International Journal of Engineering Research and Science & Technology

VOICE-ENABLED SELF INTERVIEW EVALUATION SYSTEM USING AI AND NLP

[1] Mrs M. Srimathi, [2] D.Varun, [3] Busheera Begum, [4] B.Sriya

[1]Assistant Professor in Department of CSE Teegala Krishna Reddy Engineering College Hyderabad, Telangana

[2,3,4] UG Scholars in Department of CSE Teegala Krishna Reddy Engineering College Hyderabad, Telangana

Abstract

With the increasing demand for intelligent and scalable interview preparation systems, traditional methods fail to provide personalized, unbiased, and real-time evaluation. This paper presents a **Voice-Enabled Self Interview Evaluation System** that integrates Artificial Intelligence (AI), Natural Language Processing (NLP), and speech technologies to simulate real-world interview environments. The proposed system allows users to answer interview questions through both text and voice inputs, enabling a more realistic experience. The system leverages **Sentence-BERT (SBERT)** for semantic similarity evaluation, combined with keyword matching and answer length analysis to generate accurate scores and feedback. Additionally, resume-based personalization extracts candidate skills to tailor interview questions dynamically. A MongoDB-based backend supports dynamic question retrieval, including web-scraped data, while a Flask-based architecture ensures scalability and modularity. Experimental analysis shows that the system provides reliable evaluation, reduces human bias, and improves user engagement through real-time feedback and performance analytics. The proposed solution offers a comprehensive, intelligent, and user-friendly platform for modern interview preparation.

Keywords

Artificial Intelligence, Natural Language Processing, SBERT, Speech Recognition, Interview Evaluation, Resume Parsing, Web-Based Systems

I. INTRODUCTION

In today's highly competitive job market, effective interview preparation has become essential for candidates seeking employment in both technical and non-technical domains. Traditional interview preparation methods, such as face-to-face mock interviews and written assessments, rely heavily on human evaluators. While these approaches provide valuable feedback, they are inherently subjective, time-consuming, resource-intensive, and difficult to scale for a large number of candidates. Human bias, inconsistency in evaluation, and limited availability of expert interviewers further reduce the effectiveness of such systems [1].

With the advancement of digital technologies, several online interview preparation platforms have emerged. These platforms provide coding challenges, aptitude tests, and pre-defined question banks. However, most existing systems focus primarily on static question-answer formats and lack real-time interaction, personalized learning, and intelligent evaluation mechanisms. In particular, they rely heavily on keyword-based matching techniques, which fail to capture the semantic meaning of user responses. As a result, correct answers expressed using different

wording or paraphrasing may be incorrectly evaluated, leading to inaccurate assessment outcomes [2].

Recent developments in Natural Language Processing (NLP) and Artificial Intelligence (AI) have opened new possibilities for building intelligent evaluation systems. Transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers) have significantly improved the ability of machines to understand contextual and semantic relationships in text data [3]. Further advancements like Sentence-BERT (SBERT) enable efficient sentence-level semantic similarity comparison, making them highly suitable for evaluating descriptive answers in interview scenarios [4]. These models allow systems to assess answers based on meaning rather than exact keyword overlap, thereby improving fairness and accuracy.

In addition to NLP, speech technologies have become increasingly important in simulating real-world interview environments. Modern browsers support speech recognition through APIs such as the Web Speech API, enabling real-time conversion of spoken responses into text without requiring additional infrastructure. Studies have shown that voice-based interaction improves user engagement, reduces anxiety, and closely replicates real interview conditions, thereby enhancing learning outcomes [5].

To address the limitations of existing systems, this research proposes a Voice-Enabled AI Interview Evaluation System that integrates AI, NLP, and speech technologies into a unified platform. The system enables candidates to participate in simulated interviews using both text and voice inputs. Users can customize their interview sessions by selecting categories such as technical or HR, specifying programming languages, and adjusting difficulty levels. This flexibility ensures that the system caters to diverse user requirements and skill levels.

A key contribution of the proposed system is the integration of a hybrid evaluation mechanism that combines semantic similarity analysis using SBERT, keyword matching, and response length analysis. This multi-dimensional approach ensures more accurate and balanced scoring compared to traditional methods. Furthermore, the system incorporates resume-based personalization, where user resumes are analyzed to extract relevant skills and generate tailored interview questions. This feature enhances the relevance of the interview experience and aligns it with real-world recruitment processes.

The platform also includes a comprehensive performance dashboard that provides detailed analytics such as average scores, grade distribution, and progress trends. Additionally, it supports PDF report generation, enabling users to download and review their performance for continuous improvement. An integrated admin module allows efficient management of users and question banks, ensuring system scalability and data quality.

II LITERATURE SURVEY

The rapid advancement of Artificial Intelligence (AI), Natural Language Processing (NLP), and speech technologies has significantly influenced the development of intelligent interview preparation systems. Traditional interview evaluation methods primarily relied on manual assessment by human experts, which often introduced subjectivity, inconsistency, and scalability issues. Early automated systems attempted to address these challenges using statistical approaches such as Bag-of-Words (BoW) and TF-IDF (Term Frequency–Inverse Document Frequency). While these techniques enabled basic similarity measurement between candidate responses and reference answers, they were limited in their ability to capture contextual meaning, often resulting in inaccurate evaluations when answers were paraphrased or expressed differently [1].

The emergence of transformer-based models marked a significant breakthrough in NLP. Models such as BERT (Bidirectional Encoder Representations from Transformers) enabled deep contextual understanding of language by analyzing the relationships between words in a sentence [2]. Building upon this foundation, Sentence-BERT (SBERT) was introduced to generate semantically meaningful sentence embeddings that can be efficiently compared using similarity measures such as cosine similarity [3]. These advancements have made it possible to evaluate descriptive answers more accurately by focusing on meaning rather than exact keyword matches. Furthermore, recent research suggests that combining semantic similarity with additional factors such as keyword coverage and answer length improves the robustness and reliability of automated evaluation systems [4].

In parallel, speech recognition technologies have played a crucial role in enhancing the realism of interview preparation platforms. Traditional systems largely depended on text-based input, which does not fully replicate real-world interview scenarios. The integration of speech-to-text technologies allows users to provide responses verbally, thereby improving engagement and creating a more natural interaction experience. Modern browser-based solutions, such as the Web Speech API, enable real-time transcription without requiring complex infrastructure or external services. Studies have shown that spoken responses provide valuable insights into a candidate's communication skills, fluency, and confidence, which are important factors in interview performance [5]. Additionally, automated spoken language assessment systems have demonstrated strong correlations with human evaluation, indicating their effectiveness in real-world applications [6].

Another important aspect of intelligent interview systems is resume parsing and skill extraction. Early resume parsing techniques relied on rule-based approaches and regular expressions, which lacked flexibility and struggled to handle diverse resume formats. More recent approaches utilize machine learning techniques, including Named Entity Recognition (NER), to extract structured information such as skills, qualifications, and experience [7]. Research indicates that taxonomy-based skill extraction methods, which map different variations of skill names to standardized representations, provide more reliable and interpretable results, particularly in technical domains [8]. Incorporating resume-based personalization into interview systems allows for automatic generation of relevant questions tailored to a candidate's background, thereby enhancing the effectiveness of the preparation process.

Personalization and adaptive learning have also been widely explored in educational and assessment systems. The concept of adaptive learning is rooted in theories such as Vygotsky's Zone of Proximal Development, which suggests that learners benefit most from tasks that are slightly above their current level of understanding [9]. Computerized Adaptive Testing (CAT) systems implement this concept by dynamically adjusting question difficulty based on user performance. While many existing interview platforms allow users to manually select topics or difficulty levels, automatic personalization based on user data, such as resumes, provides a more seamless and efficient approach. This ensures that users are consistently challenged while receiving relevant and meaningful practice.

Web-based interview preparation platforms have gained popularity due to their accessibility and scalability. Many platforms offer coding challenges, aptitude tests, and mock interviews; however, they often focus on specific aspects of interview preparation rather than providing a comprehensive solution. For example, coding platforms primarily emphasize algorithmic problem-solving and do

not address behavioral or HR interview skills. Additionally, several mock interview systems require peer participation or human evaluators, which limits their availability and scalability [11]. Another limitation of existing systems is the lack of detailed feedback, as many platforms do not provide insights into missing concepts or areas for improvement. Research on gamification highlights that features such as instant feedback, progress tracking, and performance visualization significantly enhance user engagement and motivation in learning environments [12].

III EXISTING SYSTEM

Current interview preparation methods are largely based on conventional practices such as mock interviews, written tests, and basic online practice platforms. In traditional settings, candidates rely on faculty members or industry professionals to conduct mock interviews and provide feedback. Although this approach can be helpful, it is not always consistent because the evaluation depends on the individual judgment of the interviewer. It also requires significant time and effort, making it difficult to provide equal opportunities for all candidates.

With the growth of online learning platforms, various digital tools have been introduced to support interview preparation. These platforms typically offer question banks, coding exercises, and practice tests. However, most of them operate in a fixed manner, where the same type of questions and evaluation patterns are used for all users. The assessment in such systems is often based on matching specific keywords or phrases with predefined answers. This method does not truly evaluate the understanding of the candidate, as it cannot recognize answers that are correct but expressed differently.

Another major limitation is the absence of natural interaction. Most existing systems require users to type their responses, which does not reflect real interview conditions where verbal communication is essential. Without voice-based interaction, candidates are unable to practice speaking skills, confidence, and clarity, which are important aspects of interviews. In addition, many platforms do not provide immediate or detailed feedback, limiting the user's ability to identify mistakes and improve performance effectively.

Personalization is also lacking in current systems. Candidates are usually required to manually select topics or areas of interest before starting practice sessions. There is no automatic mechanism to understand the user's background, such as analyzing their resume or skill set, to generate relevant questions. As a result, the practice sessions may not align with the candidate's actual knowledge or career goals.

Furthermore, many existing platforms focus mainly on technical or coding-based questions and do not adequately cover non-technical areas such as HR or behavioral interviews. This creates an incomplete preparation experience. Additionally, features like performance tracking, progress analysis, and downloadable reports are either limited or unavailable in most systems.

Another drawback is the lack of flexibility in managing content. Question banks in many platforms are fixed and not updated frequently. There is also limited support for administrative control to review, modify, or add new questions dynamically. This can lead to outdated or less relevant content being used over time.

IV PROBLEM STATEMENT

Preparing for interviews is still a difficult and often inefficient process for many candidates. Most students and job seekers depend on traditional methods like mock interviews or self-practice using online questions. While these methods are helpful to some extent, they do not always provide consistent or meaningful feedback. In many cases, the quality of evaluation depends on the

availability and perspective of the interviewer, which can vary from person to person.

At the same time, existing online platforms do not fully solve this problem. Many of them provide only a set of predefined questions and evaluate answers using simple matching techniques. This kind of evaluation does not truly reflect a candidate's understanding, especially when answers are written in different ways but still convey the correct idea. Because of this, users may receive scores that do not accurately represent their actual performance. Another important issue is that most systems do not create a real interview-like environment. Interviews in real life involve speaking, thinking on the spot, and expressing ideas clearly. However, many platforms only allow typed responses, which limits the ability of candidates to practice communication skills. As a result, users may feel unprepared when facing actual interviews. Lack of personalization is also a major concern. Every candidate has a different background, skill set, and career goal, but most systems offer the same set of questions to everyone. There is no proper mechanism to understand the user's profile and provide relevant questions based on their strengths or areas of interest. This makes the preparation process less effective and sometimes repetitive. Many platforms do not provide detailed feedback or progress tracking. Users are often unable to understand where they are going wrong or how they can improve over time. Without proper guidance and analysis, it becomes difficult to measure growth or identify weak areas.

V PROPOSED SYSTEM

The proposed system is developed as a smart and interactive interview preparation platform that aims to provide users with a more realistic and effective learning experience. It allows candidates to participate in mock interviews by selecting different categories such as technical and HR, along with customizable options like difficulty level and subject area. One of the important features of this system is the support for both text and voice-based responses. Users can answer questions by speaking, which helps them practice real-time communication, improve confidence, and get closer to actual interview conditions. The system processes voice input by converting it into text and then evaluating it, making the interaction smooth and user-friendly. It is designed with a simple interface so that users can easily navigate through questions and focus more on learning rather than handling complex tools.

The evaluation process in the proposed system goes beyond basic answer checking. It analyzes the overall meaning of the response, ensuring that answers are judged based on understanding rather than exact wording. This makes the scoring more fair and reliable. Along with evaluation, the system provides immediate feedback after each response, helping users identify mistakes and improve continuously. Another key feature is personalization, where the system can adjust the type of questions based on the user's skills or selected preferences, making practice sessions more relevant and targeted. In addition, a performance tracking module is included to record scores, monitor progress over time, and highlight areas that need improvement. Users can also generate reports to review their performance later. Overall, the proposed system focuses on combining intelligent evaluation, natural interaction, and personalized learning to create a more efficient and practical interview preparation solution.

VI METHODOLOGY

The proposed system follows a structured approach to design and implement an intelligent interview evaluation platform. The process begins with the collection and preparation of interview questions from multiple sources such as predefined datasets and online resources. These questions are organized into different categories like technical and HR, and are further classified based on

difficulty levels and topics. This structured organization helps in delivering relevant questions to users based on their selected preferences.

Once the questions are prepared, the system processes user inputs, which can be given either as text or voice. In the case of voice input, speech recognition technology is used to convert spoken responses into text format. The converted text then undergoes preprocessing steps such as removing unnecessary words, normalizing the text, and preparing it for evaluation. This ensures that the input is clean and suitable for further analysis.

The core part of the methodology lies in the answer evaluation process. Instead of relying only on keyword matching, the system analyzes the meaning of the response by comparing it with the expected answer. It checks how closely the user's answer aligns with the actual concept. Additional factors such as the presence of important keywords and the completeness of the answer are also considered. Based on these aspects, a score is generated, which reflects the overall quality of the response.

Another important step in the methodology is personalization. The system can use user-provided information, such as selected topics or skill sets, to generate more relevant questions. This helps in creating a focused learning experience tailored to individual needs. After evaluation, the system provides instant feedback, highlighting strengths and areas that require improvement.

All user interactions and scores are stored for future reference. A performance tracking mechanism is implemented to monitor progress over time. Users can view their past performance, analyze trends, and identify weak areas. This complete methodology ensures that the system not only evaluates answers effectively but also supports continuous learning and improvement.

VII IMPLEMENTATION

The system is developed as a web application that connects a user-friendly interface with a processing backend and a structured database. The interface is built using basic web technologies, allowing users to easily access the platform, choose interview types, and submit their responses. It is designed to be simple and clear so that users can focus on answering questions without any confusion. The backend handles all the internal operations, such as receiving user inputs, managing sessions, and coordinating different modules of the system. A database is used to store important information including user details, question sets, responses, and performance records, ensuring that all data is organized and easily retrievable.

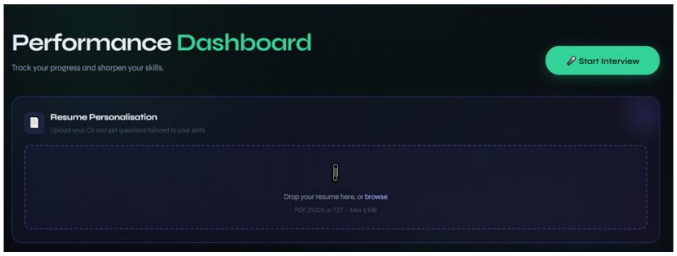
To support voice interaction, the system includes a speech input feature that converts spoken answers into text. This makes it possible for users to practice interviews in a more natural way. Once the response is available in text form, it is cleaned and prepared for evaluation by removing unnecessary words and standardizing the format. The evaluation process checks how well the response matches the expected answer in terms of meaning and relevance. It also considers whether key points are covered and whether the response is sufficiently detailed. Based on this, a score is generated and feedback is provided immediately. The system also keeps track of user performance over multiple sessions, allowing users to view their progress and identify areas that need improvement. An additional management feature is included to update questions and maintain the system content, ensuring that the platform remains useful and up to date.

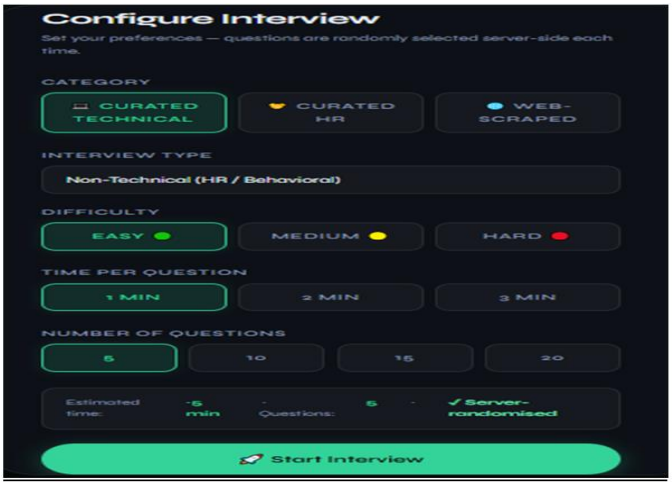
VIII RESULTS AND ANALYSIS

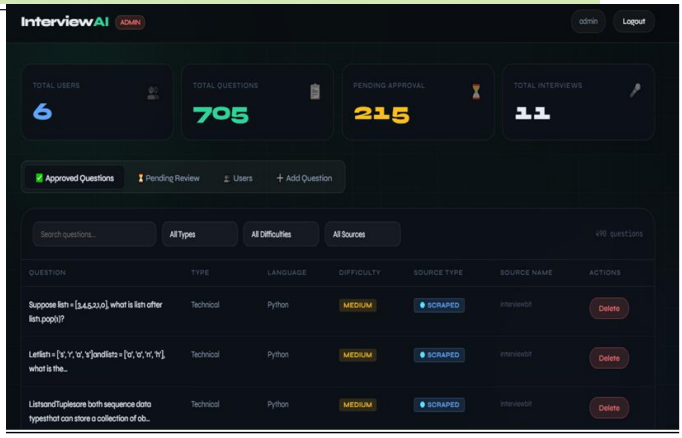
The system was tested with different users by allowing them to attempt various interview questions from both

International Journal of Engineering Research and Science & Technology

technical and non-technical areas. During testing, users provided answers through typing as well as speaking, and the system processed both types of inputs without difficulty. The responses were evaluated instantly, and scores along with feedback were displayed after each question. This immediate response helped users understand how well they performed and where they needed improvement.







While observing the results, it was noticed that the system was able to judge answers more effectively compared to simple matching-based methods. Even when users explained answers in their own words, the system could still recognize the correctness of the response if the main idea was present. This made the scoring more consistent and reduced chances of unfair evaluation. Users who used the voice feature mentioned that it felt more like a real interview, and it helped them practice speaking clearly and confidently.

The system also performed smoothly in terms of speed and handling multiple inputs. There were no major delays during evaluation, which allowed users to continue their practice without interruption. Over repeated attempts, many users showed improvement in their scores, indicating that the feedback provided by the system was useful. The progress tracking feature helped users see their development over time and focus on areas where they were weaker.

Another important observation was related to personalized practice. When users selected topics based on their interests or skills, they found the questions more relevant and engaging. This made the learning process more focused and less repetitive. Overall, the results show that the system provides a better learning experience by combining quick response, meaningful evaluation, and continuous performance monitoring.

IX CONCLUSION

A simple yet effective way to support interview preparation using technology. Instead of depending only on traditional practice methods, the system creates an environment where users can answer questions, receive feedback, and gradually improve their performance. The inclusion of both typing and speaking options makes the experience closer to real interviews and helps users become more comfortable in expressing their thoughts.

A key benefit of the system is how it checks answers by focusing on the idea behind the response rather than just looking for exact words. This makes the evaluation more meaningful and useful for learners. The feedback given after each attempt guides users in understanding their mistakes and encourages them to perform better in the next attempt. Over time, users can clearly see their progress and work on areas where they are not strong. the system offers a practical solution for candidates who want to prepare on their own without relying completely on others. It brings together ease of use, helpful feedback, and continuous tracking in one place, making the preparation process more organized and effective.

REFERENCES

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding.

[2] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT networks.

[3] Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.).

[4] Young, S., Evermann, G., Gales, M., et al. (2020). The HTK book (for speech recognition systems).

[5] Zechner, K., Higgins, D., Xi, X., & Williamson, D. (2009). Automatic scoring of non-native spontaneous speech.

[6] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition.

[7] Sayfullina, L., Malmi, E., Kannala, J., & Gionis, A. (2018). Learning to extract skills from resumes.

[8] Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*.

[9] Wainer, H. (2000). *Computerized Adaptive Testing: A Primer*.

[10] Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification.