

An Intelligent Multi-Modal AI Framework for Interview Simulation with Speech Recognition and Real-Time Feedback

Dr. Siddhartha Roy

Department of Computer Science and
Engineering (Internet of Things, Cyber
Security, Blockchain Technology)
University of Engineering and Management,
Kolkata
siddhartha_r24@yahoo.com
ORCID: 0000-0003-0278-2731

Ahona Ghosh

Department of Computer Science and
Engineering (Internet of Things, Cyber
Security, Blockchain Technology)
University of Engineering and
Management, Kolkata
ahonaghosh95@gmail.com
ORCID: 0000-0003-0498-285X

Nitu Saha

Department of Computer Science and
Engineering (Internet of Things, Cyber
Security, Blockchain Technology)
University of Engineering and
Management, Kolkata
saha.nitu.nitjsp@gmail.com
ORCID: 0009-0001-1209-2884

Dr. Susmita Biswas

Department of Computer Science and
Engineering (Internet of Things, Cyber
Security, Blockchain Technology)
University of Engineering and
Management, Kolkata
susmita.biswas@uem.edu.in

Rangon Sarkar

Department of Computer Science and
Engineering (Internet of Things, Cyber
Security, Blockchain Technology)
University of Engineering and
Management, Kolkata
sarkarrangon.2019@gmail.com
ORCID: 0009-0005-7622-1795

Tejashri Chate (Srivastava)

Department of Computer Science and
Engineering (Internet of Things, Cyber
Security, Blockchain Technology)
University of Engineering and
Management, Kolkata
tejashri.chate@gmail.com

Abstract: The creators of this system also empowered it with an intensity and versatility that can mimic the real interview situations pretty close. With the help of voice recognition tailored to the job role, exclusive question repositories, and dynamic feedback systems, the system is able to produce very credible and lively scenarios that are very helpful in the preparation and support of the candidates as well. Candidates are divided into three main categories, namely: technical skill, language skill, and emotional intelligence. The user interface frontend is developed using Next.js and the backend processes are handled by Node.js. The architecture of the system is designed in such a way that it can support a high volume of users who can access it quickly and at the same time with almost no delay, and Neon DB handles the databases at an exceptionally fast and serverless like manner for the management of system databases. Among other important findings, one significant discovery is the introduction of the Gemini API, which enables the production of mock interview questions and provides instant feedback as well. The smooth transition from one task to another has not only helped the user to take part in the platform but made the participation also more effective. User testing results together with initial evaluations pointed out that the candidates who went through practice in real interviews were more ready. As per the preliminary evaluations, the system has the potential to recreate the realistic scenario and thus, coupled with user preparedness that is increased, it offers a scalable and convenient interview solution at the same time.

Keywords: *AI in Recruitment, Multi-Modal Interview Simulation, Real-Time Feedback Generation, Speech Recognition*

I. INTRODUCTION

Artificial intelligence (AI) is still on the way to totally changing human interaction in a variety of fields, including education, recruitment, and professional development. Among them, interview preparation stands out as one of the most significant yet least used areas, which usually lacks

realistic, real-world scenarios and prompt, personalized feedback. Traditional mock interview systems usually present a limited set of static questions, very little behavioral analysis, and little customization to specific job roles. Virtual recruitment has increased tremendously in the last few years. Thus, the need for immersive AI-powered interview preparation tools that can judge candidates both on their technical knowledge and soft skills has grown accordingly. The best proposal in this regard is the Intelligent Multi-Modal Interview Simulation Platform (IIMISP), which is a mock-interview platform hosted on the cloud using the latest AI technologies to create the exact conditions of a real-world interview. The system provides Real-Time Performance Feedback (RTPF), voice recognition for spoken responses, and dynamic question generation (DQG) based on job roles through the Gemini API. Also, it incorporates behavior monitoring through camera tracking and provides secure, role-based access via Clerk, all of which are supported by a scalable cloud infrastructure powered by Drizzle ORM and Neon DB.

A. Research Gap

The majority of current AI-based interview platforms do not fully support multi-modal performance evaluation, despite recent improvements. They frequently fall short of combining behavioral feedback, personalized question creation, and speech-to-text input into a single system. Furthermore, these platforms are rarely robust or scalable enough to guarantee data security and real-time user interaction. Academic literature has not yet thoroughly examined a comprehensive solution that combines cloud-native scalability, user emotion recognition, and natural language understanding.

B. Research Objectives

The following are the study's main goals.