

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»
Факультет компьютерных наук
Департамент программной инженерии

SpeechMate: сайт и телеграм бот для транскрипции,
перевода и синтеза речи

Итоговый документ

Подп.
и
дата


Инв.
№
дубл.


Взам.
инв.
№


Подп.
и
дата

Инв.
№
подл

Исполнители:

 / Т.Д. Куликова /

 / К.С. Милорадова /

 / А.А. Насыхова /

 / Е.А. Щербакова /

«21» марта 2024 г.

Заказчик:

 /В.С. Мажаев /

ИП МАЖАЕВ ВЯЧЕСЛАВ СЕРГЕЕВИЧ

«21» марта 2024 г.

СОДЕРЖАНИЕ И СТРУКТУРА ДОКУМЕНТА

1. Введение.....	3
1.1. Наименование программы.....	3
1.2. Краткая характеристика области применения.....	3
2. Команда.....	4
2.1. Командные роли и распределение ответственности.....	4
2.2. Матрица компетенций.....	4
3. Управление проектом.....	6
3.1. Метрики проекта.....	6
3.2. Диаграмма Ганта.....	6
3.3. Использование систем управления проектом.....	6
3.4. Анализ рисков.....	7
4. Предметная область и описание продукта.....	8
4.1. Постановка проблемы.....	8
4.2. Анализ предметной области.....	8
4.3. Рамки и границы проекта.....	8
4.4. Назначение приложения.....	8
4.5. Новые предоставляемые функции.....	9
5. Описание потенциальных пользователей.....	10
5.1. Цель заказчика и ключевые стейкхолдеры.....	10
5.1.1. Матрица управления стейкхолдерами.....	10
5.2. Матрица ответственности.....	10
5.3. План управления коммуникациями.....	11
6. Анализ конкурентов.....	12
7. Функциональные требования.....	13
7.1. Требования к составу выполняемых функций.....	13
7.2. Требования к реализации.....	14
8. Нефункциональные требования к продукту.....	15
8.1. Требования к надежности.....	15
8.2. Требования к производительности.....	15
8.3. Требования к безопасности.....	15
8.4. Требования к удобству сопровождения.....	15
9. Описание архитектуры приложения и используемых технологий.....	16
Базовые.....	16
ML-сервисы.....	16
10. Первоначальный прототип.....	17
10.1. Прототипы ранней стадии планирования.....	17
10.2. Первые прототипы.....	18
11. Используемые технологии.....	19
12. Предложения по монетизации.....	20
13. Будущее развитие проекта.....	21
14. Отчеты о взаимодействии с заказчиком.....	23
15. Отзыв заказчика.....	24
16. Руководство пользователя.....	25
17. Список используемой литературы.....	35
ПРИЛОЖЕНИЕ А.....	37

1. Введение

1.1. Наименование программы

Наименование программы – «SpeechMate: сайт и телеграм бот для транскрипции, перевода и синтеза речи».

1.2. Краткая характеристика области применения

SpeechMate 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.

Сайт предоставляет пользователю следующие функции:

1. Создание транскрипта файла (видео или аудио);
2. Перевод содержимого файла (видео или аудио) на заданный язык.
3. Выбор одного из трех вариантов синтеза речи в выходном файле: с сохранением пола говорящего из источника, клонирование голоса из источника, выбор из списка доступных голосов.

Telegram-бот предоставляет пользователю дополнительные функции для обработки не только файла, но и чата, такие как:

1. Создание транскрипта переписки за выбранный период времени;
2. Озвучивание сообщений в групповом чате за выбранный период времени.

SpeechMate может быть использован для исследовательских и учебных целей, для создания контента и ведения международной коммуникации.

2. Команда

2.1. Командные роли и распределение ответственности

Распределение ролей по проекту представлено в следующей таблице.

Куликова Татьяна Дмитриевна	Менеджер проекта, дизайнер, frontend-разработчик
Милорадова Ксения Сергеевна	Аналитик, ML-инженер, тестировщик
Насыхова Анастасия Артемовна	Технический писатель, дизайнер, backend-разработчик
Щербакова Елизавета Александровна	Руководитель проекта, архитектор ПО, разработчик TG-бота

Распределение ответственности по проекту представлено в следующей таблице.

Руководство проектом	Щербакова Е. А.
Менеджер проекта	Куликова Т. Д.
Аналитика	Милорадова К. С.
Проектирование системы	Щербакова Е. А.
ML-пайплайн	Милорадова К. С.
Разработка дизайна сайта	Куликова Т. Д.
Frontend сайта	Куликова Т. Д.
Backend сайта	Насыхова А. А.
Telegram-бот	Щербакова Е. А.
Тестирование	Милорадова К. С.
Документация	Насыхова А. А.

2.2. Матрица компетенций

В следующей таблице представлена матрица компетенций разработчиков проекта. Уровень развития компетенций оценивается баллами по шкале от 0 до 3, где 3 означает полное владение навыком.

Группа компетенций	Компетенция	Куликова Татьяна	Милорадова Ксения	Насыхова Анастасия	Щербакова Елизавета
Backend	Умение проектировать и разрабатывать REST API	1	1	3	1
	Опыт интеграции с системами оплаты и другими сервисами, такими как Stripe, OAuth	1	1	3	2
	Умение работать с фреймворком веб-разработки FASTAPI и Pydantic	1	0	3	0
	Знание алгоритмов и структур данных	2	2	3	3
	Умение работать с базами данных и системами управления ими, в частности, с Firebase Firestore	3	3	3	2

Группа компетенций	Компетенция	Куликова Татьяна	Милорадова Ксения	Насыхова Анастасия	Щербакова Елизавета
Frontend	Умение работать с фреймворками и библиотеками фронтенда React	3	1	1	0
	Навыки веб-разработки (HTML, CSS, TypeScript)	3	1	1	0
	Опыт работы с анимацией и графическими эффектами	3	1	1	1
	Знание принципов адаптивного и кросс-браузерного дизайна	3	1	1	2
ML	Python	2	3	2	3
	Опыт работы с нейронными сетями и с библиотеками глубокого обучения (PyTorch / TensorFlow)	1	3	1	1
	Опыт работы с аудио- и видео-обработкой на Python	0	3	1	1
	Умение проводить оценку моделей и анализ результатов.	1	3	1	1
	Умение проводить анализ и выбор подходящих моделей для решения конкретной задачи	1	3	1	1
Telegram-бот	Опыт работы с библиотекой Telegram Bot API	0	1	0	3
	Понимание принципов интеграции с другими сервисами и API	2	2	3	3
	Опыт разработки сценариев для бота	0	1	0	3
Общие	Документация	2	2	3	2
	Git	2	2	2	2
	Умение писать понятный код	3	3	3	3

Сумма баллов по секциям:

Группа компетенций	Куликова Татьяна	Милорадова Ксения	Насыхова Анастасия	Щербакова Елизавета
Backend	8	5	15	8
Frontend	12	4	4	3
ML	5	15	6	7
Telegram-бот	2	4	3	9

3. Управление проектом

3.1. Метрики проекта

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:

1. Удовлетворенность пользователей – оценка по шкале от 1 до 5, основанная на пользовательских отзывах. Целевой показатель: средняя оценка не ниже 4.0.
2. 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.

3.2. Диаграмма Ганта

Диаграмма Ганта представлена по ссылке [21].

3.3. Использование систем управления проектом

Для управления проектом был выбран сервис Notion [10] из-за его широкого функционала, включающего инструменты для планирования задач, возможность создания заметок, комментирования и назначения задачи конкретному участнику проекта.

В проекте Notion была организована таблица с общим списком задач. Каждая задача описана следующими параметрами:

- Краткое описание
- Исполнитель
- Статус («Not started» / «In progress» / «Done» / «Canceled»)
- Номер спринта
- Модуль (Backend / Frontend / Telegram-bot / ML-pipeline)

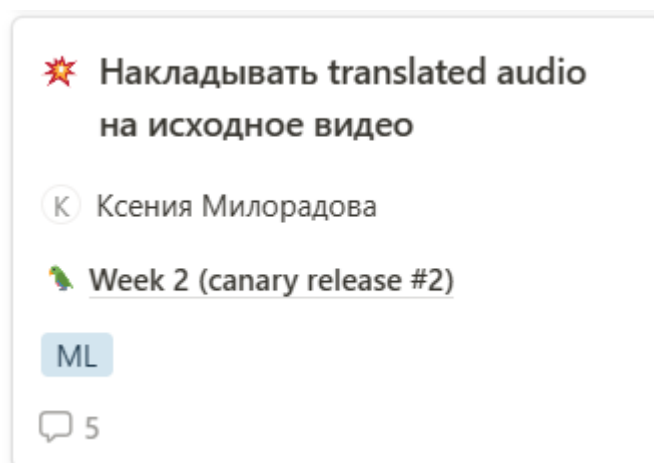


Рисунок 1 – Пример карточки с задачей в Notion

В Notion был организован календарь для фиксирования дат встреч с заказчиком и внутренних созвонов с командой. На страницах встреч для удобства фиксировались заметки по ним.

3.4. Анализ рисков

Риск	Описание	Вероятность	Влияние	Значение	Стратегия реагирования	Последствия
1. Нагрузка на учебе / основной работе	Участники проекта являются студентами выпускного курса и часть из них трудоустроены, что может вызвать проблемы с планированием	0,7	0,9	0,63	Уклонение	Сдвиг сроков выполнения задач по проекту
2. Непредвиденные обстоятельства	Различные форс-мажоры, болезнь кого-то из участников, семейные обстоятельства, поломка техники и другие факторы, мешающие участникам вести работу по проекту	0,1	0,9	0,09	Принятие	Сдвиг сроков выполнения задач по проекту
3. Загруженность заказчика	Заказчик может не найти время для обсуждения возникших у команды вопросов по функционалу	0,8	0,8	0,64	Принятие	Сдвиг сроков выполнения задач по проекту
4. Блокировка tg-бота	Поскольку Telegram заблокирован Роскомнадзором, могут возникнуть проблемы с запуском на русском удаленном сервере	0,7	0,9	0,63	Принятие	Необходимость искать подходящий проху для запуска бота на сервере

Уровни вероятности	Уровни серьезности последствий				
	Низкий (0,1-0,2)	Незначительный (0,3-0,4)	Средний (0,5-0,6)	Высокий (0,7-0,8)	Критический (0,9-1)
Часто (0,9-1)					
Возможно (0,7-0,8)				Риск №3	Риски №1 и №4
Редко (0,5-0,6)					
Маловероятно (0,3-0,4)					
Невозможно (0,1-0,2)					Риск №2

4. Предметная область и описание продукта

4.1. Постановка проблемы

В настоящее время часто возникает необходимость обращаться к контенту на иностранных языках. В частности, для просмотра образовательных материалов, создания своего собственного блога, ведения бизнеса и общения в социальных сетях. Однако многие люди не владеют иностранными языками в совершенстве, поэтому для преодоления языкового барьера им приходится пользоваться переводчиками. В то же время переводчики не предоставляют такой функционал как автоматический перевод текста из аудио- и видео- файлов. Это приводит к тому, что пользователям приходится самостоятельно воспринимать текст на слух, чтобы ввести в переводчике, что довольно проблематично и практически невозможно, если человек не знает языка текста. Кроме того, при общении на иностранных языках часто требуется не только перевести сообщение собеседника, но и прослушать его в ситуации, когда нет времени на чтение большого количества чатов.

4.2. Анализ предметной области

Предметной областью проекта является сфера разработки сервисов с использованием искусственного интеллекта для обработки и трансформации речи. В подобных сервисах информация из файла сначала проходит через транскрибацию¹ для получения текстового представления речи, а после переводится на нужный язык и накладывается на исходный аудио- или видеофайл.

Целью разработки таких приложений является масштабирование контента путем перевода речи на другой язык, а также упрощение коммуникации.

4.3. Рамки и границы проекта

Проект будет использовать внешние системы для своей работы:

- Stripe [15] в качестве системы оплаты;
- OAuth [11] в качестве системы авторизации;
- Firebase Firestore [4] в качестве системы управления базами данных;
- Для хранения файлов и медиа будет использоваться Firebase Storage [5];
- TelegramBotAPI [18] для реализации чат-бота;
- Github [6] в качестве системы контроля версий;

Границы проекта:

- Тестирование на пользователях в данном проекте не предусмотрено.
- По завершении работы команда разработчиков передает заказчику весь исходный код и документацию, включая руководство пользователя.
- Деплой проекта будет осуществлен только после полного согласия и одобрения заказчиком, а также после успешной защиты проекта.
- На данный момент не планируется дальнейшая поддержка продукта, однако это может быть обсуждено и решено в будущем.

4.4. Назначение приложения

SpeechMate, в первую очередь, предназначен для преодоления языкового барьера физическими лицами, которые регулярно взаимодействуют с аудио- и видео- контентом на разных языках.

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);

¹ Транскрибация – процесс расшифровки информации из аудио- и видеозаписи с целью дальнейшего преобразования в текст.

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:
 - a. 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.

4.5. Новые предоставляемые функции

1. Разработанная система включает в себя две платформы – веб-сайт и телеграмм-бот, что не имеют наши конкуренты.
2. Телеграм-бот дает возможность озвучивать переписки в тг-чатах с помощью клонирования голоса

5. Описание потенциальных пользователей

5.1. Цель заказчика и ключевые стейкхолдеры

Заказчик преследует несколько целей:

- Привлечение международной аудитории к продуктам компании заказчика;
- Получение дохода от монетизации разрабатываемого продукта;
- Увеличение популярности бренда компании заказчика и укрепление ее репутации на рынке.

Основными заинтересованными сторонами (стейкхолдерами) являются:

- Заказчик: предприниматель, который хочет с помощью команды разработчиков создать и продвинуть качественный продукт, который поможет ему выйти на международный рынок, получить доход и улучшить репутацию своей компании.
- Исполнители: лица, разрабатывающие продукты по заданию заказчика. Они заинтересованы в получении прибыли за выполнение проекта и получении отличной оценки по дисциплине в ВУЗе..
- Пользователи: основные пользователи системы. Заинтересованы в преодолении языкового барьера, затрудняющего их деятельность, в возможности быстро перевести необходимое аудио или видео на нужный язык, а также в удобстве общения с иностранцами в социальных сетях.

5.1.1. Матрица управления стейкхолдерами

Стейкхолдеры	Заинтересованность (-5..+5)	Влияние (0..5)
Заказчики	+5	5
Исполнители	+5	4
Пользователи	+3	0

5.2. Матрица ответственности

Задачи	Мажаев Вячеслав Сергеевич	Куликова Татьяна	Милорадова Ксения	Насыхова Анастасия	Щербакова Елизавета
	Заказчик	Дизайнер, Фронтенд-разработчик, менеджер	Аналитик, ML-разработчик Тестировщик	Бекенд-разработчик, дизайнер, технический писатель	Руководитель проекта, разработчик tg-бота, архитектор
Проведение технических интервью	A	I	I	R	I
Разработка бизнес-требований	A	I	I	R	I
Определение технических требований	C	R	R	R	R
Планирование проекта	C	R	I	I	I
Проектирование архитектуры системы	I	I	I	I	R
Разработка дизайна интерфейса сайта	I	A	-	R	-
Разработка BackEnd	I	I	I	R	-

Разработка FrontEnd	I	R	-	I	-
Разработка ML-пайплайна	I	-	R	I	I
Разработка TG-бота	I	-	I	-	R
Тестирование системы	I	I	R	I	I
Исправление ошибок	I	R	R	R	R
Разработка технической документации	I	-	-	R	-
Оценка полученных результатов	R	I	I	I	I

5.3. План управления коммуникациями

В начале каждой недели проводились онлайн-встречи с заказчиком в Google Meet (Sprint Planning). В процессе таких встреч демонстрировались результаты выполнения задач предыдущего спринта и исходя из этого формировался список задач на следующий спринт. Длительность каждого спринта – 1 неделя. Все поставленные задачи фиксировались на специальной доске в Notion. К каждой задаче прикреплялся исполнитель, исполнитель мог оставлять заметки по выполнению задачи, а также менять статус ее выполнения прямо в интерфейсе Notion. По мере возникновения проблем и вопросов заказчик устраивал дополнительные созвоны, на которые приглашались те члены команды, чьи задачи на тот момент были взаимосвязаны между собой и требовали обсуждения возникших проблем и синхронизации между членами команды. Во время каждой встречи велась запись, все записи сохранялись в специальном разделе в Notion, чтобы члены команды, не присутствовавшие на встрече, могли посмотреть ее запись и быть постоянно вовлеченными в процесс разработки. Также большая часть коммуникаций велась в рабочем чате команды в Телеграмме [17].

6. Анализ конкурентов

Сравнительный анализ конкурентов² представлен в следующей таблице.

Характеристика	SpeechMate	Video Dubber [19]	11labs [1]	rask.ai [14]
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
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

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

² Чат-бот копирует функциональность сайта, немного её дополняя, поэтому анализ конкурентов общий.

7. Функциональные требования

7.1. Требования к составу выполняемых функций

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.

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.

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.

3) Обработка Медиафайлов

- Транскрипция и перевод видео и аудио файлов.
- Озвучивание переведенного текста выбранным голосом 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.

7.2. Требования к реализации

Языки программирования и фреймворки/библиотеки:

- Для серверной части вебсайта – Next.JS [9];
- Для интерфейса вебсайта – Tailwind CSS [16];
- Для tg-бота – Python [13] и TelegramBotAPI [18];

Для совместной разработки системы была использована система контроля версий git. Разработка ML-пайплайна, веб- приложения и tg-бота проводилась в отдельных закрытых репозиториях на платформе Github [6].

8. Нефункциональные требования к продукту

8.1. Требования к надежности

Система должна соответствовать следующим требованиям к надежности (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.

8.2. Требования к производительности

Система должна соответствовать следующим требованиям к производительности (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.

8.3. Требования к безопасности

Система должна соответствовать следующим требованиям к безопасности (Security):

- доступ к сервису осуществляется по интернет-протоколу HTTP.

8.4. Требования к удобству сопровождения

Система должна соответствовать следующим требованиям к удобству сопровождения (Supportability):

- ведение журнала логов, ведение документации.

9. Описание архитектуры приложения и используемых технологий

Архитектура разрабатываемой системы является микросервисной. Такая архитектура основана на построении системы, в котором основная функциональность разделена на небольшие, автономные сервисы, каждый из которых выполняет свою специфическую задачу. Каждый микросервис может быть разработан, развернут и масштабирован независимо от других сервисов.

В системе SpeechMate сервисы можно подразделить на 2 типа: ML-сервисы и базовые.

Базовые

1. *Сервис авторизации OAuth [11]*. Отвечает за регистрацию и аутентификацию пользователей в системе, выполняет функции gateway системы;
2. *Сервис оплаты Stripe [15]*. Отвечает за оформление платной подписки на сайте и в телеграм-боте;
3. *Сервис Generator*. Отвечает за интеграцию всех ML-сервисов в одну API, а также за обработку запросов пользователей.
4. *Сервис TelegramBotAPI [18]*. Отвечает за выгрузку сообщений из чата.

ML-сервисы

1. *Сервис OpenAI Whisper [12]*. Отвечает за создание транскрипта файла и чата в текст.
2. *Сервис GPT-3.5-turbo [7] от OpenAI*. Отвечает за перевод текста на заданный язык.
3. *Сервис MicrosoftTTS [8]*. Отвечает за генерации речи по заданному тексту на заданном языке (более 80 вариантов).
4. *Сервис 11labs [1]*. Отвечает за генерацию речи по заданному тексту на заданном языке (предоставляет широкий спектр различных голосов).
5. *Сервис Coqui XTTS [2]*. Отвечает за генерацию речи по заданному тексту на заданном языке с клонированием голоса.
6. *Сервис для определения пола говорящего*.

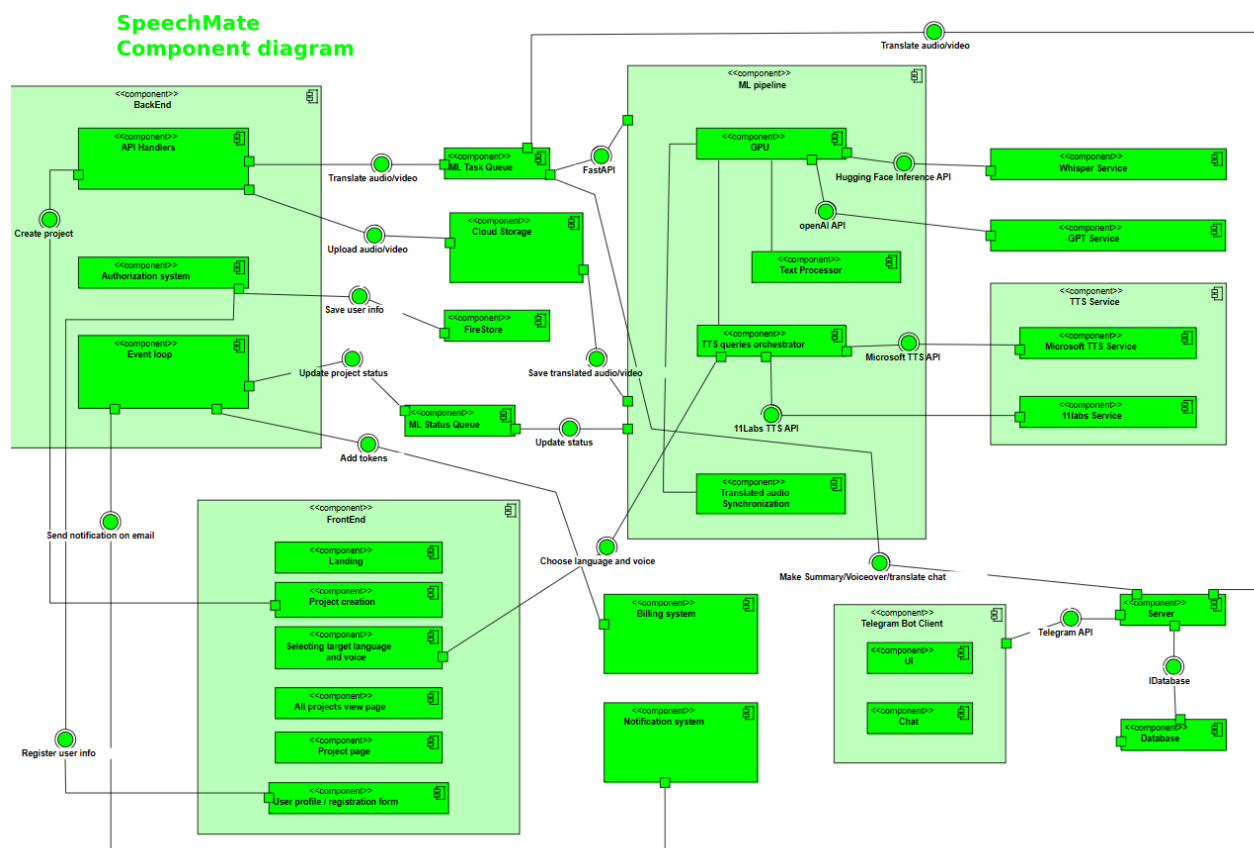


Рисунок 2 – Диаграмма компонентов

10. Первоначальный прототип

10.1. Прототипы ранней стадии планирования

На первой итерации работы над проектом дизайнерами был разработан прототип сайта в графическом редакторе Figma [3].

Первый вариант дизайна был максимально прост, в нем отсутствовали элементы навигации, но было сформировано общее расположение объектов, а также определена цветовая гамма.

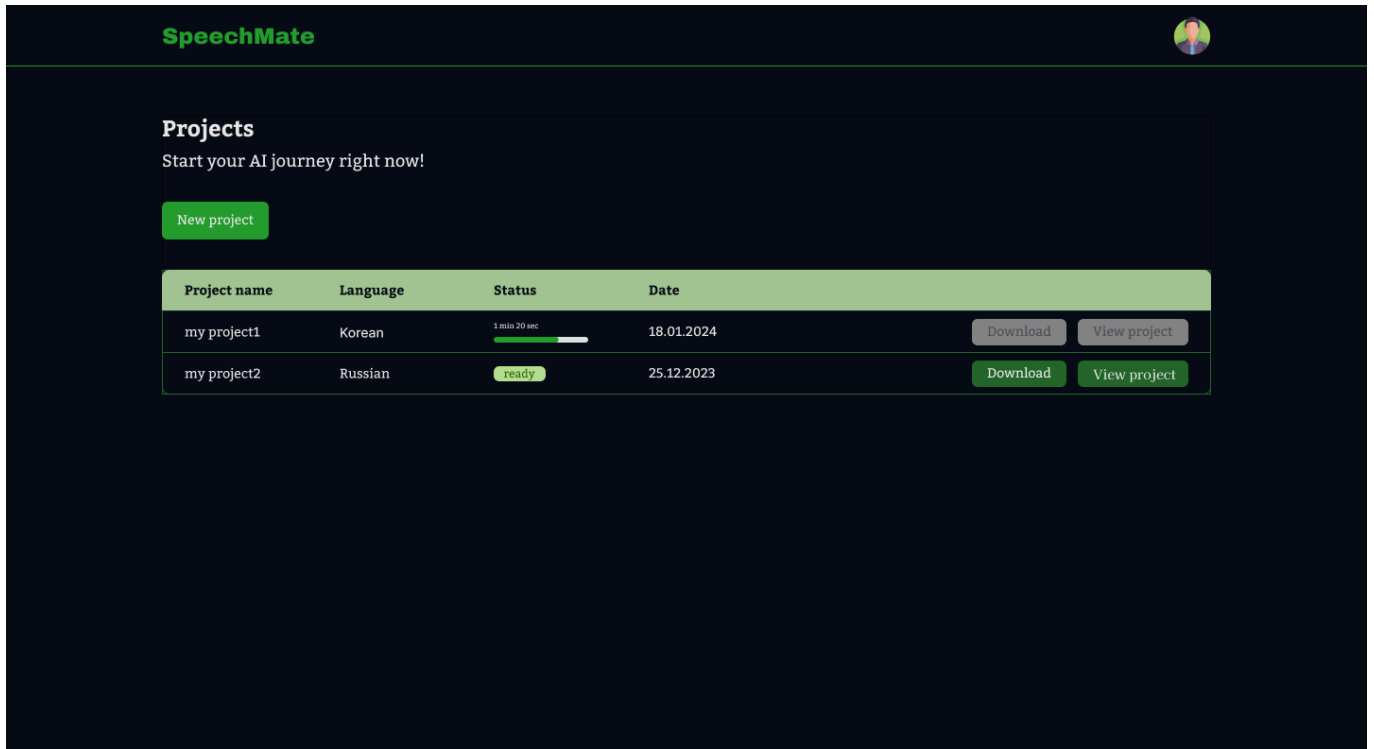


Рисунок 3 – Изначальный вид окна проектов

На второй итерации добавилось боковое меню, а также изменилась верхняя панель.

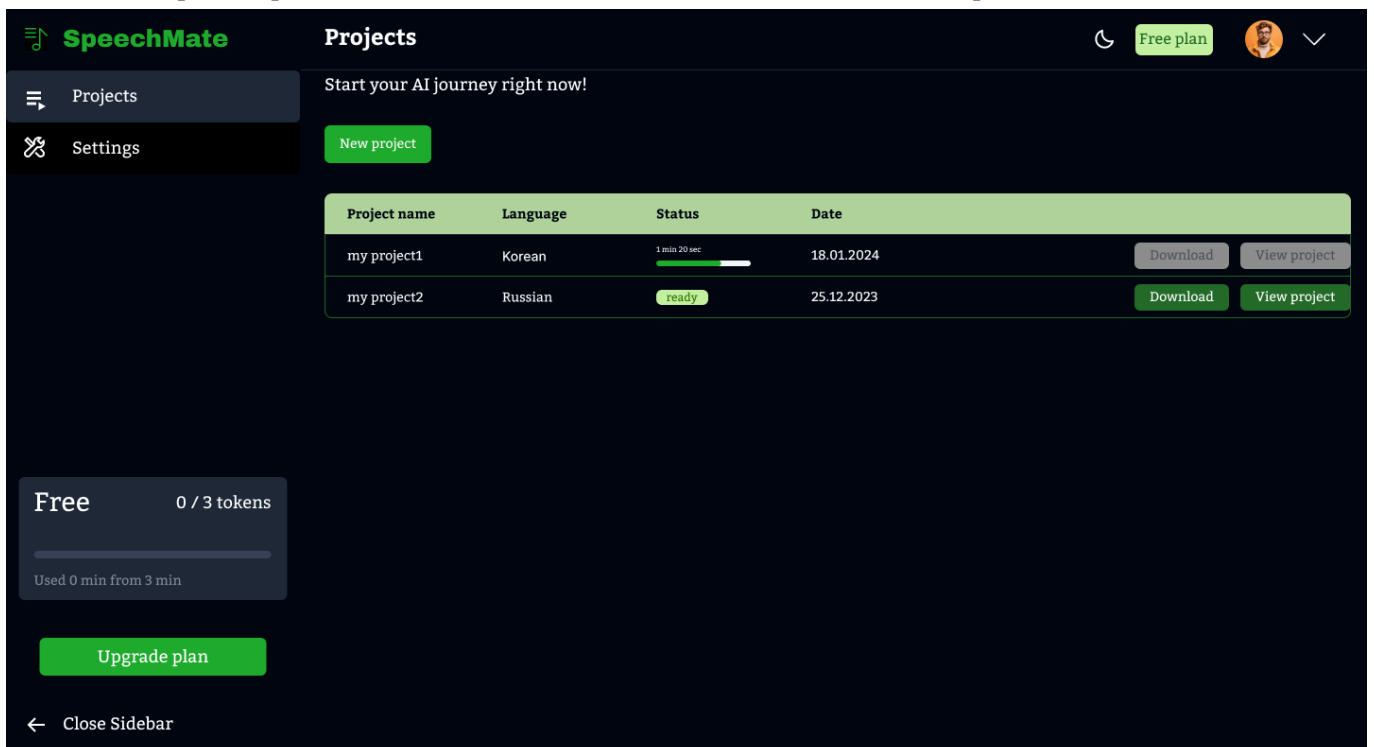


Рисунок 4 – Усовершенствованный вид окна проектов

Итоговый дизайн-макет сайта доступен по ссылке [22].

10.2. Первые прототипы

Web-сайт

После утверждения дизайна web-сайта backend и frontend разработчиками были разработаны его основные страницы. Третья итерация включала в себя:

- 1. Главную страницу:**
 - a. Описание функционала сервиса.
 - b. Презентация основных возможностей.
 - c. Призыв к действию (регистрация, загрузка видео/аудио).
- 2. Страницу регистрации:**
 - a. Форма регистрации пользователя.
- 3. Страницу загрузки видео/аудио:**
 - a. Интерфейс для загрузки файлов.
 - b. Выбор языка для перевода.
- 4. Страницу с результатами перевода:**
 - a. Скачивание переведенного файла.

Чат-бот

Первый прототип чат-бота SpeechMate был разработан с целью предоставить пользователям возможность быстро и удобно переводить видео. Он включал в себя следующие функции:

- 1. Загрузка видео:**
 - a. Пользователи могли отправить видео в чат-бот.
 - b. Поддерживались различные форматы видео.
- 2. Указание языка перевода:**
 - a. Пользователи могли выбрать язык, на который хотели перевести видео.
 - b. Поддерживались различные языки.
- 3. Скачивание видео с переводом:**
 - a. После завершения перевода пользователи могли скачать видео с озвучкой на выбранном языке.

11. Используемые технологии

Модуль	Используемые технологии
Backend	Серверная часть для веб-платформы разработана на Next.JS; В качестве системы оплаты используется Stripe, гарантируя надежность и безопасность финансовых транзакций; Для системы авторизации используется OAuth; Система управления базами данных должна реализована с использованием Firebase Firestore; Для хранения файлов используется Firebase Storage;
Frontend	Верстка интерфейса осуществляется с использованием css-фреймворка Tailwind CSS;
ML-пайплайн	ML-пайплайн разработан с использованием языка Python (версия 3.12 или выше) и библиотек FastAPI и Pydantic. 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.
Telegram-бот	Разработка Telegram-бота велась на языке Python;

12. Предложения по монетизации

С помощью сервиса Stripe в проект была встроена гибкая система тарифных планов, предоставляющая пользователям возможность выбрать оптимальное количество токенов в соответствии с их потребностями. Один токен дает возможность загрузить один файл в формате .mp4 или .mp3, wav.

Тариф Free	Тариф Creator	Тариф Standard	Тариф Producer	Тариф Enterprise
- 3 токена;	- 10 токенов в месяц;	- 30 токенов в месяц;	- 300 токенов в месяц;	- Более 1000 токенов в месяц;
- Максимальная длина видео – 1 минута.	- Максимальная длина видео – 1 минута.	- Максимальная длина видео – <u>5 минут.</u>	- Максимальная длина видео – <u>20 минут.</u>	- Максимальная длина видео – <u>90 минут.</u>

13. Будущее развитие проекта

1. Расширение функциональности:

1.1. Интеграция с API сторонних сервисов:

- 1.1.1. Добавление возможности загрузки видео с YouTube, Vimeo и других платформ.
- 1.1.2. Поддержка субтитров на разных языках и синхронизация переводов с субтитрами.

1.2. Редактирование видео:

- 1.2.1. Обрезка видео.
- 1.2.2. Добавление водяных знаков.
- 1.2.3. Склейка видео.

1.3. Расширенные возможности перевода:

- 1.3.1. Озвучивание видео на основе текста и выбранного голоса
- 1.3.2. Озвучивание видео на основе текста и аудио дорожки, сгенерированной на основе загруженного образца голоса

1.4. Создание субтитров:

- 1.4.1. Автоматическая генерация субтитров.
- 1.4.2. Редактирование субтитров.
- 1.4.3. Синхронизация субтитров с переводом.

1.5. Аналитика:

- 1.5.1. Отслеживание использования функций.
- 1.5.2. Сбор данных о пользователях.
- 1.5.3. Анализ эффективности переводов.

2. Улучшение интерфейса:

2.1. Мультиязычность:

- 2.1.1. Поддержка интерфейса на разных языках.
- 2.1.2. Автоматическое определение языка пользователя.
- 2.1.3. Улучшение системы оповещений:
- 2.1.4. Уведомления о завершении перевода.
- 2.1.5. Оповещения о новых функциях.
- 2.1.6. Персонализация оповещений.

2.2. Развитие сообщества:

2.2.1. Создание форума:

- 2.2.1.1. Обсуждение вопросов, связанных с переводом видео/аудио.
- 2.2.1.2. Обмен опытом использования сервиса.
- 2.2.1.3. Получение помощи от других пользователей.

2.3. Техническое развитие:

2.3.1. Масштабирование:

- 2.3.1.1. Увеличение производительности и пропускной способности.
- 2.3.1.2. Поддержка растущего числа пользователей.

2.4. Безопасность:

- 2.4.1. Защита от хакерских атак.

3. Исследования и разработки:**3.1. Разработка новых алгоритмов перевода:**

- 3.1.1. Повышение точности и качества перевода.
- 3.1.2. Сокращение времени перевода.

3.2. Изучение новых технологий:

- 3.2.1. Поиск новых способов улучшения сервиса.
- 3.2.2. Разработка новых функций.

14. Отчеты о взаимодействии с заказчиком

Наименование документа	Дата составления	Ссылка
ПРОТОКОЛ ОБСУЖДЕНИЯ ЗАДАЧИ С ЗАКАЗЧИКОМ 1. Основные задачи 2. Конкуренты 3. Пользователи 4. Требования к инструментам и технологиям 5. Общее описание функциональности 6. Результат работы 7. КПЭ оценки результатов	30.11.2023	[25]
ТЕХНИЧЕСКОЕ ЗАДАНИЕ 1. Функциональные требования 2. Нефункциональные требования 3. Ограничения проектирования	30.11.2023	[28]
КОМАНДА ПРОЕКТА 1. Распределение ролей 2. Распределение ответственности 3. Таблица задач по проекту	30.11.2023	[23]
АРХИТЕКТУРА ПРОГРАММНОГО РЕШЕНИЯ 1. Описание модели жизненного цикла 2. Описание используемых паттернов проектирования 3. Описание типа архитектуры	15.01.2024	[20]
ПРОЕКТИРОВАНИЕ ПРОГРАММНОГО РЕШЕНИЯ 1. Модель данных 2. Диаграмма классов 3. Диаграмма компонентов 4. Entity-relationship диаграмма 5. Диаграмма прецедентов 6. Диаграмма деятельности 7. Диаграмма потоков данных 8. Диаграмма последовательности	15.01.2024	[26]
БАГ-РЕПОРТ И АКТ-ПРИЁМКИ 1. Описание выбранных методов тестирования 2. Требования, предъявляемые к тестированию 3. Выявленные баги 4. Приемочное тестирование	13.03.2024	[27]

15. Отзыв заказчика

Работа была завершена вовремя. Функционал за исключением пары моментов реализован в полном объеме. Не реализованы опции клонирования голосов на сайте и генерации краткого пересказа в телеграм-боте.

В процессе обсуждения проекта на первоначальном этапе было обговорено не реализовывать часть опций телеграм-бота на сайте в связи с отличающимися потребностями целевых аудиторий приложений. Это решение было зафиксировано во всех документах.

Система работает стабильно и удовлетворяет все потребности пользователей на первом этапе внедрения.

Студенты подходили к заданиям с полной ответственностью. О результатах был регулярный отчет, который снабжался анализом и интерпретацией. Во время работы предлагались собственные решения к возникающим задачам и пути дальнейшего развития проекта.

Разработанные студентами продукты полностью соответствуют требованиям, зафиксированным в техническом задании. Они успешно создали сайт и телеграм-бот, которые предоставляют пользователям возможность масштабировать контент путем перевода речи на другой язык. Это отражает их хорошее понимание предметной области и способность эффективно применять технологии разработки.

Полный отзыв заказчика доступен по ссылке [24].

16. Руководство пользователя

Chat-бот

Начало работы

Добавление бота: Найдите бота в Telegram по его имени пользователя (@SpeechMateBot) и нажмите "Start" или "Начать" для активации.

Переход по ссылке: Вы также можете активировать бота, перейдя напрямую по ссылке: t.me/SpeechMateBot. Эта ссылка откроет чат с ботом в вашем Telegram.

Основное меню: После активации бота вы увидите приветственное сообщение и основное меню с доступными опциями. Используйте кнопки для навигации.

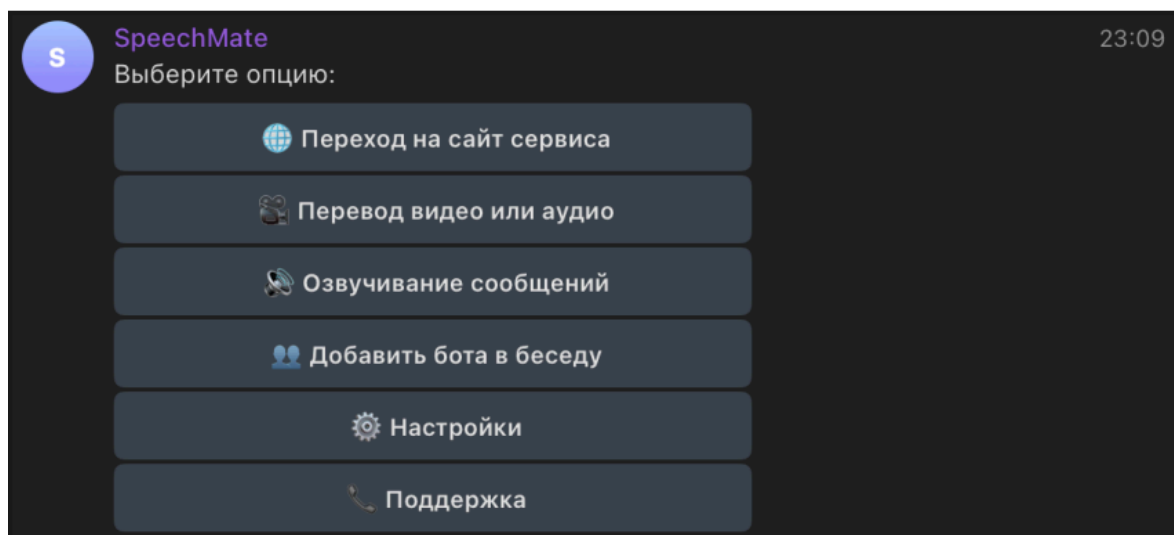


Рисунок 5 – Стартовое окно. Телеграм-бот

Функции бота

Переход на сайт

Кнопка "Перейти на сайт": Нажмите эту кнопку, если хотите перейти на официальный сайт проекта. Бот предложит открыть ссылку в браузере по умолчанию.

Озвучивание сообщений

Кнопка "Озвучивание сообщений": Нажмите, чтобы озвучить текстовые сообщения. После нажатия перешлите в бота сообщения, которые необходимо озвучить. Этот набор обязан содержать для каждого автора хотя бы одно голосовое сообщение, содержащие запись речи, длиной не менее 6 секунд для обучения.

- **Выбор языка:** После отправки сообщений вам будет предложено выбрать язык перевода из списка.
- **Результат:** ожидайте получения результата в течение нескольких минут.
-

Перевод медиа

Кнопка "Перевести медиа": Нажмите, чтобы перевести видео или аудиофайл. После нажатия отправьте медиафайл, который нужно перевести.

Выбор языка: После отправки файла вам будет предложено выбрать язык перевода из списка.

Выбор количества голосов: Затем выберите, сколько голосов вы хотите использовать для перевода.

Выбор голосов: Вы можете выбрать голоса вручную или позволить боту сделать это автоматически.

Результат: ожидайте получения результата в течение нескольких минут.

Добавление в групповой чат

Кнопка "Добавить в групповой чат": Данная функция предоставит вам подробную инструкцию по настройке бота в групповом чате.

Поддержка

Кнопка "Поддержка": Если у вас есть вопросы или вам нужна помощь, нажмите эту кнопку. Бот предоставит вам ссылку на бота поддержки. Перейдя в новый бот, нажмите кнопку "Start" или "Начать", затем напишите ваш запрос, через некоторое время команда поддержки оставит вам сообщение в этом же боте.

Озвучивание сообщений в групповом чате

Команда /voice: Используйте эту команду в групповом чате для озвучивания последних n текстовых сообщений. Например, отправьте /voice 5, чтобы озвучить последние 5 сообщений.

- **Результат:** ожидайте получения результата в течение нескольких минут.

Настройки бота

Кнопка "Настройки": Нажмите, чтобы изменить настройки бота, включая язык интерфейса.

- **Смена языка:** В меню настроек будет предложено выбрать язык интерфейса бота. На данный момент доступны два языка: английский и русский. Выбрав желаемый язык, нажмите соответствующую кнопку, и интерфейс бота автоматически изменится на выбранный язык

Web-сайт

При входе на сайт отображается главная страница-лендинг. На верхней части страницы слева направо расположены ссылки для перехода на главную страницу (SpeechMate), в блог (Blog), на страницу с тарифами (Pricing), страницу вопросов и ответов (FAQ), а также кнопка для переключения темы между светлой и темной (действие кнопки будет продемонстрировано уже со следующего скриншота), кнопка для входа и регистрации.

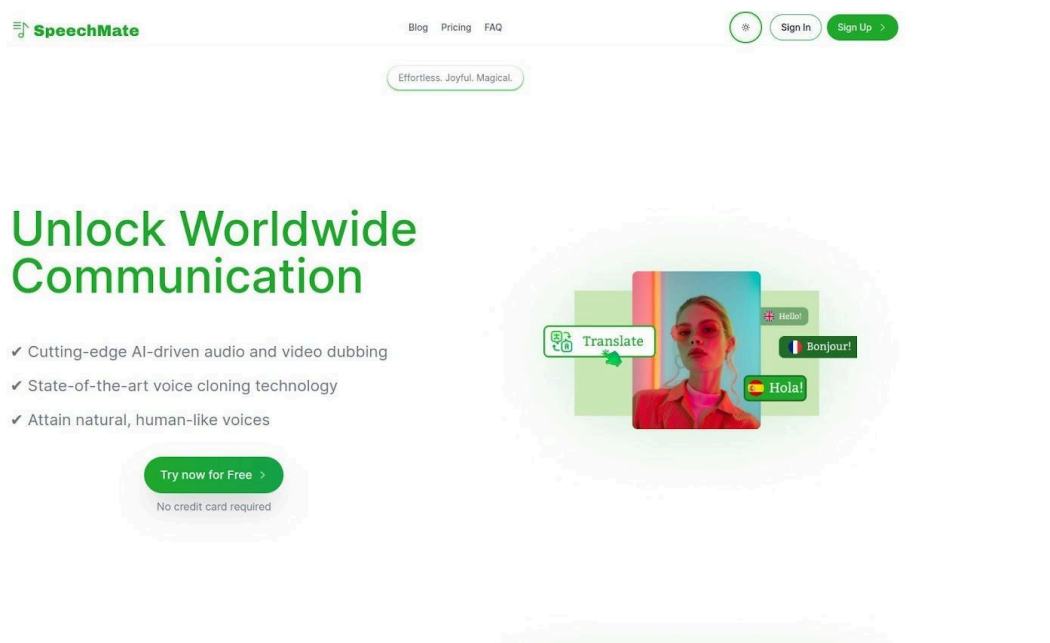


Рисунок 6 – Стартовая страница. Светлая тема

При нажатии кнопки смены темы всплывает контекстное меню с тремя выборами: светлая, тёмная, системная.

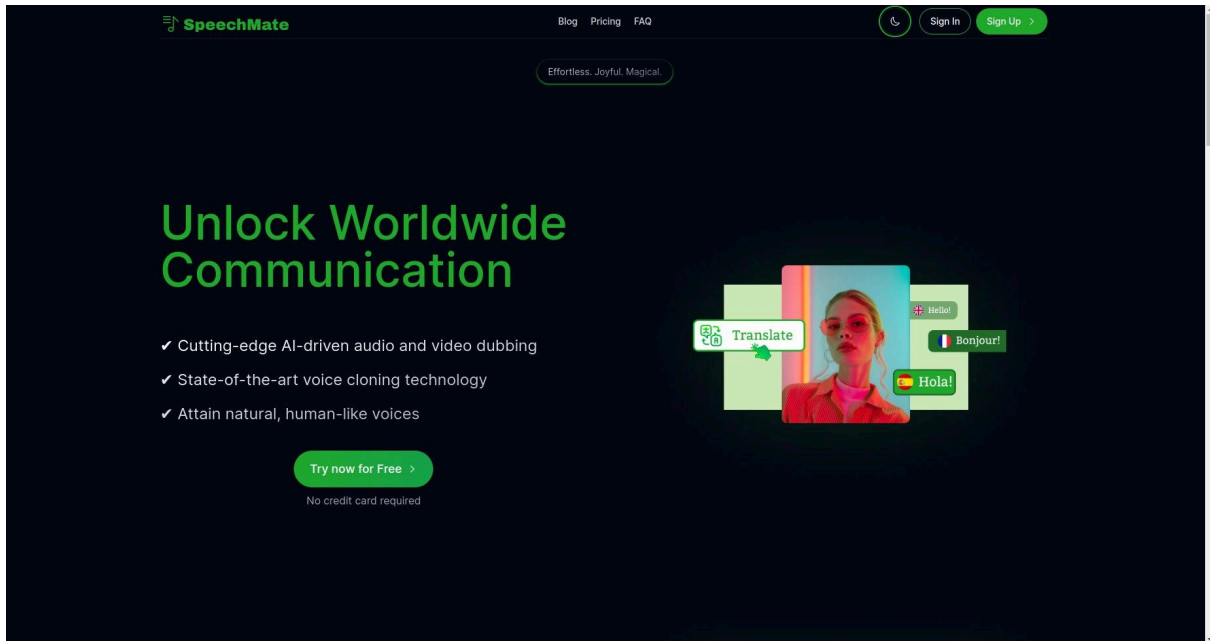


Рисунок 7 – Стартовая страница. Темная тема

Пролистав страницу вниз можно увидеть текстовое описание и видеообзор продукта, далее – отзывы и предложение попробовать продукт.

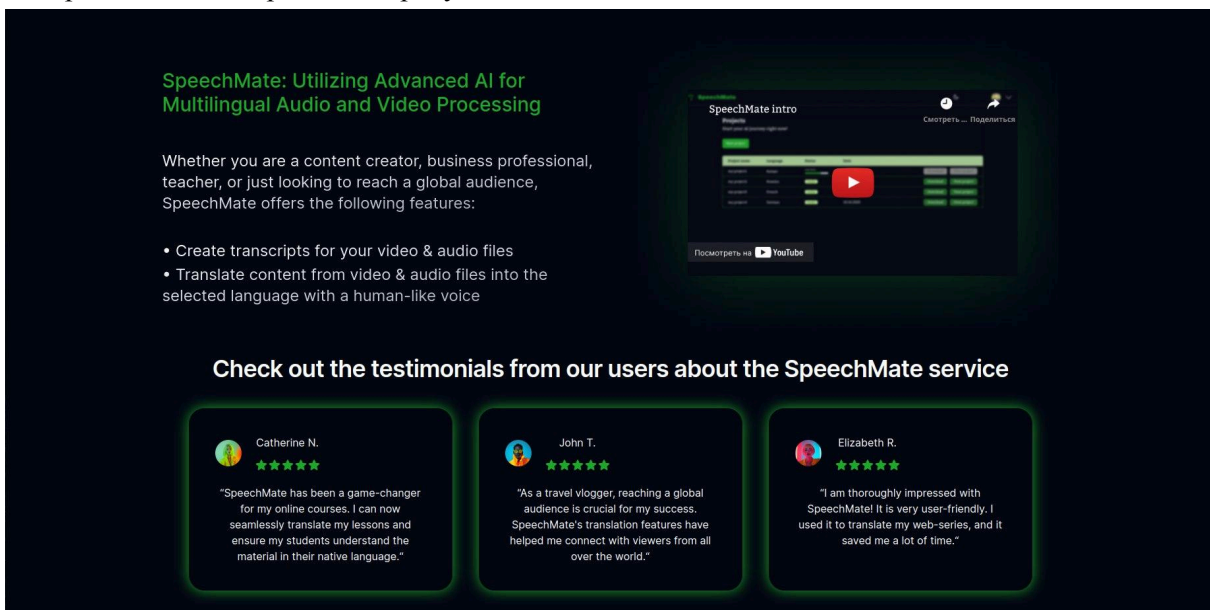


Рисунок 8 – Описание продукта и отзывы пользователей

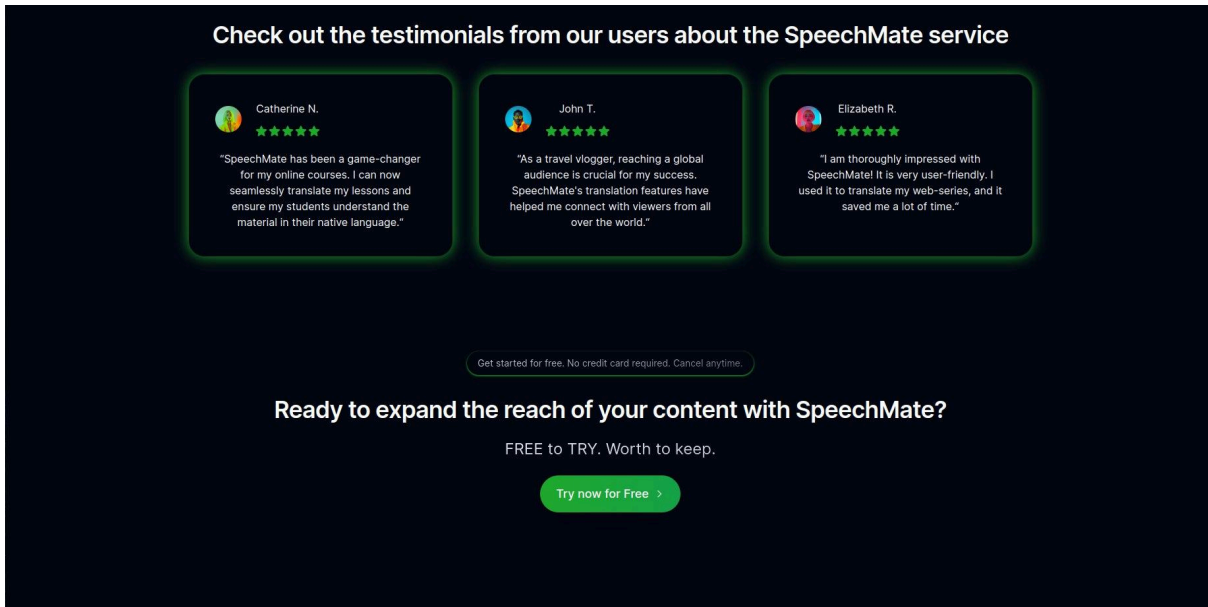


Рисунок 9 – Отзывы и предложение зарегистрироваться

По кнопке Blog осуществляется переход на окно со статьями об областях искусственного интеллекта, задействованных в реализации проекта. Статьи в будущем будут пополняться.

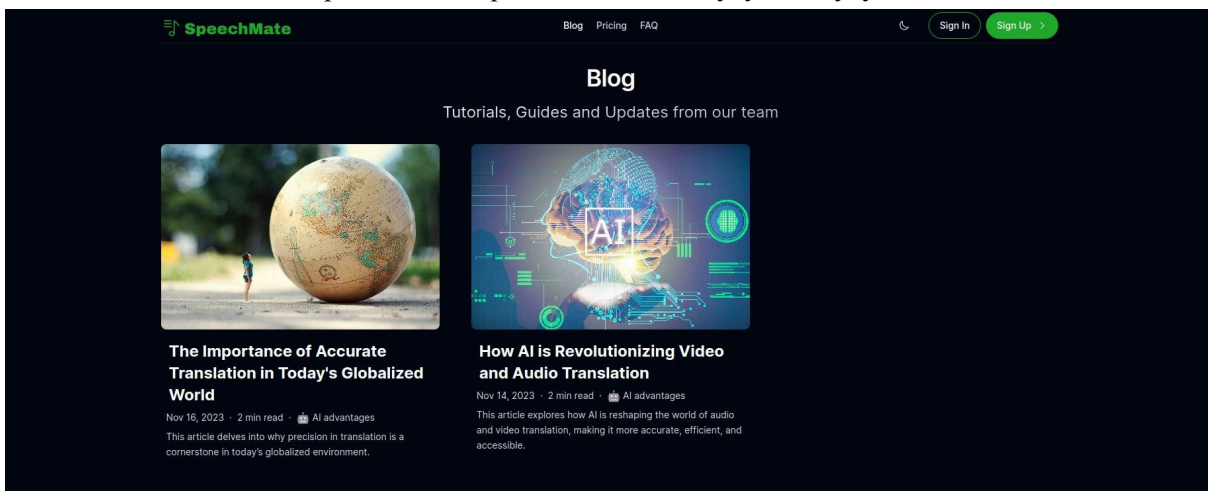


Рисунок 10 – Блог

Можно перейти в статью, кликнув по ней.



Рисунок 11 – Статья в блоге

По кнопке на Тарифы (Pricing) в верхней части страницы осуществляется переход на страницу с биллингом.

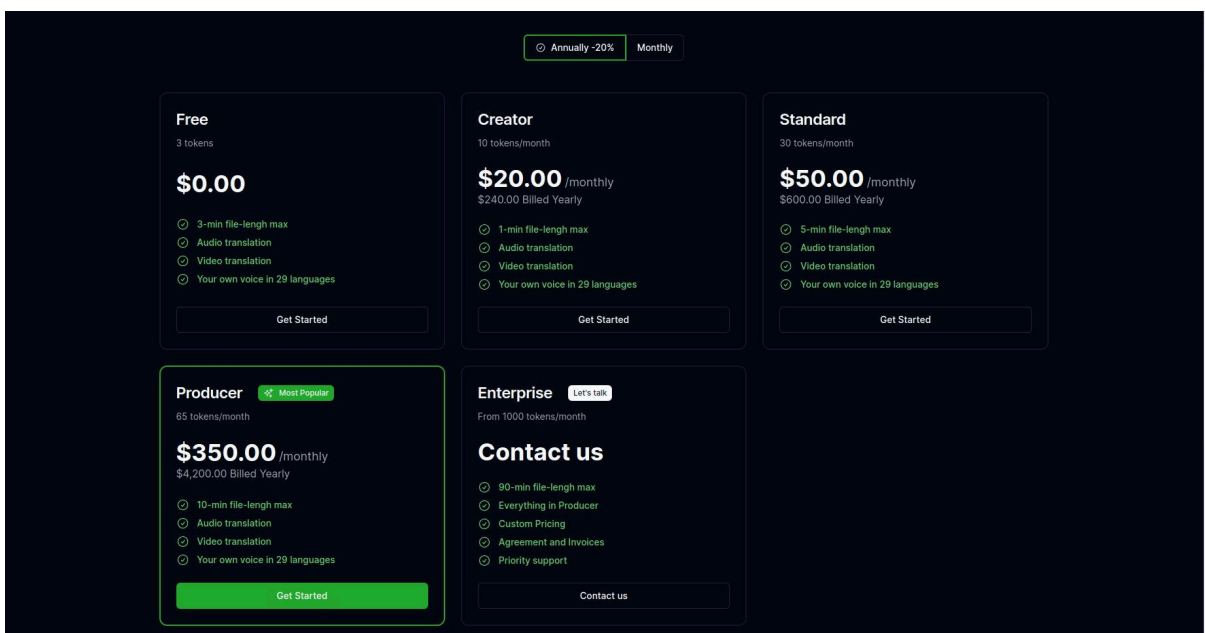


Рисунок 12 – Тарифные планы

По кнопке FAQ осуществляется переход в раздел с ответами на самые часто задаваемые вопросы.

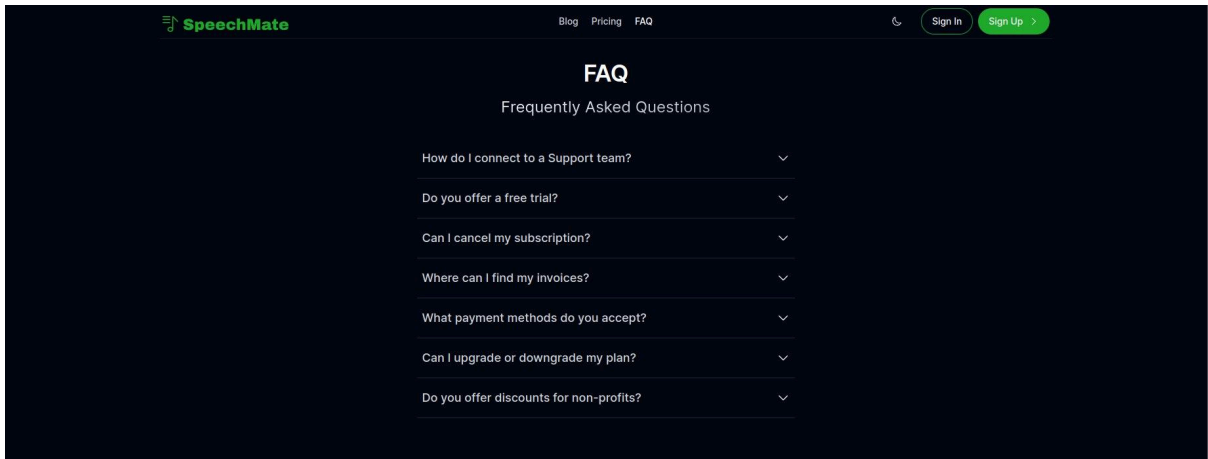


Рисунок 13 – FAQ

Для регистрации необходимо нажать кнопку Sign Up в верхнем меню сайта. Можно ввести почту и пароль или зарегистрироваться через уже существующую учетную запись Google.

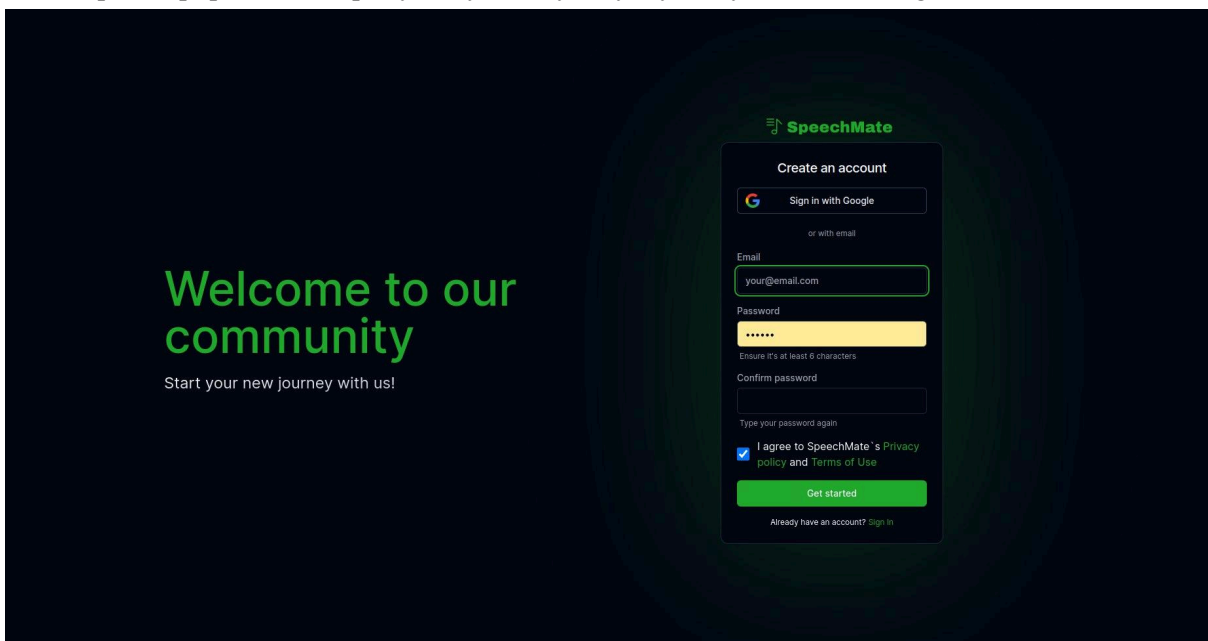


Рисунок 14 – Окно регистрации

Для авторизации нужно нажать кнопку Sign In и ввести логин и пароль или войти при помощи учетной записи Google.

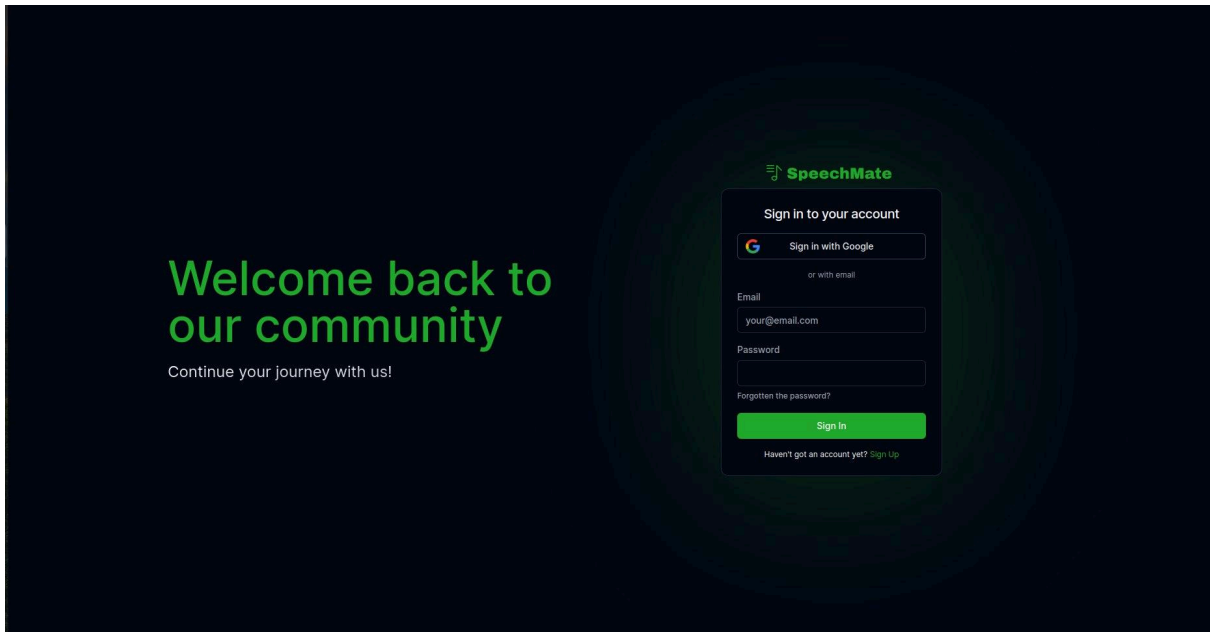


Рисунок 15 – Окно авторизации

Кнопки Sign Up и Sign In заменяются на текущий план и аватар пользователя. Справа от аватара пользователя переход в меню. Меню содержит имя аккаунта и ссылки на страницу проектов пользователя и Настройки, а также кнопку выхода из учетной записи.

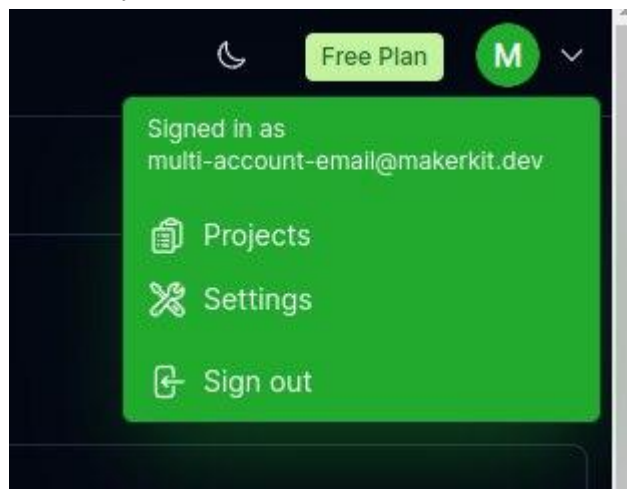


Рисунок 16 – Меню

При переходе на страницу Проекты отображается экран с кнопкой создания проекта.

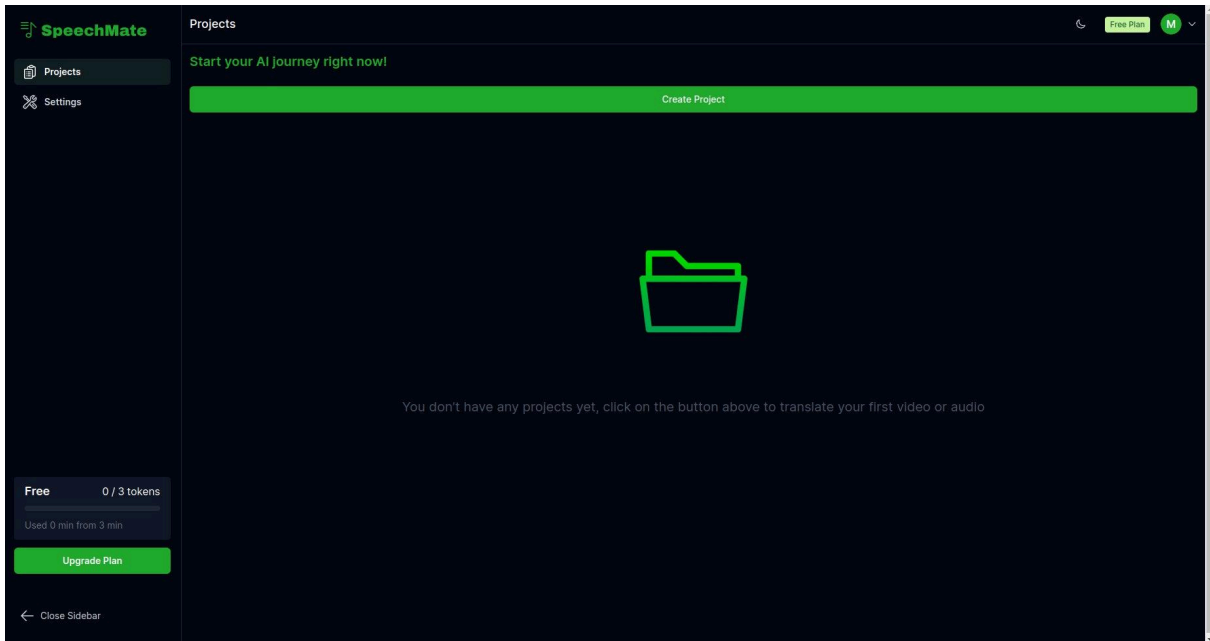


Рисунок 17 – Окно проектов

При нажатии на кнопку появляется всплывающий баннер с полями ввода имени проекта, выбором целевого языка, голосом для озвучивания, количеством говорящих в видео и окном для загрузки файла.

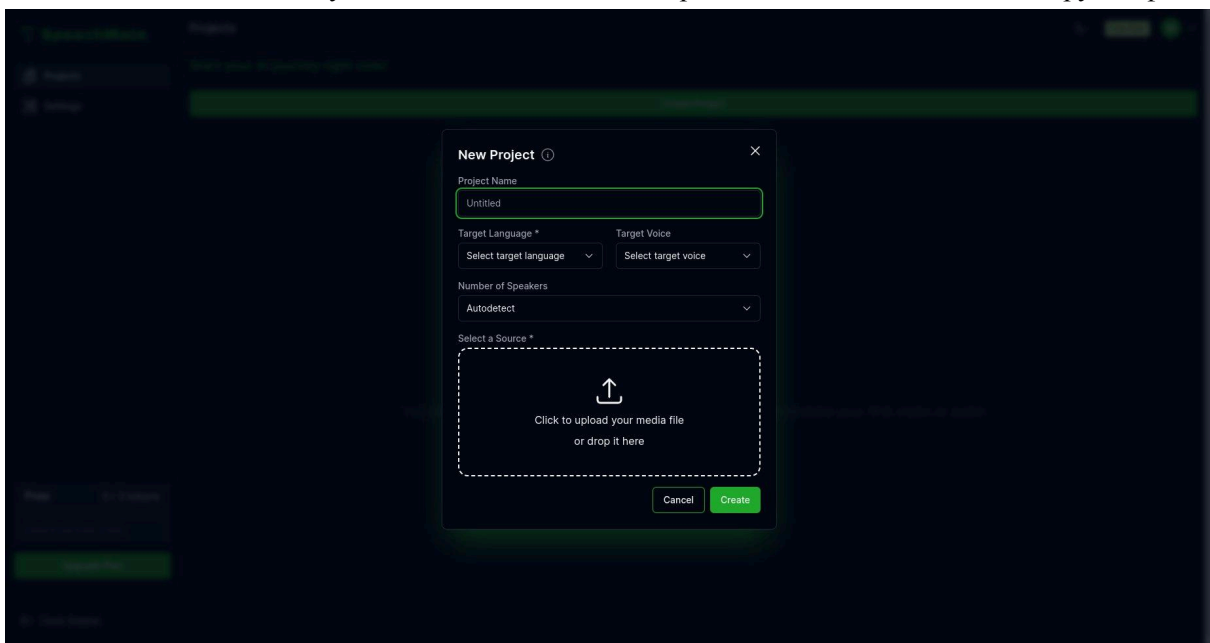


Рисунок 18 – Окно создания проекта

После заполнения полей и загрузки файла следует нажать кнопку Create.

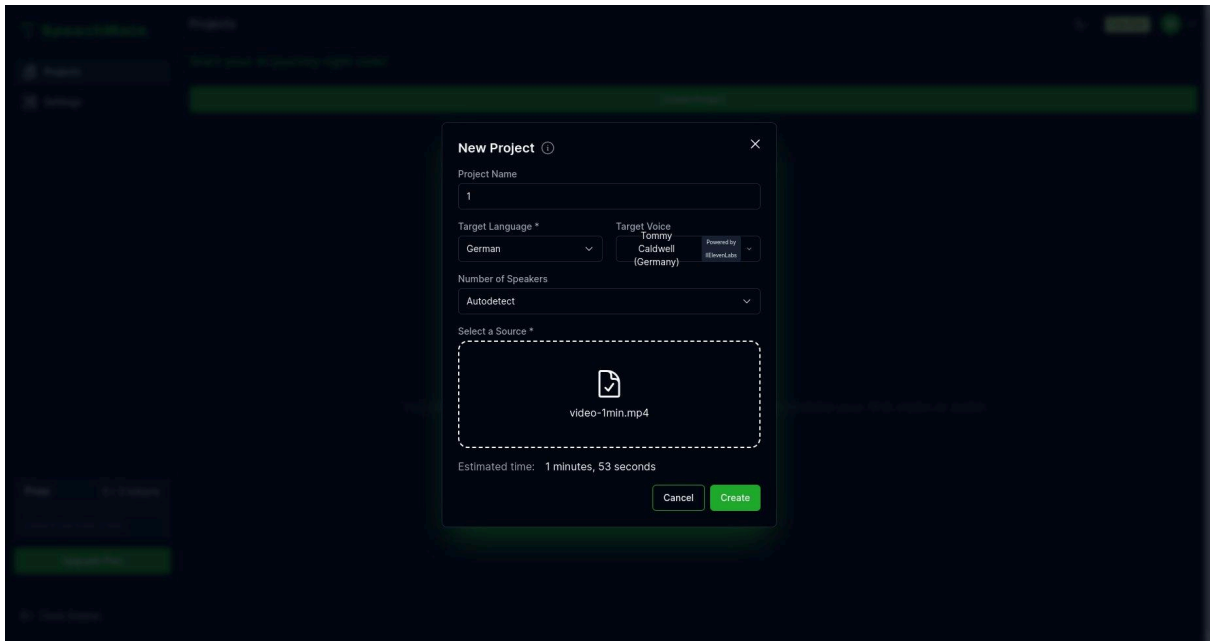


Рисунок 19 – Окно создания проекта после заполнения необходимых полей

Теперь проект появился в списке проектов и по окончании обработки результирующее видео можно скачать по кнопке Download или посмотреть результат прямо на сайте по кнопке View.

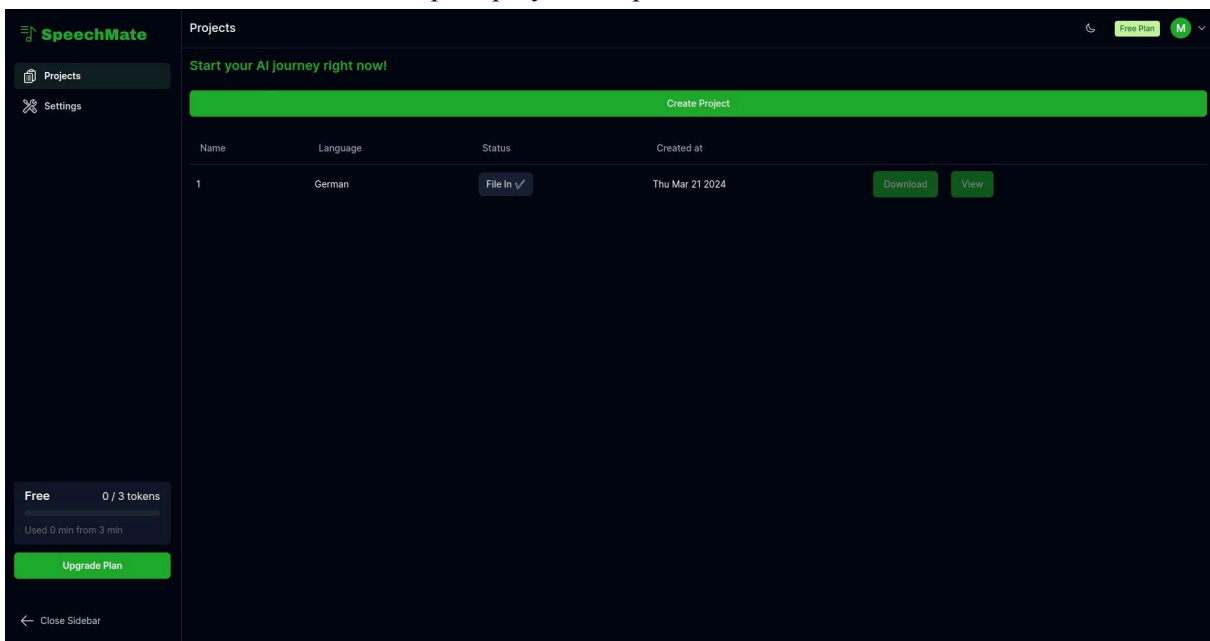


Рисунок 20 – Окно проектов после загрузки проекта

При переходе на страницу настроек (Settings) появляется окно редактирования профиля для изменения имени, аватара, электронной почты или пароля пользователя.

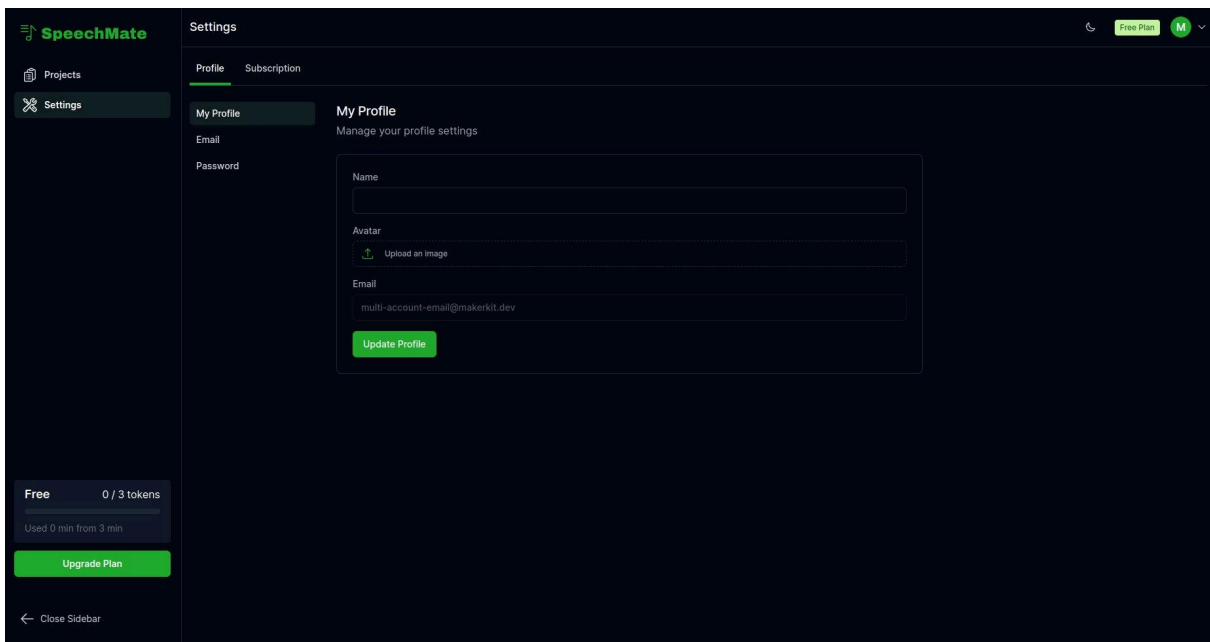


Рисунок 21 – Окно редактирования профиля

Приятного пользования!

17. Список используемой литературы

1. 11labs / [Электронный ресурс] // URL: <https://elevenlabs.io/dubbing> – Режим доступа: свободный
2. Coqui.XTTS / [Электронный ресурс] // URL: <https://huggingface.co/coqui/XTTS-v2> – Режим доступа: свободный
3. Figma / [Электронный ресурс] // URL: <https://www.figma.com/> – Режим доступа: свободный
4. Firebase Firestore / [Электронный ресурс] // URL: <https://firebase.google.com/docs/firestore?hl=en> – Режим доступа: свободный
5. Firebase Storage / [Электронный ресурс] // URL: <https://firebase.google.com/docs/storage?hl=en> – Режим доступа: свободный
6. GitHub / [Электронный ресурс] // URL: <https://github.com/> – Режим доступа: свободный
7. GPT-3.5-turbo / [Электронный ресурс] // URL: <https://openai.com/blog/gpt-3-5-turbo-fine-tuning-and-api-updates> – Режим доступа: свободный
8. MicrosoftTTS / [Электронный ресурс] // URL: <https://azure.microsoft.com/en-us/products/ai-services/text-to-speech> – Режим доступа: свободный
9. NextJS / [Электронный ресурс] // URL: <https://nextjs.org/> – Режим доступа: свободный
10. Notion / [Электронный ресурс] // URL: <https://www.notion.so/product> – Режим доступа: свободный
11. OAuth / [Электронный ресурс] // URL: <https://oauth.net/> – Режим доступа: свободный
12. OpenAI Whisper / [Электронный ресурс] // URL: <https://openai.com/research/whisper> – Режим доступа: свободный
13. Python / [Электронный ресурс] // URL: <https://www.python.org/doc/> – Режим доступа: свободный
14. Rask.AI / [Электронный ресурс] // URL: <https://www.rask.ai/> – Режим доступа: свободный
15. Stripe / [Электронный ресурс] // URL: <https://stripe.com/> – Режим доступа: свободный
16. Tailwind CSS / [Электронный ресурс] // URL: <https://tailwindcss.com/> – Режим доступа: свободный
17. Telegram / [Электронный ресурс] // URL: <https://telegram.org/> – Режим доступа: свободный
18. TelegramBotAPI / [Электронный ресурс] // URL: <https://core.telegram.org/> – Режим доступа: свободный
19. VidoDubber / [Электронный ресурс] // URL: <https://videodubber.ai/> – Режим доступа: свободный
20. Архитектура программного решения / [Электронный ресурс] // URL: https://drive.google.com/file/d/1dOpBz0ER6rNo97jKExFOGi_wleKNj2aJ/view?usp=sharing – Режим доступа: свободный
21. Диаграмма Ганта / [Электронный ресурс] // URL: <https://app.ganttpro.com/shared/token/d15260efe14cc3e63dcf5259d36fcea70ffa958c45c6eec378e0b2a71c4b183e/1386909> – Режим доступа: свободный
22. Дизайн в Figma / [Электронный ресурс] // URL: <https://drive.google.com/file/d/10dSeqm0ikusdUdwiNkOzTijOrFS0oXrF/view?usp=sharing> – режим доступа свободный
23. Команда проекта / [Электронный ресурс] // URL: https://drive.google.com/file/d/1ij6QUhnXGKTK1NwLe5uzqvaVjIZRGJ_f/view?usp=sharing – Режим доступа: свободный
24. Отзыв заказчика / [Электронный ресурс] // URL: <https://drive.google.com/file/d/1DfqGRWgaOLUtRSgOVHNq4NJjv3CabbNR/view?usp=sharing> – Режим доступа: свободный
25. Постановка задачи с заказчиком / [Электронный ресурс] // URL: <https://drive.google.com/file/d/145R2zr5YXyGIVQ4p5dMumsp808DkkXoh/view?usp=sharing> – Режим доступа: свободный
26. Проектирование программного решения / [Электронный ресурс] // URL: <https://drive.google.com/file/d/1V85nyA0MQdkmuq6l1tYsuT-Xqbx2iHJM/view?usp=sharing> – Режим доступа: свободный

27. Тестирование / [Электронный ресурс] // URL:

<https://drive.google.com/file/d/1055Ua3QZkh-CVMCAUA3KJ8moQUY2kffB/view?usp=sharing> – Режим доступа: свободный

28. Техническое задание / [Электронный ресурс] // URL:

https://drive.google.com/file/d/1B8LdL6jJ6rShIqFNaCgYtt_Et-oDLRD/view?usp=sharing – Режим доступа: свободный

Ссылки на код разрабатываемых продуктов

Ссылка на открытый репозиторий на Github с кодом сайта:

<https://github.com/kultattiana/speechmate-website>

Ссылка на открытый репозиторий на Github с кодом ML-пайплайна:

<https://github.com/Ksmiloradova/speechmate-ml?ysclid=lu274k16rh293370001>

Ссылка на открытый репозиторий на Github с кодом Telegram-бота:

https://github.com/lizashcherbakova/speech_mate