Обзор WhisperSpeech

Описание технологии

WhisperSpeech — это система преобразования текста в речь с открытым исходным кодом, созданная путем инвертирования Whisper. Ранее известна как spear-tts-pytorch.
Разработчики данной модели стремятся добиться сходств со Stable Diffusion, но для речи – сделать WhisperSpeech одновременно мощной и легко настраиваемой.
Обучение проводится только на лицензированных записях речи, а весь исходный код имеет открытый доступ, поэтому модель всегда можно будет безопасно использовать для коммерческих приложений.
В настоящее время модели обучаются на наборе данных LibreLight на английском языке. В следующем выпуске планируется использовать несколько языков.

Примеры использования

- Школы иностранных языков — для создания аудиоматериалов для изучения иностранных языков.
- Музеи — для создания аудиогидов.
- Маркетологи — для создания аудиорекламы.

Отзывы в интернете

“Инструмент очень хороший! Спасибо, за обновление.”

“Прекрасный проект, планирую тестировать его на предмет генерации речи в реальном времени и управление эмоциями.”

“Пока поддерживаются только английский, польский и французский языки. Очень жду обновление с расширением на другие языки.”

Тестовые примеры

1. Озучивание текста заданным голосом:

- Модели было необходимо озвучить фразу на английском языке голосом Илона Маска.
  Исходный текст:
  «Go back, say, 300 years. The things that we take for granted today you'd be burned at the stakeful. Know, being able to fly. That's crazy. Being able to see over long distances, being able to communicate.
  This is stuff that really would be magic, it would be considered magic in times passed.»
  Пример голоса для клонирования:

Пример голоса для клонирования

- Результат генерации:

Результат клонирования

---

2. Озучивание текста на английском языке:

- Модели было необходимо озвучить фразу на английском языке. Исходный текст:
  «A stay in Las Vegas can feel similar to a visit to many popular cities worldwide. Many of the hotels have miniature versions of important international sites and monuments.
  These famous landmarks include the Eiffel Tower, Venice, and even ancient Rome.»

Результат генерации

---

3. Озучивание текста на испанском языке:

- Модели было необходимо озвучить фразу на английском языке Исходный текст:
  «Pablo Ruiz Picasso fue un pintor y escultor español, creador, junto con Georges Braque, del cubismo. Es considerado desde la génesis del siglo XX como uno de los mayores pintores que participaron
  en los variados movimientos artísticos que se propagaron por el mundo y ejercieron una gran influencia en otros grandes artistas de su tiempo.»

Результат генерации

Возможность автономного запуска

Модель WhisperSpeech использует традиционные методы преобразования графем в фонемы (G2P). Вместо них может быть полезно использовать модель Mistral NeMo, которая является большой языковой моделью (LLM) и может быть использована как составляющая архитектуры WhisperSpeech для анализа текста и извлечения важных лингвистических особенностей — интонации,
пауз и акцентов. Однако, это не главная составляющая WhisperSpeech.

WhisperSpeech вдохновлена моделью SPEAR TTS от Google Research. Исходя из этого, для полностью автономного запуска потребуется адаптировать архитектуру Mistral NeMo и обучающие наборы
данных для достижения интересующей функциональности. 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 WhisperSpeech 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 — Ссылка на демо-версию WhisperSpeech.
В таком случае пользователю доступен полный набор функций модели.

Преимущества WhisperSpeech

- Высокое качество и естественность речи на доступных языках;
- Адаптируемость и интегрируемость модели;
- Доступна демо-версия;

Недостатки WhisperSpeech

- Нет возможности самостоятельно задавать интонацию речи;
- Нет поддержки русского языка;
- При клонировании голоса интонация и стиль речи не всегда сохраняются;

Последняя дата обновления

Последняя дата обновления на GitHub: 29.01.2024