

Интеллектуальный помощник оператора службы поддержки

RUTUBE

Zufallswald von Entscheidungsbäume

Проблемы



Медленно



Ошибаются



Дорого

Схема



Запрос

Найти подходящую
части
документации



Генерация

Сформулировать
ответ



Решение

Быстро показать
ответ специалисту

Технология



Retrieve

Synth data
Chroma_DB
Multilingual-e5-large

NDCG
Recall



Generate

Mistral Nemo,
Ollama

G-Eval



Endpoint

ТГ бот, API ...



Пробовали

LLM

Дообучение эмбеддера

Query-rewriter

Фильтрация документов

Точки роста

Расширить
синтетические
данные

Масштабирование

Multivector
retrieval



Команда



- Анастасия Юдина
LLM, капитан, @hug_asya02
- Александр Евдокимов
LLM, @snakerzr
- Алексей Бондаренко
Backend, @yabondaleksey
- Ольга Иванова
DL, @oakarabut
- Михаил Васильев
Backend, DL, @LaHundo