Personal Assistant — Project Brief

Overview

Personal Assistant is a production-grade, multimodal AI assistant prototype designed for real-time, consumer-facing interaction. It combines live speech recognition, conversational state management, and a persistent avatar into a single coordinated system capable of low-latency voice interaction and synchronized audiovisual response.

The project was developed for a commercial client as a functional prototype intended to demonstrate feasibility, scalability, and readiness for retail deployment.

Problem Addressed

Many consumer AI products fail not because of model quality, but due to:

- Unreliable audio capture and device handling
- Latency and synchronization issues between speech, reasoning, and playback
- Fragile conversational state management
- Poor integration between backend logic and front-end user interaction

Personal Assistant addresses these issues by treating audio, conversation, avatar behavior, and user interaction as parts of a single system rather than separate components.

Solution

A tightly integrated, event-driven architecture that enables:

- Continuous microphone capture with robust device handling
- Streaming speech-to-text with gating and turn control
- Scene-based conversational logic and state persistence
- Synchronized avatar playback (speech, pauses, gestures)

- Explicit user-gesture handling for browser and OS audio constraints
- Packaged Windows executable for non-technical end users

Key Contributions

- Designed and implemented end-to-end system architecture
- Built a real-time STT pipeline with low-level audio handling
- Implemented conversational state and scene management
- Coordinated avatar playback with speech timing and pauses
- Solved cross-platform audio issues through diagnostic tooling
- Delivered a standalone Windows executable suitable for demos and testing

Technical Stack

- **Languages:** Python (asyncio, WebSockets)
- **Speech-to-Text:** Deepgram (streaming)
- **Text-to-Speech:** ElevenLabs
- **Avatar Rendering:** D-ID (WebRTC)
- **Audio:** sounddevice, low-level PCM handling
- **Frontend:** Browser-based avatar interface with gesture gating
- **Deployment:** Windows standalone executable

Status

Prototype completed and delivered.

Currently undergoing final testing prior to retail release by the client.

Relevance

This project demonstrates practical experience building **real-time, user-facing AI systems** where engineering constraints, UX, and system reliability matter as much as model capability.

Source code and architectural documentation available upon request.
