

Обзор MetaVoice-1B

Описание технологии

MetaVoice-1B — это модель преобразования текста в речь (TTS), разработанная с упором на эмоциональный ритм и интонацию на английском языке. Она примечательна своей способностью клонировать голоса с минимальными обучающими данными, возможностью тонкой настройки и способностью точно клонировать голоса с американским и британским акцентом. Модель также поддерживает синтез длинных текстов и направлена на минимизацию появления искажённых слов.

Примеры использования

- Сервисы электронных и аудиокниг — для создания аудиокниг, подкастов и обучающих материалов.
- Переводчики и пользователи видеохостингов — для создания и просмотра иноязычного контента с помощью синтезированного голоса на интересующем языке.
- Разработчики игр — для создания персонажей с реалистичным голосом и диалогами.

Отзывы в интернете

“Удобен для пользователя и очень полезен для работы.”

“Метаголос может быть настолько реалистичным, что вам может быть сложно отличить настоящий голос от голоса, созданного искусственным интеллектом. Класс!”

“Вот уж действительно плохой опыт. Занимает очень много времени, и качество плохое.”

Тестовые примеры

1. Клонирование голоса для озвучивания заданного текста.

- Модели было необходимо воспроизвести голос Бенедикта Камбербетча и озвучить фразу: «The Dursleys had a small son called Dudley, and in their opinion, there was no finer boy anywhere. The Dursleys had everything they wanted, but they also had a secret, and their greatest fear was that somebody would discover it. They didn't think they could bear it if anyone found out about the Potters.»
Пример голоса для клонирования:

[Пример голоса для клонирования](#)

- Результат клонирования:

Результат генерации

2. Озвучивание заданного текста на английском языке.

- Модели было необходимо озвучить фразу:
Исходный текст: «The warrior's fame as the best of the best, of the best followed him closely! And even after seeing the legendary sword of heroes, our warrior did not blink an eye. He was such a great master!»
- Результат генерации:

Результат генерации

Возможность автономного запуска

Модель MetaVoice-1B имеет специализированную архитектуру для TTS, состоящую из текстового кодировщика, декодера и аудиокодека. Для преобразования графем в фонемы (G2P) она использует традиционные методы. Вместо них может быть полезно использовать модель Mistral NeMo, которая сама по себе является большой языковой моделью (LLM) и может быть использована для анализа текста и извлечения важных лингвистических особенностей — интонации, пауз и акцентов. Однако, это не главная составляющая MetaVoice-1B. Исходя из особенностей архитектуры MetaVoice-1B, для полностью автономного запуска потребуется адаптировать архитектуру Mistral NeMo и обучающие наборы данных для достижения интересующей функциональности. Это может включать добавление слоев, специфичных для TTS. В большинстве случаев более целесообразно использовать специализированные модели для задач генерации аудиосигналов, так как они уже оптимизированы для этих целей.

Модель распространения

Полная бесплатная версия MetaVoice-1B представлена в виде репозитория с исходным кодом на GitHub и не ограничивает пользователя в наборе инструментов.

Если нет возможности установить и использовать продукт локально, можно воспользоваться его демо-версией на replicate.com — [Ссылка на демо-версию MetaVoice-1B](#).

Данная демо-версия работает через Replicate API-key, каждый запуск стоит 0.068 USD.

*Демоверсию в Google Colab запустить не получилось.

Преимущества MetaVoice-1B

- Широкий выбор стилей речи и возможностей управления голосом;
- Высокое качество сгенерированной и клонированной речи;
- Современные методы сжатия и обработки данных делают модель быстрой и эффективной;

- Поддержка многоязычности.

Недостатки MetaVoice-1B

- Модель тяжёлая и сложная для установки и настройки (большой объём, много гиперпараметров);
- Имеет высокие вычислительные требования для использования и обучения;
- При клонировании голоса не обеспечивает точного воспроизведения интонации и эмоций говорящего;
- В основном модель ориентирована на английскую речь.

Последняя дата обновления

Последняя дата обновления на HuggingFace: 03.04.2024

From:

<http://wiki.nic.etu/docuwiki/> - REC ETU Wiki

Permanent link:

http://wiki.nic.etu/docuwiki/doku.php/ntn:llm:ai_review:20241030:metavoice

Last update: **2024/12/25 14:59**

