Grenze International Journal of Engineering and Technology, June Issue

AI-Yoda: Your AI Mentor and Interviewer

Aryan Verma[1], Ashwin Kaushik[2], Ayushi Pal[3], Dharvi Dubey[4] and Ravi Shankar Pal[5]
[1-5]ABES Engineering College, Department of Computer Science and Engineering (Data Science), Ghaziabad, India
Email: aryan.22b154111099@abes.ac.in, ashwin.22b1541185@abes.ac.in, ayushi.22b1541021@abes.ac.in, dharvi.22b1541155@abes.ac.in, ravi.shankar@abes.ac.in

***Abstract*— The problem that is addressed by this project is that many job-seekers and students typically encounter difficulties in terms of technical interviews preparation: an inadequate access to realistic, systematic, and practical practice. Conventional mock interviews are difficult to organize, they are different in terms of quality, and they are not always organized and guided by data-driven feedback. The application is based on AI that mimics the behaviour of the interviewer, it creates custom prompts to offer an experience of asynchronous practice and acts as an experiment to improve interaction with the app; consequently, users can practice in specific areas to build strengths and learn with feedback over time.**

***Keywords*— Artificial Intelligence, AI Interview System, Natural Language Processing (NLP), Speech-to-Text (STT), Text-to-Speech (TTS), Career Mentoring, Adaptive Question Generation.**

I. INTRODUCTION

Training to cope with technical interviews is both time consuming and ineffective: job seekers will have to organize with their colleagues or tutors, they will have to work with inaccurate feedback, and they will have a hard time recreating the stress and the demands of interview situations. The proposed project will fill in that gap by providing a scalable, AI-backed mock-interview application that popularizes high-quality practice on an on-demand basis.

In its essence, the app generates interviewer behaviour as it would happen in real life by using genera-tive models to generate role specific questions, follow-ups and probing prompts. One may conduct sessions asynchronously or live and every interaction is recorded to review. The system also generates un-biased, well-structured feedback on strengths, gaps (technical knowledge, communication, problem-solving), and recommendations on specific follow-up actions after a session. This converts informal practice into an iterative practice into an iterative line of learning in which progress is replicable and quantifiable. It is a platform that is accessible and extensible. It has been developed using Next.js and TypeScript, and it has a refined web interface where it is possible to conduct interviews, peruse previous sessions, and access feedback. Firebase takes care of authentication and persistence in order to enable users to store the history of their sessions and be accessible throughout the devices. A low API includes prompts and response evaluation (giving it the opportunity to be customised by role, difficulty and focus area (algorithms, system design, behavioural questions, etc).

In addition to personal preparation, the project has several structured educator and recruiter use cases: bulk practice assignments, collaborative feedback templates, and anonymized cohort-level metrics. Combining realistic simulation, playback recordings, and automated actionable feedback, the app removes the friction of the interview preparations and allows users to train to become confident and competent progressive.

Grenze ID: 01.GIJET.12.2.12_24

II. Literature Review

The latest innovations in the sphere of artificial intelligence (AI), machine learning, and natural language processing resulted in the creation of automated ap-plications that facilitate the process of interview prep-aration among the candidates. In the past, the speech recognition and NLP techniques have been used to measure communication skills, articulation and con-tent of verbal responses in real time [10], [11]. These systems have also been improved with emotion de-tection and sentiment analysis that determines the level of confidence, tone, and behaviour patterns to improve the understanding of interview performance [12], [13], [17]. Career support systems that are based on AI have also been created to examine the abilities of the candidates and suggest them to the appropriate job positions, thus providing them with a personal-ized enhancement route [14]. Conversational inter-view agents are now capable of creating context-sensitive and dynamically adaptive questions, at the same time being better at decoding user intent in a simulated interview setting with the introduction of large language models [15]. Computerized assess-ment systems are also useful in that they offer objec-tive feedback and automatic performance measure-ment, which help applicants understand their weak-nesses and strengths [16].

Although this has been made, the majority of the current systems are centered on single functionalities instead of providing the end-to-end experience of interview preparation. The recent literature focuses on interview evaluation systems based on AI with speech-to-text processing and semantic similarity analysis combined with face emotion recognition, but its applicability is largely dependent on the quality and variety of the training data, posing concerns of bias and lack of human judgment [1], [7]. The studies of human-AI collaboration emphasize the importance of large language models in adaptive questioning and rapport management and raise ethical issues concern-ing transparency, trust, and control [2]. Voice, gesture and posture-based behavioural analysis enhances both technical and soft-skill analysis, but these sys-tems are typically subject to scalability as well as overfitting in real time applications [3]. In like man-ner, AI-powered mock interviewing systems improve confidence in candidates and communication compe-tence but often do not offer cultural adaptability, sim-ulated environments, and response analysis [5], [6]. Although these tools facilitate the scalable hiring and diversity programs, the issues regarding fairness, validity, and assessment depth remain unaddressed [7] .

III. Methodology

A. Overview

In this section, the design and implementation of AI YODA, an adaptive, voice-based simulation of a mock interview platform, is introduced to introduce real-world conditions of interviews and structure mentoring. The system integrates both large language models and speech technologies and real-time com-munication to facilitate interactive interview session. The suggested methodology is based on adaptive question generation, multimodal feedback provision, low-latency interaction, safe management of user information, and sustained monitoring of user pro-gress during sessions.

B. Objectives

The main purpose of the platform is to perform simu-lated AI-based interviews that will generate domain-specific questions and re-scale their difficulty de-pending on the answers of the candidates. It can be done through fine-tuned language models and prompt-engineering strategies [2][3]. Besides, the system will also seek to offer personalized coaching via real time cues and detailed post session feedback that will cover content quality, response structure, delivery style and confidence. The text-to-speech (TTS) and speech-to-text (STT) technologies are combined to allow natural speech and audio-based responses [4][10]. To facilitate a comfortable conver-sation experience, the platform will have low interac-tion latency with streaming communication protocols like WebSockets and Vapi [2]. Lastly, the system promotes longitudinal learning, whereby interview sessions are recorded and the performance trend ana-lyzed to create tailored learning plans.

C. System Architecture

AI YODA has a layered client-server architecture. The Next.js frontend is written in TypeScript and offers the interactive user interface, session controls, and transcripts and feedback views. Communication layer is a core part of the Vapi Web SDK and Web-Socket-based protocol that real-time audio streaming and message exchange can be performed with less round-trip delay by using a specific communication layer. The backend, written in Node.js and Express, handles the coordination of the session, ingestion of transcripts, feedback generation, and analytics pro-cessing. Permanent data storage is managed with MongoDB that is used to store user profiles, interview transcripts, timestamps, model results and long-term performance indicators. The

security is implemented by means of the JWT-based authentication, the en-crypted transport with the help of HTTPS and WSS, the encrypted storage of sensitive information, and access controls basing on roles. An example of a schematic view of the audio streams, speech services, language model inference and backend analytics pipelines (not shown here) follows.

D. Fundamental Modules And Responsibilities

The AI Interviewer module will handle creating initial interview questions and adaptive follow-up prompts with the help of language-model-driven agent. Its architecture can be reused in multiple functions and in multiple difficulty levels (either by code logic, or by workflow definitions). The AI Mentor module compares the answers of the candidates and generates the feedback in the form of the structured feedback about the communication quality, problem-solving strategy, technical correctness, and tone. Session-level guidance is combined and contains a pre-determined JSON schema to maintain consistency. The speech and NLP pipeline consists of streaming STT services to do transcription, TTS elements to render spoken prompts and feedback, and NLP elements to recognize intent, semantic analysis, and dialogue state management. Qualitative indicators used by performance analytics modules include re-sponse latency, length of answer, frequency of filler-words, lexical diversity, metrics of fluency, senti-ment-based confidence proxies and domain-specific metrics of correctness.

E. Workflow

The process starts with user onboarding where the applicants will sign up, build profiles, and choose their desired position, area, and the level of difficulty. In the process of session startup, contextual infor-mation is compiled and real-time audio streaming is created by the frontend with the speech services and the backward. During the live interaction stage, the interviewer prompts can be obtained as audio, with the user answers being relayed to the STT service and the results being turned into transcripts with timestamps. The interviewer model is a continuous approach that utilizes the context of recent dialogue to identify relevant questions to be followed. The immediate feedback on a micro-level can be given after individual answers. After the session is complet-ed, transcripts and metadata are placed in a queue to be processed in the background to provide a struc-tured feedback with strengths, weaknesses and rele-vant suggested practice activities. The last phase of the review gives summaries of progress and recom-mendations to follow in the next sessions.

F. Implementation Details

Adaptive question generation Adaptive question gen-eration is performed with fine-tuned language models or prompt-conditioning models on the basis of role templates and user history [2][3]. The output of feed-back has a constant JSON structure that includes the effectiveness of communication, technical knowledge, problem-solving skills, confidence, and cultural fit, enabling the homogeneous display of the UI and analysis of the research results. The transcript entries contain the identity of the speaker, UTC time, transcribed text, and the confidence score of the STT engine, and option audio reference. The backend supports scalability with worker queues and horizon-tally scalable services that can support various simul-taneous sessions and offline analysis that can be de-ferred. External STT, TTS, and LLM are abstracted such that they can be replaced by a different provid-er.

G. Evaluation And Metrics

System level performance is measured in terms of end-to-end latency, word-error rate of the STT com-ponent, subjective mean opinion scores of TTS out-put, system uptime and load throughput. The perfor-mance at the level of the users is measured with re-spect to the frequency of sessions, the increase in the response coherence, the decrease in the use of filler-words, the decrease in response time, and the correla-tion of automated feedback and human reviewers. The human-in-the-loop validation is included with regular expert reviews, scoring threshold calibration, and user studies aimed at the evaluation of perceived realism and practical usefulness.

H. Data Security, Privacy, Ethics

Privacy settings on the platform include the ability to remove transcripts of interviews, use analytics, and retrieve personal data. The policy of data retention and encryption at rest are implemented to secure sen-sitive information. The system has access control, with JWT-based authorization, and role-based access control to allow coaches and administrators, as well as anonymity of the data used to create aggregate reports. The mitigation measures of bias include peri-odic audits of evaluation results, the use of multiple types of annotators in the model fine-tuning process, and clear explanations of automated feedbacks to lessen unwarranted evaluations.

I. Limitations of Future Works

Accent variations and noise in the background can influence the accuracy of speech-to-text and affect the quality of the transcription. The continuous miti-gation efforts involve noise-reduction methods, adap-tive speaker, and optional manual correction process-es. The model bias and assessment accuracy should be monitored continuously and controlled by people to reduce the inaccurate or unfair feedback. It is planned to be further integrated with learning man-agement systems and job portals, hand-over to hu-man coaches and enhanced lesson planning, and randomized interview scenarios in order to enhance system robustness.

J. Summary

The suggested methodology combines adaptive lan-guage-model-motivated inquiry, live speech inter-face, guided feed back mechanisms, and time-span-based performance evaluation into a safe and low-latency system structure. The platform will help facili-tate quantifiable skill acquisition with repeated and reflective practice by merging automated interview simulation with actionable mentoring.

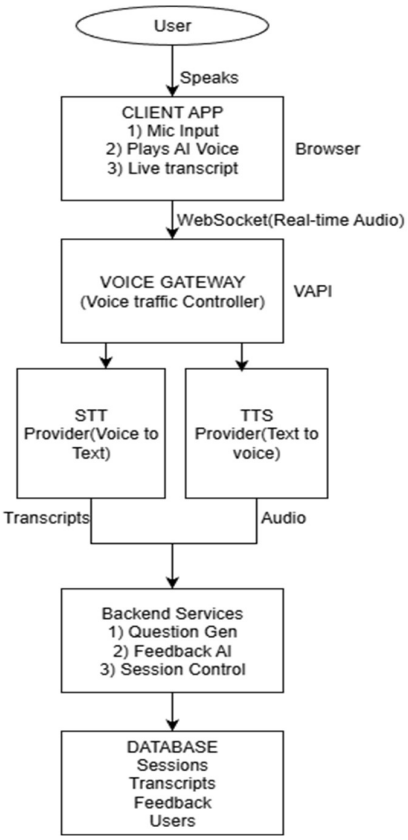

Fig. 1

IV. Design and Implementation

Design and Implementation outlines the architecture, workflow, modules, and implementation decisions of AI-YODA, a voice-driven, adaptive, and adaptable mock-interview system which combines large lan-guage models, streaming speech applications and a low-latency client-server architecture.

A. Overview and Workflow

AI-YODA deploys call interview loop: an end-to-end voice call sequence which will include:

- User onboarding
- Interview generation
- Live audio exchange
- Transcript capture

- Synthesis of automatic feedback.
- Persistence and review

B. Typical Runtime Steps

User registration and log in: Protected access of user accounts and session data is provided using secure authentication.
Interview generation: A language model uses do-main- and level-specific question sets which are the generation according to user preferences.
Session start and streaming: The client initiates a real time audio session and the session uses the commer-cial speech of the interviewer to stream and records the speech of the candidate.
Transcript capture and feedback trigger Live tran-scripts are recorded; when the session is completed, transcripts can be submitted, where they are then generated into a structured feedback.
Review and analytics Structured feedback and ses-sion metrics are saved and provided to be reviewed later longitudinally. System Architecture Its platform is based on a layered and modular architecture:

C. System Architecture

Its platform is based on a layered and modular archi-tecture:
Presentation layer Web frontend with session con-trols, live transcripts and feedback review.
The communication layer: Voice gateway that liaises between STT/TTS and transmits evented transcript messages to real-time UI updates.
Application layer: Server elements to coordinate question generation, feedback synthesis and analytics tasks.
AI layer: LLMs will generate interview content and formatted assessments; speech processing will be done by STT/TTS suppliers.
Data layer: A document store stores interview metadata, transcripts, and structured feedback.
Security layer: The user data and APIs are secured by token-based authentication and encrypted transport as well as access controls.

D. Core Modules

Key main functional modules will be:
- Interview Generator This is a tool that cre-ates questions sets given prompt condi-tioned by role, level, and technical focus via LLM prompts and templates.
- Real-Time Interviewer: Audio streams the interviewers and logs verbal answers; forces intensive follow ups depending on the latest conversation.
- Transcript Manager: Summarizes transcripts in terms of timestamps and tagging of speakers and adds STT confidence metada-ta.
- Feedback Synthesizer (AI Mentor): This model takes in transcripts and produces structured feedback (category scores, strengths, improvement items, final assess-ment) using constrained outputs of LLM software.
- Analytics Engine: Works out performance data (response time, filler rate, fluency proxies, lexical diversity) and trends be-tween sessions.

E. Implementation Choices

Dual session configurations: Both in-application in-terviewer and external workflow definition Similarly to full-fugue experiments and general deployments Supported The dual session modes allow both in-application interviewer configurations and external workflow definitions.
Structured feedback schema: A typed feedback schema is applied to give machine-readable feed-back, which could be stored and analyzed. Reactive UI: By emitting transcript events and speech events to the client, it is possible to display almost real-time and provide a responsive user interface.
Provider abstraction: STT/TTS/LLM providers STT/TTS/LLM providers Can be decoupled through configuration to allow replacement of vendors with equivalent application logic.

F. Evaluation And Metrics

Measures gathered could be under system and user:
System metrics Round-trip latency, STT word-error rate, TTS quality (MOS), throughput, and reliability.
User measures: Coherence enhancement, filler words minimization, the latency aspect of responses, and agreement between the automated feedback and the human raters.

Validation: Infrequent human-in-the-loop inspection and user research to tune automated scoring and es-timate perceived realism and usefulness.

G. Limitation and Future Work

There are a series of improvements that can be made to increase the effectiveness of the system. To en-hance the speech-to-text robustness, improvement of the accent variations and background noise can be done by speaker-adaptation and noise-reduction methods. To enhance model fairness, regular bias audits can be provided and the feedback process can be implemented using human validation to reduce the possibility of bias. Scalability can be approached to meet large scale use by implementing asynchronous worker queue and batch processing systems which can effectively deal with high concurrent load. Also, advanced changes that can be added to the system include lesson plans, easy transition to human coach-es where needed and closer integration with already present learning systems and recruitment systems.

V. CONCLUSION

AI-YODA illustrates a rational implementation of enhancing the interview preparedness through inte-grating adaptive language models, streaming speech technologies, and structured analytics to a low-latency client server set-up. The system mitigates the limitations of mock interviews (static) through a se-ries of domain-specific, context-dependent questions which are generated, natural voice conversation, and machine-version feedback that may be obtained and acted upon to aid in the iterative learning process. The methodological approach to the project is a com-bination of real-time speech processing (STT/TTS) as well as LLM-based content creation and assessment and longitudinal analytics, which create a closed feedback loop: generate - practice - transcribe - eval-uate - improve. This is a loop that is executed using modular partitions of interview generation, live inter-viewer sessions, transcript management, feedback synthesis, and analytics, that allow flexibility and independence of the providers. Its design principles are built around reproducibility and extensibility: structured feedback schemas can be stored and ana-lyzed; event-driven transcript capture can be used to enable responsive UI; provider abstraction makes it possible to swap STT/TTS/LLM vendors; scaled backend options can support future load. Individual measures of early evaluation, including the system performance (latency, STT quality), user outcomes (reduced filler usage, enhanced coherence), and hu-man-in-the-loop validation is scheduled to be carried out during calibration. Future developments1 STT is sensitive to noise and accents, automated scoring may be biased by the model, and scaled pipelines of workers are required to be used with high concurren-cy. The future work focuses on robustness (speaker adaptation, noise reduction), fairness (bias audits and calibration), more coaching features (lesson plans, live coach handoffs), and connections to learning and hiring ecosystems. To conclude, AI-YODA trans-forms ad-hoc interview practice into a quantifiable, repeatable learning system by means of combining adaptive AI, voice interactivity, and structured men-torship-assistance that enables users to gain compe-tence and confidence with time.

REFERENCES

[1] Alexander Wuttke, Matthias Aßenmacher, Chris-topher Klamm, Max M. Lang, Quirin Würschinger, and Frauke Kreuter, “AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Inter-viewers,” LMU Munich, Munich Center for Machine Learning (MCML), University of Mannheim, Univer-sity of Oxford.

[2] Zhe Liu, “Interview AI-ssistant: Designing for Real-Time Human–AI Collaboration in Interview Preparation and Execution,” University of British Columbia, Vancouver, BC, Canada.

[3] Bang Liu, Haojie Wei, Di Niu, Haolan Chen, and Yancheng He, “Asking Questions the Human Way: Scalable Question-Answer Generation from Text Cor-pus,” University of Alberta and Tencent.

[4] Hamze Muse, Sahan Bulathwela, and Emine Yil-maz, “Pre-Training with Scientific Text Improves Educational Question Generation,” Centre for Artifi-cial Intelligence, University College London, UK.

[5] Sanika Rangnath Jagtap, Vedant Kulkarni, Yashowardhan Pachorkar, Om Taur, Shivam Gupta, and Uma Pujeri, “AI-Driven Real-Time Interview Simulation App with Voice Recognition and Facial Analysis,” 2022.

[6] G. Ramachandra Rao, Bijjamula Chakradhar Reddy, Anne Srinivas Kalyan, Goli Keerthi Priya, and Kanakamedala Rajesh, “AI-Powered Mock Interview Preparation,” 2022.

[7] Alan Jones, Suzan Harkness, and Nathan Mondragon, “Acts of Meaning: How AI-Based Inter-viewing Will Transform Career Preparation in Higher Education,” 2021.

[8] R. Nithya and V. Sukirtha, “Intelligent Job Inter-view Preparation and Career Advancement,” 2021.

[9] Param Kothari, Paras Mehta, Srushti Patil, and Varsha Hole, “InterviewEase: AI-Powered Interview Assistance,” Sardar Patel Institute of Technology, Mumbai, India.

[10] A. I. Umar et al., “AI-Powered Mock Interview Coaching System Using Speech and NLP,” International Journal of Computer Science, 2023.

[11] R. Sharma and P. Gupta, “Virtual Interview As-sistant Using Natural Language Processing and Deep Learning,” International Journal of Computer Appli-cations, 2024.
[12] X. Li et al., “Emotion Recognition-Based Candi-date Evaluation for AI Interviews,” IEEE Access, vol. 10, pp. 112345–112356, 2022.
[13] T. Mahmood et al., “Speech and Sentiment Analysis for Interview Preparedness Using Deep Learning,” Artificial Intelligence Review, Springer, 2023.
[14] M. Hossain et al., “AI-Based Career Recommen-dation Using Resume and Skills Analysis,” in Proc. IEEE ICCIT, pp. 331–336, 2022.
[15] J. Kim et al., “LLM-Based Conversational Inter-view Simulation Platform,” arXiv preprint arXiv:2308.14227, 2023.
[16] N. Mondragon et al., “AI-Driven Interview Feedback: A Study on Automated Interview Analy-sis,” Journal of Employment Technology, 2021.
[17] D. Bui and K. Singh, “Hybrid Facial-Speech Emotion Recognition for Interview Assess-ment,” Journal of Machine Vision & Emotion AI, 2023.