

Encode

Visualizing the Pipeline and Features

Encode aims to revolutionize customer engagement through an advanced AI phone agent. This presentation will provide an overview of our approach, focusing on the project's pipeline and key features that set it apart in the market.

Problem Statement

Overview of Challenges: The primary goal is to build an AI phone agent capable of:

- **Cold Calling Customers:** Initiating calls to potential clients effectively.
- **Holding Meaningful, Real-Time Conversations:** Engaging customers in natural dialogue.
- **Closing Sales with Natural Voice Interactions:** Utilizing conversational techniques to finalize sales.
- **Handling Dynamic Function Calls and Multiple Concurrent Calls:** Managing several calls simultaneously while adapting to varying customer needs.

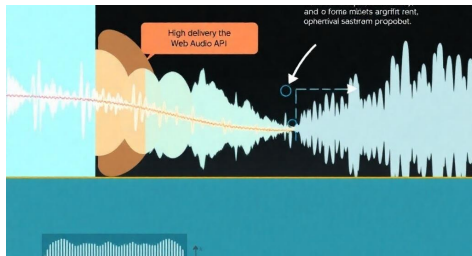


Pipeline Overview



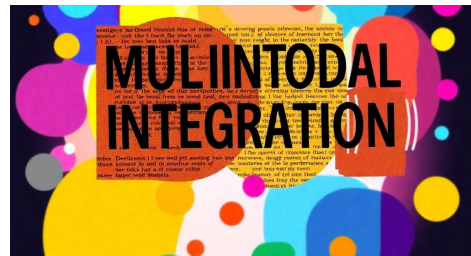
Step 1: Real-Time Interaction Management

Establishes a WebSocket connection to Google's Gemini API, facilitating the sending and receiving of real-time audio and text data, and emits events for content updates and tool calls to ensure fluid communication.



Step 2: Audio Processing

Converts and buffers PCM16 audio for seamless playback, managing the lifecycle of audio streams using the Web Audio API to ensure high-quality audio delivery.



Step 3: Multimodal Integration

Effectively handles simultaneous text and audio streams to provide robust and engaging interactions.

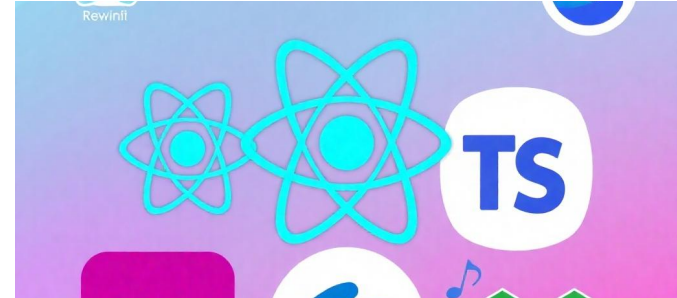
Key Features and Technologies



Distinguishing Features

The AI phone agent is equipped with unique features enabling superior performance:

Real-Time Bidirectional Communication: Utilizes WebSockets for instant interaction. **Multimodal Data Processing:** Integrates both text and audio for a comprehensive communication experience. **Custom Audio Pipeline with Dynamic Worklets:** Ensures adaptable and efficient audio handling.



Underpinning Technologies

The project leverages cutting-edge technologies, including: **React:** For intuitive user interface development. **Google's Gemini Multimodal Live API:** Facilitates AI interactions. **Web Audio API:** Manages audio processing tasks. **TypeScript and Node.js:** Forms the backbone of the implementation stack.

Data Flow in the System

Data Flow Diagram: The system's data flow is structured as follows:

- **Input:** User Interaction (audio/text) is captured and sent to the Gemini API via WebSocket.
- **Processing:** Real-time interaction events are generated (e.g., tool calls, content updates).
- Audio streams are buffered and processed using the Web Audio API.
- **Output:** Processed audio/text is displayed in the user interface, ensuring a seamless interaction experience.



Summary

Key Takeaways: The Encode Project demonstrates:

- **Low-Latency, Real-Time Communication:** Ensures instant interactions with customers.
- **Integration of Advanced AI Tools:** Leverages multimodal interactions for enhanced engagement.
- **Robust Architecture:** Combines front-end technologies with sophisticated audio pipelines, positioning the project for success in the competitive landscape.

