Tender Hack 2025 Команда РуВуе

Реализация ИИ-ассистента для анализа поисковых запросов пользователей в режиме чата

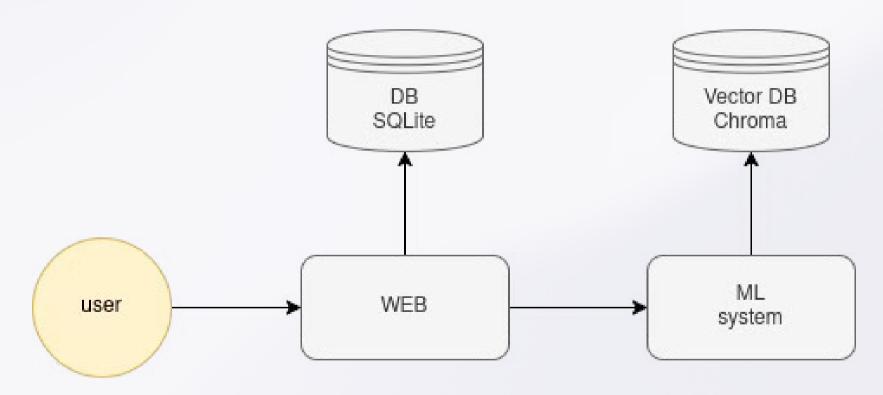




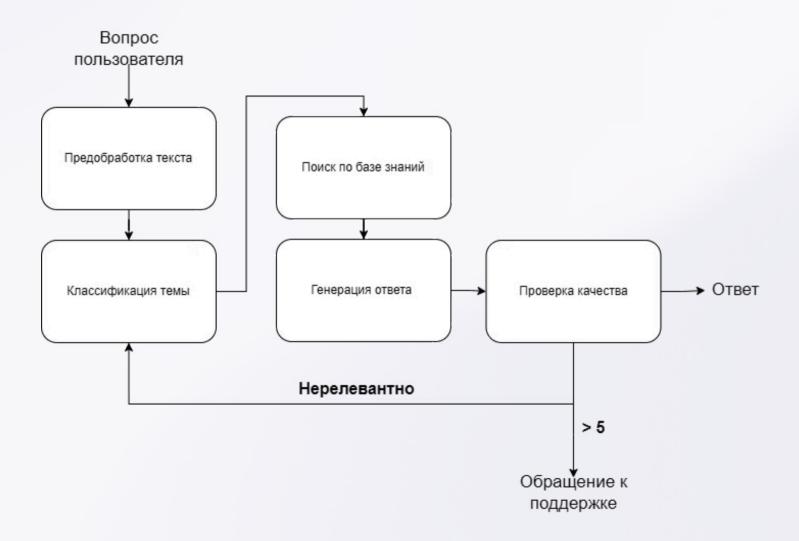




Архитектура проекта



Пайплайн ассистента



Классификация запроса

Категории: вопрос о функционале, жалоба, техническая поддержка, другое



Модель: MoritzLaurer/mDeBERTa-v3-base-mnli-xnli

Кол-во параметров: 279 млн. параметров

Преимущества:

Zero-shot подход

Поддержка русского языка

Взаимодействие с базой знаний



Модель: sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2

Кол-во параметров: 118 млн. параметров

Преимущества:

Многоязычность (включая русский)

Скорость

Формирование ответов



Mодель: yandex/YandexGPT-5-Lite-8B-instruct

Кол-во параметров: 8 млрд. параметров

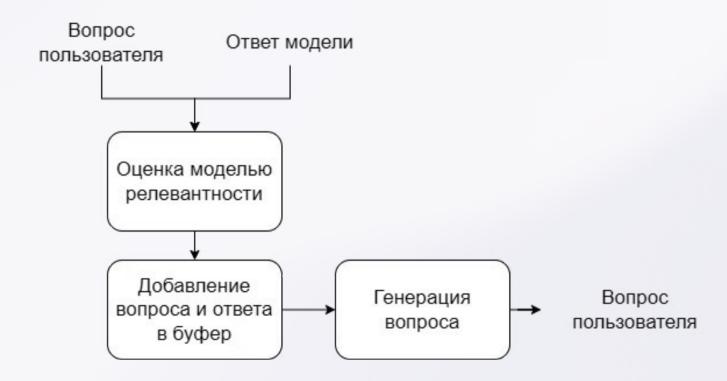
Контекстное окно: 32 тыс. токенов

Преимущества:

Многоязычность (включая русский)

Скорость

Генерация релевантных вопросов



Интерфейс системы

Чат

Возможность получать быстрые, релевантные ответы

Все диалоги сохраняются, обеспечивая возможность вернуться к истории переписки в любой момент

Аналитика

Разбивка по категориям

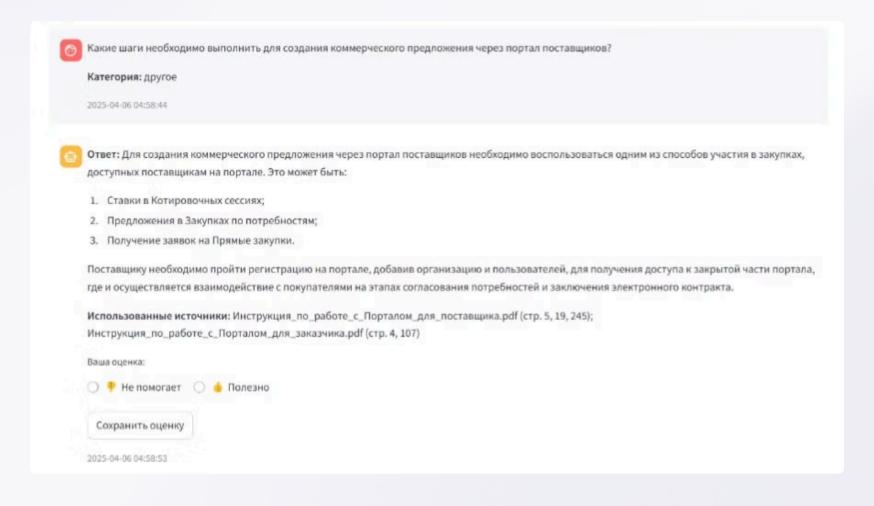
Процентное соотношение положительных и отрицательных оценок

Подсчёт общего количества оценок

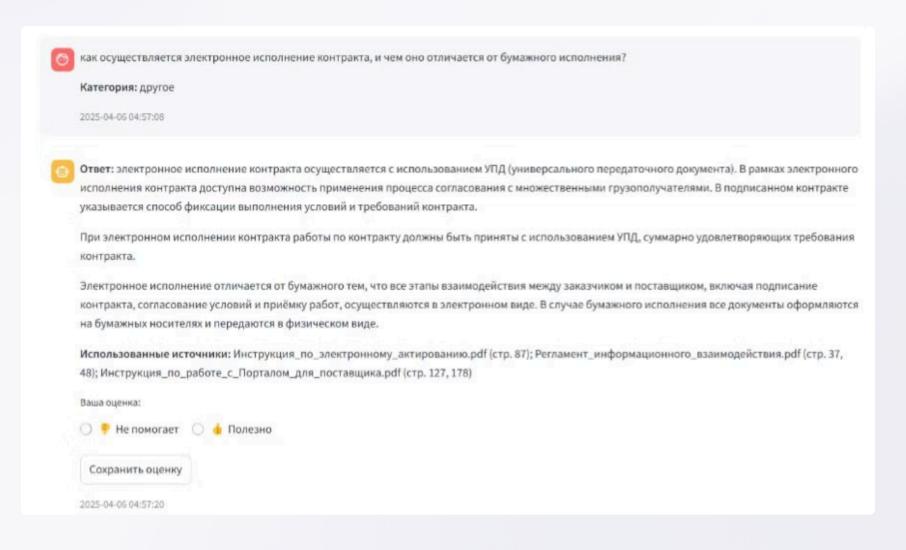
Визуальный стиль

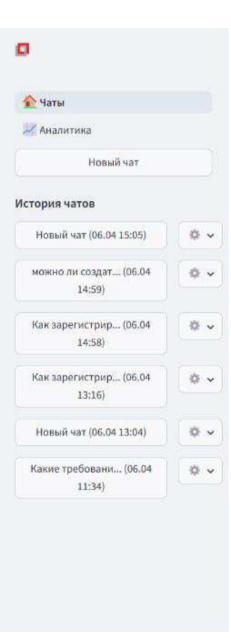
Интерфейс приложения выдержан в едином стиле с Порталом поставщиков — это обеспечивает визуальную целостность и узнаваемость

Пример ответа ассистента



Пример ответа ассистента

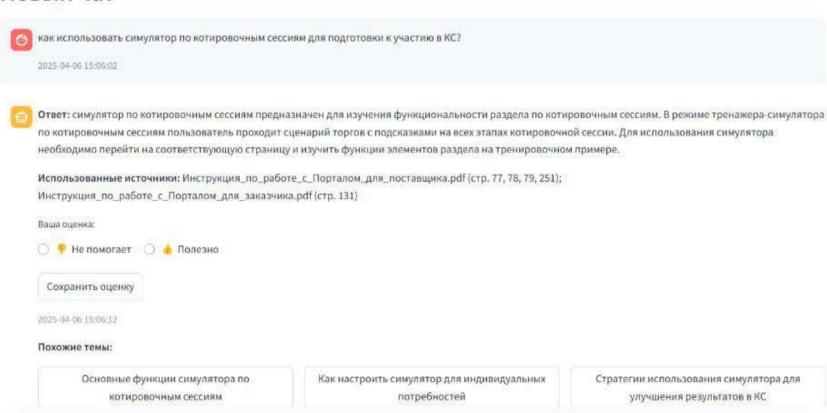






Новый чат

Взедите сообщение...



Перспективы проекта

Интеграция в инфраструктуру заказчика

Улучшение качества генерации и расширение возможностей языковой модели

Динамическая длинная память

Онлайн-обновление базы знаний

Расширение логики диалога и контекста

Обработка изображений и документов

Интеллектуальные рекомендации