Обзор Fish Speech V1.4

Описание технологии

Fish Speech V1.4 — это ведущая модель преобразования текста в речь (TTS), обученная на 700 тысячах часов аудиоданных на нескольких языках. В настоящее время поддерживаются английский, японский, корейский, китайский, французский, немецкий, арабский и испанский языки. Модель обладает широкими возможностями обобщения и не использует фонемы для преобразования текста в речь. Она может обрабатывать текст на любом языке.

Fish Speech имеет простой использовании веб-интерфейс на базе Gradio, совместимый с Chrome, Firefox, Edge и другими браузерами, а также предлагает графический интерфейс PyQt6, который легко интегрируется с сервером API. Поддерживает Linux, Windows и macOS.

Примеры использования

- Сервисы электорнных и аудиокниг — для создания аудиокниг и подкастов.
- Онлайн-школы и преподаватели — для создания обучающего видео-контента.
- Разработчики игр — для создания персонажей с реалистичным голосом и диалогами.

Отзывы в интернете

“Воспроизведение моего голоса составляет 90%. Действительно хорошо.”

“Я всё ещё жду полноценный туториал по этой модели. Она достойна внимания.”

“Лучший сервис TTS со сверхнизкой задержкой!”

Тестовые примеры

1. Перевод аудиозаписи в текст.

- Модели было необходимо преобразовать заданный аудиофайл с речью Илона Маска в текст на наглийском языке. Исходный файл:

Аудиофайл для преобразования

- Результат обработки:
«Go back, say, 300 years. The things that we take for granted today would be 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. Having effectively with the Internet, a group mind of sorts and having access to all the world's information instantly from almost anywhere on the earth. This is stuff that really would be magic, it would be considered magic in times passed.»

2. 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 would be 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. Having effectively with the Internet, a group mind of sorts and having access to all the world's information instantly from almost anywhere on the earth. This is stuff that really would be magic, it would be considered magic in times passed.»
- Результат генерации модели:

Результат перевода

Возможность автономного запуска

Модель Fish Speech V1.4 использует большие языковые модели (LLMs) для анализа текста и извлечения важных лингвистических особенностей — интонации, пауз и акцентов.
Модель Mistral NeMo сама по себе является большой языковой моделью (LLM), поэтому может быть использована как составляющая архитектуры Fish Speech V1.4.
Однако, Fish Speech V1.4 состоит не только из LLM:

- Архитектура Dual Autoregressive (Dual-AR) использует два авторегрессивных процесса — быстрый и медленный для стабилизации процесса генерации последовательностей.
- Метод Grouped Finite Scalar Vector Quantization (GFSQ) помогает эффективно обрабатывать кодовые книги (codebooks) и улучшает качество выходного сигнала.

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 речи, так как они уже оптимизированы для этих целей.

Модель распространения

Бесплатная версия Fish Speech V1.4 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-версиями:

- Ссылка на демо-версию Fish Speech V1.4 (Текст в речь)
- Ссылка на демо-версию Fish Speech V1.4 (Речь в текст)

В таком случае пользователю доступен неполный набор функций модели, и исходный текст должен быть не длиннее 500 символов. Кроме этого к ограничениям относится, например,

невозможность использовать свой голос для озвучивания.

Для расширения возможностей демо-версии существует подписка Premium. Варианты подписки отличаются длительностью — Premium: 10 USD/месяц или 80 USD/год. Подписка включает всебя все инструменты модели и ограничивает исходный текст 5000 символов.

Преимущества Fish Speech V1.4

- Высококачественная, естественно звучащая речь на выходе.
- Высокая скорость вывода.
- Доступна демо-версия.
- Многоязычная поддержка.

Недостатки Fish Speech V1.4

- Нет возможности самостоятельно задавать интонацию речи.
- Требуются значительные вычислительные ресурсы для обучения и тонкой настройки.
- Есть погрешности в произношении, интонации, ударениях.
- Есть граничения в обработке определенных произношений или специальной лексики;

Последняя дата обновления

Последняя дата обновления на GitHub: 10.09.2024