

ZazuAI: On-Premise AI Meeting Assistant for Automated Minutes

Overview

ZazuAI is an advanced, self-hosted meeting intelligence platform that automatically records, transcribes, and summarizes discussions into actionable minutes. Designed for organizations that prioritize data security, ZazuAI leverages a powerful dual-pipeline architecture built on state-of-the-art open-source models. It delivers real-time transcription with live speaker identification during meetings and generates high-accuracy, diarized transcripts for post-meeting analysis. By automatically extracting key decisions, action items, and topics without ever requiring cloud connectivity, ZazuAI eliminates manual note-taking, ensures privacy, and significantly enhances meeting productivity and accountability.

Solution & Implementation

1. Engineered a dual-pipeline architecture to simultaneously provide low-latency real-time transcription and high-accuracy post-processed analysis, all hosted on-premise for maximum data security.
2. Integrated Whisper.cpp and PyAnnote's DIART for the real-time pipeline, delivering instant, speaker-identified transcripts during live meetings.
3. Utilized WhisperX with forced alignment and PyAnnote in the post-processing pipeline to generate a definitive, word-level accurate transcript with superior speaker segmentation.
4. Developed a custom Llama.cpp-powered engine with engineered prompts to automatically generate structured meeting minutes, including summaries, action items, and topics.

TechStack

- Python
- Whisper.cpp / WhisperX
- PyAnnote.audio & DIART
- Llama.cpp (Open-source LLM)
- FastAPI, gRPC, Web Sockets
- Docker
- PyAudio (Audio Processing)

Application Showcase



Customer Profile

A leading enterprise client from the UAE seeking a secure, on-premise AI solution.

Industry: Enterprise Technology

Country: United Arab Emirates (UAE)



Results

ZazuAI was successfully developed and deployed as a custom on-premise solution for the UAE-based client. The system achieves 96% transcription accuracy and 80% diarization accuracy on full audio processing, meeting critical performance thresholds while providing superior data sovereignty. The client now benefits from a secure, automated tool that captures meeting intelligence, ensures no critical detail is lost, and tracks every action item, all within their own private infrastructure. This project demonstrates deep expertise in building full-stack, privacy-focused AI applications for discerning international clients.