Название программы: Методы искусственного интеллекта в задачах обработки результатов дистанционного зондирования Земли

Название команды: Unspoken Team

Team lead (ФИО, tg): Кириллова Елена Константиновна @KirillovaElenaK

Ментор (ФИО, tg): Буреева Полина Сергеевна @orlando\_plbre

Паспорт проекта

*«Создание системы хранения и обработки датасетов»*

1. Общая информация

* **Краткое описание проекта**:  
  *В рамках выполнения проектной работы разрабатывается система (приложение) для хранения и обработки наборов данных (датасетов), используемых в процессе создания дипломных работ студентами Цифровой кафедры Московского авиационного института, создание которой актуально для повышения удобства студентов в процессе поиска и получения датасетов, необходимых в их разработках.*
* **Команда**:

|  |  |  |  |
| --- | --- | --- | --- |
| *№* | *ФИО* | *Группа по ООП* | *Роль в команде* |
| *1* | *Кириллова Елена Константиновна* | *М8О-203Б-23* | *Team lead, frontend- разработчик* |
| *2* | *Борисов Денис Сергеевич* | *М8О-209Б-23* | *Backend-разработчик* |
| *3* | *Ветошкина София Владимировна* | *М8О-203Б-23* | *Frontend-разработчик* |
| *4* | *Власко Михаил Михайлович* | *М8О-208Б-23* | *Технический писатель, Backend-разработчик* |
| *5* | *Жуховицкий Александр Дмитриевич* | *М8О-203Б-23* | *Backend-разработчик* |
| *6* | *Михайлов Александр Денисович* | *М8О-203Б-23* | *Fullstack-разработчик* |
| *7* | *Никитцев Антон Валерьевич* | *М8О-203Б-23* | *TechLead, Fullstack-разработчик* |
| *8* | *Слободин Никита Алексеевич* | *М8О-203Б-23* | *ML-инженер* |
| *9* | *Штыхно Илья Алексеевич* | *М8О-209Б-23* | *ML-инженер* |
| *10* | *Юсупов Артём Маратович* | *М8О-209Б-23* | *ML-инженер* |
| *11* | *Яковлев Вадим Дмитриевич* | *СМ-31* | *Frontend-разработчик* |

2. Цель проекта

* **Цель проекта**:  
  *Разработать веб-приложение для хранения датасетов (наборов данных) различных типов.*
* **Ожидаемые результаты**:  
  *Комплексная система, позволяющая пользователю проходить регистрацию и авторизацию, осуществлять загрузку, поиск и выгрузку датасетов.*

3. Задачи и процесс работы

* **К работе (только цифрами)**:

4. Прогресс и результаты

* **Текущий статус проекта**:

*На данный момент на review находятся блоки, отвечающие за организации, приглашения, уведомления, а также deploy-манифесты для развёртки системы внутри docker-compose на сервере.*

*В работе сейчас находятся endpoint’ы, связанные с датасетами, практически завершено подключение RAG-системы к векторной базе данных Weaviate, вспомогательной LLM GigaChat и основной базе данных Postgres. Начата работа над encoder’ом и preprocessor’ом RAG-системы.*

* **Достигнутые результаты по задачам**:  
  *Начата разработка всех основных компонентов приложения:* 
  + *выбран стек технологий*
  + *определены общие принципы и архитектура приложения*
  + *проведён анализ данных для последующего обучения и функционирования RAG-системы*
  + *разработана API-спецификация сервиса для взаимодействия с RAG-системой, а также поднят веб-сервис на FastAPI*
  + *разработаны концепции пользовательского интерфейса*
  + *реализованы страницы входа, главная страница, страницы взаимодействия с датасетами и страницы взаимодействия с организациями*
* **Риски и препятствия**:  
  *Основной сложностью в процессе работы над проектом стала необходимость поиска датасетов как можно большего числа типов, что необходимо для полноценного обучения и работы системы поиска датасетов. Тяжёлая коммуникация внутри команды, вызванная некомпетентностью тимлида. Смена темы проекта вызывала сильное волнение членов команды. Однако на данный момент проблемы, волновавшие команду, решены и атмосфера стабилизировалась.*

5. Ресурсы и материалы проекта

* **Используемые инструменты и технологии**:
  + Backend:
    - Dotnet 9 (C#)
    - EF Core 9
  + Frontend:
    - TypeScript
    - Next.js
    - Tailwind
    - Shadcn
  + ML
    - Python
    - FastAPI
    - Langchain
    - Transformers
  + Docker compose
  + Github actions
* **Ссылки на внешние ресурсы**:
  + <https://github.com/orgs/UnspokenTeam/repositories> - репозитории разработки сервисов приложения;
  + https://dbdiagram.io/d/Dataset-storage-67b6f107263d6cf9a0d5fe4f - схема базы данных приложения
* **Данные**
  + Hugging Face (https://huggingface.co)

6. Комментарии и мысли команды

* **Комментарии**:  
  *Проект уверенно движется к реализации: выполнен анализ данных и выбор подходящих наборов для тестирования RAG-системы, разработана архитектура с использованием таких технологий, как Kafka, Minio, PostgreSQL, Weaviate, Prometheus, Grafana, KeyCloak, Caddy, Vault, а также продвигается работа над фронтендом и ключевыми бэкенд-сервисами.*

*В ближайших планах завершение разработки backend-сервисов, интеграция RAG-системы для поиска датасетов. Также в процессе полная пользовательская и техническая документация для обеспечения удобства внедрения и эксплуатации приложения. Мы стремимся создать продукт, который станет значимым и полезным инструментом в деле дистанционного обучения студентов Цифровой кафедры.*