

Обзор Mini-Omni

Описание технологии

Mini-Omni — первая open source мультимодальная языковая модель, позволяющая вести диалог голосом с минимальной задержкой ответа и без использования внешних text-to-speech моделей.

Метод «Any Model Can Talk» позволяет интегрировать речевые возможности в любые языковые модели, улучшая их производительность в задачах, связанных с речью. Существует Mini-Omni2 — полностью интерактивная модель. Она может распознавать вводимые изображения, аудио, текст и вести сквозной голосовой диалог с пользователями.

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

“Отличная штука для ботов поддержки клиентов.”

“Это впечатляюще!”

“Интересная модель, хотя я заметил, что она не так хорошо распознаёт неродные английские голоса. Возникают огромные проблемы с моим голосом”

Тестовые примеры

1. Задача нейросети — ответить на запрос «Привет! Дай мне пожалуйста краткую информацию по африканским львам.» Запись диалога приведена ниже:

[Диалог с Mini-Omni: Русский язык](#)

2. Аналогичная П.1 задача — «Hello! Tell me short information about african lions.» Запись диалога приведена ниже:

[Диалог с Mini-Omni: Английский язык](#)

Возможность автономного запуска

Модель Mini-Omni основана на трансформерах, как и Mistral NeMo, поэтому автономный запуск возможен. Однако, Mini-Omni использует Whisper Small Encoder для распознавания речи.

Исходя из этого, потребуется адаптировать архитектуру Mistral NeMo и обучающие наборы данных для достижения интересующей функциональности.

В большинстве случаев более целесообразно использовать специализированные модели, такие как Whisper, для задач распознавания речи, так как они уже оптимизированы для этих целей.

Модель распространения

Полная бесплатная версия Mini-Omni представлена в виде репозитория с исходным кодом на GitHub и не ограничивает пользователя в функциональности.

Если нет возможности установить и использовать продукт локально, можно воспользоваться его демо-версией — [Ссылка на демо-версию Mini-Omni](#).

В таком случае пользователю доступна возможность записать свою речь и получить ответ от нейросети тоже в формате аудио (только английский язык).

Преимущества Mini-Omni

- Поддержка голосового взаимодействия в реальном времени;
- Решает проблему задержки, характерную для современных языковых моделей в обработке голоса;
- Архитектура Mini-Omni может быть адаптирована для других исследований и Llm-моделей;
- Открытый исходный код;

Недостатки Mini-Omni

- Сложность и ресурсоёмкость модели;
- Источник данных может иметь ограничения. (VoiceAssistant-400K, синтезированный с помощью GPT-4o);
- Не поддерживает русский язык. Модель обучена и выдаёт ответы исключительно на английском языке;
- У модели есть проблемы с качеством информации в ответах нейросети;

Последняя дата обновления

Последняя дата обновления на GitHub: 13.10.2024

From:

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

Permanent link:

http://wiki.nic.etu/docuwiki/doku.php/ntn:llm:ai_review:20241030:mini-omni

Last update: **2024/12/18 16:11**

