



Faculty of Computer Science

Software Engineering

Moscow, 2024

SpeechMate: сайт и телеграмм-бот для транскрипции, перевода и синтеза речи

Заказчик

ИП Мажаев Вячеслав Сергеевич,
Мажаев Вячеслав Сергеевич

Выполнили студенты:

Куликова Татьяна Дмитриевна БПИ206
Милорадова Ксения Сергеевна БПИ206
Насыхова Анастасия Артемовна БПИ203
Щербакова Елизавета Александровна БПИ201

Цели и задачи 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.

Задачи:

- Проведение технических интервью с заказчиком;
- Планирование проекта с помощью диаграммы Ганта;
- 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;





Распределение ролей

ФИО	Роли
Куликова Татьяна Дмитриевна	Разработчик дизайна сайта, разработчик Frontend, менеджер проекта,
Милорадова Ксения Сергеевна	Разработчик ML-пайплайна, тестировщик, аналитик
Насыхова Анастасия Артемовна	Разработчик Backend, разработчик технической документации
Щербакова Елизавета Александровна	Руководитель проекта, разработчик tg-бота, архитектор



Управление проектом

Коммуникация с заказчиком и между членами команды:

- Онлайн-встречи с заказчиком в Google Meet (Sprint Planning)
- Длительность спринта - 1 неделя
- Дополнительные созвоны между членами команды по мере возникновения проблем
- Во время встреч велась запись

Использование систем управления проектом:

- Для управления проектом был выбран сервис Notion
- Была организована таблица с общим списком задач
- Создан календарь для фиксирования дат встреч с заказчиком и внутренних созвонов с командой



Пример карточки с задачей в Notion

Анализ конкурентов



	SpeechMate	Video Dubber	11labs	rask.ai
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

Архитектура приложения

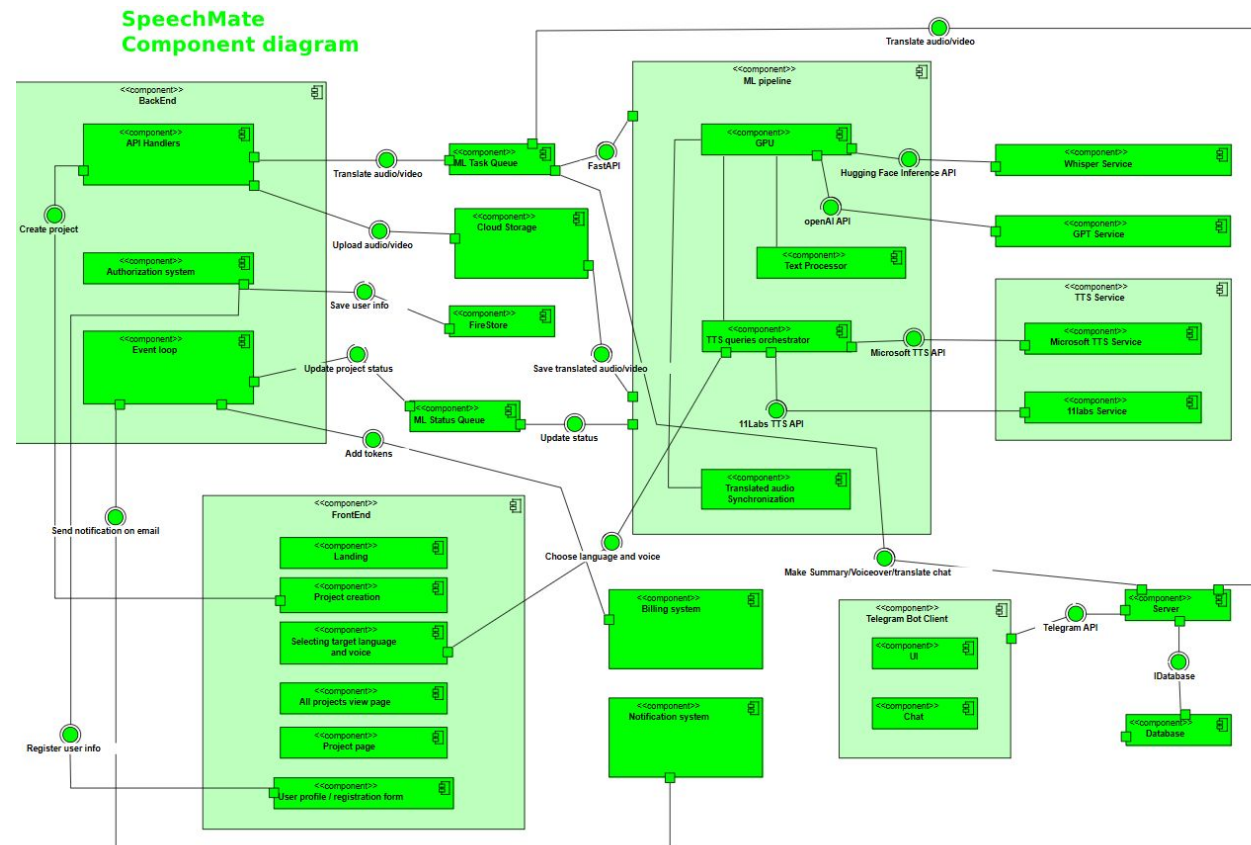


Диаграмма компонентов



Frontend

Исполнитель - Куликова Т. Д.

- Лендинг
- Страницы регистрации / авторизации
- Страница с таблицей созданных проектов
- Страница создания нового проекта
- Страница просмотра, прослушивания и загрузки переведенных аудио/видео
- Страница настроек профиля пользователя
- Страница с тарифными планами
- Страницы Blog, FAQ
- Страницы Privacy Policy, Terms of Use
- Основные обработчики событий взаимодействия с интерфейсом и изменения статуса проекта
- Реализация двух цветовых схем

SpeechMate

Blog Pricing FAQ

Sign In Sign Up

Effortless. Joyful. Magical.

Unlock Worldwide Communication

- ✓ Cutting-edge AI-driven audio and video dubbing
- ✓ State-of-the-art voice cloning technology
- ✓ Attain natural, human-like voices

Try now for Free

No credit card required



SpeechMate

Blog Pricing FAQ

Sign In Sign Up

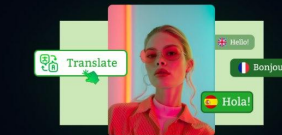
Effortless. Joyful. Magical.

Unlock Worldwide Communication

- ✓ Cutting-edge AI-driven audio and video dubbing
- ✓ State-of-the-art voice cloning technology
- ✓ Attain natural, human-like voices

Try now for Free

No credit card required



Доступны две цветовые схемы - светлая и темная



Frontend

Исполнитель - Куликова Т. Д.

New Project ⓘ

Project Name

4

Target Language *

French

Target Voice

Leona Adams (France)

Powered by ElevenLabs

Number of Speakers

Autodetect

Select a Source *

video-1min.mp4

Estimated time: 1 minutes, 53 seconds

Cancel

Create

Страница создания проекта

SpeechMate

Projects

Settings

Free0 / 3 tokens

Used 0 min 21 sec from 3 min

Upgrade Plan

Close Sidebar

Projects

Start your AI journey right now!

Create Project

Name	Language	Status	Created at	
2	French	Translation Complete ✓	Wed Mar 27 2024	<div>DownloadView</div>
1	French	Translation Complete ✓	Wed Mar 27 2024	<div>DownloadView</div>

Таблица проектов

Полный дизайн интерфейса в фигме [1]







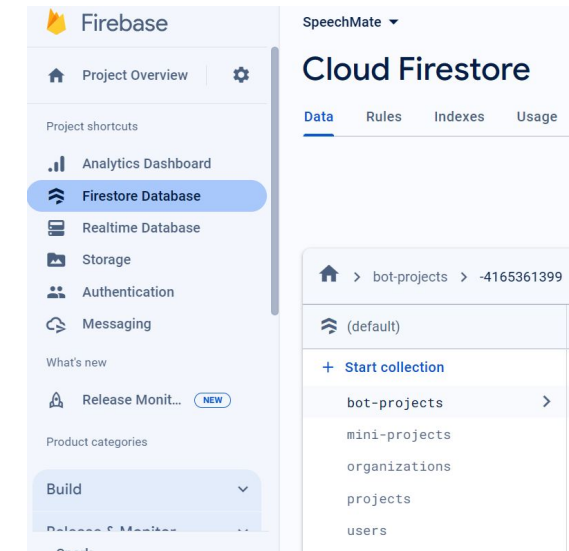
Backend

Исполнитель - Насыхова А. С.

- Регистрация / Авторизация / Выход из аккаунта
- Регистрация с Google
- Восстановление пароля / Обновление пароля
- Обновление настроек профиля
- Создание проекта, Обновление статуса проекта
- Фильтрация голосов по языку
- Обертка пайплайна в FastAPI
- Интеграция с Firebase Firestore, Storage, Stripe, OAuth, ML-пайплайн
- Создание формы в Typeform

Products

NAME	
 Creator	\$20.00 USD / month, \$192.00 USD / year
 Standard	\$50.00 USD / month, \$480.00 USD / year
 Free	\$0.00 USD / month, \$0.00 USD / year
 Producer	\$350.00 USD / month, \$3,360.00 USD / year



My workspace / SpeechMate

Content

- 1 Hello, what's your company name or...
- 2 First and Last name of contact person
- 3 Contact email
- 4 Contact phone number
- 5 How many hours of audio/video do...

Create Connect Share Results



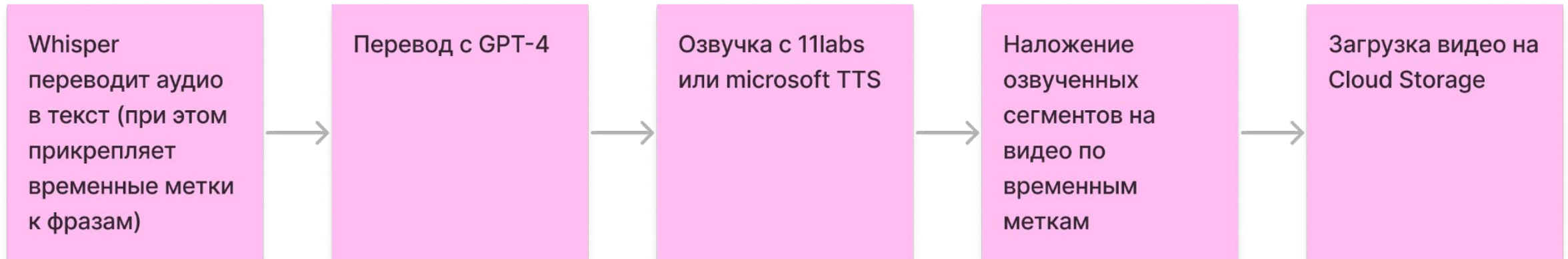
ML-pipeline

Исполнитель - Милорадова К. С.

- Распознавание речи с помощью модели OpenAI Whisper
- Разработка алгоритма дробления текста на чанки перед подачей в GPT
- Запрос к GPT на перевод текста
- Обработка переведенного текста перед подачей в модель для озвучки (Microsoft TTS, 11labs)
- Разработка алгоритма наложения аудио-перевода на видео

Модели:

- OpenAI Whisper [2]
- GPT [3]
- Microsoft TTS [4]
- 11Labs [5]

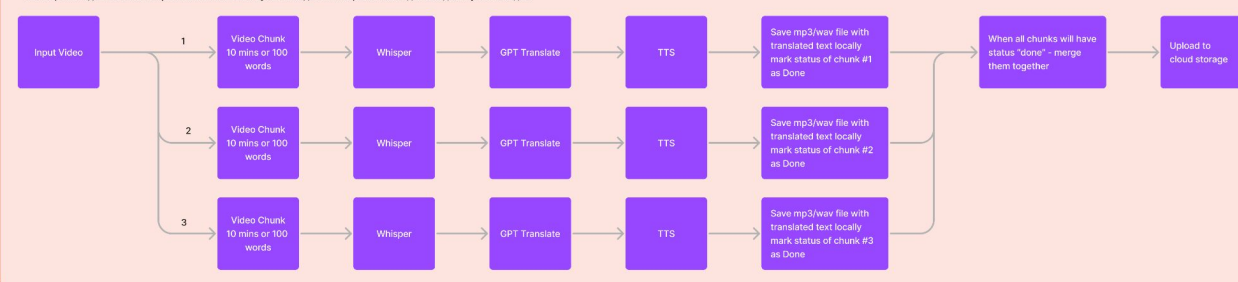




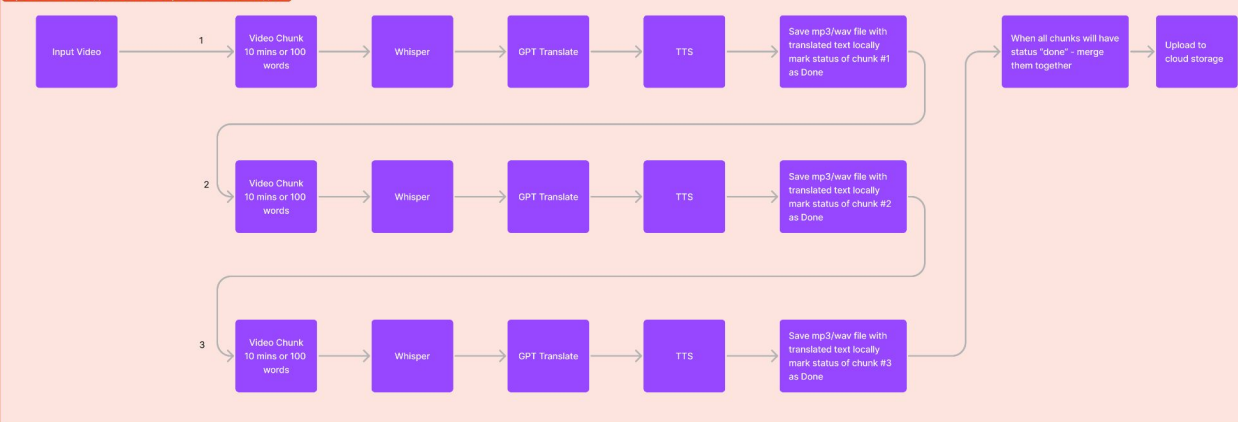
ML-pipeline: варианты конфигурации

Исполнитель - Милорадова К. С.

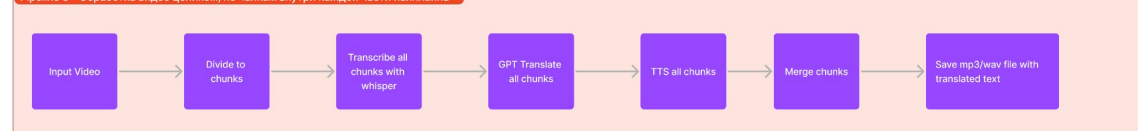
Pipeline 1 - Параллельная обработка чанков видео



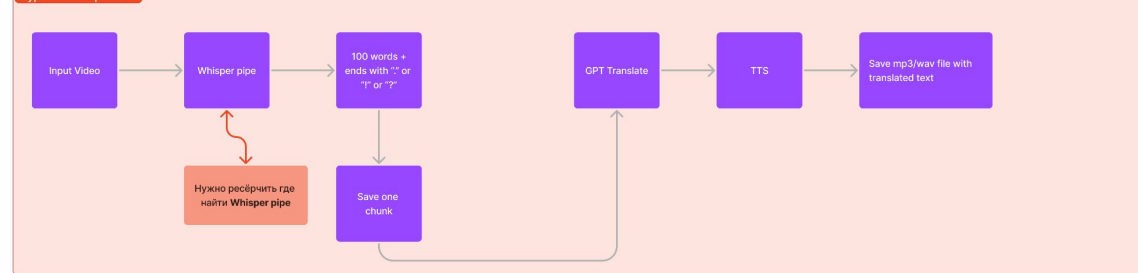
Pipeline 2 - Последовательная обработка чанков видео



Pipeline 3 - Обработка видео целиком, по чанкам внутри каждой части пайплайна



Hypothesis - Pipeline 4





ML-pipeline: решения для перевода

Исполнитель - Милорадова К. С.

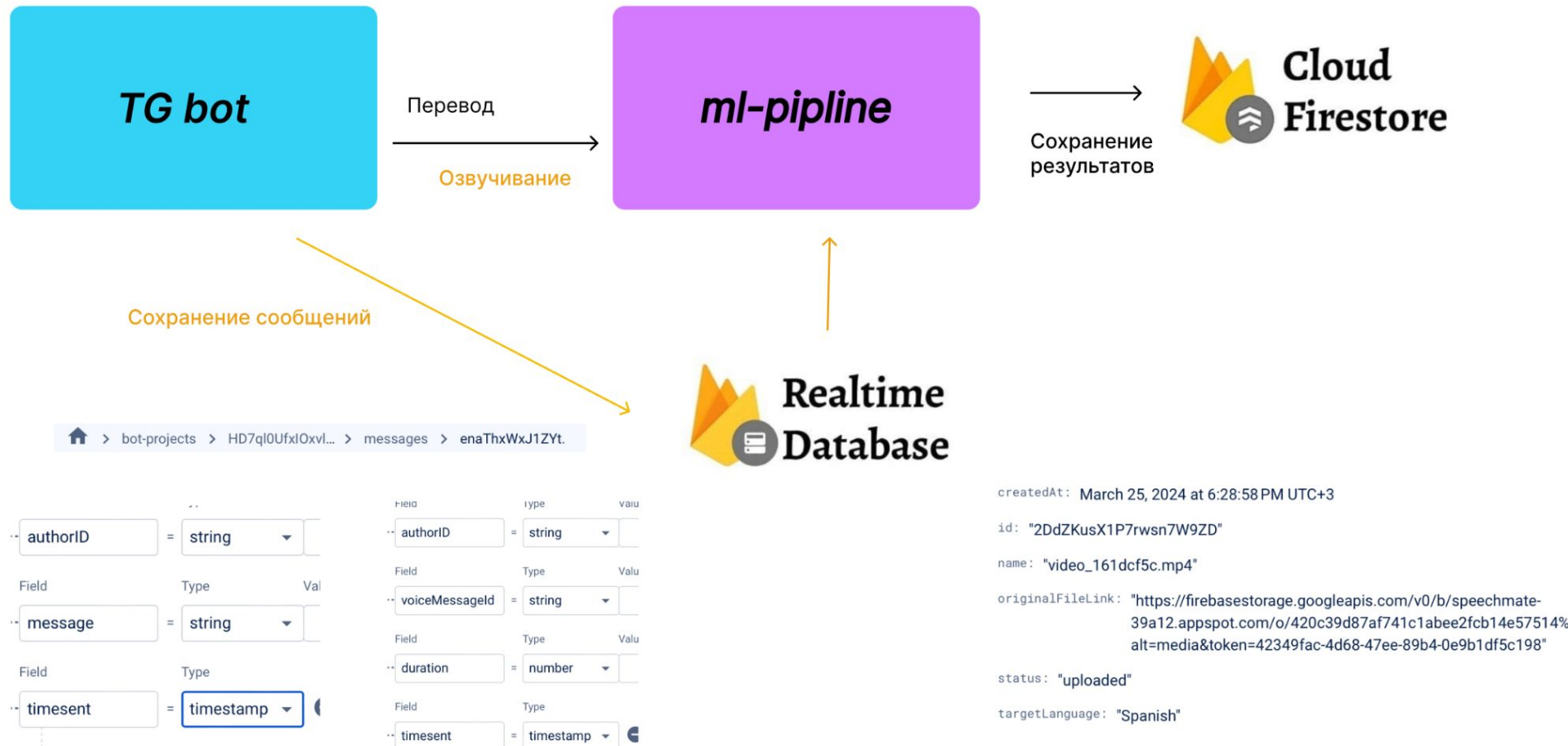
Название	Ограничение	Ссылка на документацию/инструкцию
DeepL	500,000 characters/month for free	https://pypi.org/project/deepl/1.3.0/
GPT	\$ / 1K tokens (платная, ищем замену)	https://medium.com/@caglarlaledemir/op-enai-gpt-3-5-turbo-in-action-building-a-simple-chatbot-interface-41300696094b
MarianMT	Модели можно бесплатно запускать на своих ресурсах	https://huggingface.co/docs/transformers/model_doc/marian
Google Translate	free and unlimited	https://pypi.org/project/googletrans/



Telegram-bot

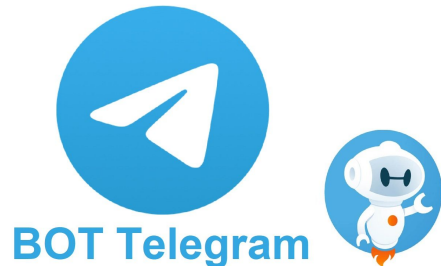
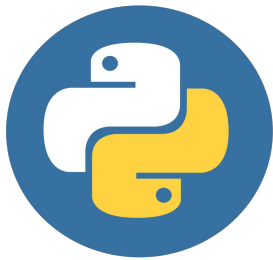
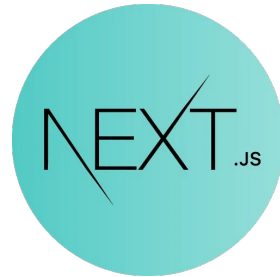
Исполнитель - Щербакова Е. А.

Модель для клонирования
голоса - **Coqui.XTTS** [6]





Технологии и инструменты реализации





Тестирование

Некоторые из выявленных проблем:

- Отсутствие обновления статуса проекта в таблице
- Ошибка при обработке ботом не видео файлов
- Ошибка при создании проекта при отсутствии доступных токенов
- Ограничение на длину текста у gpt

ML-пайплайн	Unit-тестирование; Интеграционное тестирование; Ручное end-to-end тестирование; Приемочное тестирование.
Backend	Unit-тестирование; Интеграционное тестирование; Тестирование совместимости; Ручное end-to-end тестирование; Приемочное тестирование.
Frontend	Тестирование совместимости; Ручное end-to-end тестирование; Приемочное тестирование.
TG-бот	Интеграционное тестирование; Ручное тестирование; Приемочное тестирование.

Основные результаты

- Разработан прототип интерфейса сайта в Figma
- Разработан ML-пайплайн
- Разработан основной функционал веб-сайта
- Разработан основной функционал телеграм-бота
- Проведено тестирование
- Разработана техническая документация.



Перспективы развития проекта

- Добавление функции краткого пересказа переписки чата тг-ботом
- Добавление возможности загрузки видео с YouTube, Vimeo и других платформ.
- Поддержка субтитров на разных языках и синхронизация переводов с субтитрами.
- Возможность редактирования видео
- Автоматическая генерация субтитров
- Расширенные возможности перевода
- Добавление аналитики (анализ эффективности переводов и т.д.)
- Создание мультязычного интерфейса





Демонстрация

Список источников

- 1) Дизайн в Figma / [Электронный ресурс] // URL: <https://drive.google.com/file/d/10dSeqm0ikusdUdwiNkOzTijOrFS0oXrF/view?usp=sharing> – режим доступа свободный
- 2) OpenAI Whisper / [Электронный ресурс] // URL: <https://openai.com/research/whisper> – Режим доступа: свободный
- 3) GPT-3.5-turbo / [Электронный ресурс] // URL: <https://openai.com/blog/gpt-3-5-turbo-fine-tuning-and-api-updates> – Режим доступа: свободный
- 4) MicrosoftTTS / [Электронный ресурс] // URL: <https://azure.microsoft.com/en-us/products/ai-services/text-to-speech> – Режим доступа: свободный
- 5) 11labs / [Электронный ресурс] // URL: <https://elevenlabs.io/dubbing> – Режим доступа: свободный
- 6) Coqui XTTS / [Электронный ресурс] // URL: <https://huggingface.co/coqui/XTTS-v2> – Режим доступа: свободный

