Обзор Parler-TTS Large v1

Описание технологии

Parler-TTS Large v1 — это лёгкая модель преобразования текста в речь (TTS), способная генерировать высококачественную, естественно звучащую речь в заданном стиле.
Стилем речи можно управлять с помощью простой текстовой подсказки, указав, например, пол, фоновый шум, скорость речи, высоту тона и реверберацию.

Parler-TTS — это воспроизведение работы из статьи Natural language guide of high-fidelity text-to-speech with synthetic annotations Дэна Лита и Саймона Кинга из Stability AI и Эдинбургского университета соответственно. Модель основана на трансформерах и генерирует аудиотокены в причинно-следственном порядке.

Примеры использования

- Сервисы электорнных и аудиокниг — для создания аудиокниг, подкастов и обучающих материалов.
- Музеи — для создания аудиогидов.
- Разработчики игр — для создания персонажей с реалистичным голосом и диалогами.

Отзывы в интернете

“Когда модель работает, она кажется хорошим решением. Но, к сожалению, она слишком перегружена для использования. Создает длительные периоды молчания в середине и занимает слишком много времени по сравнению с существующими решениями tts”

“Я бы сказал, что этот Parler хорош.”

*“Очень странно. Текстовое описание голоса совпадает с несколькими реальными голосами, поэтому каждая генерация с одним и тем же запросом выдает разные голоса.
С такой задержкой и невозможностью всегда использовать один и тот же голос, это непригодно для ассистентов и чат-ботов.”*

Тестовые примеры

Модели было необходимо озвучить фразу на английском языке разными голосами и с разной интонацией:
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!»

- Промпт 1: «Gary's voice is loud and emotional. He speaks clearly and expressive, without

noise.»

Результат генерации 1

- Промпт 2: «Gary's voice is quiet and uncertain. He whispers. Clear audio.»

Результат генерации 2

- Промпт 3: «Lea's voice is quiet and uncertain. She whispers. Clear audio.»

Результат генерации 3

Возможность автономного запуска

Модель Parler-TTS Large v1 имеет специализированную архитектуру для TTS, состоящую из текстового кодировщика (Flan-T5), декодера и аудиокодека (DAC).
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 Parler-TTS Large v1.

Исходя из особенностей архитектуры Parler-TTS Large v1, 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 Parler-TTS Large v1 представлена в виде репозитория с исходным кодом на GitHub и HuggingFace и не ограничивает пользователя в наборе инструментов.
Если нет возможности установить и использовать продукт локально, можно воспользоваться его демо-версией — Ссылка на демо-версию Parler-TTS Large v1.
В таком случае пользователю доступен полный набор функций модели, но генерируемая аудиозапись ограничена 15 секундами.

Преимущества Parler-TTS Large v1

- Доступна демо-версия;
- Большой выбор голосов для генерации;

- Способность модели обеспечивать единообразие голоса во всех генерациях;
- Есть возможность самостоятельно задавать скорость, интонацию и стиль речи;

Недостатки Parler-TTS Large v1

- Модель поддерживает только английский язык;
- Единообразие голоса обеспечивается только для голосов из предоставленного разработчиками списка;
- Заданная интонация и стиль речи не всегда соблюдаются;

Последняя дата обновления

Последняя дата обновления на GitHub: 08.08.2024