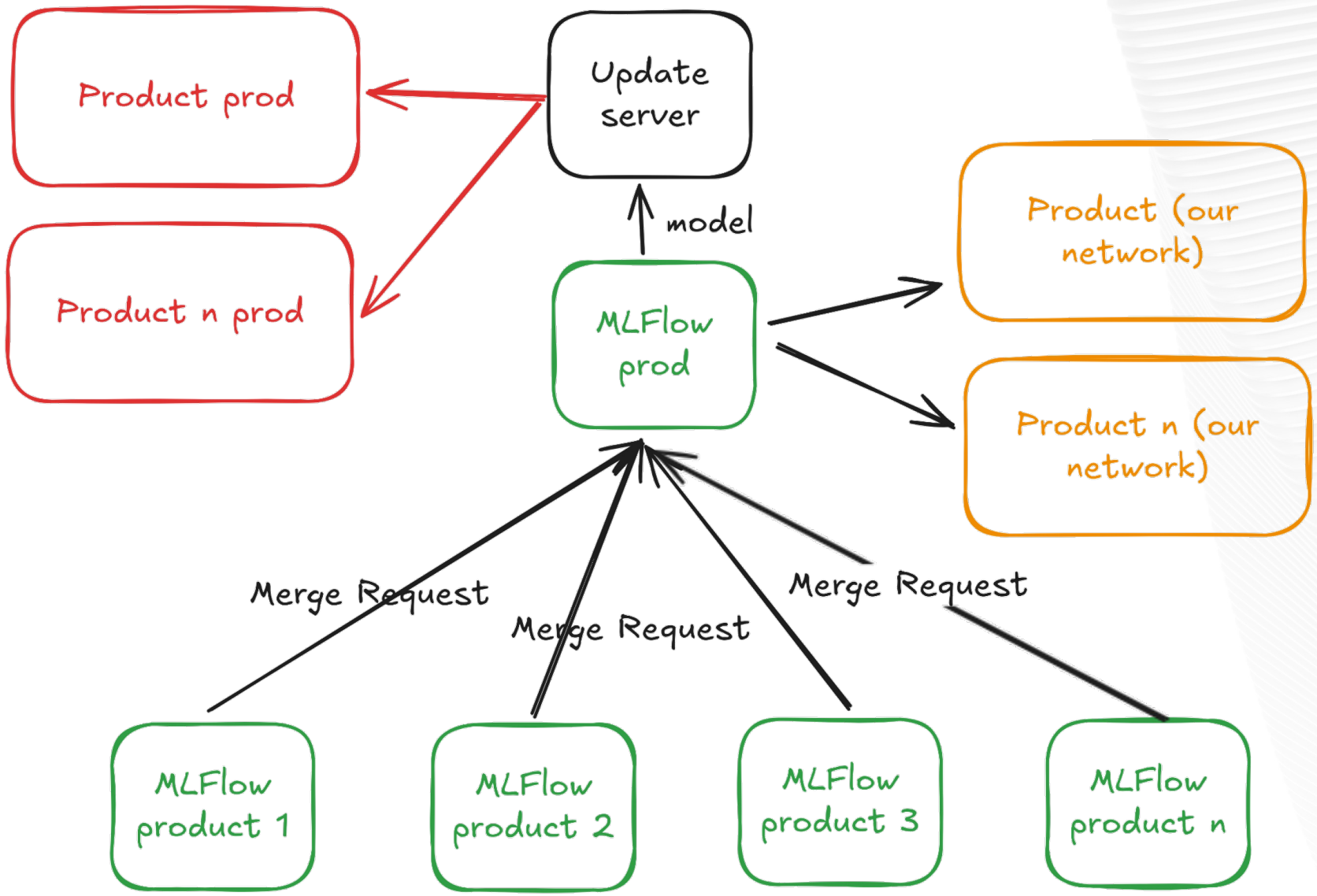
- Очень хотим делать AB, пока строим ML телеметрию



Как попадем в прод издалека?



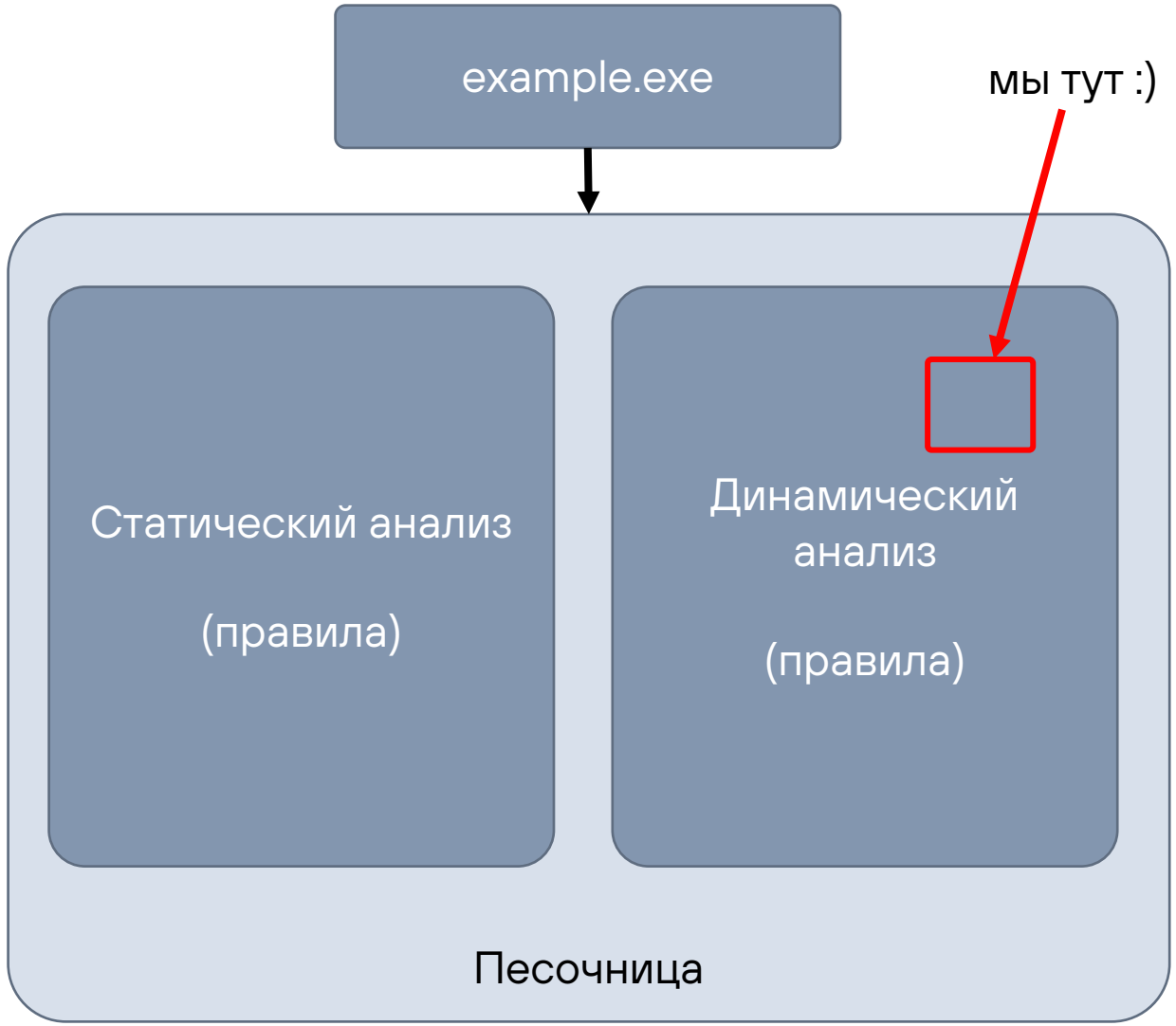
PoC->MVP. Зачем нам много инстансов mlflow?



MVP->Prod.
Как плохую модель не отправить клиенту?

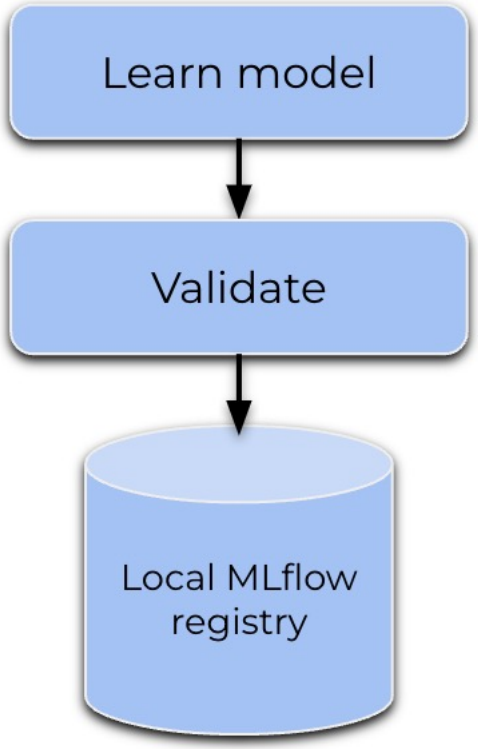
- **Поток с внутренней сети** (наблюдаем данные полностью)
  - Онлайн доразметка экспертами и выявление уникальных детектов
  - Интеграция с экспертными системами
  - Переобучение модели каждую неделю
- **Телеметрия с клиентов** (не видим данные)
  - Расхождение соотношений детектов модели и экспертных эвристик (скриптов)
  - SLA 24 часа на фикс модели в случае значимого расхождения
- **Клиенты с фолзами** (надо валидировать и исправлять)
  - Обращения от клиентов с жалобой на некорректный вердикт модели
  - SLA 24 часа на фикс модели

Пример с PT Sandbox



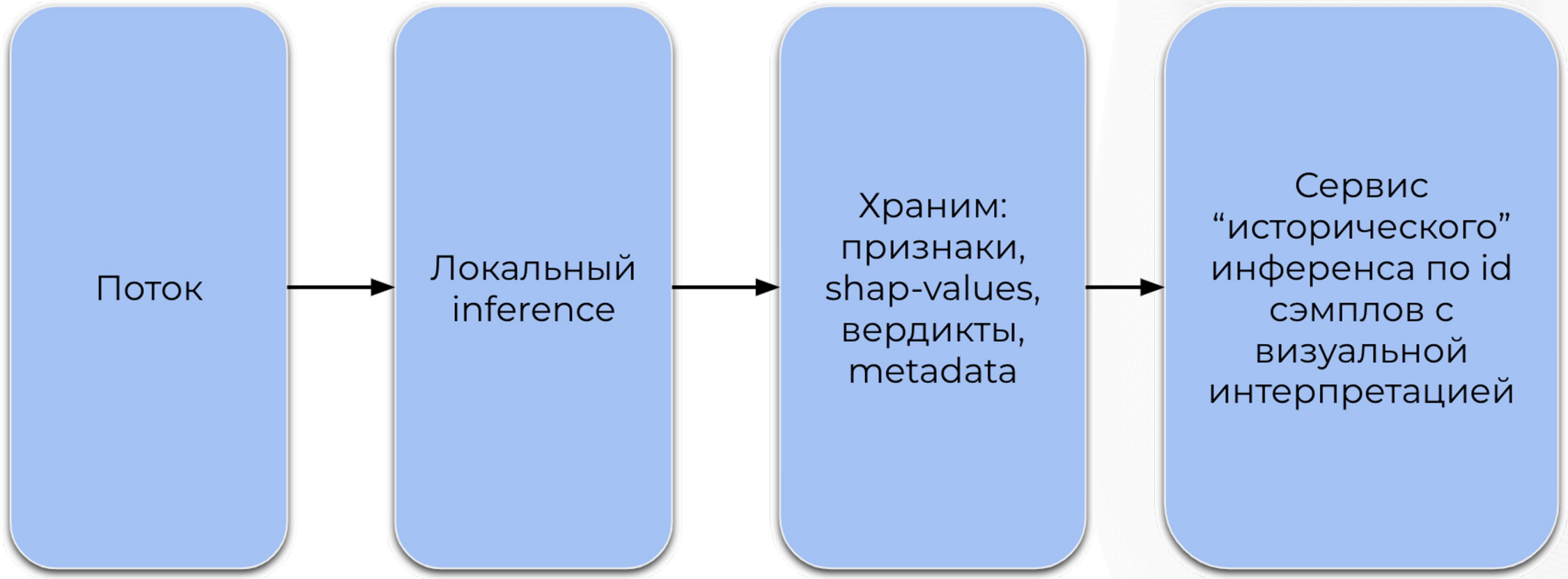
Positive Security Days 2024
Про модель анализа http-сессий

Переход к внедрению. Эталонная выборка

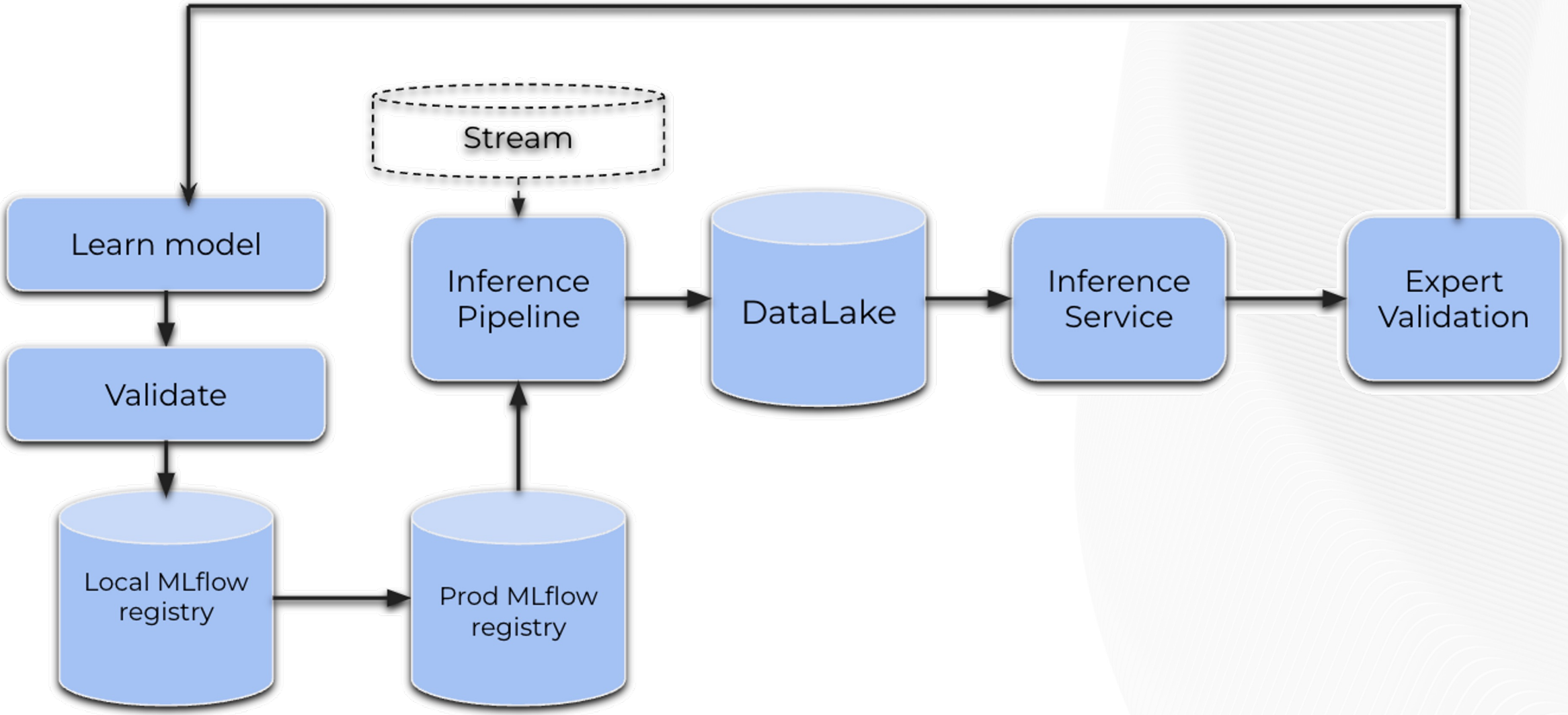


- Выборка размеченная экспертами
- Требуется 100%-точность модели
- Перед каждым обновлением модель валидируется

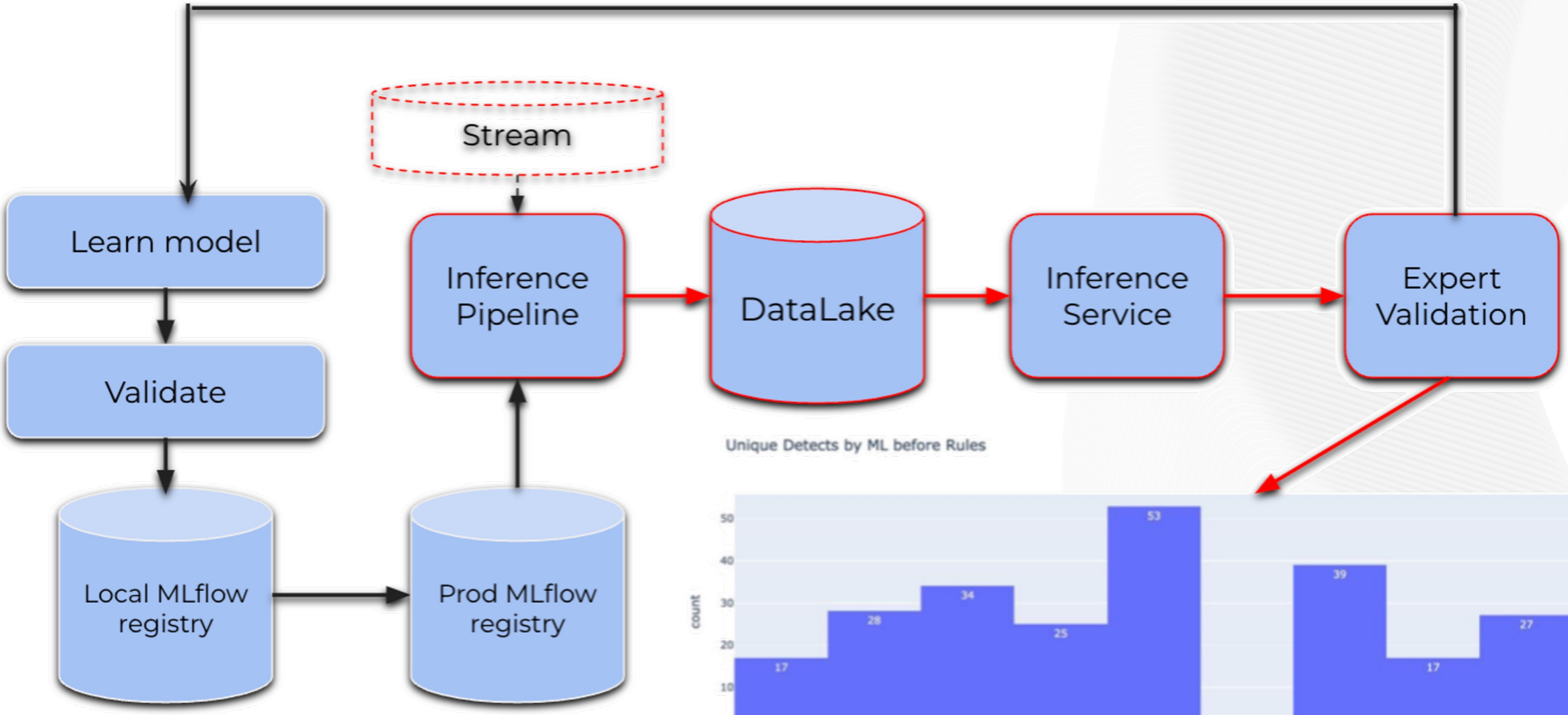
Переход к внедрению. Интерпретируемость



Переход к внедрению. Общая картина

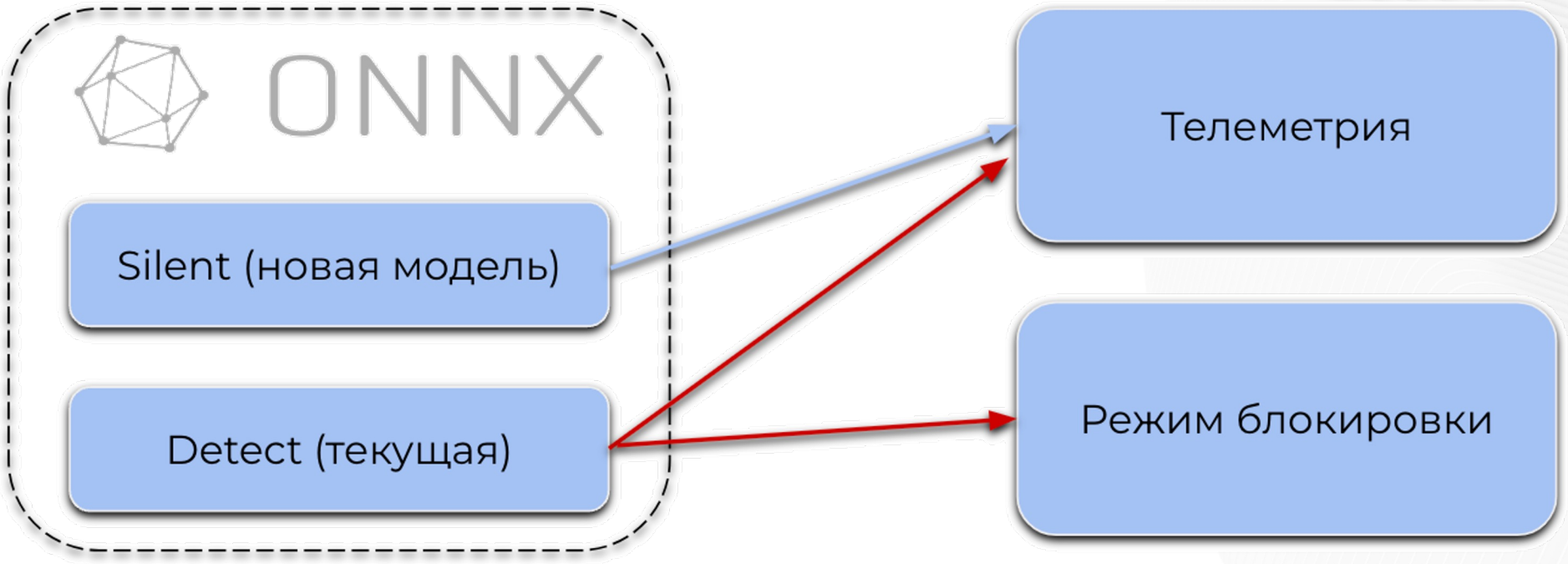


Переход к внедрению. Общая картина

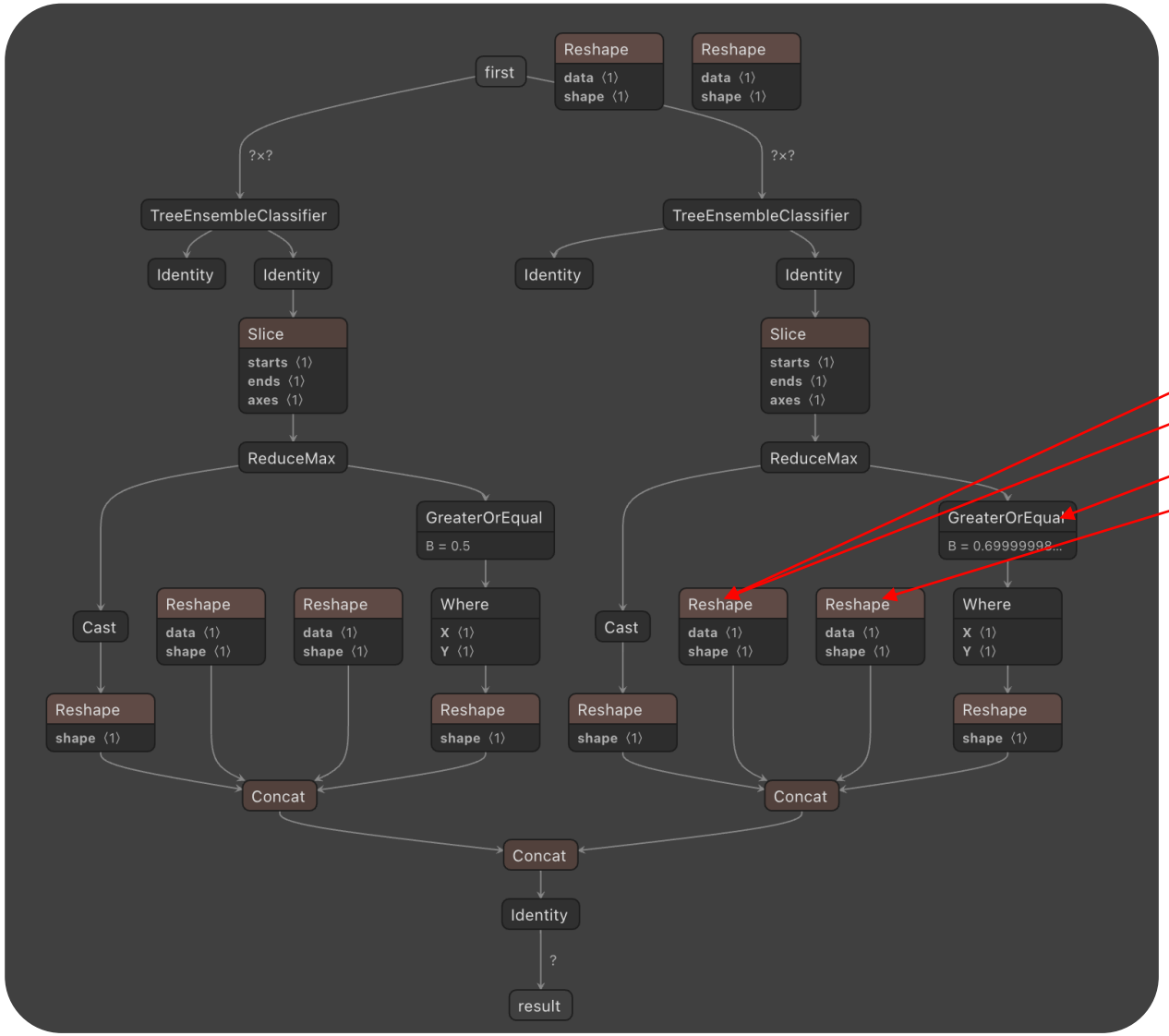


Переход к внедрению. Поставка обновления

- Нужна уверенность в качестве новой модели на данных клиентов
- Реализуем пост-процессинг на стороне ONNX (пороги, агрегации и тд)
- Побочно: хотим ускорить time-to-market



Переход к внедрению. Поставка обновления



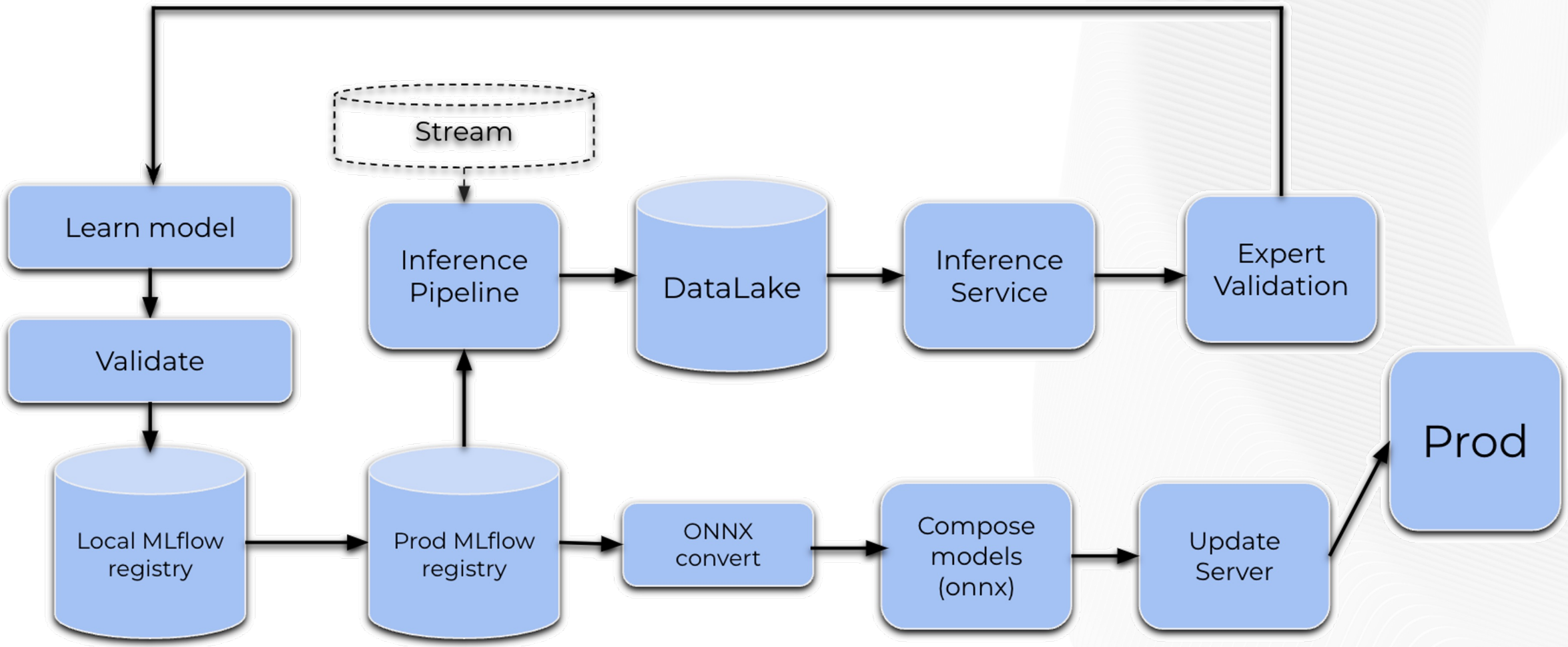
build_config.yml

```yaml
models:
- registry_name: "Model_LGBM"
  version: 1
  threshold: 0.7
  is_silent: "true"

- registry_name: "Model_LGBM"
  version: 2
  threshold: 0.5
  is_silent: "false"
```

Визуализация: https://netron.app/

Внедрение. Общая картина



А что в итоге?

- В ИБ-специфике цена ошибки модели велика
- Важно выстраивать процессы непрервывной экспертной разметки и валидации
- Онпрем поставки не должны быть привязаны к релизам продукта
- ONNX позволяет управлять пост-процессингом и делать AB-тесты
- Правильная работа с телеметрией позволяет оценивать качество моделей без доступа к данным

Спасибо и до новых встреч!

Про ML команду в Positive Technologies

AI & Security Community