Minor Project (ARP 455)

SYNOPSIS

EduVi : AI Voice Agent for Education

Submitted by:
Naveen Sharma (20719051622) AI&ML-B2
Ankit Sharma (20319051622) AI&ML-B2

Under the Supervision of
Mr. Anupam Lakhanpal
Assistant Professor

University School of Automation and Robotics
East Campus, GGSIP University
Surajmal Vihar, New Delhi - 110092

INTRODUCTION

The rise of artificial intelligence has created new paradigms for personalized and accessible education. This project involves the development of a full-stack Software as a Service (SaaS) application that functions as an AI-powered voice agent. The primary objective is to create an interactive platform where users can engage in real-time, two-way voice conversations with an AI to achieve various educational goals, such as conducting mock interviews, learning new topics, practicing for Q&A sessions, or mastering a new language.

The application is built on a modern technology stack, featuring a Next.js and React frontend for a responsive user experience. The backend logic and database are managed by Convex, which supports real-time data synchronization. The core AI capabilities are integrated through specialized third-party services: AssemblyAI for streaming speech-to-text, OpenRouter for access to powerful Large Language Models (LLMs) like Gemini, and Amazon Polly for lifelike text-to-speech conversion. The final product is a deployed, cloud-native application that offers users a dynamic and effective learning tool.

PROBLEM STATEMENT

Traditional educational methods and self-study tools often lack the immediate, interactive feedback crucial for effective learning and skill development. Students and professionals seeking to practice skills like interviewing or conversational language often face significant barriers, including:

- High Cost and Inaccessibility: Access to human tutors, coaches, and language partners can be expensive and logistically challenging.
- Lack of On-Demand Practice: Scheduling sessions with experts is often inconvenient, preventing spontaneous or frequent practice.
- Impersonal Learning Experience: Standard e-learning platforms may not offer the conversational practice that builds real-world confidence and proficiency.

This project addresses these challenges by creating an AI voice agent that provides a free, on-demand, and personalized conversational learning experience. It simulates interactions with a human expert, offering a safe and accessible environment for users to practice, learn, and receive instant feedback at their own pace.

LITERATURE SURVEY

As this is a development-oriented project based on the implementation of existing technologies, a formal academic literature survey was not conducted. However, the project's architecture and functionality are informed by the current state of AI in educational technology.

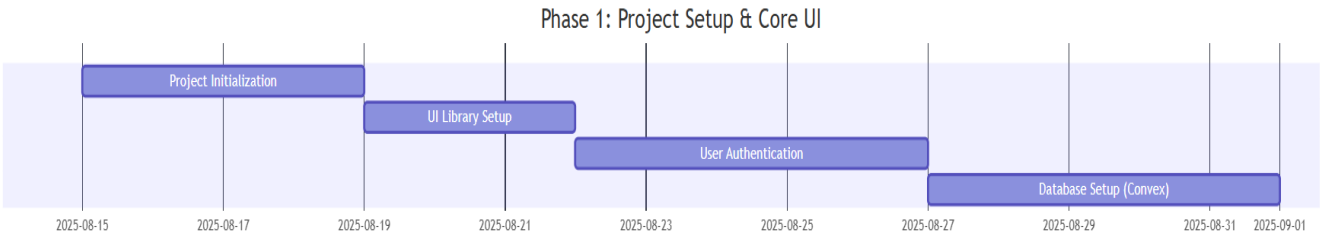
The project leverages several key technological advancements:

- **Large Language Models (LLMs):** The core conversational intelligence is powered by models like Google's Gemini, accessed via platforms like OpenRouter. These models have demonstrated a profound ability to understand context, generate relevant information, and simulate human-like dialogue.
- **Real-time Speech Recognition:** Services like AssemblyAI's streaming speech-to-text API are critical. Their high accuracy and low latency allow for seamless conversion of spoken words into text, making real-time interaction possible.
- **Cloud-based Text-to-Speech (TTS):** Platforms such as Amazon Polly provide high-quality, natural-sounding voices that are essential for creating an engaging and believable conversational agent.

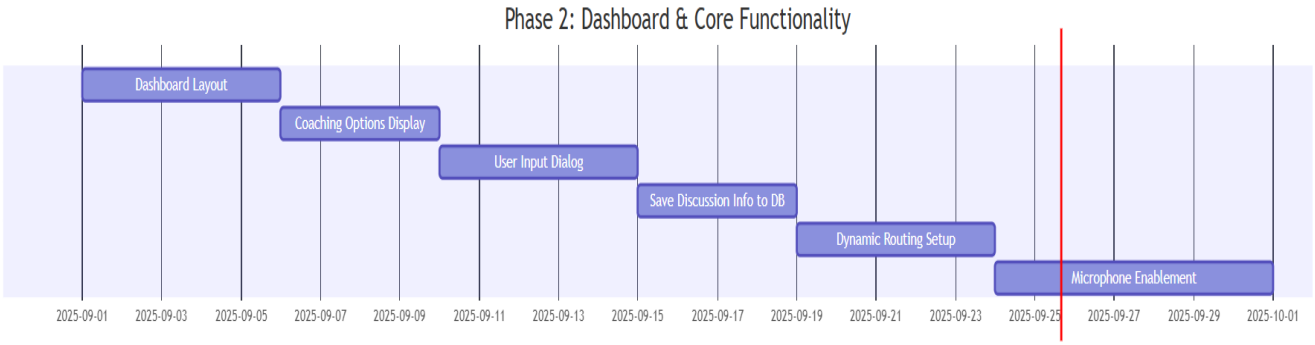
This project integrates these technologies into a singular framework, similar in concept to specialized applications like the language-learning app Duolingo or dedicated interview preparation software. Its unique value lies in offering a versatile, multi-purpose educational agent built on a modern, replicable tech stack.

GANTT CHART

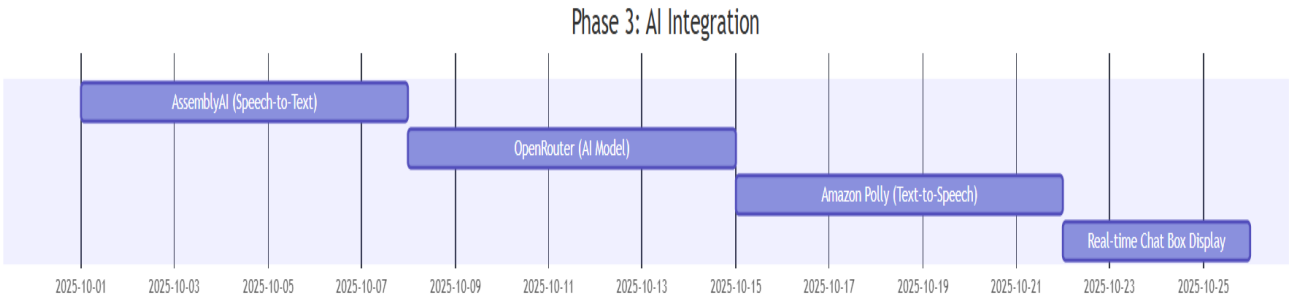
Phase 1: Project Setup & Core UI:



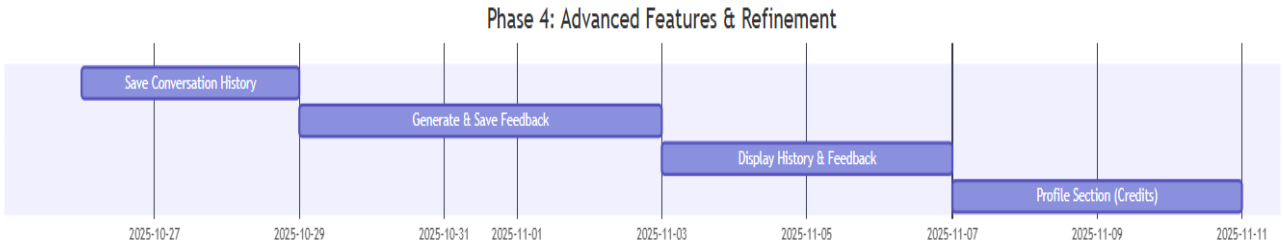
Phase 2: Dashboard & Core Functionality:



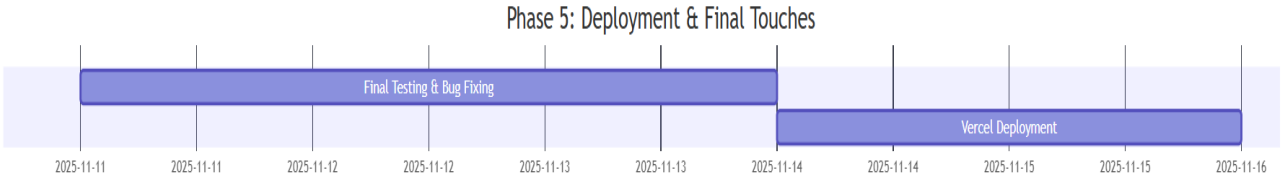
Phase 3: AI Integration:



Phase 4: Advanced Features & Refinement:



Phase 5: Deployment & Final Touches:



HARDWARE & SOFTWARE REQUIREMENTS

Software Requirements:

- **Development Environment:**
  - Operating System: Windows, macOS, or Linux
  - Code Editor: Visual Studio Code (or similar)
  - Runtime Environment: Node.js (v18 or later)
  - Version Control: Git
  - Web Browser: Google Chrome, Firefox, or Safari
- **Application Stack:**
  - **Frontend:** React, Next.js 15
  - **Styling:** Tailwind CSS v4, ShadCN UI
  - **Database:** Convex
  - **Authentication:** Stack
  - **AI Services:** AssemblyAI, OpenRouter, Amazon Polly
  - **Libraries:** Axios, Moment.js, React Markdown, React Webcam

Hardware Requirements:

- **For Development:**
  - A standard development computer with at least 8 GB RAM and a modern processor.
- **For End-User:**
  - A computer, tablet, or smartphone with a modern web browser.
  - A functional microphone (essential for voice input).
  - A stable internet connection.
  - A webcam (optional, for video display).