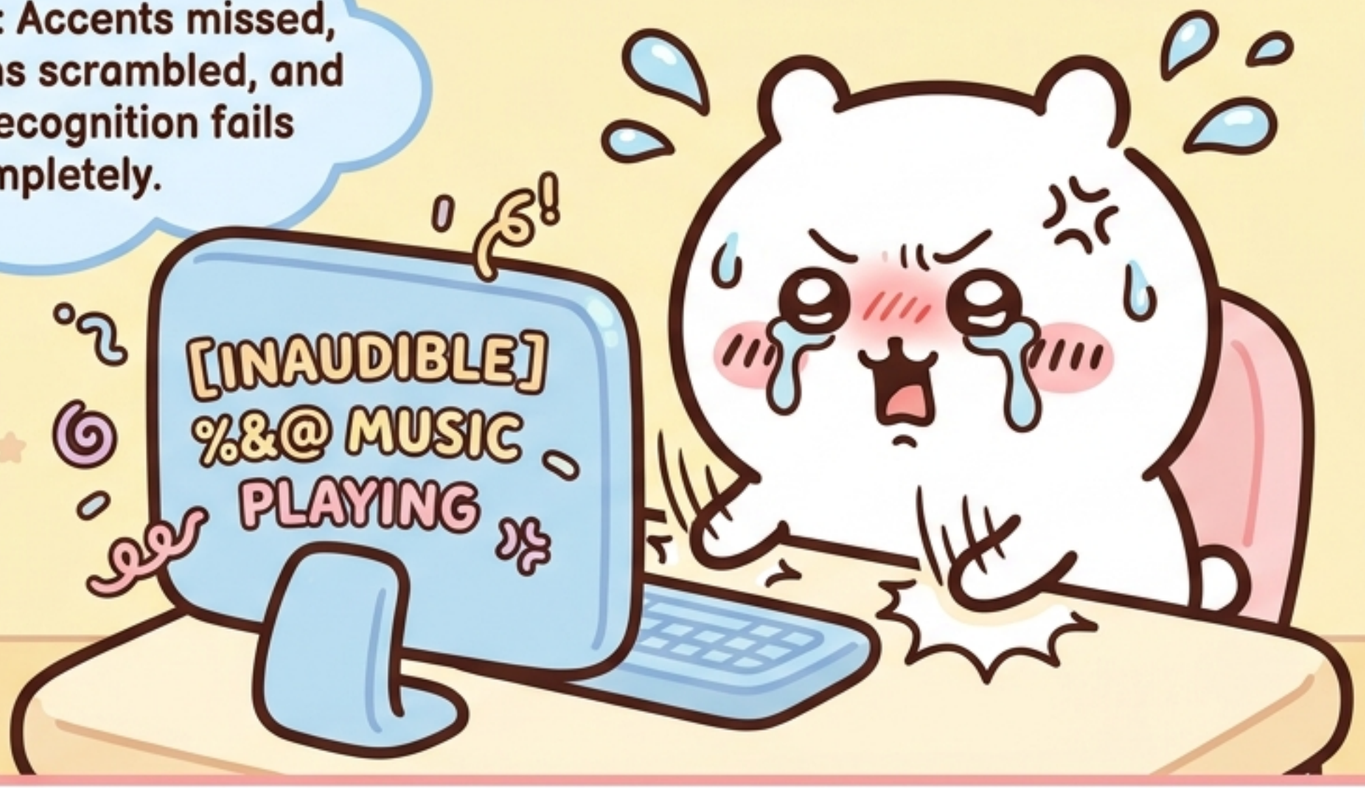
# The 8 open-source meeting recorders are completely failing users

Broken ASR: Accents missed, domain terms scrambled, and Chinese recognition fails completely.

Bad Summaries: Notes do not read like prose; half of the key meeting points simply disappear.

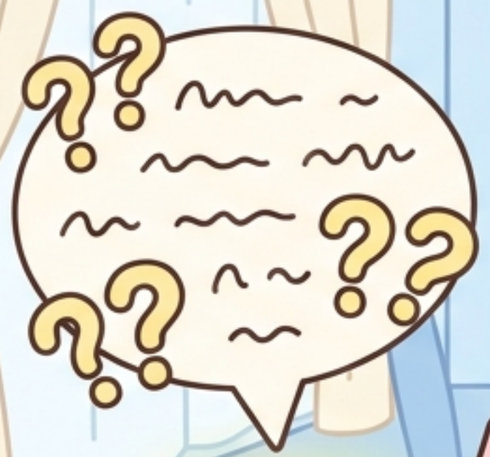


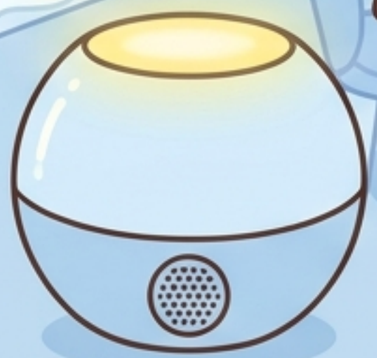

Dire UI/UX: Constant crashes, freezing, and friction reminiscent of a 1998 Linux desktop setup.

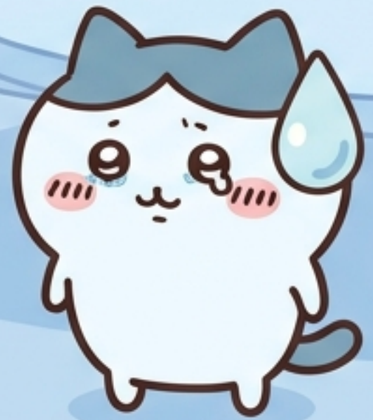
**The verdict after testing 11.6k-star projects down to 3-star toys? It is not a glitch. It is a structural failure.**

# Whisper has been outpaced by newer models for over a year

**The Hallucination Bug:**
Whisper invents complete sentences during dead air. Newer models fix this with built-in noise training and Voice Activity Detection (VAD).







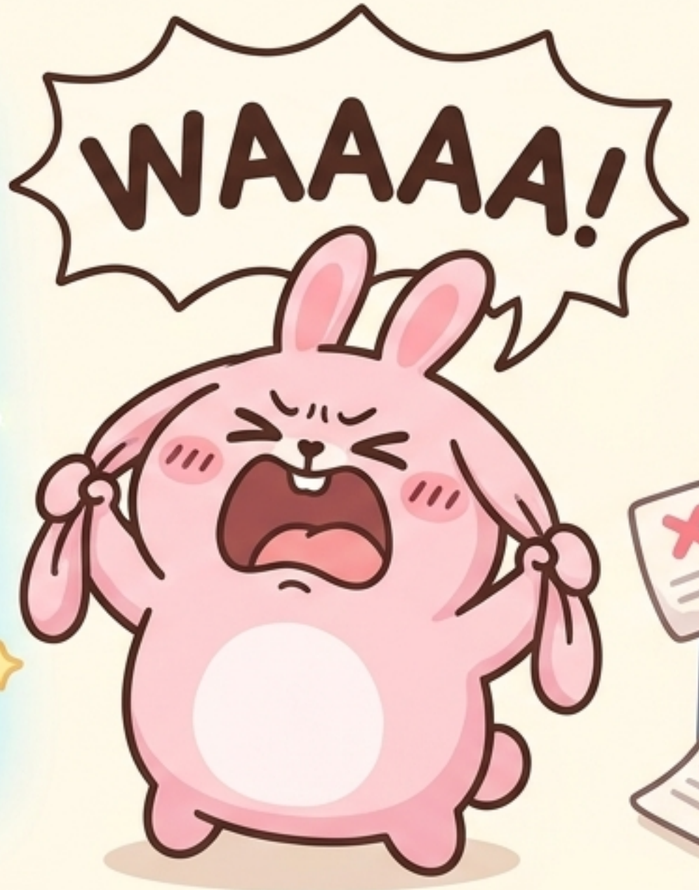
## Mandarin Character Error Rates Comparison

| | |
|---|---|
| Whisper-Large-v3 (OpenAI): | 5.14% |
| FireRedASR2-LLM (Xiaohongshu) | 2.89% |
| SenseVoice-Large (Alibaba): | 2.09% |
| Fun-ASR (Alibaba): | 1.22% |

# SOTA backends exist, but zero polished frontends can actually mount them

## README (Marketing Promise)

- Bring your own LLM!
- Incredible new models available: Voxtral 2, FireRedASR2, Streaming Sortformer v2.

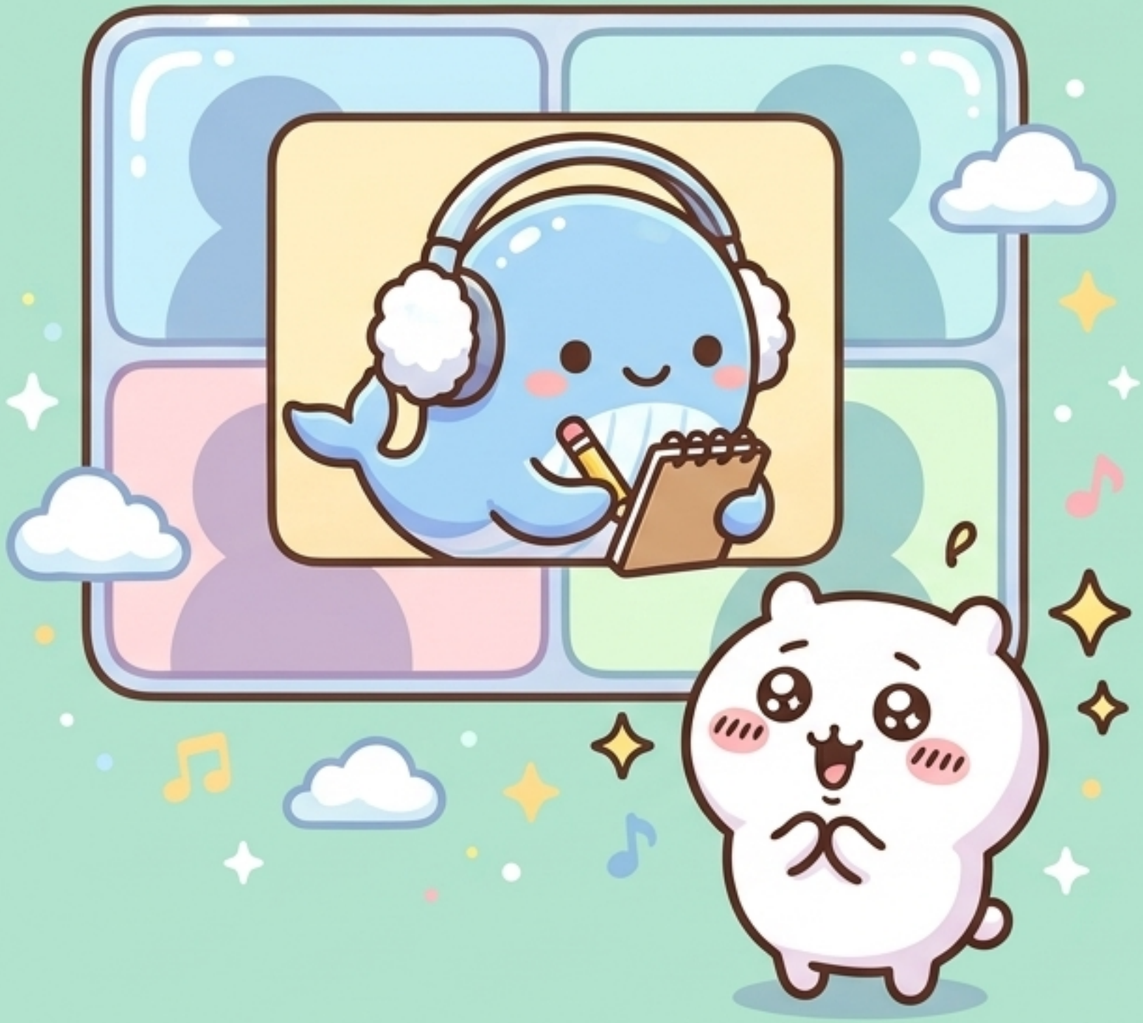



## ISSUES (Reality)

- Hyprnote: Chinese isn't on the roadmap; closed.
- Meetily: Hardcoded to Whisper. Selecting other models causes hard crashes.
- Recap: Not actually an open-source license.

**The Missing Layer: The ecosystem desperately needs ASR proxy middleware so frontends become backend-agnostic.**

# Digging past the GitHub trending page reveals four better paradigms



## Vexa (The Bot Paradigm)

A self-hosted Docker bot that joins the meeting (Zoom, Teams, Meet) rather than relying on local mic recording. Features real-time WebSocket transcripts.

## Handy (The Dictation King)

An excellent multi-backend tool (SenseVoice, Parakeet) perfect for daily voice typing, but strictly for dictation, not meetings.

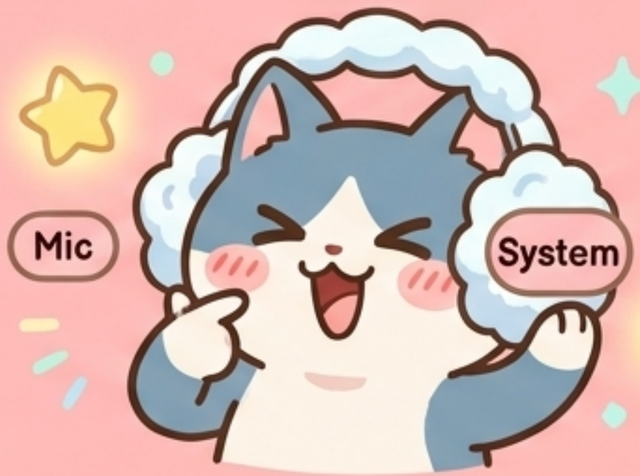
## Muesli (The Mac Native)

Multi-backend Swift implementation. Runs Qwen3-ASR (52 languages) and real pyannote diarization directly on the Apple Neural Engine.

## Voxtral 2 (The Raw Engine)

A drop-in Whisper replacement model with native diarization and 13 languages. It is just waiting for frontends to adopt it.

# Bypassing algorithmic diarization entirely with a physical audio split




Left Channel: Captures only YOUR microphone.

Right Channel: Captures everyone else (System Audio).

No Mixing

Save directly as a stereo .m4a file.

Prompt Gemini 3 Pro

Channel L is me, Channel R is everyone else. Transcribe and label.

100% Accuracy on Speaker Separation + Transcript + Summary.

This method bypasses old Whisper models and immature middleware.
Total cost for a 3.5-hour meeting? $1.42. DIY currently beats picking an app.