



# Pratibimb: A Real-Time AI Twin for Conversational Intelligence and Digital Presence

Tanishq Kokane<sup>1</sup>, Sagar Karegaonkar<sup>2</sup>, Yogesh Landge<sup>3</sup>, Pratik Khandge<sup>4</sup>, Prof. Shalu Saraswat<sup>5</sup>

<sup>1,2,3,4</sup>Student, <sup>5</sup>Professor Department of Information Technology  
PDEA's College of Engineering, Pune, India

**Abstract**— Pratibimb is an AI-powered digital twin designed to replicate human conversational behavior in real-time. With the growing demand for intelligent assistants that can operate autonomously in meetings, calls, and online interactions, Pratibimb aims to bridge the gap between human communication and artificial intelligence. The system integrates real-time speech recognition, natural language processing, contextual memory, and text-to-speech synthesis to create a seamless conversational experience. Unlike traditional assistants, Pratibimb supports interruption handling (barge-in), dynamic context switching, and personalized response generation.

The architecture consists of microphone input processing, streaming speech-to-text via WebSockets, LLM-based reasoning, and low-latency speech output. It is designed to function as a virtual participant in platforms such as Microsoft Teams, capturing transcripts, summarizing discussions, and responding intelligently. This paper presents the system design, implementation methodology, challenges, and potential applications of Pratibimb. Experimental results indicate improved conversational naturalness and reduced latency compared to conventional AI assistants.

**Keywords**— AI Twin, Conversational AI, Real-Time Systems, Speech Recognition, Text-to-Speech, Human-AI Interaction, Virtual Assistant, Meeting Intelligence

## I. INTRODUCTION

The evolution of artificial intelligence has led to the development of conversational systems that can interact with users in natural language. However, most existing systems lack real-time responsiveness, personalization, and contextual awareness. The concept of an AI Twin extends beyond traditional assistants by creating a digital replica capable of acting, speaking, and responding like its human counterpart.

Pratibimb is designed as a next-generation AI Twin that enables users to delegate communication tasks such as attending meetings, answering queries, and summarizing discussions. The system mimics human conversational patterns and supports real-time interaction with minimal delay. This paper explores the architecture and implementation of Pratibimb, focusing on its ability to deliver a natural, human-like conversational experience.

## II. MOTIVATION

The rapid advancement of artificial intelligence has significantly improved the way humans interact with machines, yet most existing systems still rely on rigid, turn-based communication that lacks the fluidity of natural human conversation. Voice assistants today often fail to provide a seamless experience due to latency, lack of contextual continuity, and inability to handle real-time interruptions. This creates a gap between human expectations and system capabilities. The motivation behind Pratibimb is to bridge this gap by developing an AI Twin that can communicate as naturally as a human, understanding speech in real time, responding intelligently, and adapting dynamically to user interruptions. In an era where digital presence is becoming increasingly important, there is a growing need for systems that can represent individuals, assist in tasks such as meetings and decision-making, and function as intelligent companions. Pratibimb aims to address these challenges by combining real-time speech processing, advanced language models, and low-latency communication into a unified system. This work is driven by the vision of creating AI that does not just respond, but truly interacts, making human-computer communication more intuitive, efficient, and human-like.

## III. OBJECTIVE

The primary objective of this work is to design and develop Pratibimb, a real-time AI Twin capable of enabling natural, human-like interaction through seamless voice communication. The system aims to integrate continuous speech recognition, intelligent language processing, and real-time speech synthesis into a unified low-latency pipeline. A key objective is to implement a barge-in mechanism that allows users to interrupt the system during response generation, thereby enhancing conversational fluidity and realism. Additionally, the system seeks to maintain contextual awareness across interactions, ensuring coherent and meaningful responses.



Another important goal is to build a scalable and modular architecture that can be extended for applications such as virtual meeting assistants, personal digital companions, and productivity tools. Overall, the objective is to bridge the gap between human communication patterns and machine interaction by creating an AI system that behaves more like a conversational partner than a traditional assistant

#### IV. RELATED WORK

Several conversational AI systems have been developed, including virtual assistants like Siri, Alexa, and Google Assistant. While these systems provide voice interaction, they are limited in contextual understanding and real-time adaptability.

Recent advancements in large language models (LLMs) such as GPT and Claude have significantly improved language understanding. Additionally, speech processing technologies have enabled real-time transcription and synthesis. However, integrating these components into a seamless AI twin remains a challenge.

Pratibimb builds upon these advancements by combining:

- Real-time streaming speech recognition
- Context-aware LLM processing
- Interrupt-driven interaction (barge-in)
- Personalized conversational memory

#### V. PROPOSED SYSTEM

The proposed system, Pratibimb, is a real-time AI Twin designed to enable seamless and natural human–AI interaction through an integrated voice-based architecture. The system combines continuous audio capture, real-time speech recognition, intelligent language processing, and natural speech synthesis into a unified pipeline. Unlike conventional systems that operate in a turn-based manner, Pratibimb follows a streaming approach where audio data is processed continuously, allowing immediate response generation with minimal latency.

The architecture is modular, consisting of distinct components for microphone input, streaming speech-to-text conversion, large language model-based response generation, and text-to-speech output. These components communicate through a WebSocket-based layer, ensuring efficient and low-latency data transmission.

A key feature of the proposed system is the implementation of a barge-in mechanism, which enables the system to detect user interruptions during speech output and immediately halt the ongoing response. This significantly enhances the naturalness and responsiveness of interaction, making conversations more dynamic and human-like. Additionally, the system maintains contextual continuity across interactions, allowing it to generate coherent and relevant responses. The proposed system is designed to be scalable and adaptable, supporting integration with external platforms such as virtual meetings and productivity tools. Overall, Pratibimb represents a shift from static voice assistants to interactive AI companions capable of real-time, intelligent communication.

#### VI. SYSTEM ARCHITECTURE

The Pratibimb system is composed of multiple interconnected modules:

**Microphone Input Module:** Captures real-time audio using hardware devices and converts it into Base64-encoded chunks for streaming.

**Speech-to-text Engine:** Uses WebSocket-based streaming (e.g., Sarvam AI) to convert speech into text with minimal latency.

**Language Processing Unit:** Processes the transcribed text using LLMs such as Claude or Gemini to generate context-aware responses.

**Text-to-Speech Module:** Converts generated responses into natural-sounding audio using TTS systems.

**Memory and Context Manager:** Maintains conversation history and user-specific preferences to ensure personalized interactions.

**Output Delivery System:** Streams audio responses in real-time while allowing interruption (barge-in capability)

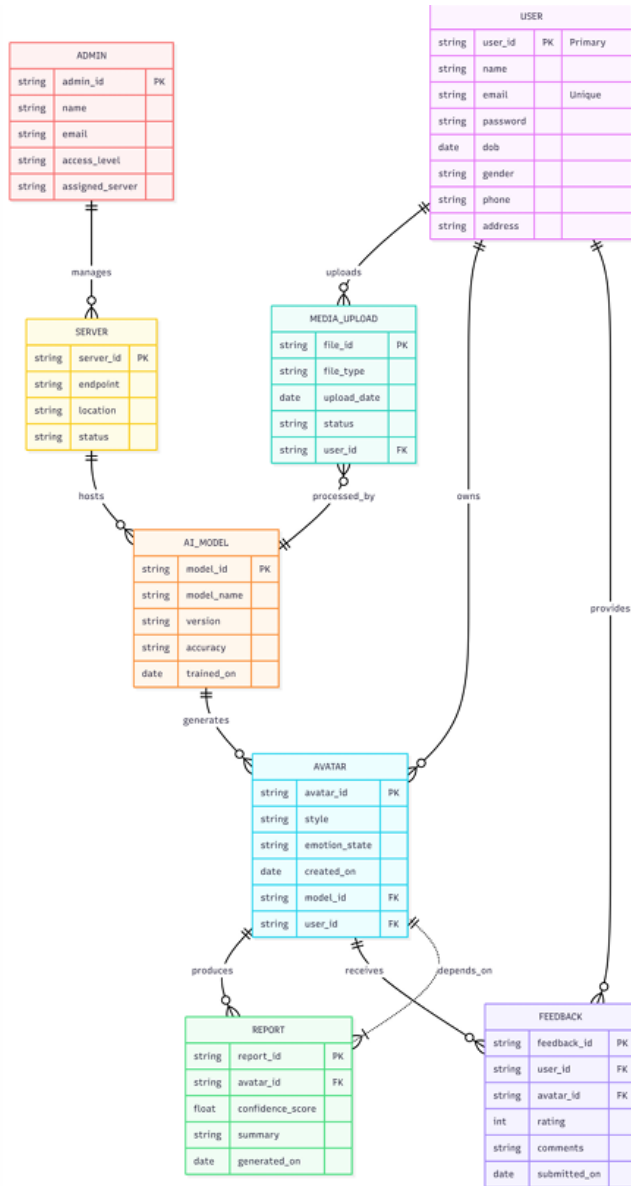


Fig. 1. High-Level System Architecture Diagram of Pratibimb

### VI.A Methodology

- **Real-Time Streaming:** The system uses asynchronous communication to ensure continuous data flow between modules.
- **Barge-In-Handling:** Pratibimb supports interruption detection. When the user starts speaking:
  1. Current audio output is Stopped
  2. Input stream is prioritized
  3. New response is generated instantly

**Context Retention:** Conversation history is stored and used to generate meaningful responses, enabling continuity.

### VII. CONCLUSION

Pratibimb represents a significant step toward the development of AI twins capable of real-time, human-like interaction. By integrating speech processing, language understanding, and personalized memory, the system provides a seamless conversational experience. Its ability to handle interruptions and maintain context makes it a powerful tool for modern communication needs.

### VIII. REFERENCES

- [1] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [2] OpenAI, "GPT Models and Applications," Available: <https://openai.com>(<https://openai.com>)
- [3] Anthropic, "Claude: AI Assistant for Safe and Scalable Language Models," Available: <https://www.anthropic.com>(<https://www.anthropic.com>)
- [4] Google, "Gemini: Multimodal Large Language Models," Available: <https://deepmind.google>(<https://deepmind.google>)
- [5] RFC 6455, "The WebSocket Protocol," Internet Engineering Task Force (IETF), 2011.
- [6] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd Edition, Pearson, 2023.
- [7] J. Allen, "Natural Language Understanding," Benjamin/Cummings Publishing, 1995.
- [8] K. Richmond, R. Clark, and S. Fitt, "Robust Text-to-Speech Synthesis: A Review," *IEEE Transactions on Audio, Speech, and Language Processing*, 2019.
- [9] H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition," *INTERSPEECH*, 2014.
- [10] M. McCandless, E. Hatcher, and O. Gospodnetić, *Lucene in Action*, Manning Publications, 2010.