



CLIP + LLM в проде:
мультимодальный «Поиск
по фото» для маркетплейса



Романов Никита

Tech Lead продуктов «Поиск по фото»
и «Похожие по фото»

6+ лет опыта в CV

План выступления

Описание и архитектура сервиса

Векторный индекс

Модель распознавания

Текстовые Теги

Уточнение текстом

Результаты

Описание и архитектура сервиса

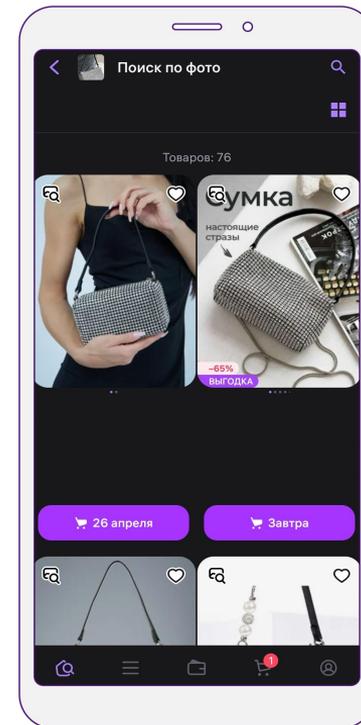
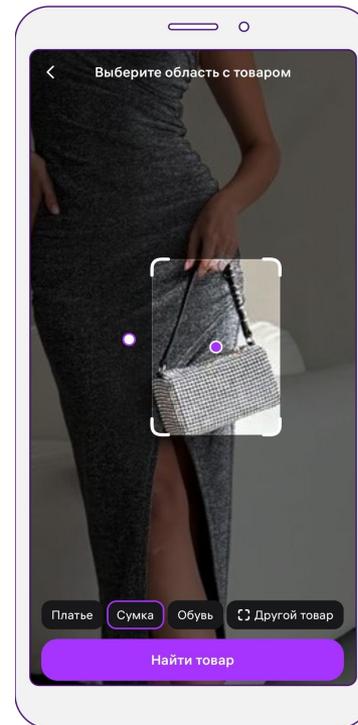
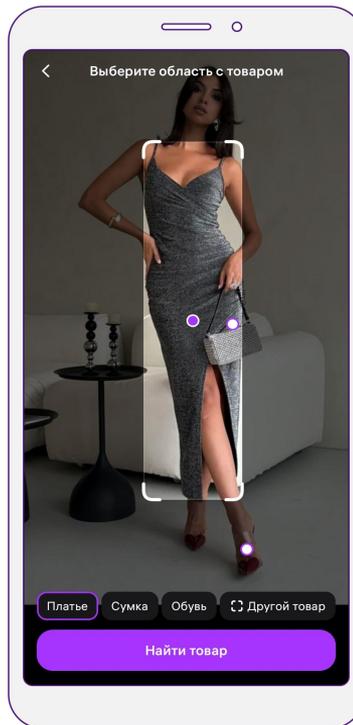
Описание сервиса



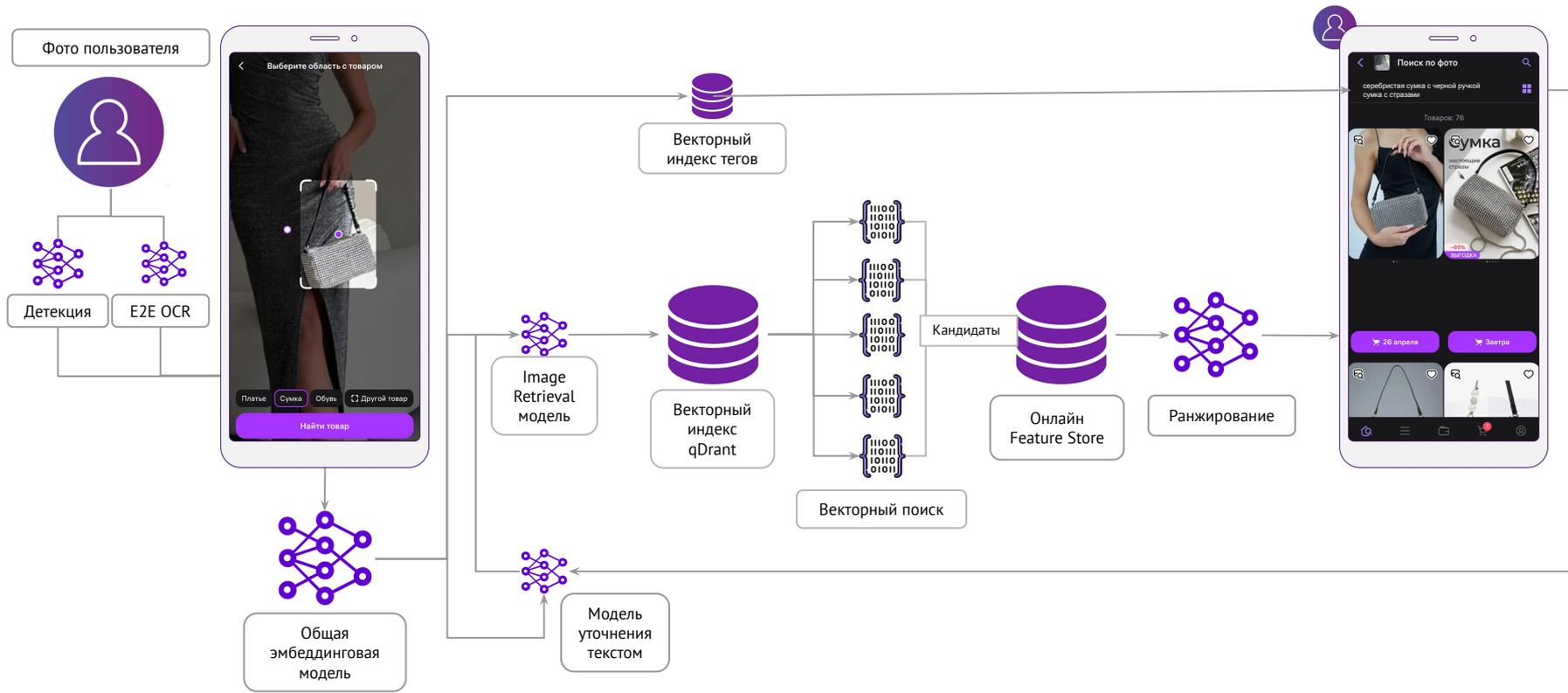
Поиск по фото — сервис, который позволяет пользователям находить товары, основываясь на изображении. Продукт делает процесс покупок проще и быстрее

Какие преимущества?

- Специфичный запрос
- Сложно подобрать описание
- Пользователь не знает название / производителя
- Долго / неудобно вводить текст



Архитектура сервиса



Векторный индекс и нагрузка



Векторная БД



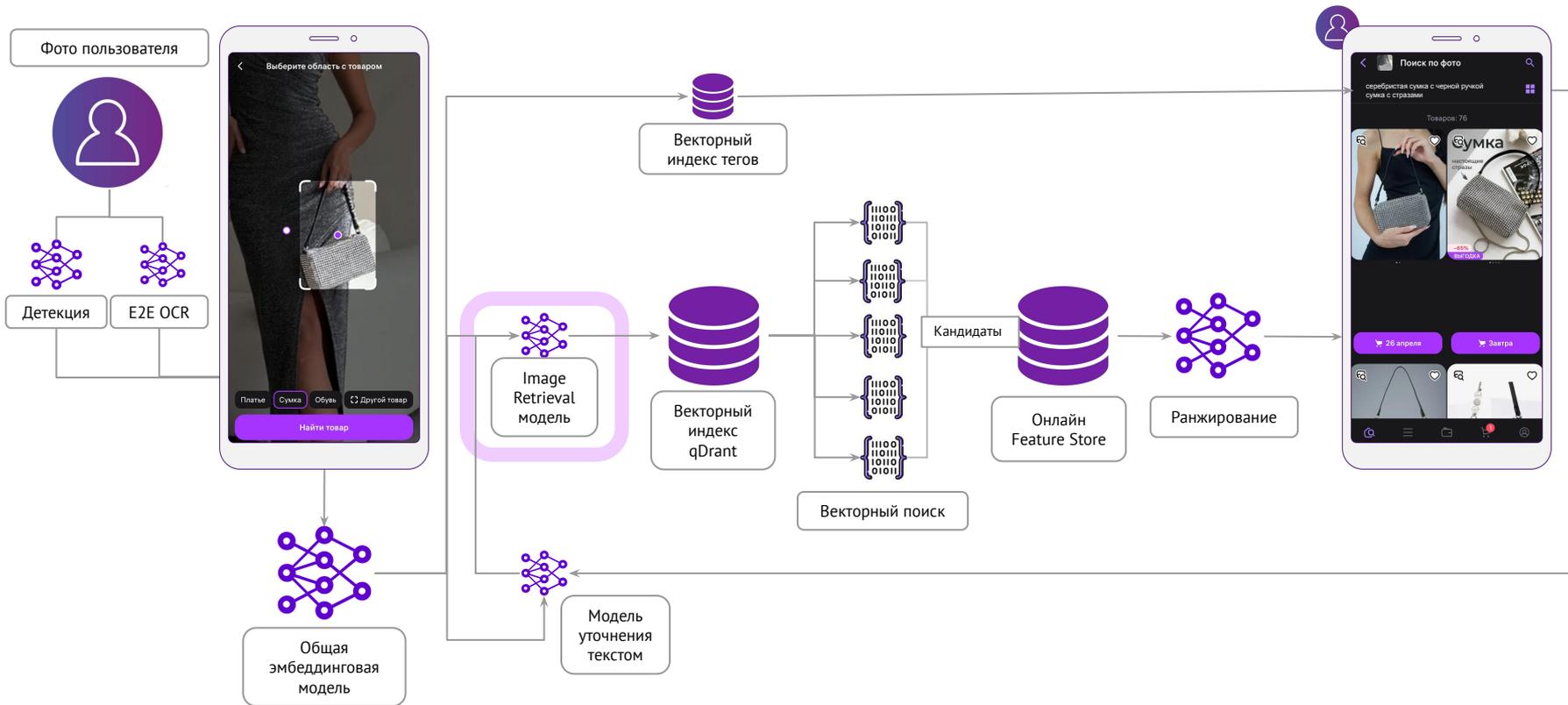
- qDrant
- Квантованные вектора 256 fp16
- *Время ответа всего сервиса 250ms*
- Количество товаров 350млн
- Обновление коллекции 5 раз в день



Модель распознавания



Архитектура сервиса



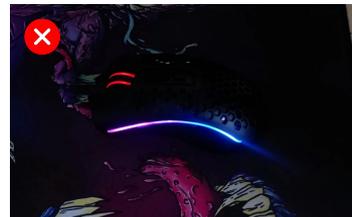
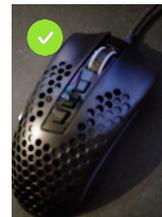
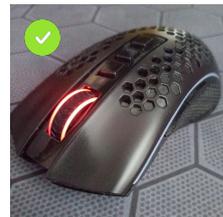
Данные



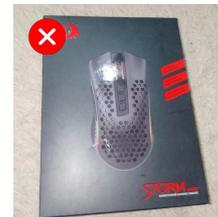
- Датасет WB 200M
- Галерейные фото КТ (Карточка Товара) & фото отзывов КТ
- Реальные фото пользователей

Предобработка

- Фильтр дубликатов — CLIP-like
- Фильтрация от шумов/выбросов — CLIP-like
- Фильтрация дефектных фото — команда разметки
- Детекция главного товара КТ



слишком темно



упаковка



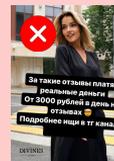
не то



не основной объект

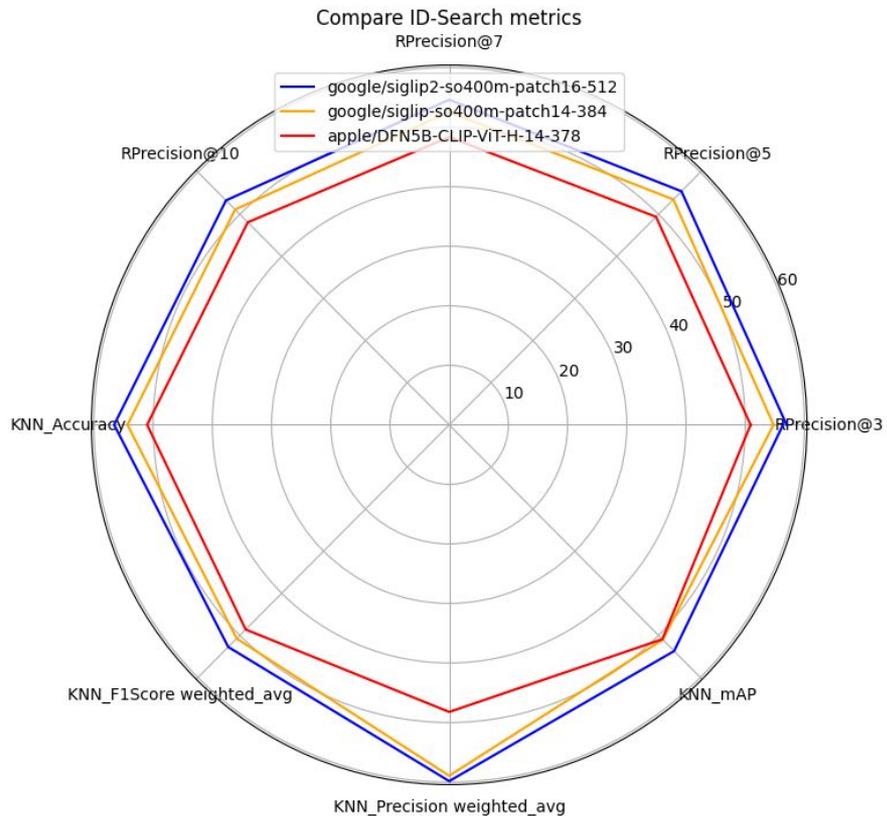


слишком маленький фрагмент



объект перекрыт спам

Метрики моделей



Основные метрики:

- RPrecision@k
- F1
- mAP

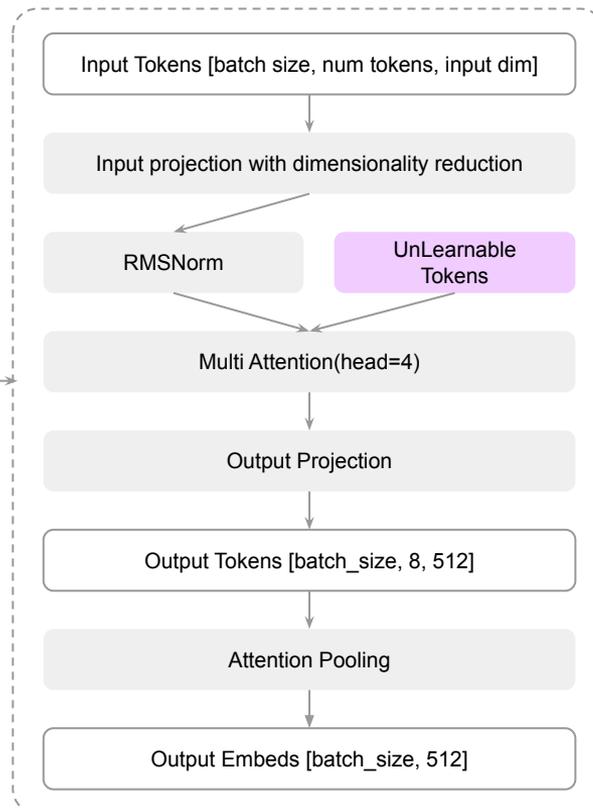
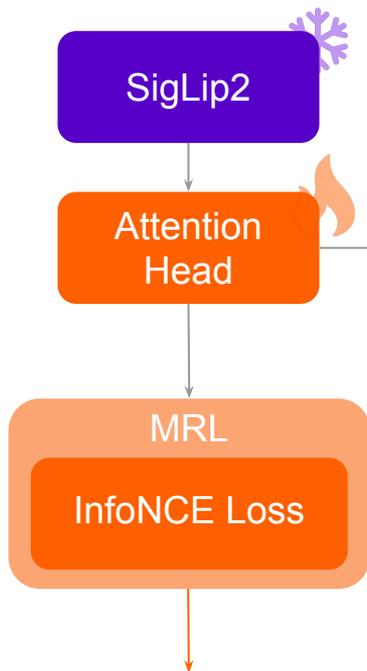
Архитектура ImageRetrieval Модели



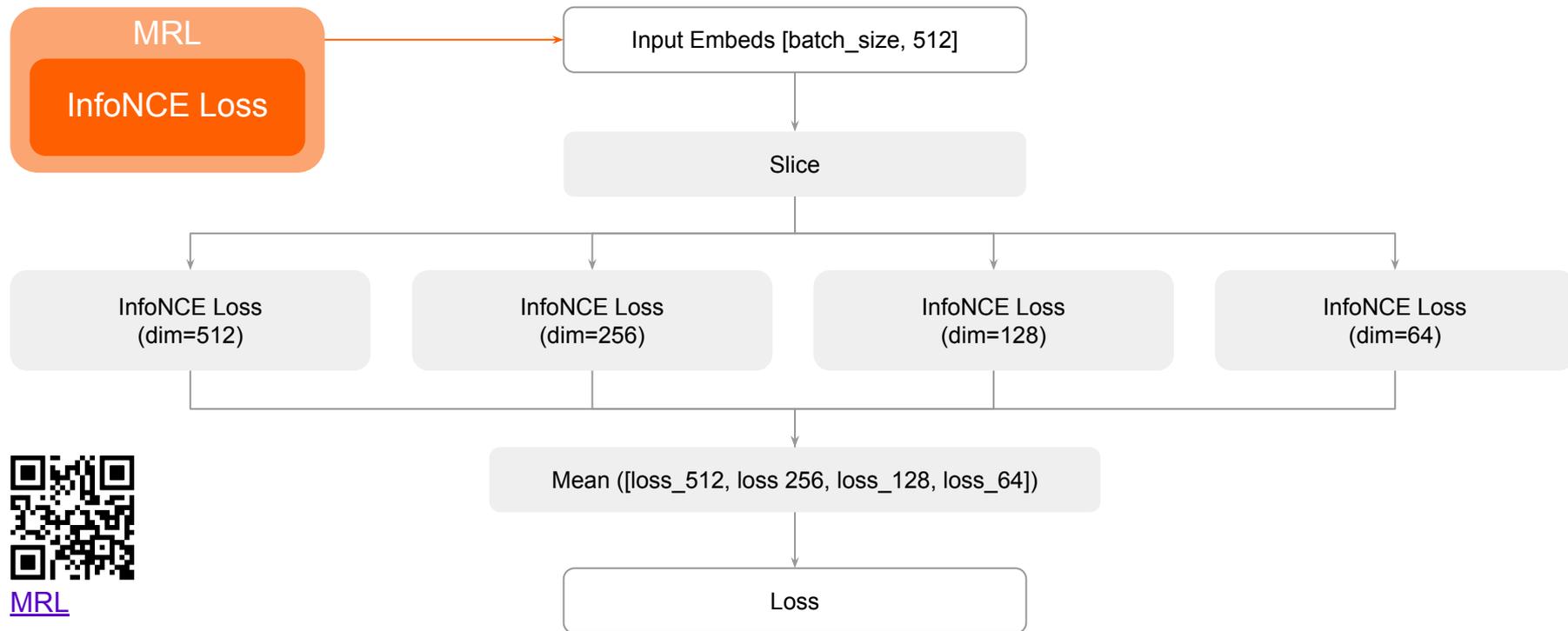
google/siglip2-so400m-patch14-384

Attention Head

MRL



MRL Архитектура обучения



[MRL](#)

Метрики обучения с MRL и без MRL

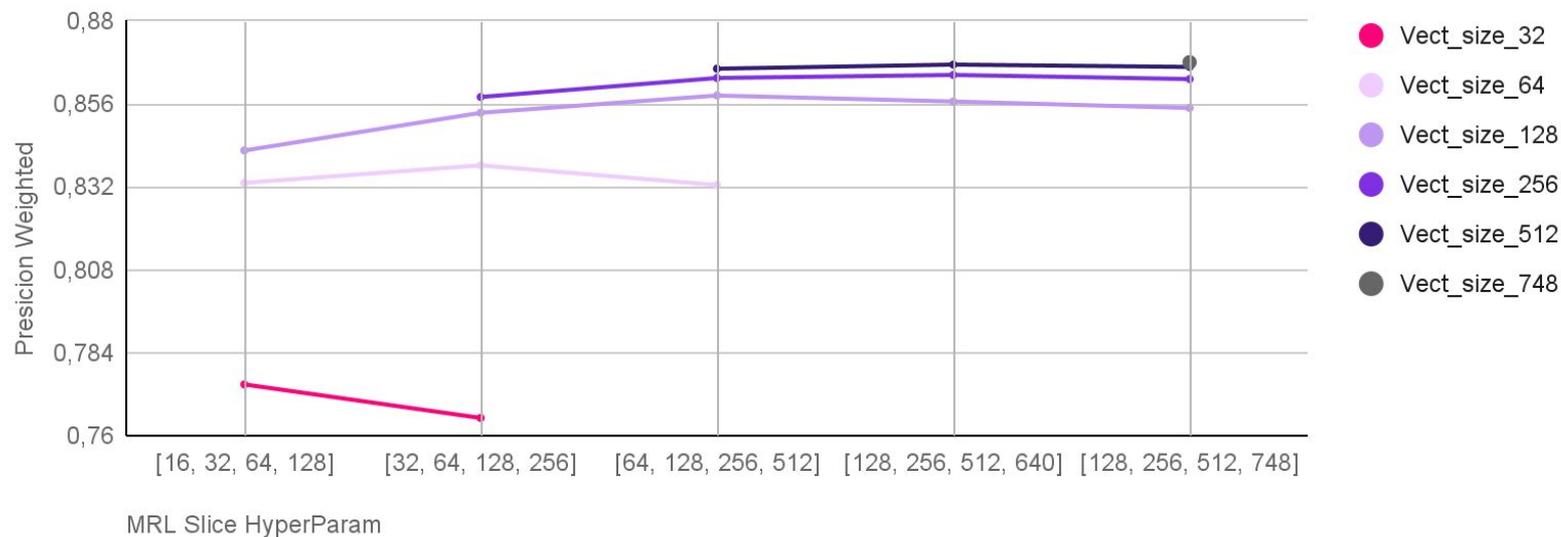


Metrics	without MRL <i>dim=1024</i>	with MRL [1024 ...128] <i>dim=1024</i>	Δ
RPrecision 3 / 5 / 10	0,7352 / 0,7087 / 0,6883	0,7598 / 0,7324 / 0,7116	3,39% / 3,34% / 3,34%
Accuracy	0,7498	0,7761	3,50%
F1Score / weighted_avg	0,7298	0,7571	3,73%
Precision / weighted_avg	0,8120	0,8385	3,26%
Recall / weighted_avg	0,7498	0,776	3,49%
mAP / 10	0,7758	0,795	2,48%

MRL Метрики



MRL Final metrics

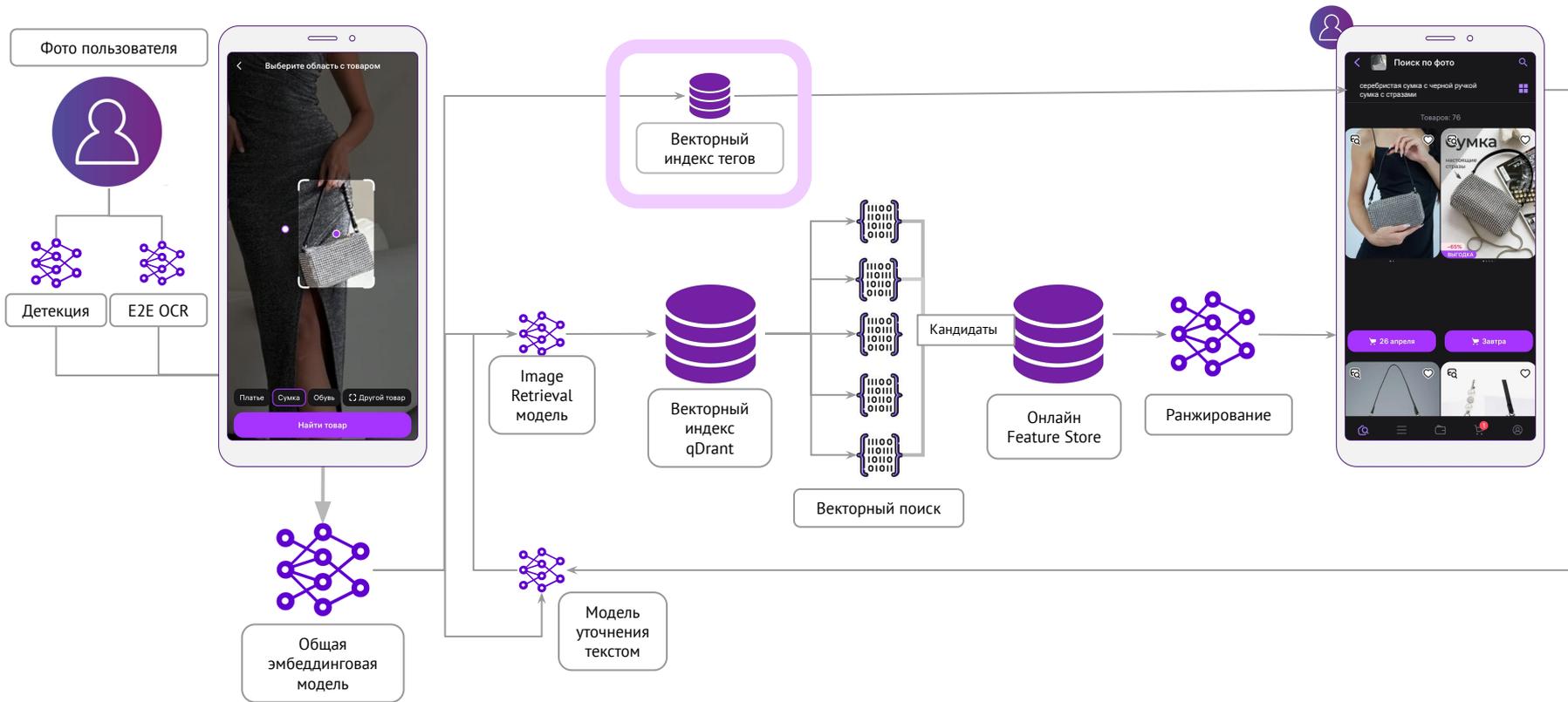


- Ускорение сходимости модели
- Сжатие векторного пространства
- Ранжирование фичей

Текстовые теги

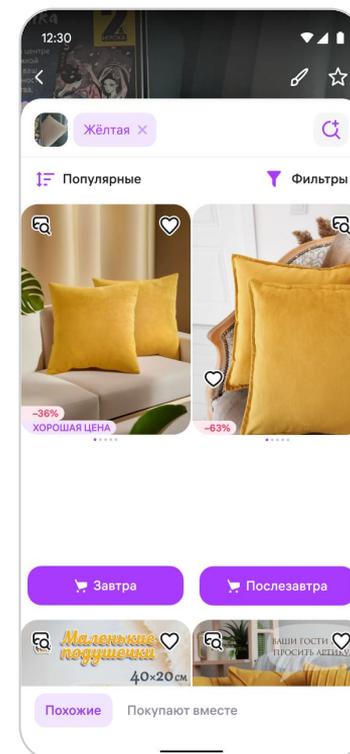
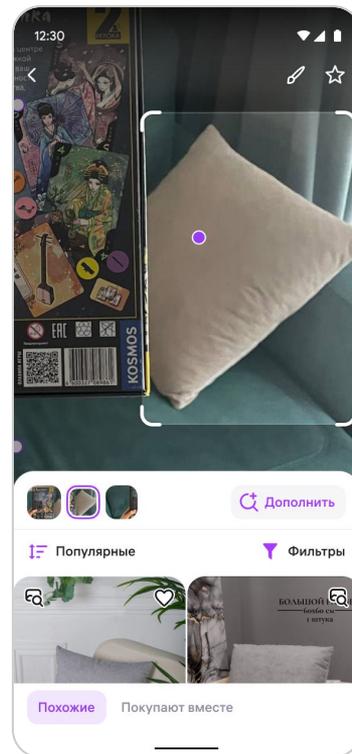


Архитектура сервиса

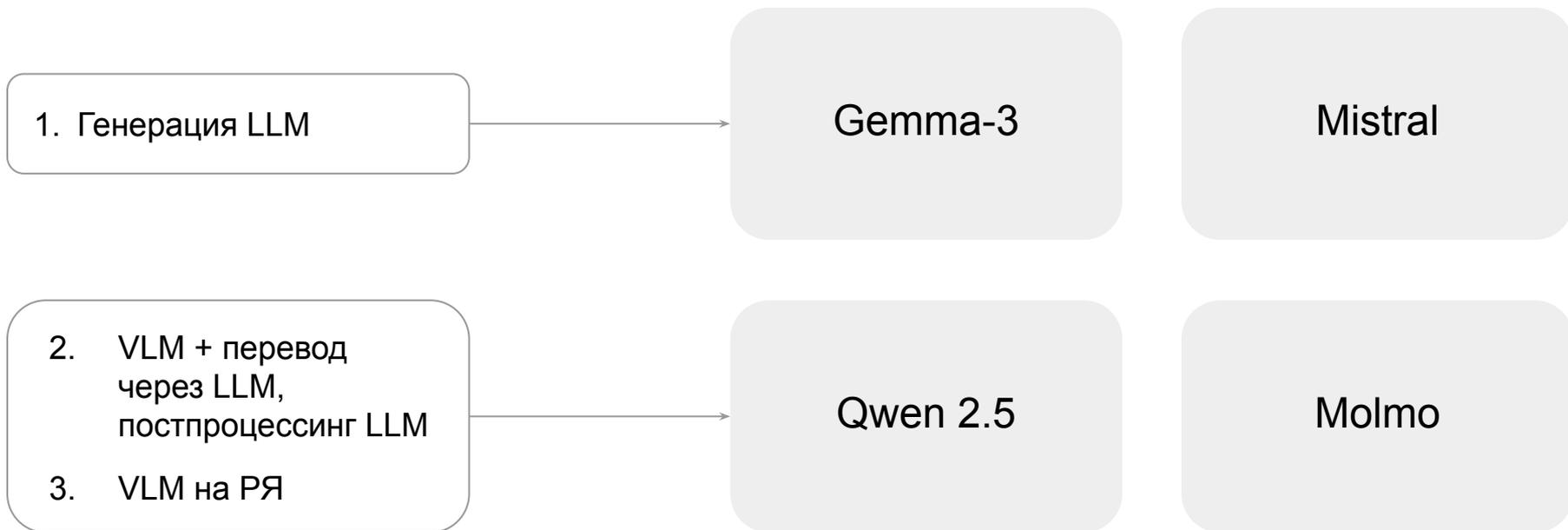


Текстовые теги

- Выделение ключевых характеристик товара
- Уточнение запроса
- Разнообразиие



Подходы в генерации тегов



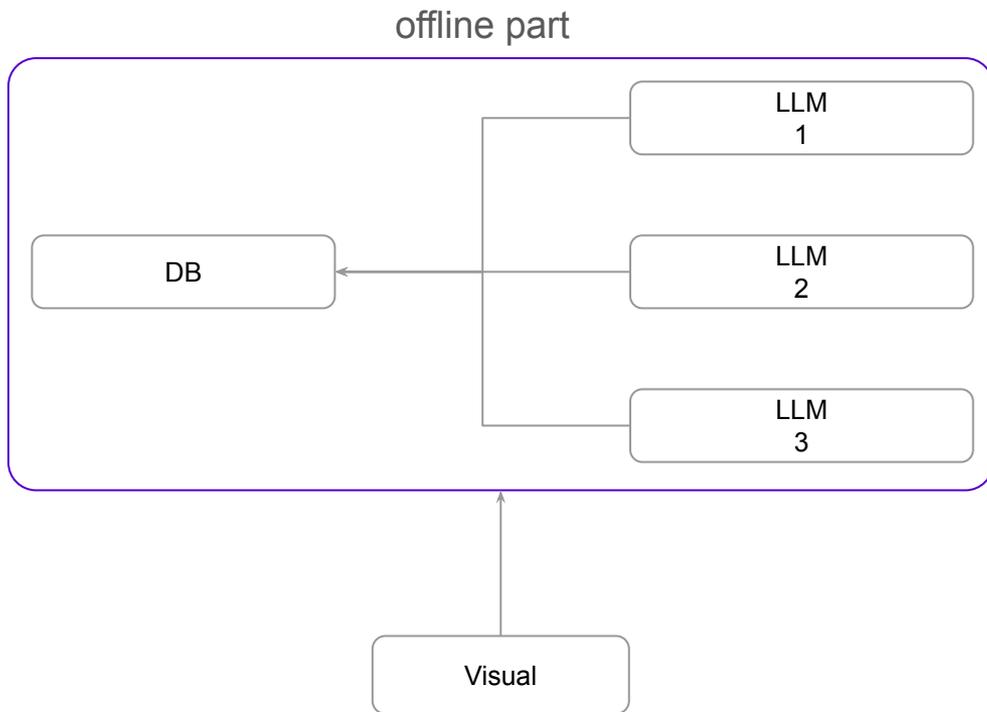
Данные



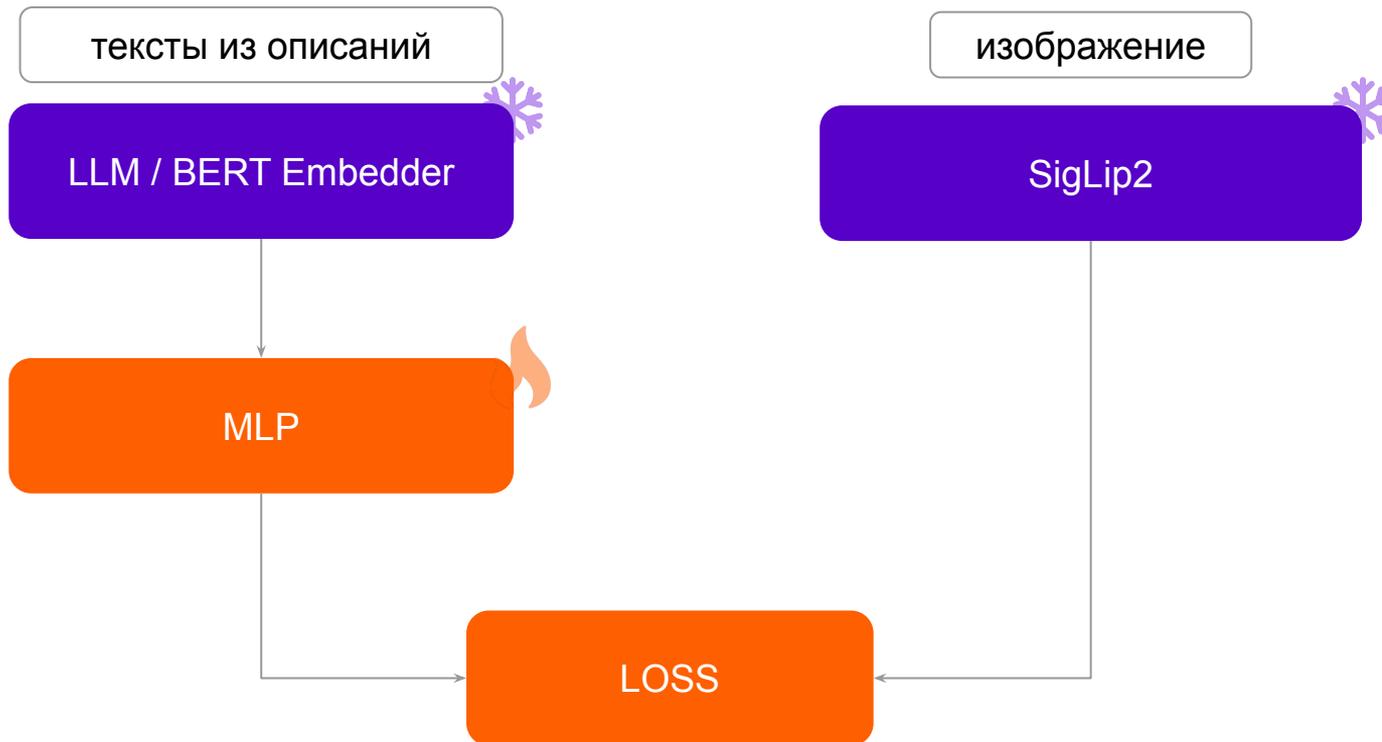
- Эмоциональное описание
- Семантическая несогласованность
- Опечатки
- Слова на другом языке

Плохой тег	Хороший тег
пуловер из веселого полотна	футболка-поло с круглым вырезом
туник с отрезными рукавами	брюки с завышенной талией
полосатая t-shirt	сарафан с открытыми плечами
кардиган с приближенными к шее	футболка с рисунком рыбки
велосипедки с тремя точками длины	водоотталкивающая шапка

Асимметричный сетап



Архитектура обучения модели

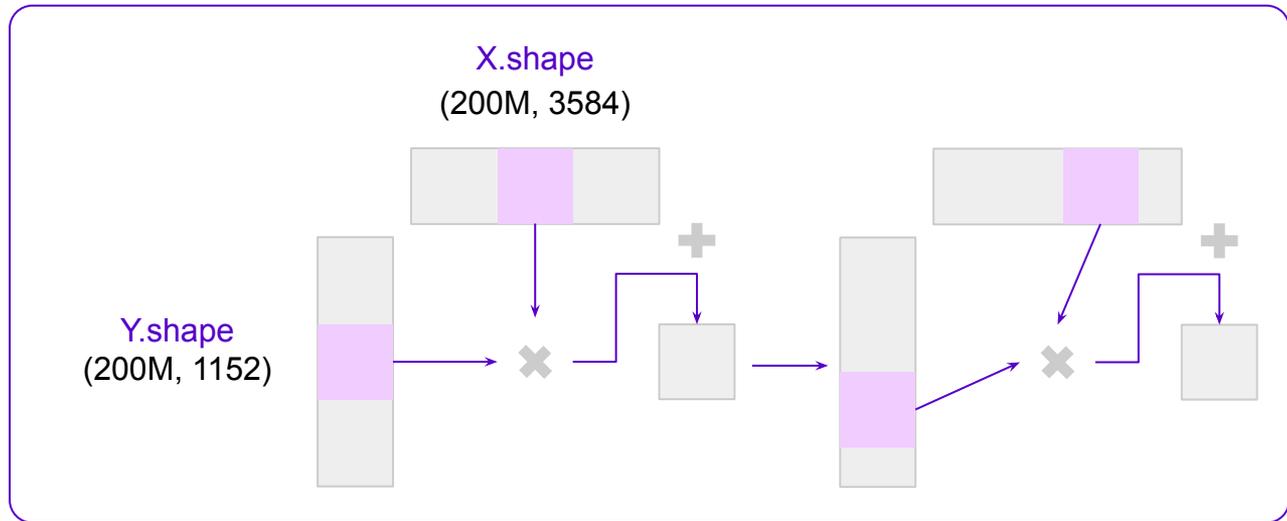


Линейные модели: как быстро тренировать

Closed-form решение,
неплохой бейзлайн

$$W^* = (X^T X)^{-1} X^T Y$$

Можно накопить $X.T@X$
и $X.T@Y$ батчи, считать
решение перемножением



Обучение MLP поверх линейных моделей

Ускоряет сходимость модели 4x

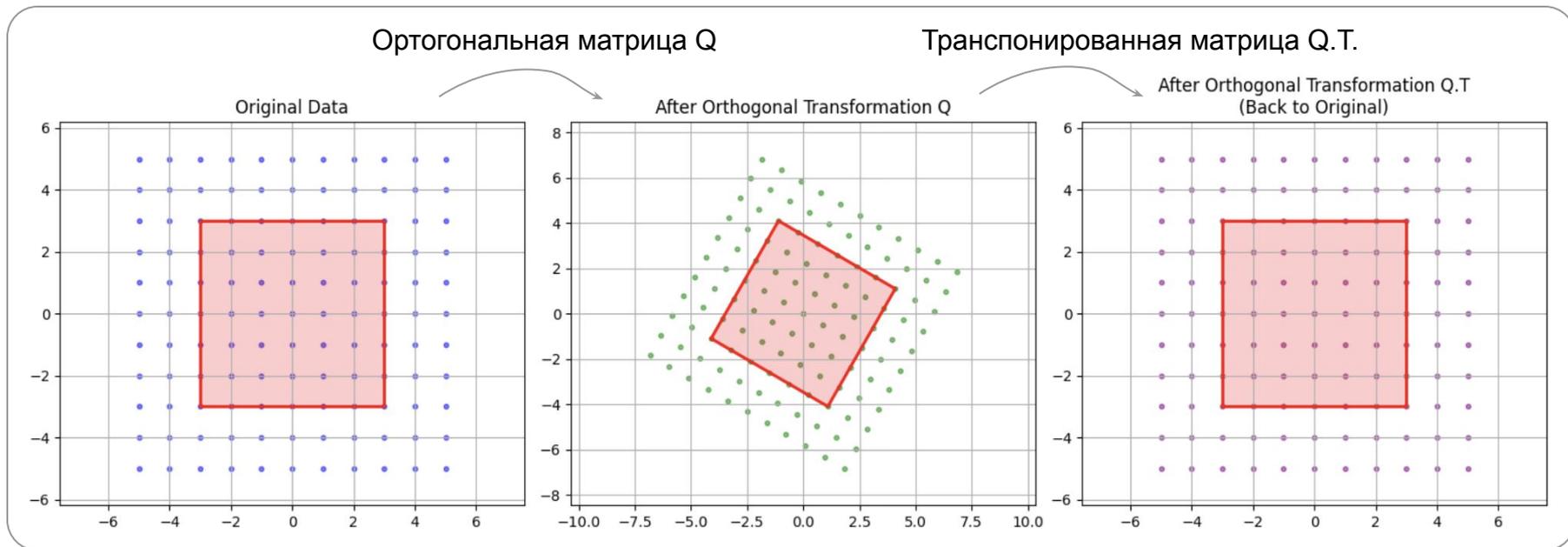


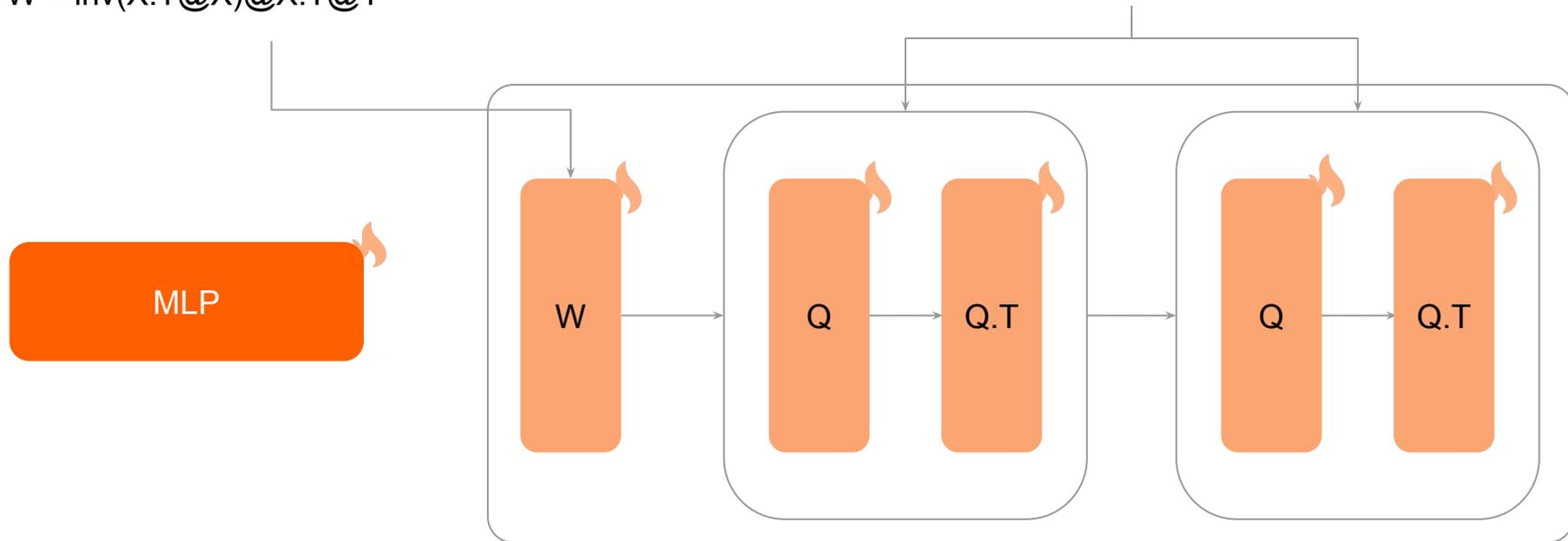
Схема инициализации MLP



Предобученный closed-form слой

$$W = \text{inv}(X.T@X)@X.T@Y$$

Пары инициализированных слоев



Метрики



- Word Recall @ k

Минус в том, что хвалим за более длинные теги

- Word Error Rate @ k

- Retrieval Metrics

Минус в том, что штрафует за дополнительные уточнения

- Hitrate @ k

- MRR @ k

- NDCG @ k

- ...

Примеры



Теги

- кружевной белый топ без рукавов, укороченный
- белая кружевная блузка без рукавов
- белый кружевной топ без рукавов с цветочным узором
- серебряные кольца, аксессуары



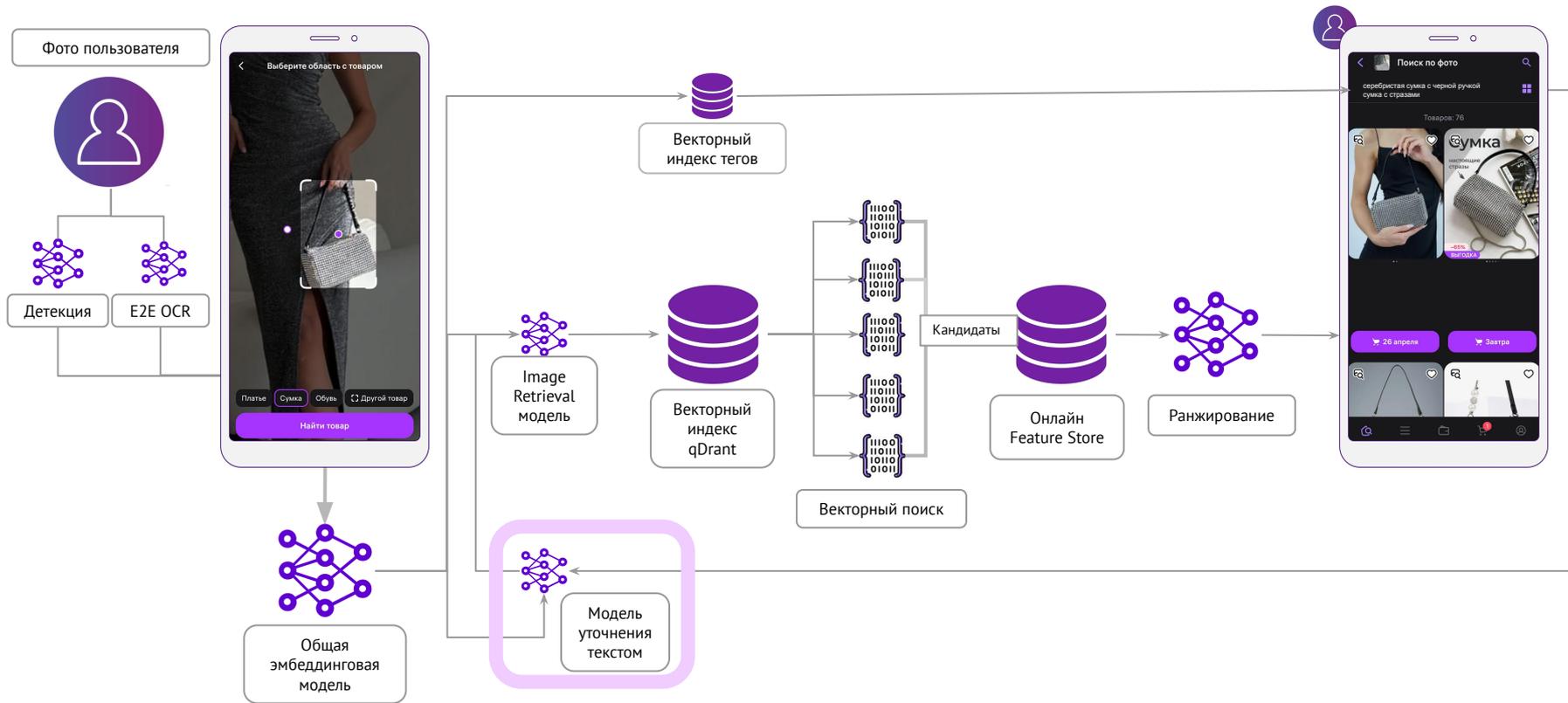
Теги

- топ с длинным рукавом с tribal узором
- синий топ с белым и красным узором
- бикини с завязками, высокая талия

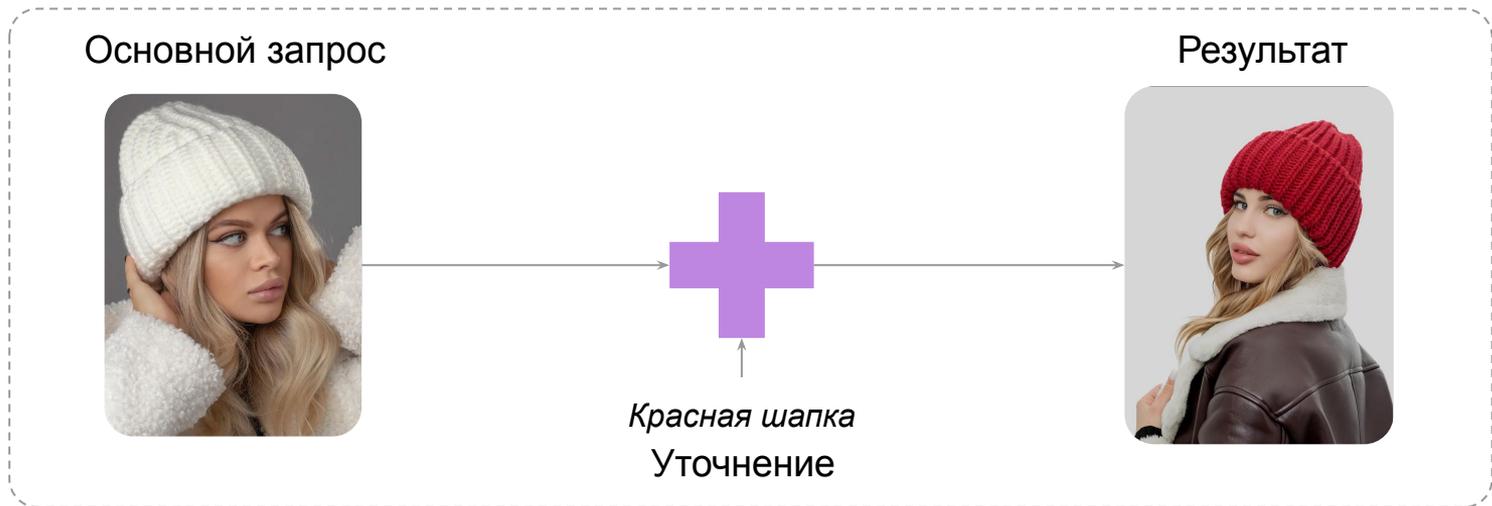
Уточнение ТЕКСТОМ



Архитектура сервиса

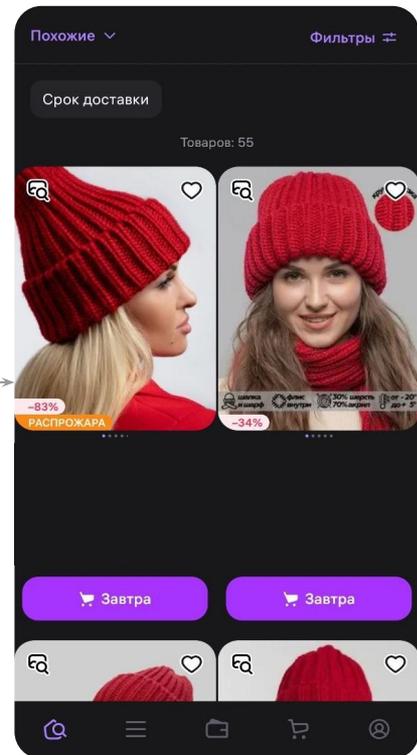
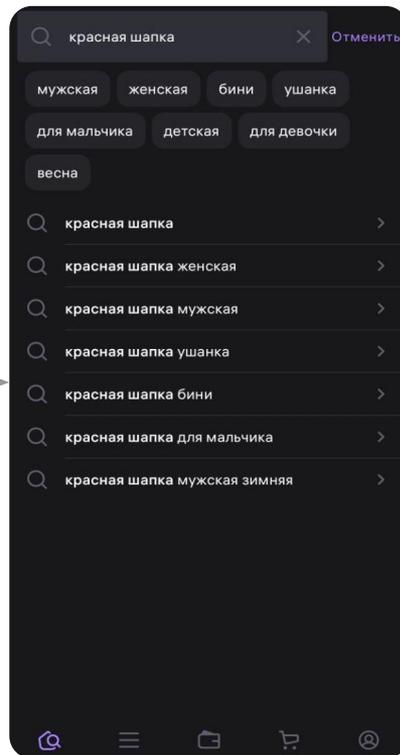
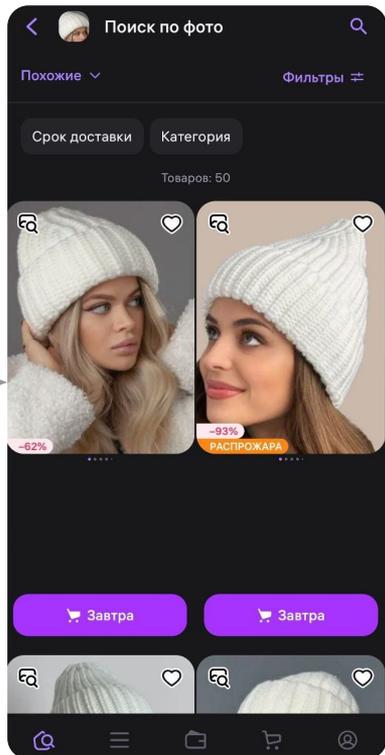
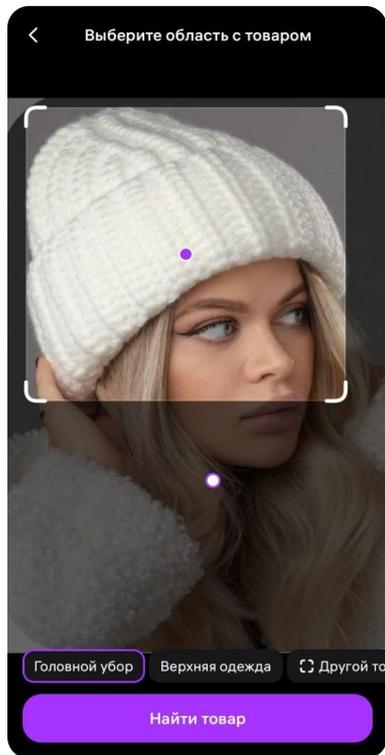


Задача



Что хочет
пользователь *

Задача



Данные

Картинка — текст — картинка

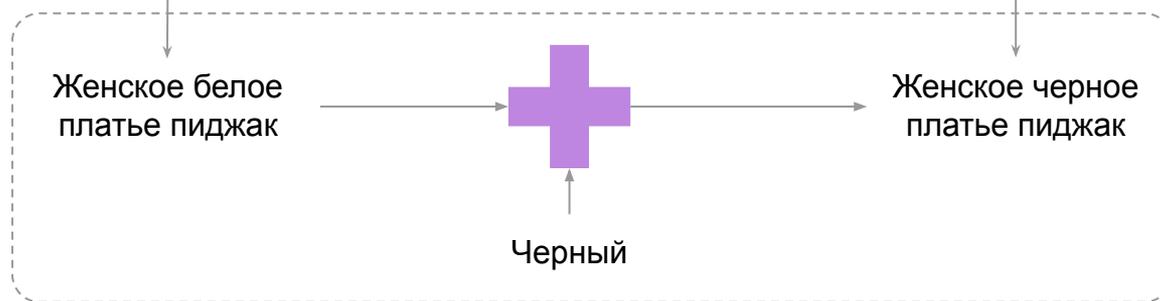


Черный

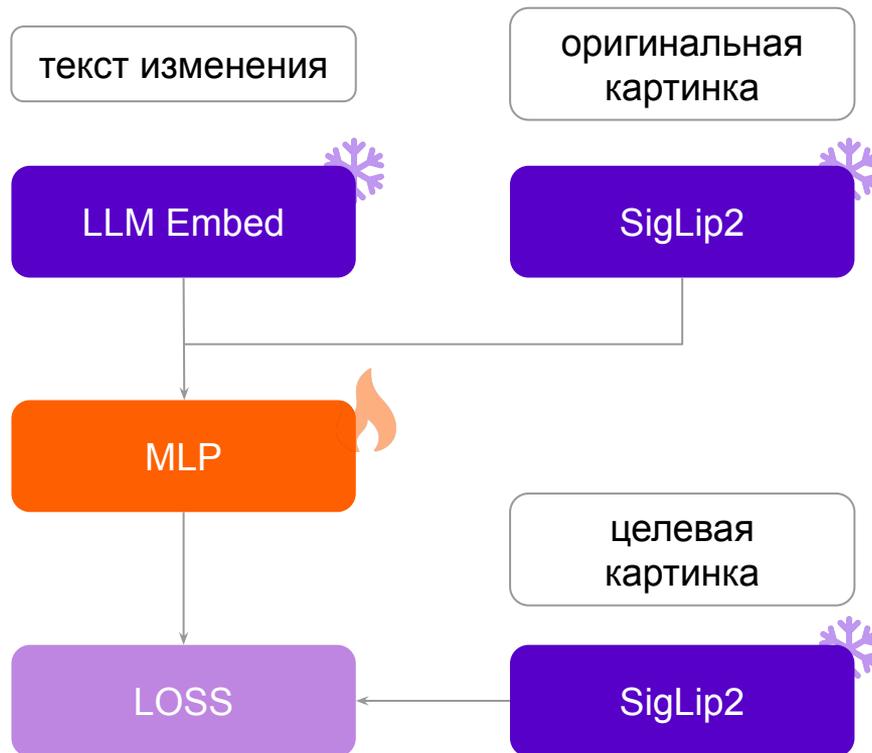
Датасет на основе нескольких «цветов» с одной КТ + VLM



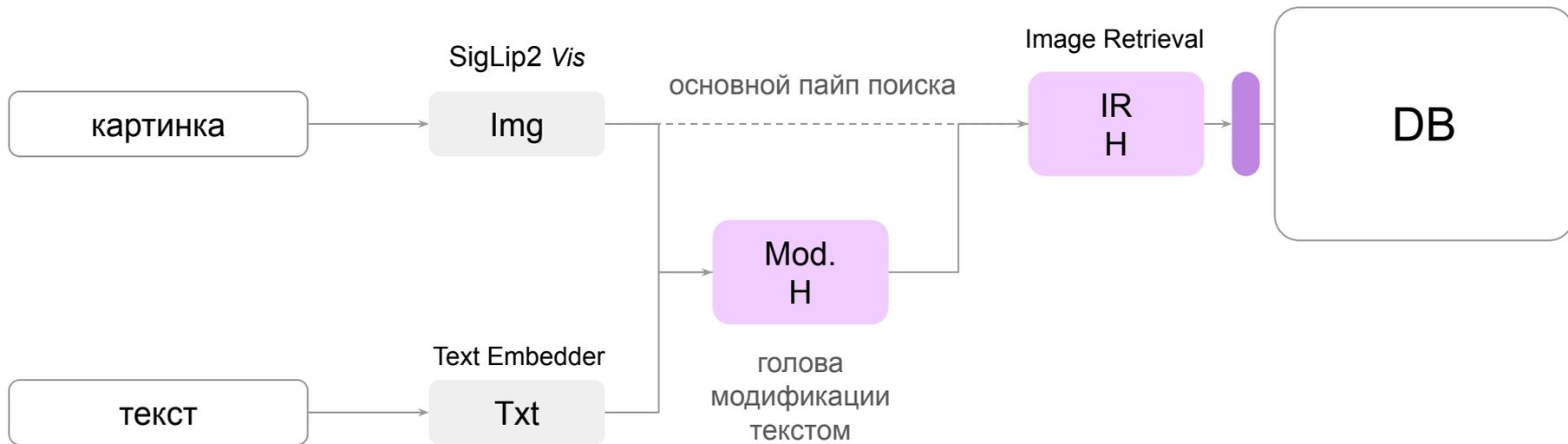
Данные — триплеты текста



Обучение модели уточнения текстом



Архитектура уточнения текстом



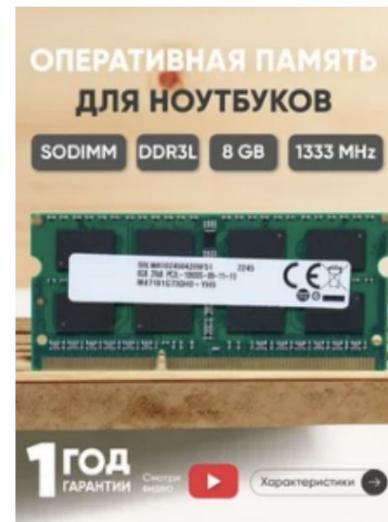
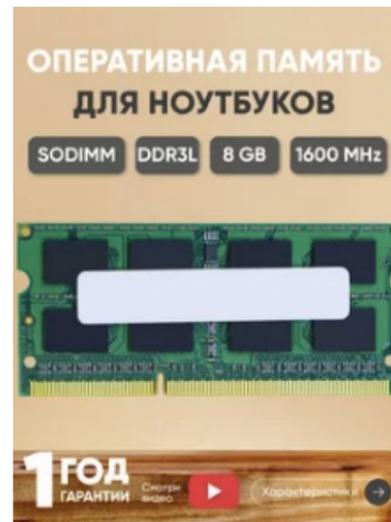
Примеры & Метрики

Фигурка Эльзы из Frozen



- Hitrate @ k
- MRR @ k
- NDCG @ k

Частота 1333 mhz



Результаты



Метрики ImageRetrieval модели



Метрика / Размер вектора fp16	Android		iOS	
	128	256	128	256
GMV	+14,47%	+14,95%	+18,11%	+18,27%
Orders	+10,21%	+10,52%	+16,32%	+16,56%
ARPPU	+5,93%	+6,01%	+4,01%	+4,04%
CR (View-Order)	+1,24%	+1,31%	+1,67%	+1,74%

Спасибо за внимание и до встречи!



@Nikita_R_omanov



@MakovEvgeny



@o6m0o9m



Telegram-канал
WB Space



Блог WB Tech
на Habr