

Паспорт проекта: MAI's Hive

1. Общая информация

- **Название проекта:** Интерактивная система анализа электронного документа
- **Краткое описание проекта:** Проект упрощает обработку PDF-документов для бухгалтеров и секретарей. Система позволяет определить тип документа на основе содержимого и проверить наличие определенных критериев: подпись, печать. Для удобства использования система представлена в виде веб-приложения. Бэкенд-часть приложения использует большую языковую модель для анализа содержимого.
- **Дата начала проекта:** 11.11.2024
- **Дата завершения проекта:** 25.05.2025
- **Команда:**
 - **Мишин Сергей Алексеевич:** Team lead
 - **Барменков Артемий Сергеевич:** UX/UI-дизайнер
 - **Бондаренко Виктория Арсеновна:** Технический писатель
 - **Борзова Дарья Максимовна:** Frontend-разработчик
 - **Иванченко Макар Дмитриевич:** Backend-разработчик
 - **Казец Полина Олеговна:** Аналитик
 - **Савинов Никита Олегович:** Frontend-разработчик
 - **Сергеев Владимир Андреевич:** Data-инженер
 - **Федоров Ярослав Артемович:** QA-инженер
 - **Шведов Александр Иванович:** Backend-разработчик

2. Цель проекта

- **Основная цель проекта:** Разработка интерактивной системы анализа электронных документов
- **Ожидаемые результаты:** Интерактивная система анализа электронных документов, обеспечивающая автоматизацию процессов и повышение эффективности работы персонала

3. Задачи и процесс работы

- **К работе (только цифрами):**
 - В процессе: 0 задач
 - Завершено: 5 задач

- Заблокировано: 0 задач
- **Список задач (по статусам):**
 - **Завершено:**
 1. Ресёрч-док
 2. Диаграмма юз-кейсов
 3. Диаграмма контекстов, компонентов
 4. Презентация проекта
 5. Паспорт проекта

4. Прогресс и результаты

- **Текущий статус проекта:** Проект находится на ранней стадии разработки. Нами была разработана предварительная архитектура на основе микросервисов и выбран стек разработки.
- **Достигнутые результаты по задачам:** Наша команда работала над анализом существующих решений и созданием различных диаграмм, отражающих концепцию проекта. Для организации эффективной работы оформлен репозиторий и трекер задач на GitHub. К защите проекта была подготовлена презентация и оформлен паспорт проекта.
- **Риски и препятствия:** Поскольку проект использует большую языковую модель для анализа содержимого документов, для развёртывания потребуются значительные вычислительные мощности. К тому же, система должна быть разработана с учётом больших объёмов данных и должна легко масштабироваться. Не стоит забывать про неточность системы распознавания текста — возможно, придётся искать другие решения.

5. Ресурсы и материалы проекта

- **Используемые инструменты и технологии:**
 - Фронтенд-фреймворк: **React JS**
 - Бекенд-фреймворк: **FastAPI**
 - Связь между микросервисами: **gRPC**
 - Распознавание текста: **Tesseract OCR**
 - Анализ текста документа: **Ollama, Llama 3.3**
 - Анализ подписи, печати: **OpenCV**
- **Ссылки на внешние ресурсы:**
 - Репозиторий проекта: <https://github.com/oryce/mai-ck>
 - Трекер задач: <https://github.com/users/oryce/projects/1>

6. Комментарии и мысли команды

- **Комментарии:** Нам будет крайне интересно поработать с большой языковой моделью типа Llama, свободно доступной и отлично функционирующей на суперкомпьютерах.