



Aeroclub Challenge 2023

12 - 27 мая



Название проекта

Трек: Создание обработчика эмейлов

Название команды: Team42



Уриэль Кайя

AI Энтузиаст



Описание задачи

Сортировка писем по характеру их содержания с целью дальнейшего извлечения релевантных запросу сущностей для дальнейшей автоматизации задачи.

- Определение сути письма
- Поиск ключевых сущностей
- Составление JSON

Подход к реализации - исследование возможностей современных решений на базе технологий

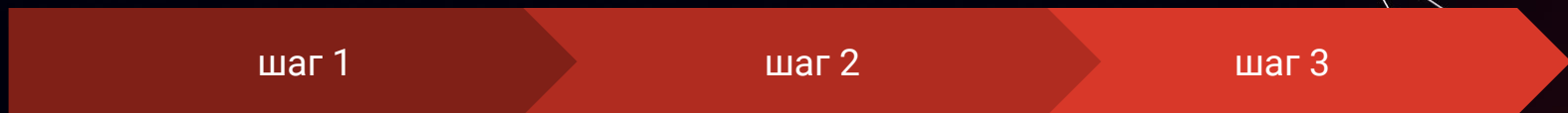
- LowCode
- AI на базе LLM (GPT-3.5-Turbo, Davinci)
- Python 3.11



Основные шаги

- Чистка запроса.
 - удаление лишнего из содержания письма: пунктуация, подписи, повторяющиеся блоки и тд
- Определение типа запроса. API Backend
 - создание мини резюме содержания
 - Классификация на запрос или другое
- Выделение сущностей
 - По типу запроса определение соответствующей JSON схемы

Архитектура



Python Jupiter
обработка и
подготовка письму.

Формирование Get
запроса в
разработанный сервис
на базе LowCode API
Backend XANO.
Обработка запроса с
помощью LLM
моделей OpenAI

Python Jupiter
получение и обработка
вызова API



Результаты

Скорость работы:

~ от 4 до 10 сек/письмо (без учета предобработки)

Выходные данные:

Классификация запрос / не запрос

Определение типа запроса и формата перемещения

JSON код, где требуется



Выводы

Плюсы

- Высокая скорость разработки сервиса с поддержкой API
- Быстродействие решений
- Вычленение сущностей на высоком уровне и формирование json и классификация текста

Минусы

- Сравнительно низкий score модели остался загадкой.

Дополнительно

Для тестирования модели необходимо предоставить данные в excel файле с обязательными колонками

- ID порядковый номер
- text (содержание письма)

На выходе будут добавлены колонки

- label [0,1]
- type пример { "Type": "booking", "Mean": "flight" }
- json только структура под соответствующий запрос (mean)