

Видеочат с переводом на английский язык

Описание проекта

Этот проект представляет собой видеочат с функцией перевода речи в реальном времени на английский язык. Решение использует:

- **Flask** для backend-части и веб-интерфейса
- **Whisper** (от OpenAI) для распознавания и перевода речи
- **WebRTC** для организации видеосвязи между пользователями

Преимущества:

- Работает полностью локально, без сторонних сервисов
- Поддержка перевода в реальном времени
- Простой интерфейс для пользователей

Технологии

- Python 3.8+
- Flask
- Whisper
- WebRTC
- HTML5, JavaScript

Доступ к сервису

Рабочий экземпляр развернут по адресу:

`https://5.35.100.220`

Требуется современный браузер с поддержкой: - WebRTC

- MediaStream API

- HTTPS-соединений

Инструкция по использованию

1. Инициализация сеанса

- Разрешите доступ к камере/микрофону при запросе браузера
- Создайте новую комнату (автогенерация ID)

2. Подключение собеседника

- Передайте 4-значный номер комнаты
- Второй участник вводит номер в поле “Join room”

3. Процедура перевода

[Start audio recording] → Ведение диалога → [Stop audio recording]

- Обработка занимает 10-20 сек (зависит от длины записи)
 - Результат появляется в текстовом блоке под видео
- ## Особенности реализации

- Аудио захватывается порциями по 1-10 секунд
- Whisper обрабатывает весь аудиоблок целиком
- Поддерживается только направление → Английский

Проект в разработке

Текущее состояние:

Проект находится в активной разработке. На данный момент реализована базовая функциональность перевода аудиозаписей, но для достижения максимальной эффективности и качества требуется доработка с использованием GPU-ускорения.

Запланированные улучшения:

1. Оптимизация для мобильных устройств

- Устранение ошибок интерфейса (адаптация под разные размеры экранов, исправление “вылетов” при повороте устройства).
- Улучшение обработки аудио на слабых устройствах (оптимизация алгоритмов для работы с ограниченными ресурсами).
- Исправление задержек при записи и воспроизведении (минимизация лагов на Android/iOS).
- Поддержка сенсорного ввода (корректная работа с тапами, свайпами, долгими нажатиями).

2. Кросс-браузерная совместимость

- Тестирование и исправление ошибок в Chrome, Firefox, Safari, Edge, мобильных браузерах.
- Устранение проблем с Web Audio API (некоторые браузеры могут по-разному обрабатывать аудиопоток).
- Поддержка старых версий браузеров (по возможности без потери функциональности).

3. Инфраструктурные обновления

Покупка доменного имени

- Выбор и регистрация профессионального домена
- Настройка DNS для корректной работы сервиса
- Получение доверенного HTTPS-сертификата
- Обеспечение “зеленого замочка” в браузере для безопасности пользователей

4. Улучшение качества перевода

- Интеграция более совершенных моделей машинного перевода (например, на основе трансформеров).
- Постобработка текста для исправления грамматических и контекстных ошибок.
- Использование нейросетевых методов для учета интонации и смысловых нюансов.

5. Уменьшение времени на перевод

- Оптимизация алгоритмов обработки аудио и текста.
- Распараллеливание вычислений с помощью CUDA (NVIDIA GPU).
- Кэширование часто используемых фраз для ускорения перевода.

6. Увеличение максимальной длины аудиозаписи

- Реализация потоковой обработки длинных записей без потери качества.
- Улучшение сегментации аудио для эффективного перевода больших файлов.
- Оптимизация использования оперативной памяти и видеопамати GPU.

7. Добавление новых языков перевода

- Поддержка популярных языков (китайский, арабский, французский и др.).
- Возможность автоматического определения исходного языка.
- Настройка моделей под специфику каждого языка (идиомы, сокращения, диалекты).

Требования для реализации улучшений:

- **Графический процессор (GPU)** с поддержкой CUDA (NVIDIA) для ускорения нейросетевых вычислений.
- Увеличенные вычислительные ресурсы для обработки длинных аудиофайлов.