Voice AI

Gen Audio

- Stability AI
  - Stable Audio
  - HF: https://huggingface.co/stabilityai/stable-audio-open-1.0
  - Stability AI Launches Open-Source Model to Generate Audio (itsfoss.com)
- FunAudioLLM - Voice Understanding and Generation Foundation Models for Natural Interaction Between Humans and LLMs
  - GitHub: https://github.com/FunAudioLLM

Instant voice cloning

- OpenVoice

Text to Speech (TTS)

- ChatTTS
  - 6drf21e/ChatTTS_colab: ? ???????????????? ChatTTS ????????????????????????????????? (github.com)
- MARS 5
  - GitHub: https://github.com/Camb-ai/MARS5-TTS
  - HF: https://huggingface.co/CAMB-AI/MARS5-TTS
- edge-tts - An Python module that allows you to use Microsoft Edge's online text-to-speech service from within your Python code or using the provided edge-tts or edge-playback command.
  - GitHub: https://github.com/rany2/edge-tts
- fish-speech
  - GitHub: https://github.com/fishaudio/fish-speech
- ???? (?????????)
  - https://github.com/TaiwanAILabs-Yating
- Qwen-TTS (???? TTS)
  - Qwen3-TTS vs ElevenLabs : Voice Cloning & Real-Time Streaming - Geeky Gadgets

- VibeVoice (Microsoft)
  - HF: https://huggingface.co/microsoft/VibeVoice-1.5B
- Pocket TTS
- Kitten TTS - ???? GPU????????

ASR - Automatic Speech Recognition

- FrogBase - OpenAI ??????????
- InstantID - ?????? AI?????????
- WhisperDesktop - ??????????For Windows Only
  - [Video] ????Whisper ?????????????????????????? ?????????
- OpenAI Whisper
- Whisper WebUI - ??????
- WhisperX - ? whisper large-v2 ? 70 ?
- Fast Whisper - ? OpenAI Whisper ???????????
- Vosk - Offline speech recognition API for Android, iOS, Raspberry Pi and servers with Python, Java, C# and Node
- Handy - A free, open source, and extensible speech-to-text application that works completely offline.

MTK Breeze 3

???????MediaTek Research????? MediaTek Research Breeze 3??? MR Breeze 3 ????????????? Breeze ASR 26????????? BreezyVoice 26?????????? AI ???????? Breeze Guard 26?

- MediaTek Research Breeze 3?? AI ?????????????

Vibevoice (Microsoft)

Microsoft VibeVoice ??????? AI ??????? TTS???????? ASR????????????? 7.5Hz ???????? tokenizer??? next-token diffusion ?????????? 90 ???????????? 60 ??????TTS ???? 4 ??????ASR ??????????????????????

- https://github.com/microsoft/VibeVoice

Revision #25
Created 2024-04-25 19:43:31 CST by A-Lang (Admin)
Updated 2026-05-19 12:30:05 CST by A-Lang (Admin)