Есина Екатерина

[kate19esina@gmail.com](mailto:kate19esina@gmail.com)

**Сопроводительный отчет к тестовому заданию «Умный ассистент»**

Тестовое задание включает в себя создание автономного умного ассистента для поиска в интернете новостей для мам и генерации на их основе текста для блога и сопроводительного изображения.

Папка проекта включает в себя следующие файлы:

* news\_agent
  + api\_keys.json
  + main.py — код проекта
  + README.md
  + requirements.txt — список необходимых библиотек для установки через терминал (*pip install -r requirements.txt*)
* Примеры генераций — с использованием API
  + Мировые новости
    - blog\_post\_{номер поста}.docx
    - image\_{номер поста}.png
    - image\_prompt\_{номер поста}.txt
  + Специализированный сайт
* Отчет.docx

В рамках выполнения задания было сделано предположение о том, что поиск новостей и генерация не должны требовать много ресурсов, в связи с этим было принято решение использовать API для получения наилучшего результата. Для использования необходимы ключи, которые являются индивидуальными и требуют регистрации, в связи с этим ключи предоставлены не будут. При желании пользователь может получить собственные ключи и указать их в файле api\_keys.json.

В качестве альтернативы API в проекте реализована генерация текста и изображения с помощью локальных моделей, однако в этом случае для генерации изображения требует много места и времени (в случае запуска на CPU). В случае не указания ключей API будут скачены и запущены локальные модели.

**Поиск новостей**

Поиск новостей осуществляется двумя способами:

* NewsAPI — получение новостей с новостных порталов всего мира на английском языке. Перед генерацией осуществляется перевод на русский. Проблемой такого подхода является трудность фильтрации новостей. Были опробованы разные варианты запросов, однако, по причине того, что поиск происходит по ключевым словам, например, запрос «обучение детей математике» вернет строку «помимо обучения детей математике, ему еще очень нравилось препарировать лягушек», что не совсем соответствует тематике.
* Скрапинг сайтов, ориентированных на мам. В качестве примера был выбран сайт <https://libertymag.ru/novosti-mamy-deti>.

Отбираются только новости за последние 7 дней, в связи с этим данных достаточно мало.

Результаты этого этапа можно улучшить, осуществив парсинг большего количества сайтов для мам, однако стоит учитывать, что не все сайты разрешают скрапинг. Проблему фильтрации мировых новостей можно решить, обучив (или используя существующую) модель классификатор и отбирая только новости соответствующей категории.

Также важно определить, что является более приоритетным: свежесть новости или целевая аудитория. На мой взгляд свежие статьи для мам не совсем попадают под классификацию «новости».

**Генерация текста поста для блога и промпта**

Генерация текста поста осуществляется двумя способами:

* Запрос по API к модели Mistral — позволяет быстро получить качественный текст, но требует регистрации ключа.
* Запуск локальной модели — для примера была выбрана модель d0rj/rut5-base-summ с huggingface. Такой подход позволяет напрямую ограничить количество токенов при генерации и даёт хорошие результаты, но иногда в тексте оказываются слова на английском и повторения.

Длина текста для блога составляет около 200 слов. Генерация промпта для создания изображения на следующем этапе осуществляется путем подачи на вход модели текста для блога и ограничения генерации до примерно 20 слов. Промпт для генерации изображения потенциально можно улучшить, скорректировав запрос к текстовой модели.

**Генерация изображения**

Генерация изображения также осуществляется двумя способами:

* Запрос на сайт для генерации <https://pollinations.ai/> — позволяет сгенерировать красочную достаточно реалистичную картинку, не требует регистрации. Минусом является не очень большое разрешение изображения – 768x768 пикселей, и вотермарка с названием сайта в правом нижнем углу.
* Запуск локальной модели — на huggingface не так много небольших моделей, принимающих промпты на русском языке, для примера была выбрана модель openskyml/lexica-aperture-v3-5. Позволяет достаточно быстро генерировать изображение (при условии использования GPU), но скачивание требует около 4-5 Гб свободного места. Также реалистичность изображения оставляет желать лучшего.

Результаты данного этапа можно улучшить, выбрав большую модель, что возможно только при наличии ресурсов. Также возможным является повышение качества изображения и обрезка вотермарки (вероятно, что это не соответствует правилам пользования сайта, этот пункт требует дальнейшего изучения).

**Сохранение результатов**

Для сохранения результатов генерации на рабочем столе пользователя автоматически создается папка blog\_files, в которую добавляются файлы:

* blog\_post\_{i}.docx — текст для блога
* image\_prompt\_{i}.txt — текст для генерации изображения
* image\_{i}.png — сгенерированное изображение

Где i – порядковый номер новости.