Some Open Source Tools and Services for creating AI Voice Assistants

Ayesha Noman

11/06/2025

Health Tune

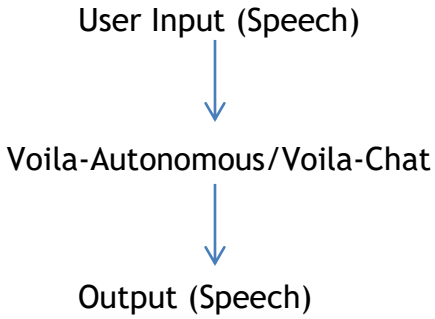
1. Voila

Voila family of large voice-language foundation models aims to improve interactions between humans and AI. Breaking away from the constraints of traditional voice AI systems—high latency, loss of vocal nuances, and mechanical responses—Voila employs an innovative end-to-end model design and a novel hierarchical Transformer architecture. With a latency of just 195 milliseconds, this method enables rich, autonomous, and real-time voice interactions that outperform typical human response times. Voila excels in a variety of audio tasks, from ASR and TTS to speech translation across six languages, and offers persona-driven engagements that are customizable. It also combines advanced voice and language modeling.

License: Open-Source (Models Available on Hugging Face)

Use Case: Perfect for real-time patient-doctor conversations as it is direct speech-to-speech model.

Implementation:



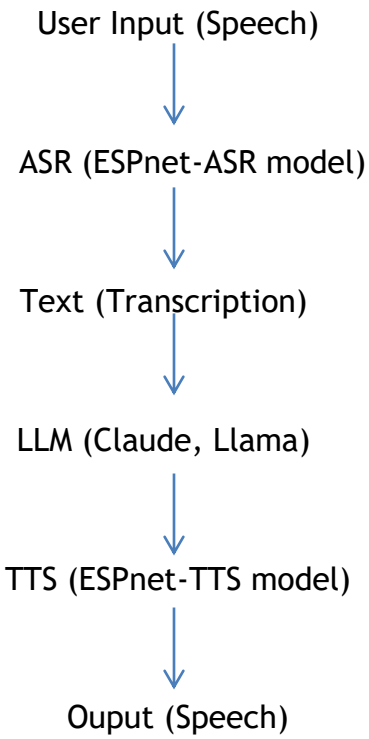
2. ESPnet

ESPnet is an end-to-end speech processing toolkit covering end-to-end speech recognition, text-to-speech, speech translation, speech enhancement, speaker diarization, spoken language understanding, and so on. ESPnet uses pytorch as a deep learning engine.

License: Open Source (Apache 2.0)

Use Case: ESPnet toolkit for speech processing is built for ASR and TTS. It works offline or in your own servers. Additionally, it is highly modular and can be used effectively in healthcare system by further fine-tuning as it is already being used in research and medicine.

Implementation:



3. Pipecat

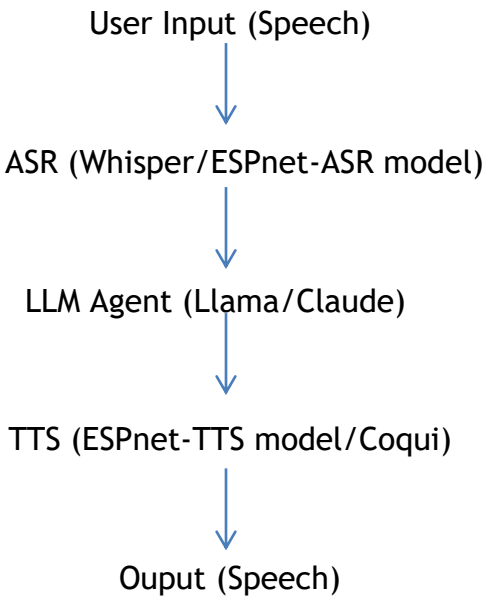
Pipecat is an open-source Python framework for building real-time voice and multimodal conversational agents. Orchestrate audio and video, AI services, different

transports, and conversation pipelines effortlessly—so you can focus on what makes your agent unique.

License: BSD 2-Clause "Simplified" License

Use Case: PipeCat is best suited for modular healthcare systems. It allows for multi-model experimentation such as using different ASR and TTS.

Implementation :



4. OpenVoiceOS/Mycroft (OVOS)

Open Voice OS (OVOS) is a flexible voice platform that goes beyond traditional voice assistants. It provides the foundational tools and frameworks for integrating voice interaction into a wide range of projects.

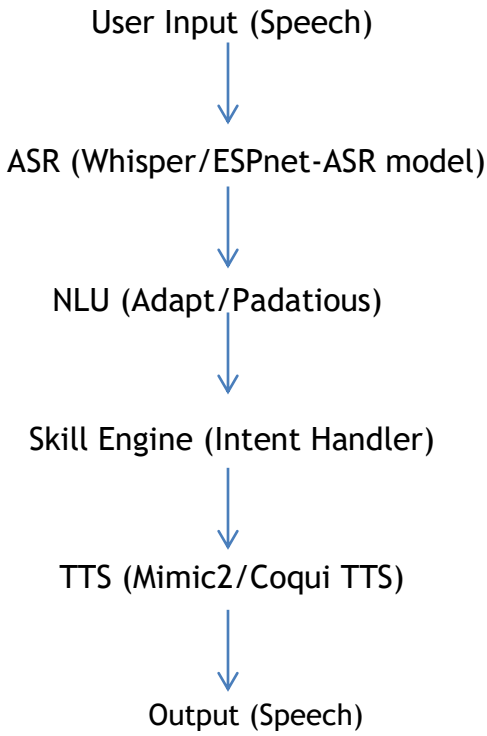
OVOS is designed to work wherever voice interfaces are needed — whether that's on a local device or in the cloud.

License: Open Source

Use Case: While OVOS can power a "Hey Mycroft…"-style assistant, it is not limited to that use case. As a voice operating system, OVOS is highly customizable and has been used in:

- Robots and automation systems
- Smart furniture and mirrors
- Cloud-based voice services
- Embedded devices and smart TVs

Implementation:



5. Vosk

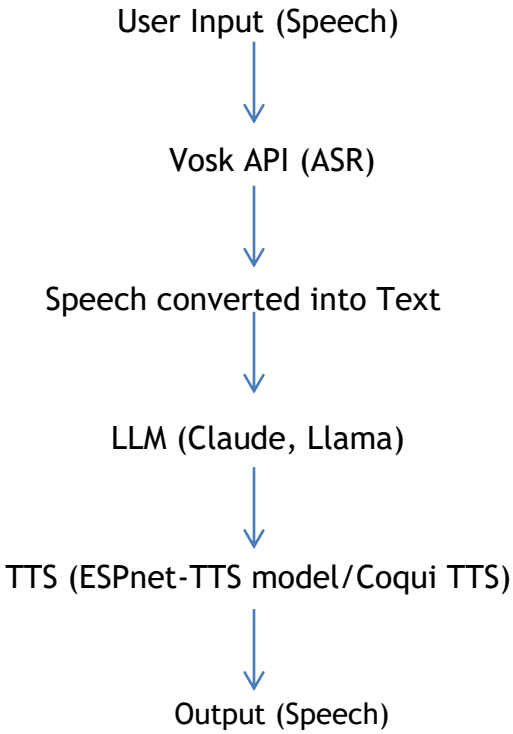
Vosk is an offline open source speech recognition toolkit. It enables speech recognition for 20+ languages and dialects. Vosk models are small (50 Mb) but provide continuous large vocabulary transcription, zero-latency response with streaming API, reconfigurable vocabulary and speaker identification.

Use Case: Vosk supplies speech recognition for chatbots, smart home appliances, virtual assistants. It can also create subtitles for movies, transcription for lectures

and interviews. We can utilize it for speech recognition between doctor and patient and then produce transcripts.

License: Open Source

Implementation:



We can just use the transcriptions for our use case but can also employ Vosk API for speech-to-speech by including TTS models.