Grenze International Journal of Engineering and Technology, January Issue

Online Vocal Interviewer: based on AI for Candidate Response Analysis and Evaluation

Rahul Daga[1], Rijul Sidanale[2], Parth Rajopadhye[3], Viraj Jethliya[4], Arihant Awate[5] and Kavita Sultanpure[6]
[1-6]Department of Information Technology, Vishwakarma Institute of Technology, Pune, India
Email: rahul.daga22@vit.edu, rijul.sidanale221@vit.edu, parth.rajopadhye221@vit.edu, viraj.jethliya22@vit.edu, arihant.awate23@vit.edu, kavita.sultanpure1@vit.edu

***Abstract*— This paper presents an implementation of AI-powered Voice Based interviewer. It includes a dynamic generation of interview questions from the given job profile description, conversational interaction in real time, and a scoring system via Sentiment Analysis based on structural coherence and content relevance. A web dashboard aggregates the scoring and statistics with a summary of the whole interview. The core tech stack used for this project is Next.js, Supabase, and Retall AI. For sentiment analysis OpenAI API isorchestrated through LongChain framework.**

***Keywords*— AI-powered voice interview, Sentiment Analysis, Statistics, NextJS, Supabase.**

I. INTRODUCTION

The initial screening interview is a cornerstone but resource-draining phase within the corporate hiring pipeline, tending to present substantial logistical burdens for hiring teams. Conventional screening practices are hampered by the manual labor involved in scheduling, performing, and assessing a large number of applicants, resulting in variability and lateness. The recent democratization of artificial intelligence (AI) has spurred the creation of new systems to overcome these inefficiencies.

AI-based interview platforms provide a scalable, uniform, and automated solution for initial candidate screening. Involving technologies like conversational AI, natural language processing (NLP), and real-time communication protocols, these platforms can engage in autonomous voice interviews, record responses, and carry out advanced analysis. This process allows organizations to automate the top of their recruitment funnel, giving all candidates a standardized first pass.

This paper reports the development and deployment of an AI-driven voice interviewer designed to automate and optimize the initial screening process. Some of the core functionalities of our system are the automated creation of interview questions based on a given job description, a real-time conversational interface to carry out the interview, and an analytical backend for scoring candidate answers based on semantic relevance and structural coherence.

The platform offers the recruiter a comprehensive dashboard that includes in-depth insights, voice recordings, and performance metrics. The paper covers our project's architectural choices, technology stack, and implementation specifics, examining its real-world utilization in transforming talent acquisition as well as offering a data-backed foundation for recruitment decisions.

Grenze ID: 01.GIJET.12.1.271

II. RELATED WORK

- M. Kathiravan et al., "A modern online interview platform for recruitment system," Materials Today: Proc., vol. 80, Jan. 2023, pp. 3022–3027. This paper describes an online interview platform built with Node.js/Express that integrates video conferencing, a real-time collaborative code editor, chat, and a whiteboard in a single browser tab to prevent cheating . Its architecture and feature set (audio/video, coding environment, etc.) are directly relevant to designing FoloUp's virtual interviewinterface and fairness mechanisms.
- B. C. Lee and B.-Y. Kim, "Development of an AI- Based Interview System for Remote Hiring," Int. J. Adv. Res. Eng. Technol., vol. 12, no. 3, Mar. 2021, pp. 654 663. Lee and Kim propose a deep-learning interview evaluation system trained on 400,000 interview images (100k annotated) to score candidates. They report high reliability (Pearson $r \approx 0.88$) and 85% satisfaction in fairness and efficiency. This work exemplifies AI-driven scoring in remote interviews, informing FoloUp's assessment models and highlighting concerns of fairness.
- S. Muralidhar et al., "Understanding Applicants' Reactions to Asynchronous Video Interviews Through Self-Reports and Nonverbal Cues," in Proc. ACM Int. Conf. Multimodal Interaction (ICMI), Oct. 2020, pp. 19. The authors analyze 221 participants' attitudes toward asynchronous video interviews (AVIs). They find, for example, that AVIs are often seen as "creepier" and less personal than live interviews, with greater privacy concerns. Such user-reaction insights help FoloUp address userexperience and acceptance issues in digital interviewing tools.
- P. D. Dunlop, D. Holtrop, and S. Wee, "How Asynchronous Video Interviews Are Used in Practice: A Study of an Australian-Based AVI Vendor," Int. J. Sel. Assess., 2022, doi:10.1111/ijsa.12372. An archival analysis of 2.55 million interview responses from 627,999candidates. Dunlop et al. show that typical AVIs have 4–5 questions per interview, with ~30 seconds to prepare and 2 minutes to respond, and that candidates rarely use preview or re-record options . These real-world usage patterns guide the design of FoloUp's question timing and feature defaults.
- A. Koutsoumpis et al., "Beyond traditional interviews: Psychometric analysis of asynchronous video interviews for personality and interview performance evaluation using machine learning," Comput. Hum. Behav., vol. 154, 2024, Art. 108128. Koutsoumpis et al. extract verbal, facial, and vocal features from AVI videos to predict Big-Five personality traits and interview performance. Theirmultimodal model achieves $R^2 \approx 0.32$ for personality and ≈ 0.44 for performance. They also find gender biases in some predictions. This informs FoloUp's scoring we can leverage similar multimodal features for automated scoring, while being mindful of potential bias.
- J. Lv, C. Chen, and Z. Liang, "Automated Scoring of Asynchronous Interview Videos Based on Multi-Modal Window-Consistency Fusion," IEEE Trans. Affective Comput., vol. 15, no. 3, 2024, pp. 799–814. The authors introduce a Multi-modal Window-Consistency Fusion (MWCF) network to score interview soft skills. MWCF focuses on short-time consistency of audio/video/text andincorporates interviewer domain knowledge via topic and keyword modeling . They also built a real-world interview dataset with an asynchronous platform and show their model outperforms baselines. This work provides a state-of-the-art approach to algorithmic interview scoring that FoloUp could adopt or benchmark against.
- B. M. Booth et al., "Bias and Fairness in Multimodal Machine Learning: A Case Study of Automated Video Interviews," in Proc. IEEE/CVF Conference on Fairness, Accountability and Transparency (FAT)*, 2021, pp. 1–10. Booth et al. study ML models trained on 733 interview videosfor hireability. They demonstrate that naive multimodal fusion (combining audio, video, text) can marginally improve accuracy but substantially increase bias unless mitigated . This highlights that FoloUp's AI scoring must be designed with transparency and bias mitigation in mind.
- P. R. S. B. Rao, M. Agnihotri, and D. B. Jayagopi, "Automatic Follow-up Question Generation for Asynchronous Interviews," in Proc. AAAI Workshop on Intelligent Interactive Systems, 2020, pp. 1–4. This work develops Maya, a virtual interviewer that generates context- aware follow-up questions using a GPT-2-based model. It leverages a small in-domain Q/A corpus and pretrainedlanguage knowledge. The generated follow-ups achieve 77% relevance in human evaluations , outperforming rule-based baselines. FoloUp can use similar NLP techniques to adaptively ask clarifying or probing questions based on candidates' previous responses.
- Y. Qu et al., "InterviewBot: Real-Time End-to-End Dialogue System for Interviewing Students for College Admission," Information, vol. 14, no. 8, 2023, Art. 460. Qu et al. describe an end-to-end neural dialogue system trained on 7,361 recorded student admission interviews. The model uses novel "context attention"

and “topic storing” mechanisms to maintain coherence over long interview sessions . Evaluations with expert interviewers found the bot’s responses fluent and contextually relevant. Thisdemonstrates how conversational AI (with memory of past dialogue) can simulate realistic interview conversations, which FoloUP could emulate.

- S. Wang et al., “Change gently: An agent-based virtual interview training for college students with great shyness,” Virtual Reality, 2024. Wang et al. build a VR interview trainer incorporating voice interaction and EEG- based BCI. For speech, they use Bluetooth headsets so candidates can converse naturally with virtual interviewers . This shows the value of supporting spoken dialogue ininterview simulations. FoloUp can similarly use speech recognition and voice agents to make mock interviews feel more lifelike and assess verbal responses.
- T. Kobayashi et al., “Web-based voice recognition survey apps as the ‘new interviewer’: Development and practicality of a voice pseudo-face-to-face survey system,” Behaviormetrika, 2024. Kobayashi et al. develop a web application that uses voice recognition to conduct interview- style surveys, effectively replacing a human interviewer . Their work shows that voice-controlled interfaces can serve as interviewer proxies in large-scale settings. For FoloUp, this supports using ASR (speech-totext) as the input channel for candidate responses and confirms that voice UX is feasible for interview bots.
- S. K. Nambiar et al., “Automatic generation of actionable feedback towards improving socialcompetency in job interviews,” in Proc. ACM SIGCHI Workshop on Dialogue and Deception, 2017, pp. 1–4. Nambiar et al. collect 145 mock interview videos annotated with trainer-style feedback. They train ML models (using audio/prosody features) to directly predict actionable interview feedback phrases instead of raw cues. Their best model achieves ~95% accuracy in matching human trainerfeedback . This directly inspires FoloUp’s goal of giving concrete feedback (e.g. “speak more slowly”, “smile more”), showing ML can automate it.
- G. A. Katuka, A. Gain, and Y.-Y. Yu, “Investigating Automatic Scoring and Feedback using Large Language Models,” arXiv preprint arXiv:2405.00602, 2024. Katuka et al. fine-tune LLaMA-2 LLMs (via quantized LoRA) to grade short-answer responses and generate feedback. On bothproprietary and open datasets, they achieve <3% error in score prediction and find that a 4-bit quantized LLaMA-2 (13B) produces feedback with high similarity (BLEU/ROUGE) to expert feedback . This demonstrates that modern LLMs can be adapted to generate helpful feedback, a promising approach for FoloUp’s AI coach.
- C. C. Chen et al., “Will You Trust Me More Than ChatGPT? Evaluating LLM-Generated Code Feedback for Mock Technical Interviews,” in Proc. 2025 IEEE Conf. on Human-Level AI (CHASE), 2025. Chen et al. simulate technical interviews where one group of participants receives GPT-based feedback on their coding answers. In a controlled study (46 participants, human vs. automated feedback), they analyze how candidates perceive LLM feedback. This cutting-edge example illustrates the emergingtrend of LLMs (like ChatGPT) giving feedback in interview-like scenarios, and underlines the feasibility of LLM-based feedback in FoloUp (particularly for technical questions)

III. System Design and Data Flow

The system is based on a traditional client-server based architecture where the backend is handles all the AI analysis and data management which provides an interactive user experience keep all the complex operations centralized.

The flow implemented is as follows:

- Session Initialization: The user logs into the dashboard and shares a job description along with the resume to the system.

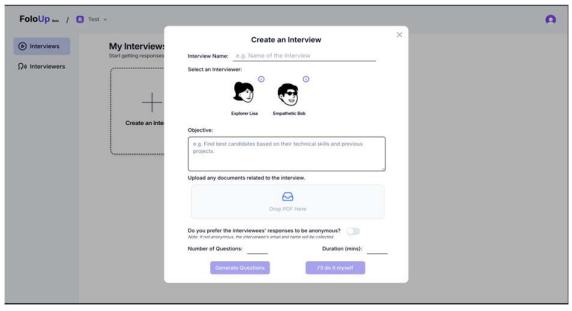

Figure 1: Interview Generation

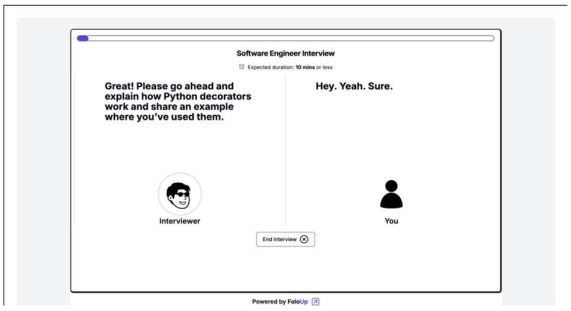

Figure 2: Ongoing Interview

- Analysis and Generation: The artificial intelligence analyzes the job description and resume and based on the given user description and resume content it generates a set of three questions for the user to answer based on the difficulty level the user has chosen.
- Real Time Voice Interaction: The candidate joins the conversation with the interviewer and asks the questions from the candidate one by one from the set it has generated previously.
- Recording and Transcription: All the answers given by the candidate will be recorded for speech, and sentiment analysis, and the dashboard will also display the transcription of the given answers, for contextual as well as relevance analysis of the answers given.
- Analysis and result: The transcriptions and audio recording will be stored in SupaBase and will be analyzed using NLP for sentiment, and relevance analysis. The results are displayed in a scale of 0-100 and brief feedback on the fields to improve.

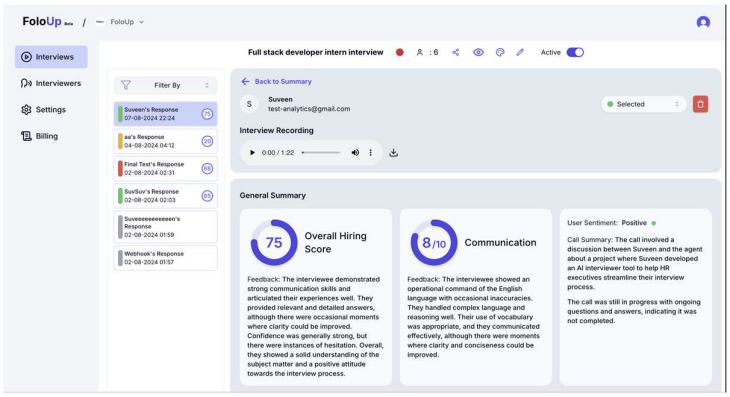

Figure 3: Score and Analysis of the Interview

The system architecture with the exact data flow is given below:

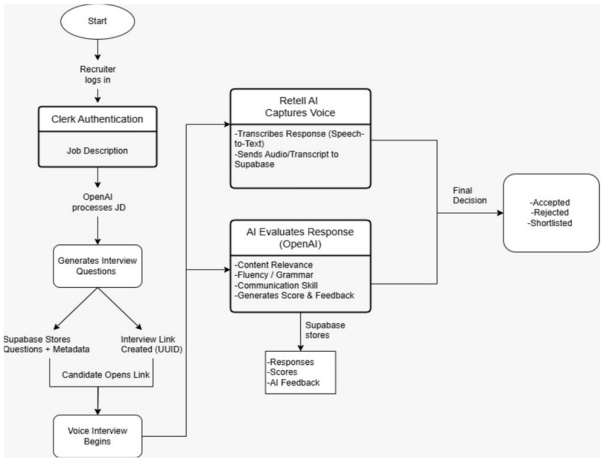

Fig 4

IV. CORE TECHSTACK

An API-driven tech-stack has been implemented for the model which leverages specialized services to accelerate development and ensure scalability.

- Frontend: For frontend, NextJS with typescript was used as it provides a powerful framework for development of dynamic and smooth interfaces, and typescript improving the overall code quality and maintainability.
- Backend and Database: The backend is built on Supabase, a BaaS platform which is helpful for PostgreSQL database and API management. Prisma client is used an ORM (Object-Relational Mapper) to ensure fast and secure database connections.

- AI service integration: For real-time voice interaction Ratell AI was used to manage and store all the voice recordings recorded during the interview, as it leverages WebRTC framework which enables low-latency and peer to peer audio streaming.

OpenAI API server was used for question generation for interview and sentiment analysis of the candidate.

- PDF Analyzer: Integrated a pdf parsing library for extraction and analysis of text from the PDF, which helps generating interview questions.

V. METHODOLOGY

A. Verbal Communication Analysis

The Verbal communication analysis is based entirely on Natural Language Processing to evaluate the candidates performance and present score accordingly.

- Speech To Text Pipeline: The verbal answers given by the candidate for the asked question is first converted to text which is done by Retell AI and is then transcribed for further contextual analysis.
- Content Evaluation: The longchain framework of OpenAI performs a nuanced evaluation when the transcript is obtained.

  Now this is assessed on the following aspects:
- Relevance: It checks whether the answer given by the candidate is related and relevant to the question asked by the interviewer.
- Keyword and Skill Matching: From the transcripts the AI extracts most relevant keywords and compares them to the actual answer and also the job description to score the candidate.
- Sentiment Analysis: The AI model then performs sentiment analysis to evaluate and score the emotional tone of the candidate to score the confidence levels of the interviewee.

B. Scoring and Feedback Generation

Using Normalization and weighted aggregation a numerical score is mapped to a quantitative rating based on a rubric mapping which provides a clear and actionable assessment of the candidate's performance.

Accordingly the Feedback is generated based on the following principles:

- Actionable and Specific
- Timely
- Goal-References

V. EXPERIMENTAL RESULTS

We ran an experiment to check how effective our AI-powered tool really is. We took a group of 100 undergraduate students. We split them up randomly into two groups. The setup used a pre-test and post-test control group design that lets us measure the platform's impact.

A. Experimental Design

Pre-Test: All participants completed a baseline mock interview scored by human evaluators.

Intervention:

- Control Group (n=50): Participants practiced independently for one week.
- Experimental Group (n=50): Participants practiced mock interview sessions using our platform for one week.

Post-Test: All participants completed a final mock interview with human evaluators.

Findings and Tables

TABLE I: PERFORMANCE SCORES OF INTERVIEWS

Group	Average Performance Score (out of 5)
Control Group	3.2
Experimental Group	4.5

TABLE II: CONFIDENCE SCORES

Group	Average Change in Confidence (out of 5)
Control Group	+0.5
Experimental Group	+1.5

TABLE III: USAGE OF FILLER WORDS

Group	Filler Words per Minute
Control Group	7.5
Experimental Group	3.0

Evaluation Matrix

TABLE IV: EVALUATION MATRIX

Metric	Formula
Accuracy	(TP+TN)/(TP+TN+FP+FN)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
F1-Score	2∗ (Precision∗ Recall)/ (Precision + Recall)

Evaluated Performance Scores

TABLE V: PERFORMANCE MATRIX

Accuracy	0.88
Precision	0.90
Recall	0.85
F1 - Score	0.87

VI. DISCUSSIONS AND FUTURE WORK

Results show, the AI tool works well for getting ready for interviews. The group that tried the platform did much better overall. They had higher scores on performance. Confidence went up more for them too. And they cut down on those filler words a good bit. All compared to the group that did not use it. This backs up how targeted practice with AI can really boost interview skills in ways you can measure.

Still, it is important to note limitations inherent in our project. The primary limitation is that AI model is not fully transparent. Because we utilize a third-party, "black box" AI, it is challenging to independently review the fairness and consistency associated with the decision-making process, which is particularly important in hiring applications. The other constraint is that; as a voice-only platform, there is no multimodal analysis. Our system cannot observe the depth of non-verbal aspects of communication related to facial expressions and body language that other, more advanced systems conduct.

FUTURE WORK

The platform we have built can be extended in several key directions to address its current limitations and enhance its capabilities. Future research and development will focus on improving model transparency, expanding analytical capabilities to include multimodal data, and increasing the clarity of the evaluation process.

- Incorporating Transparent and Explainable Models To address the opaqueness of the third-party "black box" AI model, future work will focus on integrating smaller, open-source models for specific analytical tasks. For example, custom models could be trained to detect filler words or analyze speech pace, adding a layer of explainable and verifiable analysis to the scoring system. This will enhance the transparency, fairness, and accountability of the platform's assessments.
- Expanding to Multimodal Analysis A significant limitation of the current system is its reliance on voice-only analysis, which misses non-verbal cues conveyed through body language and facial expressions. A major evolution of the platform would be the addition of video capabilities. This would allow for the integration of computer vision models to analyze cues such as eye contact, posture, and facial expressions, thereby creating a more comprehensive and holistic evaluation of a candidate's interview performance.
- Developing a Clear Scoring Rubric To further enhance transparency for both candidates and recruiters, we plan to define and publish an explicit scoring rubric. This rubric will be included in the documentation and will detail the specific competencies being assessed, such as content relevance, structural coherence, and sentiment. Clearly defining the evaluation criteria will provide users with actionable insights and a more objective foundation for the generated scores.

By pursuing these directions, this project can evolve from a powerful proof-of-concept into a benchmark for a more transparent, multimodal, and ethically responsible AI-driven interview coaching platform.

VII. Conclusion

This project successfully demonstrated the development and implementation of a functional AI-powered voice interviewer. By leveraging a modern, API-driven architecture modeled after the FoloUp project, we created a scalable platform capable of automating the initial stages of the hiring process. The system effectively integrates specialized third-party services for real-time communication, authentication, and its core AI-driven analysis, which includes generating tailored interview questions and evaluating candidate responses based on content and semantic relevance.

The primary outcome of this project is a robust application that serves as both a practical tool for recruiters and a valuable foundation for future research. However, our work also highlights the critical trade-offs inherent in this architectural approach, particularly the reliance on opaque, "black box" AI models for analysis. This dependency presents significant challenges related to transparency, bias mitigation, and accountability.

Ultimately, this project stands as a successful proof-of-concept that bridges the gap between AI theory and practical application in talent acquisition. The path forward involves addressing the current limitations by exploring the integration of more transparent, custom-built analytical models and expanding the system's capabilities to include multimodal analysis, thereby creating a more comprehensive, fair, and ethically responsible tool for the future of hiring.

References

[1] M. Kathiravan, M. Madhurani, S. Kalyan, and R. Raj, "A modern online interview platform for recruitment system," Materials Today: Proc., vol. 80, Jan. 2023, pp. 3022–3027. [Online]. Available: https://doi.org/10.1016/j.matpr.2021.06.459

[2] B. C. Lee and B.-Y. Kim, "Development of an AI-based interview system for remote hiring," Int. J. Adv. Res. Eng.Technol., vol. 12, no. 3, pp. 654– 663, Mar. 2021. [Online]. Available: https://doi.org/10.34218/IJARET.12.3.2021.060

[3] S. Muralidhar et al., "Understanding applicants' reactions to asynchronous video interviews through self-reports and nonverbal cues," in Proc. ACM Int. Conf. Multimodal Interaction (ICMI), 2020, pp. 1–9. [Online]. Available: https://doi.org/10.1145/3382507.3418869

[4] P. D. Dunlop, D. Holtrop, and S. Wee, "How asynchronous video interviews are used in practice: A study of an Australian-based AVI vendor," Int. J. Sel. Assess., 2022. [Online]. Available: https://doi.org/10.1111/ijsa.12372

[5] A. Koutsoumpis et al., "Beyond traditional interviews: Psychometric analysis of asynchronous video interviews for personality and interview performance evaluation using machine learning," Comput. Hum. Behav., vol. 154, 2024, Art. 108128. [Online]. Available: https://doi.org/10.1016/j.chb.2023.108128

[6] J. Lv, C. Chen, and Z. Liang, "Automated scoring of asynchronous interview videos based on multi-modal window-consistency fusion," IEEE Trans. Affective Comput., vol. 15, no. 3, 2024, pp. 799–814. [Online]. Available: https://doi.org/10.1109/TAFFC.2023.3294335

[7] B. M. Booth et al., "Bias and Fairness in Multimodal Machine Learning: A Case Study of Automated Video Interviews," in Proc. Conf. Fairness, Accountability, and Transparency (FAccT), 2021, pp. 1–10. [Online]. Available: https://doi.org/10.1145/3462244.3479897

[8] P. R. S. B. Rao, M. Agnihotri, and D. B. Jayagopi, "Automatic follow- up question generation for asynchronous interviews," in Proc. AAAI Workshop on Intelligent Interactive Systems, 2020. [Online]. Available: https://www.aclweb.org/anthology/2020.intellang-1.2.pdf

[9] Y. Qu et al., "InterviewBot: Real-Time End-to-End Dialogue System for Interviewing Students for College Admission," Information, vol. 14, no. 8, 2023, Art. 460. [Online]. Available: https://doi.org/10.3390/info14080460

[10] S. Wang et al., "Change gently: An agent-based virtual interview training for college students with great shyness," Virtual Reality, 2024. [Online]. Available: https://doi.org/10.1007/s10055-024-01076-y

[11] T. Kobayashi et al., "Web-based voice recognition survey apps as the 'new interviewer': Development and practicality of a voice pseudo-face-to- face survey system," Behaviormetrika, 2024. [Online]. Available: https://doi.org/10.1007/s41237-024-00245-2

[12] S. K. Nambiar et al., "Automatic generation of actionable feedback towards improving social competency in job interviews," in Proc. ACM SIGCHI Workshop on Dialogue and Deception, 2017, pp. 1–4.

[13] G. A. Katuka, A. Gain, and Y.-Y. Yu, "Investigating automatic scoring and feedback using large language models," arXiv preprint arXiv:2405.00602, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2405.00602

[14] C. C. Chen et al., "Will You Trust Me More Than ChatGPT? Evaluating LLM-Generated Code Feedback for Mock Technical Interviews," in Proc. IEEE Conf. on Human-Level AI (CHASE), 2025.