

## ФУНКЦИОНАЛЬНЫЕ И НЕФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ / ТЕХНИЧЕСКОЕ ЗАДАНИЕ

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

### Способы взаимодействия с системой:

- Web-сайт
- Чат-бот

## ТЕХНИЧЕСКОЕ ЗАДАНИЕ

### WEB-САЙТ

1. Вводы системы. Система должна давать возможность пользователю выбирать язык перевода и голос из выпадающего списка, которым он хочет озвучить переведенный текст, а также загружать видео в формате .mp4 или аудио в формате .mp3 в соответствующее окно на сайте одним из следующих способов:

- путем выбора нужного файла из проводника;
- функцией drag-and-drop;

Система должна автоматически генерировать транскрипт файла и давать возможность пользователю скачать его путем нажатия на соответствующую кнопку.

2. Выводы системы. Система должна генерировать видео в формате .mp4 или аудио в формате .mp3 с наложенной на него переведенной аудиодорожкой в зависимости от ввода пользователя.

Система должна автоматически генерировать транскрипт видео или аудио в виде текстового файла.

3. Функции системы. Система должна реализовывать следующие функции:

- 1) Создание транскрипта видео и аудио с указанием спикера и временных границ сказанного предложения;
- 2) Перевод транскрипта видео и аудио на заданный язык с сохранением разделения текста по спикерам;
- 3) Озвучивание видео выбранным голосом или с сохранением пола говорящего;
- 4) Озвучивание видео голосом спикера в оригинальном источнике – Voice Cloning;

4. Атрибуты системы. Система должна удовлетворять следующим атрибутам качества:

- 1) Запрос по переводу система должна выполнять в течение часа;
- 2) Одновременно сервис должен выдерживать не менее 100 запросов в один момент времени;
- 3) Доступ к сервису осуществляется по интернет-протоколу HTTP;
- 4) Особые требования к надежности не предъявляются.

5. Атрибуты системной среды. Система должна быть кроссбраузерной и функционировать при одновременном наличии подключения к интернету, достаточного количества RAM, достаточного количества ROM.

### ЧАТ-БОТ

1. Вводы системы. Система должна уметь функционировать в двух режимах: личная беседа с пользователем и интеграция с групповыми чатами пользователя. Выбор режима осуществляется с помощью выбора соответствующей опции в телеграм-боте.

### Личная беседа

Система должна отправлять пользователю запрос на ввод данных путем отправки сообщения с описанием информации, которую должен ввести пользователь.

Система должна давать возможность пользователю выбирать требуемый язык и голос, которым он хочет озвучить переведенный текст, а также загружать видео в формате .mp4 или аудио файл в формате .mp3, wav одним из следующих способов:

- путем выбора нужного файла из проводника;
- функцией drag-and-drop.

Выбор языка и голоса должен быть осуществлен путем отправки отдельных текстовых сообщений.

Выбор задачи, которую хочет решить пользователь (транскрипт, перевод, озвучивание чата, создание пересказа), а также функции клонирования голоса должен быть осуществлен с помощью выбора соответствующей опции в телеграм-боте.

В конце система также должна отправить пользователю запрос на необязательный параметр ввода: степень уменьшения громкости оригинальной дорожки на отрывках наложения перевода, которая указывается в виде числа типа Float в диапазоне от 0 до 100, где 0 – исходный громкости уровень оригинальной дорожки, 100 – отсутствие оригинальной дорожки в переведенном файле.

### Интеграция в групповые чаты

Система должна иметь возможность интегрироваться в групповые чаты.. Для этого она должна запросить у пользователя ссылку на чат в телеграмме.

При присоединении участника в чат система должна отправлять запрос на сохранение голоса пользователя в библиотеке бота путем записи голосового сообщения для последующей озвучки сообщений, при несогласии пользователя – использовать встроенный голос.

Активация функции озвучивания, транскрипта, перевода чата или создания краткого пересказа происходит в случае отправки сообщения с нужной командой и корректным вводом параметров. Требования к параметрам:

- Даты временного окна: “ДД.ММ.ГГГГ - ДД.ММ.ГГГГ”;
- Язык перевода: текстовое название языка;

Система должна предусматривать алгоритм обработки некорректного ввода.

2. Выводы системы. Система должна генерировать видео в формате .mp4 или аудио файл в формате .mp3 с наложенной на него переведенной аудиодорожкой в зависимости от ввода пользователя.

В случае, когда пользователем была выбрана функция транскрибации видео или аудио, выводом системы является текстовое сообщение.

В случае, когда пользователем была выбрана функция транскрибации видео- или аудио- сообщений, выводом системы является текстовое сообщение, в котором каждому транскрибированному текстовому отрывку через знак “:” сопоставляется уникальный идентификатор распознанного спикера.

3. Функции системы. Система должна реализовывать следующие функции:

- 1) Создание транскрипта видео и аудио с указанием спикера и временных границ сказанного предложения;
- 2) Перевод транскрипта видео и аудио на заданный язык с сохранением разделения текста по спикерам;
- 3) Озвучивание видео выбранным голосом или с сохранением пола говорящего;
- 4) Озвучивание видео голосом спикера в оригинальном источнике – Voice Cloning;
- 5) Возможность задать степень уменьшения громкости оригинальной дорожки на отрывках наложения перевода;
- 6) Озвучивание переписки в групповом чате за выбранный период времени с сохранением пола на основе голосов участников беседы, сохраненных в библиотеке бота;

- 7) Генерация краткого пересказа переписки за выбранный период времени в виде текста или аудио;
- 8) Возможность добавления бота в групповые чаты;
- 9) Возможность сбора сообщений в чате за определенный период для их автоматического перевода и/или озвучивания, а также их транскрипции и суммации.

4. Атрибуты системы. Система должна удовлетворять следующим атрибутам качества:

- 1) Запрос по переводу система должна выполнять не более часа.
- 2) Одновременно сервис должен выдерживать не менее 100 запросов в момент времени.
- 3) Доступ к боту осуществляется через защищенные протоколы взаимодействия с Telegram API.
- 4) Особые требования к надежности не предъявляются.

5. Атрибуты системной среды. Бот должен корректно функционировать на всех платформах, где доступно приложение “Telegram”, включая десктопные и мобильные устройства на различных операционных системах (iOS, Android, Windows, macOS, Linux). Все функции системы должны быть совместимы с текущей версией Telegram Bot API, обеспечивая стабильность работы и возможность внедрения новых функций при обновлениях платформы.

## ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ

### WEB-САЙТ

Система должна соответствовать следующим функциональным требованиям:

#### *Регистрация и авторизация*

- 1) Возможность регистрации и авторизации через Google и через email/password.
- 2) Возможность восстановления пароля через email.

#### *Работа с Проектами*

- 1) Просмотр пустого списка проектов с туториалом после регистрации/авторизации.
- 2) Создание нового проекта, который включает:
  - a) Загрузка аудио/видео файлов.
  - b) Выбор языка перевода.
  - c) Выбор голосов перевода (в том числе функции voice cloning).
- 3) Процессинг проекта (транскрипция, перевод, озвучка аудио/видео) и генерация результата.
- 4) Прослушивание результата.
- 5) Скачивание файла с результатом.

#### *Монетизация*

В проекте присутствуют следующие тарифные планы:

- Тариф **Free**
  - 3 токена
  - Максимальная длина видео – 1 минута
- Тариф **Creator**
  - 10 токенов в месяц
  - Максимальная длина видео – 1 минута
- Тариф **Standard**
  - 30 токенов в месяц
  - Максимальная длина видео – 5 минут
- Тариф **Producer**
  - 300 токенов в месяц

- Максимальная длина видео – 20 минут
- Тариф **Enterprise**
  - Более 1000 токенов в месяц
  - Максимальная длина видео – 90 минут

Токен – возможность загрузить видео в формате .mp4 или аудио файл в формате .mp3.

#### *Работа с тарифными планами*

- 1) Система должна предоставлять возможность просмотра и управление текущим тарифом.
- 2) В системе должна быть возможность покупки минут в рамках текущего тарифа.
- 3) Возможность апгрейда/даунгрейда плана.
- 4) Автоматическое списание средств по тарифу.
- 5) Система должна отслеживать потраченное количество минут конкретным пользователем на перевод видео/аудио и блокировать желание выйти за пределы тарифа.

#### *Уведомления и События*

Система должна уметь распознавать события и отправлять уведомления на почту по различным событиям (регистрация, готовность проекта, статусы платежей).

#### *Дополнительный контент и Поддержка*

- 1) FAQ, Team – разделы на сайте для дополнительной информации и поддержки пользователей;
- 2) Privacy Policy и Terms of Use страницы для юридической информации.

#### *Профиль и Настройки пользователя*

- 1) Сайдбар для навигации с основными пунктами:
  - a) Мои проекты.
  - b) Биллинг и подписка.
  - c) Настройки профиля и т.д.
- 2) Управление профилем:
  - a) изменение email;
  - b) изменение пароля.

#### *ML-core*

- 1) ML-core должен уметь переводить аудио/видео, делать транскрипт;
- 2) ML-core должен принимать на вход файлы MP4, MP3, WAV;
- 3) ML-core должен возвращать аудио- или видео- файл на целевом языке;
- 4) Голос в выходном файле должен совпадать с голосом во входящем файле либо с голосом, выбранным пользователем;
- 5) Формат результирующего файла должен соответствовать принятому формату.

### **ЧАТ-БОТ**

Система должна соответствовать следующим функциональным требованиям:

#### *1) Взаимодействие с Пользователем*

- Возможность подключения к Telegram боту авторизованного пользователя в Telegram.
- Возможность выбора языка перевода и голоса для озвучивания.
- Возможность загрузки видео и аудио файлов непосредственно в чат.
- Возможность отправки ссылок на медиа-контент с платформ YouTube, Vimeo, TikTok, Twitter, Google Drive.

#### *2) Работа с Текстовыми и Голосовыми Сообщениями*

- Автоматический перевод текстовых сообщений на выбранный язык.

- Озвучивание текстовых сообщений голосом собеседника или голосом по умолчанию при отсутствии голоса собеседника.
- Возможность озвучивания переписки в групповых чатах за выбранный период времени.
- Генерация кратких пересказов переписки в виде текста или аудио за выбранный период времени.

### *3) Обработка Медиафайлов*

- Транскрипция и перевод видео и аудио файлов.
- Озвучивание переведенного текста с функцией Voice Cloning.
- Генерация видео и аудио файлов с наложенной переведенной аудиодорожкой.

### *4) Управление и Настройки*

- Настройка пользовательских предпочтений для транскрибирования, перевода и озвучивания.
- Возможность включения и отключения бота в групповых чатах.
- Управление уведомлениями от бота.

### *5) Монетизация и Тарифные Планы*

- Возможность выбора различных тарифных планов пользователями.
- Опция для покупки дополнительных минут/услуг в рамках выбранного тарифа.
- Возможность просмотра и управления текущим тарифным планом.

### *6) Интеграция и Поддержка*

- Интеграция с ML-core для обработки аудио и видео.
- Предоставление FAQ и раздела поддержки пользователей.

### *7) Профиль Пользователя*

- Возможность управления профилем пользователя в рамках Telegram-бота (изменение настроек, персональной информации и т.д.).

## **НЕФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ К ПРОГРАММНОМУ ОБЕСПЕЧЕНИЮ**

Система должна соответствовать следующим нефункциональным требованиям:

- Надежность (Reliability):
  - должен быть обеспечен механизм восстановления после падения;
  - исключения, возникшие в коде, должны быть обработаны;
  - одновременно сервис должен выдерживать не менее 100 запросов в момент времени.
- Производительность (Performance):
  - 95%-квантиль времени запросов должен быть не больше 10 минут;
  - сервис должен выдерживать индексацию минимум 1.000 запросов в день;
  - запрос по переводу система должна выполнять в течение часа.
- Удобство сопровождения (Supportability):
  - ведение журнала логов, ведение документации.
- Безопасность (Security): доступ к сервису осуществляется по интернет-протоколу HTTP.

### *Организационные требования.*

- Еженедельный отчет по проделанной работе;
- Фиксация статуса в Notion.

### *Внешние требования.*

- Безопасное хранение персональных данных.

## ОГРАНИЧЕНИЯ ПРОЕКТИРОВАНИЯ

- 1) Клиентская часть для веб-платформы разработана на Next.JS;
- 2) Верстка интерфейса осуществляется с использованием css-фреймворка Tailwind CSS;
- 3) В качестве системы оплаты используется Stripe;
- 4) Для системы авторизации используется OAuth;
- 5) В качестве системы управления базами данных будет использоваться Firebase Firestore;
- 6) Для хранения файлов используется Firebase Storage;
- 7) TG-бот должен быть разработан на языке Python;
- 8) ML-pipeline должен быть разработан с использованием языка программирования Python (3.12 или выше) с использованием библиотек FastAPI и Pydantic;
- 9) Использование ML-моделей:
  - a) OpenAI's Whisper для транскрипции;
  - b) GPT-3.5-turbo для перевода;
  - c) Модели от Azure и 11labs для озвучивания;
  - d) Coqui.XTTS для клонирования голосов.

## Приложение 1. АНАЛИЗ КОНКУРЕНТОВ В ЧАСТИ ФУНКЦИОНАЛА

	SpeechMate	<a href="#">Video Dubber</a>	<a href="#">11labs</a>	<a href="#">rask.ai</a>
Возможность управлять проектами пользователя	Есть	Есть	Есть только для платной подписки и только для text to speech перевода	Есть
Количество языков	148	12	29	Более 130
Клонирование голоса	Есть	Нет (только встроенные голоса)	Нет (только встроенные голоса)	Есть
Возможность переводить аудио и видео	Есть	Нет (можно переводить только видео)	Есть	Есть
Возможность определения количества спикеров в видео и их озвучка разными голосами	Есть	Нет	Есть	Есть
Наличие телеграм-бота	Есть	Нет	Нет	Нет
Возможность подключать телеграм-бот к телеграм-чату и озвучивать его	Есть	Нет	Нет	Нет
Возможность генерации краткого пересказа переписки в телеграм-чате или аудио/видео	Есть	Нет	Нет	Нет
Максимальная длительность переведенного аудио/видео	Максимум 3500 минут в месяц, одно видео не более 90 минут	Максимум 1800 минут перевода в месяц	максимум 2400 минут перевода в месяц	Максимум 3000 минут в месяц
Максимальный размер аудио / видео	Нет ограничений	Нет ограничений	Одно аудио не более 50 MB, одно видео не более 100MB	Нет ограничений


Примечание: Чат-бот копирует функциональность сайта, немного её дополняя, поэтому анализ конкурентов общий

Таким образом, планируется, что наш продукт по функциональной наполненности и возможностям, предоставляемым пользователям, будет превосходить рассмотренных конкурентов. Кроме того, плюсом нашего проекта будет стабильность ML-системы, так как конкуренты, как правило, для решения задачи транскрибирования и перевода аудио/видео используют свои модели. Мы же будем использовать готовые модели глубокого обучения, произведенные и протестированные крупными компаниями, и выстраивать из них свой пайплайн обработки аудио и видео. Также наше решение будет отличаться своей масштабируемостью, так как будет возможность легко в любой момент заменить какую-либо из моделей в пайплайне на более крупную, эффективную и способную обрабатывать больше запросов.

Заказчик

 / Мажаев В.С.

Ответственный по проекту

 / Щербакова Е.А.

Дата: 25.11.2023