

AI-Powered Mock Interview System with Multi-Accent Support

Mrs. Archana, Potharaju Bhoomika, Mulukalla Anusha, Guduru Vimala Bhavana, Kande Shreecharani
Department of Artificial Intelligence and Data Science at Vignan's Institute of Management and Technology
for Women, Telangana, India.

ABSTRACT

The increasing competition in today's job market has made interview preparation an essential skill for students and job seekers. Traditional mock interview methods often depend on human trainers, which can be time-consuming, costly, and inconsistent. To address these challenges, this paper presents an AI-Powered Mock Interview System with Multi-Accent Support that provides an automated platform for practicing interview questions and receiving instant feedback. The system allows candidates to respond using either voice or text input. Spoken responses are converted into text using Automatic Speech Recognition (ASR), while Natural Language Processing (NLP) techniques are used to evaluate the semantic relevance, clarity, and completeness of the answers. The proposed system supports multi-accent speech recognition to improve fairness and usability for diverse users. It also separates content evaluation and delivery analysis to provide transparent scoring and actionable feedback. The system is implemented using Python, Streamlit, Whisper, SentenceTransformers, and spaCy. Experimental testing shows that the system can effectively simulate interview practice, provide instant feedback, and help users improve their communication skills and confidence. The proposed solution is cost-effective, accessible, and suitable for students, job seekers, and educational institutions.

Keywords: Mock Interview, Artificial Intelligence, Speech Recognition, Natural Language Processing, Whisper, Semantic Analysis, Multi-Accent Support, Interview Preparation

INTRODUCTION

The interview process is one of the most critical stages in recruitment, where candidates are assessed not only on their technical knowledge but also on their communication skills, confidence, and ability to present ideas clearly. Many students and job seekers face difficulties during interviews due to lack of proper practice, nervousness, and limited access to personalized feedback. Traditional mock interviews usually require human interviewers such as mentors, trainers, or faculty members, which may not always be available or affordable.

With the rapid growth of artificial intelligence technologies, it is possible to automate interview practice and provide intelligent feedback to candidates. Speech recognition and natural language processing can be used to simulate real interview environments, analyze responses, and generate suggestions for improvement. However, many existing systems rely heavily on keyword-based

evaluation and often suffer from speech recognition inaccuracies, especially for non-native speakers or regional accents.

This paper proposes an AI-Powered Mock Interview System with Multi-Accent Support that enables candidates to practice interview questions through speech or text input. The system uses ASR to convert speech into text and NLP techniques to evaluate the semantic quality of responses. It provides instant feedback, transparent scoring, and a flexible web-based platform for continuous interview practice.

LITERATURE SURVEY

1. InterviewAI (2023)

InterviewAI is an AI-powered mock interview assistant that provides users with simulated interview practice and automated feedback on their responses. The platform evaluates candidate answers and offers suggestions for improvement. However, the system has limitations in terms of transparency, as the scoring mechanism is not

clearly explained to users. This makes it difficult for candidates to understand how their performance is being assessed and which specific aspects require improvement.

2. Whisper: Robust Speech Recognition via Large-Scale Weak Supervision (2022) – Alec Radford et al.

The work by Alec Radford and colleagues introduced Whisper, a powerful speech recognition model trained on large-scale weakly supervised data. Whisper is designed to support multilingual and multi-accent speech recognition with improved robustness compared to conventional ASR systems. This research is highly relevant to the proposed system because accurate transcription of spoken interview responses is essential for fair evaluation. Whisper's ability to handle diverse accents and speaking styles makes it a strong foundation for reducing transcription errors in AI-based mock interview platforms.

3. Google Interview Warmup (2021)

Google Interview Warmup is an interview preparation tool developed by Google that helps candidates practice answering common interview questions. It uses speech-to-text technology and provides keyword-based feedback by identifying frequently used terms in the candidate's response. While the tool is useful for basic interview preparation, its feedback is primarily based on keyword detection rather than deeper semantic understanding. As a result, it may not fully capture the relevance, clarity, or completeness of the candidate's answer.

4. Accent Bias in Automatic Speech Recognition (2020) – IEEE/ArXiv Researchers

Several research studies published by IEEE and ArXiv researchers have highlighted the issue of accent bias in Automatic Speech Recognition (ASR) systems. These studies found that traditional ASR models often produce higher word error rates for non-native speakers and individuals with regional accents. This creates challenges in systems that rely on speech-based input, as incorrect transcription can negatively affect the fairness and accuracy of evaluation. These findings strongly support the need for multi-accent robust ASR in the proposed AI-powered mock interview system.

5. AI-Powered Mock Interview System for Automated Skill Assessment (2025) – Dr. Vijayant Verma et al.

This research work proposed a virtual mock interview system that integrates Natural Language Processing (NLP), speech recognition, and semantic similarity algorithms to assess candidate responses and generate automated feedback. The system focuses on evaluating the quality of answers and supporting skill development through repeated practice. This work is closely related to the proposed system and validates the effectiveness of combining AI techniques for interview preparation. However, the proposed system further improves upon such approaches by incorporating dual input modes, multi-accent support, and transparent scoring mechanisms.

The literature survey indicates that traditional mock interview methods are highly dependent on human trainers and lack scalability. Existing AI-based interview tools often rely on keyword-based evaluation, which may not accurately assess the semantic quality of candidate responses. Several systems also provide limited support for speech input and may not effectively handle accent variations, leading to fairness issues in evaluation. In addition, most available tools do not clearly separate content assessment from delivery analysis. These limitations highlight the need for a more inclusive, intelligent, and transparent interview preparation system, which motivates the development of the proposed AI-Powered Mock Interview System.

PROBLEM STATEMENT

Many candidates lack access to effective and affordable mock interview practice. Traditional preparation methods do not simulate real interview conditions and fail to provide intelligent evaluation of candidate responses. Existing AI-based interview tools often suffer from accent bias, inaccurate speech recognition, and keyword-based scoring that does not truly understand the meaning of answers. Therefore, there is a need for a fair, inclusive, and intelligent system that supports multi-accent speech recognition, semantic response evaluation, transparent scoring, and actionable feedback to improve interview readiness.

OBJECTIVES

- To develop an AI-based mock interview platform for students and job seekers
- To support both speech and text-based responses
- To implement multi-accent speech recognition using Whisper
- To evaluate answers using semantic NLP techniques instead of only keyword matching
- To generate instant, transparent, and actionable feedback
- To separate content scoring and delivery scoring and to store interview sessions and track user performance over time .

PROPOSED METHODOLOGY

The proposed system is developed as a web-based application using Streamlit for the frontend interface and Python-based backend processing. The methodology includes the following stages:

1. **Question Selection** – The user selects the type of interview such as HR or technical, and the system displays appropriate questions.
2. **Response Capture** – The user answers the question either through speech or text input.
3. **Speech-to-Text Conversion** – If speech mode is selected, the audio response is converted into text using the Whisper ASR model.
4. **Transcript Review** – The user can review and edit the transcript to correct transcription errors.
5. **Semantic Analysis** – The final response text is processed using NLP techniques such as SentenceTransformers and spaCy to measure semantic similarity and response quality.
6. **Delivery Analysis** – Optional speech-based analysis evaluates pace, fluency, and silence ratio.
7. **Feedback Generation** – The system produces performance scores and suggestions for improvement.
8. **Data Storage** – Session details, transcripts, scores, and feedback are stored for future review.

SYSTEM ARCHITECTURE

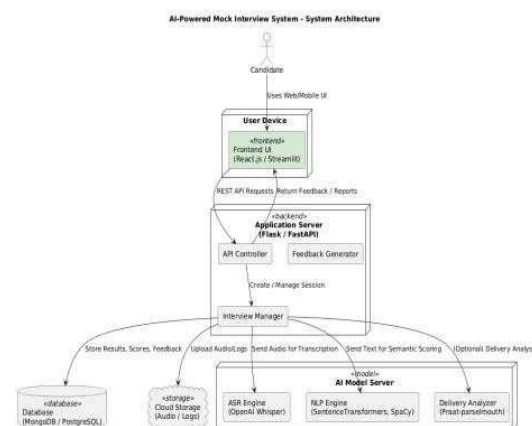


Fig. 1. System Architecture of the Proposed AI-Powered Mock Interview System with Multi-Accent Support

The system architecture of the proposed **AI-Powered Mock Interview System with Dual Input and Multi-Accent Support** is designed as a layered and modular framework that enables efficient interaction, response processing, evaluation, and feedback generation. The architecture integrates user interaction, speech processing, natural language understanding, scoring mechanisms, and data storage into a unified web-based platform. The system is developed to simulate a realistic mock interview environment while ensuring flexibility, fairness, and ease of use for candidates.

At the **user layer**, the candidate interacts with the system through a web-based interface developed using **Streamlit**. The user can select the type of interview, such as HR or technical, and choose the preferred mode of response input. The platform supports **dual input modes**, allowing candidates to answer interview questions either through **speech input using a microphone** or through **direct text entry**. This dual-input design improves accessibility and enables users to practice according to their convenience.

The **frontend layer** is responsible for displaying interview questions, capturing candidate responses, and presenting feedback and performance reports. It provides an interactive and user-friendly environment where candidates can navigate through interview sessions smoothly. In the case of voice input, the recorded audio is forwarded to the speech processing component, while text input is directly passed to the evaluation pipeline.

The **Automatic Speech Recognition (ASR) module** plays a crucial role in handling spoken responses. This module uses the **Whisper model** to convert audio responses into text. Whisper is selected because of its strong capability to recognize speech from speakers with different accents and speaking styles. This improves the inclusiveness of the system and reduces the limitations commonly found in traditional speech recognition systems. After transcription, the system provides a **transcript review mechanism**, allowing users to verify and correct the converted text before final evaluation. This step helps improve fairness and reduces the impact of possible transcription errors.

Once the final response text is available, it is processed by the **Natural Language Processing (NLP) module**. This module uses technologies such as **SentenceTransformers** and **spaCy** to analyze the semantic quality of the candidate's answer. The system compares the user response with the expected or ideal answer and evaluates factors such as **relevance, clarity, completeness, and contextual meaning**. Unlike keyword-based systems, the semantic evaluation approach allows the system to understand the actual intent and quality of the response, leading to more accurate and meaningful assessment.

In addition to content evaluation, the system may optionally perform **delivery analysis** for spoken responses. This component examines communication-related aspects such as **speech fluency, pace, pauses, and silence ratio**. By separating **content analysis** from **delivery analysis**, the system provides more transparent scoring and enables candidates to understand whether improvement is needed in the substance of their answers or in their speaking style.

The **feedback and scoring module** generates the final performance evaluation based on the outputs from the NLP and delivery analysis components. It provides **instant, transparent, and actionable feedback** in the form of scores, suggestions, and improvement recommendations. This helps users identify their strengths and weaknesses and supports repeated self-practice for continuous enhancement of interview skills.

Finally, the **database layer** stores important session details such as selected interview type, candidate responses, transcriptions, scores, and generated feedback. This storage mechanism allows users to review previous interview sessions, track progress

over time, and analyze performance trends. The modular architecture of the proposed system ensures scalability, maintainability, and future extensibility, making it suitable for further enhancements such as multilingual support, resume-based question generation, and advanced behavioral analysis.

TOOLS AND TECHNOLOGIES

- **Programming Language:** Python
- **Frontend:** Streamlit
- **Speech Recognition:** Whisper
- **NLP Libraries:** SentenceTransformers, spaCy
- **Audio Processing:** Librosa, Pydub
- **Data Handling:** NumPy, Pandas
- **Database:** MongoDB / PostgreSQL (or local storage based on implementation)

IMPLEMENTATION DETAILS

The proposed system is implemented as a web-based application using Python and Streamlit. The frontend provides an interactive interface for selecting interview categories and capturing user responses through either text or microphone input. For spoken responses, the Whisper model is used to convert audio into text. The transcribed or directly entered text is then processed using SentenceTransformers and spaCy to evaluate semantic relevance, clarity, and completeness. Optional delivery analysis is performed for speech responses by examining parameters such as fluency, pace, and silence ratio. The system generates scores and personalized feedback, while user responses and session details are stored for future review and progress tracking.

RESULT AND DISCUSSION

The developed AI-Powered Mock Interview System was successfully tested as a functional prototype for interview practice and automated response evaluation. The system enables users to select interview categories, respond through either speech or text input, and receive immediate feedback on their performance. During testing, the speech recognition module demonstrated effective transcription of spoken responses, including improved tolerance toward accent variations when compared with conventional speech-to-text systems. This contributed to a more inclusive and userfriendly experience for candidates from different linguistic backgrounds.

The Natural Language Processing (NLP) module was able to analyze candidate responses based on semantic relevance, clarity, and completeness rather than relying solely on keyword matching. This resulted in more meaningful evaluation and better alignment with the expected answer quality. The transcript review mechanism further enhanced system fairness by allowing users to correct any transcription inaccuracies before final assessment.

The system generated content-based scores, optional delivery-related analysis, and actionable suggestions that helped users identify both strengths and areas requiring improvement. Overall, the experimental results indicate that the proposed system provides a practical, accessible, and intelligent solution for interview preparation. It supports repeated self-practice, reduces dependence on human interviewers, and improves candidate confidence and readiness for real-world interviews.

OUTPUTS

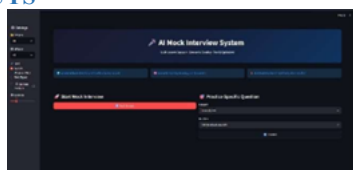


Fig.2. Out put of AI-Powered Mock Interview System with Multi-Accent Support



Fig.3. The picture represents options available while attempting the interview with AI



Fig.4. Shows the breakdown and output after completion of the interview with AI

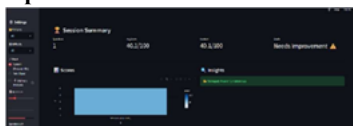


Fig.5. This is the output of session summary that is evaluated by AI after the Interview

ADVANTAGES OF THE PROPOSED SYSTEM

The proposed AI-Powered Mock Interview System offers several significant advantages over traditional mock interview methods and existing automated interview preparation tools. One of the major strengths of the system is its ability to support both speech and text-based inputs, allowing users to choose the most convenient mode of interaction. This dual-input capability increases accessibility and ensures that candidates can practice in a flexible manner according to their comfort and technical availability.

Another important advantage is the integration of multi-accent speech recognition using advanced Automatic Speech Recognition (ASR) technology. Unlike conventional systems that may struggle with regional or non-native accents, the proposed system is designed to improve fairness and usability for a diverse range of users. This makes the platform more inclusive and suitable for students and job seekers from different linguistic backgrounds.

The system also provides a more intelligent form of evaluation by using Natural Language Processing (NLP) and semantic analysis rather than relying only on keyword-based matching. This enables the platform to understand the actual meaning and relevance of candidate responses, resulting in more accurate and meaningful feedback. As a result, users receive better guidance on how well their answers align with the expected response.

A further advantage is the system's ability to generate instant, transparent, and actionable feedback. Candidates are not only given performance scores, but are also provided with suggestions that highlight their strengths and areas requiring improvement. The separation of content evaluation and delivery analysis further enhances the reliability of the feedback by giving users a clearer understanding of

whether they need to improve the quality of their answers or their communication style.

In addition, the proposed system includes a transcript review and correction feature, which allows users to verify and edit speech-to-text outputs before final evaluation. This feature helps reduce the impact of occasional transcription errors and ensures a fairer assessment process. The web-based implementation of the platform makes it cost-effective, scalable, and easily accessible, reducing dependency on human interviewers and enabling repeated practice sessions at any time.

Overall, the proposed system offers a practical, intelligent, and user-friendly solution for interview preparation. By combining automation, fairness, flexibility, and personalized feedback, it significantly enhances the learning experience and improves candidate readiness for real-world interviews.

CONCLUSION

The proposed **AI-Powered Mock Interview System with Multi-Accent Support** offers an effective and intelligent solution for interview preparation. By integrating speech recognition, semantic text analysis, and automated feedback generation, the system provides a realistic and accessible platform for candidates to improve their interview skills. It reduces dependency on human trainers, supports repeated practice, and enables users to track their progress over time. The system is especially beneficial for students and job seekers preparing for placements and professional interviews. In the future, the platform can be extended with emotion analysis, resume-based question generation, and multilingual support to further enhance its capabilities.

FUTURE SCOPE

- Integration of emotion and facial expression analysis.
- Resume-based personalized question generation.
- Multilingual interview support.
- Advanced scoring dashboards.
- Adaptive question difficulty levels.

REFERENCES

[1] A. Radford et al., "Whisper: Robust Speech Recognition via Large-Scale Weak Supervision,"

arXiv preprint arXiv:2212.04356, 2022. [2] OpenAI, "Whisper Documentation," Available:

<https://github.com/openai/whisper> [3] N. Reimers and I. Gurevych, "Sentence-BERT:

Sentence Embeddings using Siamese BERTNetworks," *Proceedings of EMNLP-IJCNLP*, 2019.

[4] Explosion AI, "spaCy Documentation,"

Available: <https://spacy.io/>

[5] Google, "Interview Warmup," Available:

<https://grow.google/certificates/interview-warmup/> [6] V. Verma et al., "AI-Powered

Mock Interview System for Automated Skill Assessment," 2025.

[7] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. *arXiv preprint arXiv:2212.04356*.

[8] Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. *arXiv preprint arXiv:1904.08779*.

[9] Park, D. S., Zhang, Y., Chiu, C. C., Chen, Y., Li, B., Chan, W., & Le, Q. V. (2019). SpecAugment on Large Scale Datasets. *arXiv preprint arXiv:1912.05533*.

[10] Yazdani, N., Mahajan, A., & Ansari, A. (2025). Zara: An LLM-based Candidate Interview Feedback System. *arXiv preprint arXiv:2507.02869*.

[11] OpenAI. (2022). Whisper: Automatic Speech Recognition System. Available at: *OpenAI Whisper Repository / Documentation. (Use this as a tool/implementation reference if needed along with the Whisper paper.)* [12] Reimers, N., & Gurevych, I. (2019). SentenceBERT: Sentence Embeddings using Siamese BERTNetworks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*.

[13] Honnibal, M., & Montani, I. (2017). spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing. *To appear / software library reference*.

- [14] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*.
- [15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS 2017)*.
- [16] Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed. draft). Pearson / Online.