

AeroDoc Assistant

ИИ-ассистент по технической документации в авиастроении

Интеллектуальный агент на основе LLM для семантического поиска и анализа технических документов авиастроительных предприятий

Ключевые проблемы отрасли

Авиастроительные предприятия работают с тысячами технических документов: конструкторская документация (ЕСКД), технологические процессы (ЕСТД), руководства по ремонту, сертификаты материалов, отчеты об испытаниях, отраслевые стандарты.

Потеря контекста

Инженеру требуются часы, чтобы найти все документы, связанные с конкретным узлом – чертежи, спецификации, технологические карты, требования к материалам

Риск ошибок

Неучет актуальной редакции документа или связанного ограничительного бюллетеня может привести к критическим последствиям

Низкая скорость onboarding

Новому специалисту требуются месяцы, чтобы погрузиться в корпус документов предприятия

Ручной труд

Ответы на стандартные запросы требуют ручного поиска в документации

Задача хакатона

Разработать прототип интеллектуального агента на основе Large Language Models (LLM), который обеспечивает семантический поиск, анализ и интеллектуальный диалог с корпусом технических документов авиастроительного предприятия.

Агент должен быть не просто чат-ботом, а **экспертной системой с пониманием контекста авиастроения**

01

Умный семантический поиск

Поиск документов и фрагментов по смыслу, а не только по ключевым словам

03

Создание связей

Автоматическое построение связей между документами с визуализацией

02

Диалоговый Q&A

Ответы на вопросы с обязательной ссылкой на конкретный документ, раздел и страницу

04

Проверка актуальности

Выявление устаревших ссылок и потенциальных противоречий между требованиями

Образ результата

Обязательный функционал (MVP)

- Семантический поиск по смыслу запроса
- Ответы с цитированием источников
- Граф связей между документами (Graph RAG)
- Проверка актуальности и противоречий

Продвинутые функции

- Мультиmodalность: анализ чертежей с помощью CV
- Голосовой интерфейс для цеха
- Авто-суммаризация отчетов об испытаниях
- Прогнозирующие подсказки

Ключевое ограничение: Все ответы агента должны иметь явные ссылки на обработанные документы. Галлюцинации недопустимы.

Технический стек

1 Архитектура RAG

Retrieval-Augmented Generation – обязательна для достоверности и избегания галлюцинаций

2 Ядро LLM

OpenAI GPT, Llama 3, Mixtral, GigaChat или YandexGPT на выбор участников

3 Векторизация

Эмбединги и векторная БД: Chroma, Qdrant или Pinecone

4 Бэкенд

Python с FastAPI или Flask

5 Фронтенд

Чат-интерфейс на Streamlit, Gradio или React (опционально)

Участникам предоставляется симулированный датасет: структурированные документы (JSON/XML), неструктурированные (PDF/DOCX), метаданные и словарь терминов авиастроения.

Критерии оценивания

Жюри будет оценивать прототип по демо-сессии с заранее подготовленным сценарием запросов.

Точность и достоверность

Процент правильных ответов с корректными ссылками. Штраф за галлюцинации

Глубина контекста

Способность связать информацию из нескольких документов в единый ответ

Инновационность

Реализация продвинутых функций: граф знаний, анализ противоречий, мультимодальность

Юзабилити

Интуитивность, скорость работы, ясность вывода и визуализация

Архитектура кода

Модульность, возможность развертывания, обработка ошибок

Обоснованный ML

Оценивается глубина анализа при выборе финального алгоритма или архитектуры.