

ПРОТОКОЛ ОБСУЖДЕНИЯ ЗАДАЧИ С ЗАКАЗЧИКОМ

Заказчик:	ИП МАЖАЕВ ВЯЧЕСЛАВ СЕРГЕЕВИЧ, Мажаев Вячеслав Сергеевич
Название проекта:	SpeechMate
Исполнители	Куликова Татьяна Дмитриевна, Милорадова Ксения Сергеевна, Насыхова Анастасия Артемовна, Щербакова Елизавета Александровна

ОПИСАНИЕ ПРОЕКТА

Цель: разработать сайт и телеграм-бот, которые дают возможность пользователям масштабировать контент путем перевода речи на другой язык, а также упростить коммуникацию путем озвучивания и генерации краткого пересказа чатов.

Сайт должен предоставлять пользователю следующие функции:

1. Создавать транскрипт файла (видео или аудио);
2. Переводить содержимое файла (видео или аудио) на заданный язык с сохранением пола и голоса (опционально) говорящего.

Телеграм-бот должен предоставлять пользователю дополнительные функции для обработки не только файла, но и чата, такие как:

1. Создавать транскрипт переписки за выбранный период времени;
2. Озвучивать сообщения в групповом чате за выбранный период времени с одной из трех опций синтеза речи;
3. Генерировать краткий пересказ переписки за выбранный период времени в виде текста или аудио.

КОНКУРЕНТЫ

По мнению заказчика, ближайшим к нам конкурентом является Rask. AI. Это один из самых мощных и успешных инструментов для перевода и озвучки аудио/видео с помощью искусственного интеллекта.

Плюсы Rask.AI	Минусы Rask.AI
<ul style="list-style-type: none">- Поддержка более 130 языков;- Возможность клонирования голоса для озвучки или выбора одного из встроенных голосов;- Возможность определять количество спикеров в аудио/видео и озвучивать их разными голосами;- Наличие бесплатного пробного периода;- Возможность управления проектами пользователя.	<ul style="list-style-type: none">- Нет возможности интеграции с мессенджерами и озвучки чатов.

Следующий конкурент – 11ElevenLabs – технология для создания реалистичной озвучки на широком спектре языков.

Плюсы 11ElevenLabs	Минусы 11ElevenLabs
<ul style="list-style-type: none">- Возможность озвучки текстов, аудио и видео;- Интеграция с большим количеством медиа-платформ, с которых можно загружать видео;	<ul style="list-style-type: none">- Поддерживает мало языков по сравнению с предыдущим конкурентом;- Нет возможности клонирования голосов;- Нет возможности управлять проектами пользователя без платной подписки, кроме

Плюсы 11ElevenLabs	Минусы 11ElevenLabs
<ul style="list-style-type: none"> - Возможность озвучивания огромных текстов по блокам (например, озвучка книги по главам); - Возможность выбора определенного промежутка времени из аудио/видео для озвучки. 	<p>того, такая возможность реализована только для озвучки текстов;</p> <ul style="list-style-type: none"> - Нет возможности интеграции с мессенджерами и озвучки чатов.

Еще один конкурент - Video Dubber – инструмент для озвучки и перевода видео и текстов.

Плюсы VideoDubber	Минусы VideoDubber
<ul style="list-style-type: none"> - При покупке платной подписки нет ограничений на размер и длительность видео; - Есть возможность управлять проектами пользователя. 	<ul style="list-style-type: none"> - Нет возможности переводить аудио-контент; - Поддерживается мало языков; - Не реализовано клонирование голосов; - Нет возможности интеграции с мессенджерами и озвучки чатов.

ОСНОВНЫЕ ЗАДАЧИ

- Проведение технических интервью с заказчиком с целью сбора требований с фиксацией результатов в протоколе;
- Планирование проекта с помощью диаграммы Ганта;
- Исследование существующих решений для транскрибирования, перевода и синтеза речи;
- Сравнительный анализ конкурентов;
- Проектирование микросервисной архитектуры сайта и телеграмм-бота с интеграцией ML-пайплайна;
- Разработка прототипа интерфейса сайта в Figma;
- Разработка ML-пайплайна;
- Разработка backend и frontend частей сайта;
- Разработка телеграм-бота;
- Тестирование;
- Разработка технической документации;
- Оценка полученных результатов.

ПОЛЬЗОВАТЕЛИ

SpeechMate, в первую очередь, предназначен для преодоления языкового барьера физическими лицами, которые регулярно взаимодействуют с аудио- и видео- контентом на разных языках. В частности, среди них можно выделить студентов, занимающихся исследованиями, журналистов и блогеров, работающих с международной аудиторией, переводчиков, которым необходимо транскрибировать или переводить речь и других людей в возрасте от 18 до 60 лет, активно использующих социальные сети и мессенджеры для работы, учебы, общения и активно потребляющих контент из средств массовой информации разных стран.

SpeechMate предоставляет пользователям возможность транскрибировать аудио и видео в текст, переводить его на различные языки и синтезировать речь, что особенно полезно для быстрой

обработки и усвоения информации. Телеграм-бот SpeechMate дополнительно предоставляет возможность озвучивать и переводить телеграмм-чаты, что упрощает общение и позволяет экономить время на чтение и перевод переписок. Наши продукты могут быть использованы для учебных целей, исследований, создания контента и ведения международной коммуникации.

Разрабатываемый проект способен обрабатывать до 100 запросов одновременно и до 1000 запросов ежедневно.

ТРЕБОВАНИЯ К ИНСТРУМЕНТАМ И ТЕХНОЛОГИЯМ

1. Клиентская часть для веб-платформы должна быть разработана на Next.JS, верстка интерфейса должна осуществляться с использованием css-фреймворка Tailwind CSS;
2. Должны поддерживаться такие браузеры как Google Chrome, Yandex, Safari.
3. В качестве системы оплаты используется Stripe, гарантируя надежность и безопасность финансовых транзакций;
4. Для системы авторизации используется OAuth;
5. Система управления базами данных должна реализоваться через Firebase Firestore;
6. Для хранения файлов и медиа используется Firebase Storage;
7. Разработка Telegram-бота должна быть на языке Python;
8. Приложение-сервер должно быть разработано с использованием языка Python (версия 3.12 или выше) и библиотек FastAPI и Pydantic;
9. Используются ML-модели, включая OpenAI's Whisper для транскрипции, GPT-3.5-turbo для перевода, модели от Azure и 11labs для озвучивания, а также coqui.XTTS для клонирования голосов.

ОБЩЕЕ ОПИСАНИЕ ФУНКЦИОНАЛЬНОСТИ

WEB-САЙТ

1. Система должна обеспечивать возможность авторизации клиента через Google и email/password, а также предоставлять функцию восстановления пароля;
2. Пользователи могут создавать проекты, загружать аудио/видео файлы, выбирать язык перевода и голос для озвучивания (включая функцию Voice Cloning);
3. Результатом процессинга проекта является аудио или видео файл, в зависимости от формата файла-источника, с переведенной речью.
4. Пользователю должна быть предоставлена возможность прослушать файл в системе;
5. Система должна иметь возможность предлагать различные тарифные планы, в том числе и бесплатные, с возможностью управления текущим тарифом и покупкой дополнительных минут;
6. Backend сайта и telegram-бота должны быть интегрированы с ML-core для перевода аудио/видео и генерации результата в соответствии с выбранными пользовательскими настройками;
7. Уведомления и поддержка клиентов должны происходить через разделы FAQ и Team на сайте, также должно быть предусмотрено предоставление информации в разделах Privacy Policy и Terms of Use;
8. Управление профилем пользователя включает в себя настройки профиля и тарифного плана.

ЧАТ-БОТ

1. Система должна обеспечивать взаимодействие с пользователем в личной беседе и в групповых чатах;
2. Система должна реализовывать следующие функции:
 - а. транскрипция и перевод видео и аудио файлов с возможностью выбора языка перевода и других параметров;

- b. озвучивание чатов за выбранный период;
 - c. генерация краткого пересказа переписки за выбранный период.
3. Управление настройками и уведомлениями от бота, а также возможность включения и отключения бота в групповых чатах;
 4. Система должна предоставлять пользователю раздел FAQ и раздел поддержки пользователей.

РЕЗУЛЬТАТ РАБОТЫ

Результатом работы является телеграм-бот и web-сайт. Web-сайт представляет из себя web-интерфейс с реализованной логикой взаимодействия с ним пользователя (фронтенд), интегрированный с бэкэндом. В сайт и телеграм-бот должна быть встроена ML-система, способная обрабатывать аудио, видео и тексты в соответствии с поставленной задачей. Исходный код web-сайта, ML-системы, tg-бота и документация к системам будут переданы заказчику. Тестирование на реальных пользователях не предусмотрено.

КПЭ ОЦЕНКИ РЕЗУЛЬТАТОВ

Критериями оценивания работоспособности ML-пайплайна являются следующие показатели:


1. Время обработки запроса на перевод файла системой – ожидаемое время ответа на запрос не должно превышать час.
2. Процент правильно транскрибированных слов от общего числа слов – целевой показатель: не менее 95% точности.

Для оценки качества сайта были предложены следующие метрики:


3. Удовлетворенность пользователей – оценка по шкале от 1 до 5, основанная на пользовательских отзывах. Целевой показатель: средняя оценка не ниже 4.0.
4. Количество пользователей, дошедших до этапа получения результата перевода, за месяц – целевой показатель на первой стадии деплоя не менее 30.

Для оценки полноты реализации будет использоваться приемочное тестирование – должно быть реализовано не менее 90% функций из запланированного функционала.

Заказчик

 / Мажаев В.С.

Ответственный по проекту

 / Щербакова Е.А.

Дата: 30.11.2023