IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155140

AI-POWERED SMART INTERVIEW ASSISTANT WITH REAL-TIME FEEDBACK AND SCORE PREDICTION

Dr. Umma Habiba[1], Jeni Pragashini J[2], Kayathri A[3], Yogalakshmi M[4]

Dean of SOAC, Department of Artificial Intelligence and Data Science,

Dhanalakshmi Srinivasan College of Engineering and Technology, Mamallapuram[1]

Department of Artificial Intelligence and Data Science,

Dhanalakshmi Srinivasan College of Engineering and Technology, Mamallapuram[2,3,4]

Abstract: The rapid growth of Artificial Intelligence has created new opportunities to improve traditional interview preparation methods. Many students struggle with confidence, communication and evaluating the quality of their responses during job interviews. This paper proposes an AI-Powered Smart Interview Assistant that evaluates candidate responses and provides real-time feedback and score prediction using machine learning and natural language processing techniques. The proposed system allows candidates to answer interview questions through text or voice input. The system analyzes the responses using natural language processing methods to evaluate relevance, grammar, semantic similarity and communication quality. A machine learning model then predicts the interview score based on extracted linguistic and contextual features. The system was tested with different categories of answers including relevant, irrelevant, short and filler-based responses. Experimental results show that the proposed system achieves model accuracy between 88 percent and 92 percent while maintaining an average response time of less than three - five seconds. The speech recognition component achieves approximately 90 percent accuracy in converting spoken responses into text. The system provides immediate suggestions that help candidates improve communication skills, reduce interview anxiety and enhance overall interview readiness. The proposed framework demonstrates the potential of artificial intelligence in building intelligent training tools that support career preparation and professional development.

Keywords: Artificial Intelligence, Interview Assistant, Natural Language Processing, Machine Learning, Speech Recognition, Score Prediction

INTRODUCTION

Interviews play a crucial role in the recruitment process across most industries. Employers use interviews to evaluate a candidate's communication ability, subject knowledge, problem solving skills and overall personality. However many students and job seekers experience anxiety and lack of preparation when facing real interviews. Traditional preparation methods such as mock interviews conducted by trainers or mentors are helpful but often limited by time, cost and availability of experts. With the rapid development of artificial intelligence technologies, intelligent systems can now assist candidates in preparing for interviews in a more accessible and scalable way. Artificial Intelligence and Natural Language Processing have enabled machines to understand and analyze human language. These technologies can be used to evaluate interview answers, measure response quality and provide automated feedback to users. Intelligent interview preparation platforms can simulate interview scenarios and help candidates practice answering questions while receiving guidance. The proposed AI-Powered Smart Interview Assistant aims to create a realistic interview practice environment where users can answer questions through voice or text. The system evaluates the response content, predicts a performance score and provides suggestions for improvement. The goal of the system is to enhance interview readiness, improve communication skills and build candidate confidence before attending real interviews.

RELATED WORKS

Automated interview assessment systems have gained significant attention over the past two decades, evolving from rule-based evaluation methods to sophisticated artificial intelligence (AI) driven platforms capable of analyzing speech, language, and behavioral cues. Early research primarily focused on text-based evaluation using classical Natural Language Processing (NLP) techniques. For instance, Christopher D. Manning and Hinrich Schütze (1999) introduced foundational NLP methods for linguistic feature extraction and text classification, which later influenced automated

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155140

assessment systems [1]. Similarly, Thorsten Joachims (1998) demonstrated the effectiveness of Support Vector Machines (SVM) for text categorization tasks, establishing an early machine learning approach for evaluating candidate responses in interview scenarios.As machine learning techniques matured, researchers began applying statistical models to analyze candidate responses and predict employability outcomes [2]. Leo Breiman (2001) introduced Random Forest classifiers, which were later used in recruitment analytics for predicting candidate success based on structured and unstructured data [3]. Additionally, Tomas Mikolov et al. (2013) proposed Word2Vec, a distributed representation model that significantly improved semantic understanding in textual responses.

These word embeddings enabled automated systems to capture contextual relationships in candidate answers, improving evaluation accuracy [4]. The rapid development of deep learning marked a significant paradigm shift in automated interview analysis. The breakthrough work of Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton (2012) on deep convolutional neural networks (CNNs) demonstrated the superior capability of deep architectures in learning hierarchical representations from large datasets [5]. Although originally applied to computer vision tasks, these architectures inspired researchers to adopt deep neural networks for multimodal interview analysis, including facial expressions, speech signals, and textual responses. In the field of speech processing, the emergence of deep learning significantly improved speech recognition systems. Geoffrey Hinton et al. (2012) introduced deep neural network-based acoustic models that substantially enhanced automatic speech recognition (ASR) accuracy [6].

Later, Alec Radford et al. (2022) developed the Whisper speech recognition model under OpenAI, capable of robust multilingual speech transcription. Such advancements enabled AI-based interview systems to analyze spoken responses with high precision, making real-time voice-based evaluation feasible [7]. Natural language understanding models further advanced with the introduction of transformer architectures. Ashish Vaswani et al. (2017) [8] proposed the Transformer architecture in the seminal work *"Attention Is All You Need."* This innovation enabled large-scale contextual language modeling and subsequently led to the development of powerful models such as BERT by Jacob Devlin et al. (2019) .BERT demonstrated remarkable performance in text classification, question answering, and sentiment analysis tasks, making it highly suitable for automated interview response evaluation [9].

Recent studies have focused on multimodal AI systems that integrate text, speech, and visual cues to simulate human interviewer evaluation. Louis-Philippe Morency et al. (2011) introduced multimodal sentiment analysis techniques that combine facial expressions, speech patterns, and linguistic features [10]. These approaches later influenced automated interview platforms that assess candidate confidence, tone, and communication skills.

In addition to prediction accuracy, interpretability has become a crucial requirement for AI-based recruitment systems. Explainable Artificial Intelligence (XAI) techniques have been proposed to provide transparency in machine learning decisions. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin (2016) introduced Local Interpretable Model-agnostic Explanations (LIME), which explains model predictions by approximating them with interpretable surrogate models [11]. Similarly, Ramprasaath R. Selvaraju et al. (2017) proposed Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize decision regions in deep neural networks [12].

Further advancements in explainability include the development of SHAP (SHapley Additive exPlanations) by Scott Lundberg and Su-In Lee (2017) [13], which applies cooperative game theory to determine feature importance. These techniques are increasingly integrated into AI interview assistants to justify automated scoring decisions and improve transparency for recruiters [14]. Despite substantial progress, existing automated interview systems often focus on either speech analysis or textual evaluation independently. Few systems combine deep learning-based NLP models, speech recognition, and explainable AI techniques within a unified architecture capable of providing real-time feedback. Moreover, many current solutions lack interpretability mechanisms that allow recruiters and candidates to understand how evaluation scores are generated. Therefore, the development of an integrated AI-based interview assistant that combines deep learning-driven NLP analysis, speech recognition, and explainable AI frameworks such as LIME and Grad-CAM represents an important research direction. Such systems can improve the fairness, transparency, and reliability of automated interview evaluation platforms while enabling scalable recruitment processes.

METHODOLOGY

This section describes the detailed methodology of the proposed AI-Powered Smart Interview Assistant system. The proposed framework is designed to automatically analyze candidate responses during interviews and provide real-time feedback along with performance score prediction. The workflow consists of several modules including data acquisition, data preprocessing, feature extraction, model construction, speech processing, prediction, explainability, and performance evaluation. Each module performs a specific function that contributes to the overall functionality of the intelligent interview assessment system.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155140

A. Data Acquisition Module

The data acquisition module is responsible for collecting interview-related datasets consisting of questions, candidate responses, and evaluation labels. The dataset includes both textual answers and speech responses provided by candidates during interview simulations.

Algorithm:

1. Start the data acquisition process.
2. Collect interview questions from publicly available interview preparation datasets and recruitment platforms.
3. Gather candidate responses in two formats:
   a. Text-based responses
   b. Voice recordings of spoken answers.
4. Assign evaluation labels to each response based on parameters such as relevance, communication quality, confidence level, and correctness.
5. Verify the dataset to ensure that each response contains corresponding evaluation labels.
6. Store the collected dataset in a structured database for further analysis and processing.
7. End of data acquisition module.

B. Data Preprocessing Module:

The raw interview responses require preprocessing to improve machine learning performance and ensure consistency across the dataset. This module prepares both text and audio data for analysis.

Algorithm:

1. For text responses:
   a. Convert text into lowercase format.
   b. Remove punctuation marks and special characters.
2. Remove stop words that do not contribute to meaningful analysis.
3. Perform tokenization to split sentences into individual words.
4. Apply stemming or lemmatization to convert words into their root forms.
5. For speech responses:
   a. Convert voice recordings into text using speech recognition models such as Whisper.
6. Remove incomplete, noisy, or corrupted responses from the dataset.
7. Organize the cleaned data into training, validation, and testing sets.
8. End of preprocessing module.

C. Feature Extraction Module:

The feature extraction module converts candidate responses into numerical representations that can be processed by machine learning models.

Algorithm:

1. Initialize the feature extraction process.
2. Convert each candidate response into numerical feature vectors.
3. Use semantic embedding models such as BERT to capture contextual meaning.
4. Extract linguistic features such as:
   a. word frequency
   b. sentence length
   c. keyword relevance
5. Combine extracted features into a unified feature vector.
6. Store feature vectors for training the prediction model.
7. End of feature extraction module.

D. Machine Learning Model Construction Module:

This module constructs the machine learning architecture used to evaluate interview responses and predict performance scores.

Algorithm:

1. Initialize the machine learning model.
2. Input feature vectors extracted from candidate responses.
3. Train multiple classification algorithms including:
   a. Logistic Regression
   b. Decision Tree
   c. Random Forest.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155140

4. Compare model performance using validation datasets.
5. Select the best-performing model based on accuracy and F1-score.
6. Configure model hyperparameters for optimal performance.
7. Compile the final model for training and prediction.
8. End of model construction module.

E. Model Training Module:
The training module enables the system to learn patterns from interview responses and predict evaluation scores accurately.

Algorithm:
1. Initialize training parameters such as:
    a. batch size
    b. learning rate
    c. number of iterations.
2. Input training data into the machine learning model.
3. Train the model using labeled interview responses.
4. Compute classification errors using appropriate loss functions.
5. Update model parameters through iterative optimization.
6. Validate the model using validation datasets.
7. Save the best-performing model configuration.
8. End of model training module.

F. Prediction Module:
The prediction module evaluates candidate responses during interview sessions and generates real-time performance feedback.

Algorithm:
1. Load the trained evaluation model.
2. Input a candidate response into the system.
3. If the response is speech-based, convert it into text using speech recognition.
4. Extract features from the processed response.
5. Feed the feature vector into the trained model.
6. Obtain predicted evaluation scores and response classification results.
7. Display feedback and performance score to the candidate.
8. End of prediction module.

G. Explainability Module:
Explainable Artificial Intelligence (XAI) techniques are integrated to provide transparency in model predictions and highlight the factors influencing the evaluation.

Algorithm:
1. Input the candidate response into the trained prediction model.
2. Generate perturbed variations of the response text.
3. Evaluate predictions for each variation.
4. Fit a local surrogate model to interpret prediction behavior.
5. Identify the most influential words or phrases contributing to the prediction.
6. Highlight these words in the response explanation interface.
7. Display the explanation to the user along with the predicted score.
8. End of explainability module.

H. Performance Evaluation Module:
This module evaluates the effectiveness of the proposed AI-based interview assistant using standard machine learning metrics.

Algorithm:
1. Calculate classification metrics including:
    a. Accuracy
    b. Precision
    c. Recall
    d. F1-score.
2. Generate confusion matrices to analyze prediction performance.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155140

3. Evaluate speech recognition accuracy for voice inputs.
4. Measure system response time during real-time evaluation.
5. Compare performance across different machine learning models.
6. Store evaluation results and performance graphs for analysis.
7. End of performance evaluation module.

PROPOSED MODEL

The primary objective of the proposed AI-Powered Smart Interview Assistant is to automatically evaluate candidate responses during interview practice sessions and provide intelligent feedback that helps improve communication and answer quality. The system analyzes candidate responses using Natural Language Processing (NLP), speech recognition, and machine learning algorithms to determine the relevance, clarity, and effectiveness of answers. In addition to predicting performance scores, the system integrates Explainable Artificial Intelligence (XAI) techniques to provide transparent feedback highlighting the words and phrases that influenced the evaluation.

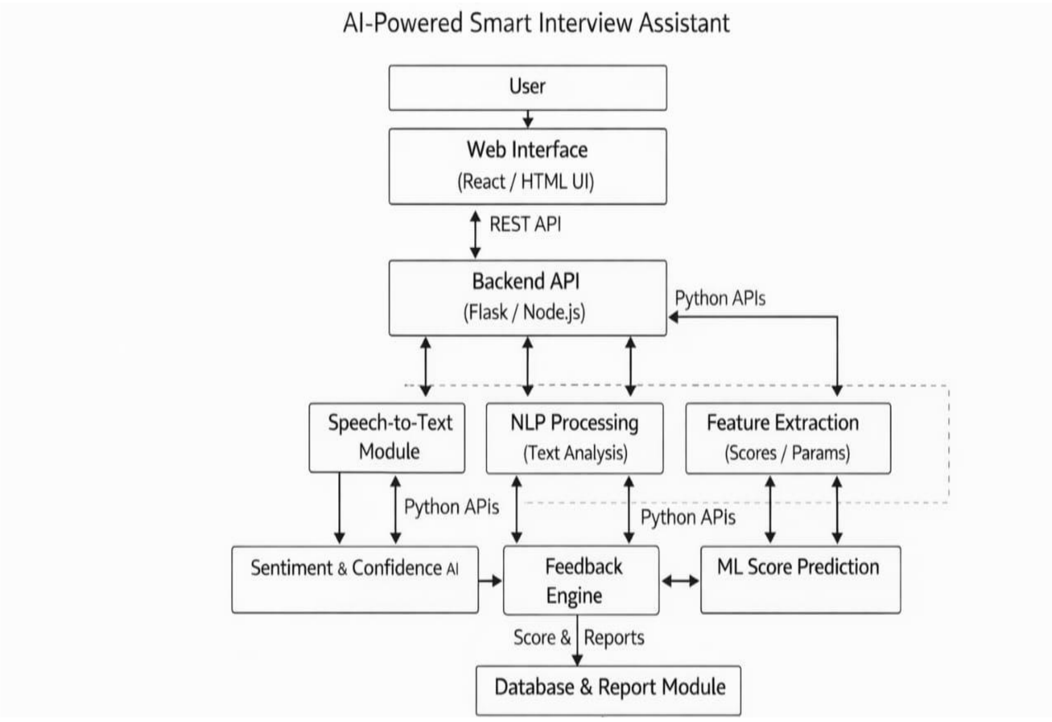

FIGURE 1**:** illustrates the overall architecture of the proposed framework. The architecture demonstrates the sequential workflow of the system beginning from data acquisition and preprocessing, followed by machine learning-based interview response evaluation, and finally explainable feedback generation and visualization.

In Phase I, The system collects interview-related datasets consisting of interview questions, candidate responses, and corresponding evaluation labels. The responses can be provided either in textual format or through speech recordings. Speech responses are converted into text transcripts using automatic speech recognition systems such as Whisper. The converted transcripts are then processed using Natural Language Processing techniques to prepare them for machine learning analysis. The preprocessing stage includes several operations such as text cleaning, tokenization,stop-word removal, and lemmatization. These operations help standardize candidate responses and remove unnecessary linguistic variations.To capture the contextual meaning of sentences, textual responses are transformed into semantic embeddings using transformer-based language models such as BERT. These embeddings represent candidate answers as numerical vectors that preserve semantic relationships between words.The dataset is then divided into training, validation, and testing subsets, enabling reliable model training and evaluation.Start the data collection process.Collect interview questions and candidate responses from interview datasets and preparation platforAccept candidate responses in two formats:Textual answers,Speech recordings.Convert speech responses into textual transcripts using speech recognition technology.Perform text preprocessing operations including:Lowercase conversion,Removal of punctuation and special characters,Stop-word removal.Apply tokenization and lemmatization to normalize the responses.Convert processed text

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155140

into contextual embeddings using transformer models.Verify dataset consistency and remove incomplete or noisy responses.Split the dataset into training, validation, and testing sets.End of Phase I.

In Phase II, the system evaluates candidate responses using machine learning models trained to classify answer quality and predict interview performance scores.The extracted feature vectors are provided as input to several classification models including Logistic Regression, Decision Tree, and Random Forest. Among these models, the **Random Forest classifier is selected as the final model due to its high prediction accuracy and robustness.Random Forest is an ensemble learning technique that constructs multiple decision trees during training and combines their predictions to improve generalization performance. This approach helps reduce overfitting and improves prediction stability across different interview response patterns.The model is trained using labeled interview responses and optimized through hyperparameter tuning. Cross-validation techniques are used to evaluate the model and ensure reliable performance across unseen data.initialize the machine learning classification framework.Input feature vectors extracted from candidate responses.Train multiple classification algorithms including:Logistic Regression,Decision Tree,Random Forest.Evaluate model performance using validation datasets.Optimize model hyperparameters to improve accuracy and generalization.Select the best-performing model based on evaluation metrics.Train the selected model using the complete training dataset.Save the trained model for real-time prediction.End of Phase II.

In Phase III Although machine learning models can provide accurate predictions, their decision-making process is often difficult to interpret. To enhance transparency and user trust, the proposed system incorporates Explainable Artificial Intelligence (XAI) techniques.The system uses Local Interpretable Model-Agnostic Explanations (LIME) to identify the most influential words or phrases in a candidate's response. LIME generates multiple perturbed variations of the input text and observes how these variations affect model predictions. Based on these observations, a local surrogate model is trained to estimate the contribution of each word or phrase.The explanation results are presented visually by highlighting important keywords in the candidate's answer. This allows users to understand which parts of their responses were evaluated positively or negatively by the AI system.Input the candidate response into the trained evaluation model.Generate predicted performance scores and classification labels.Apply LIME to interpret the prediction results.Generate multiple perturbed versions of the input response.Evaluate predictions for each perturbed response.Train a local surrogate model to approximate prediction behavior.Identify the most influential words or phrases affecting the evaluation.Highlight important words in the response explanation interface.Display feedback and explanation to the user.End of Phase III.

IMPLEMENTATION

The implementation of the proposed AI-Powered Smart Interview Assistant involves several stages, including data preparation, feature extraction, machine learning model training, prediction generation, and explainable feedback visualization. The main objective of the system is to automatically evaluate candidate responses during interview practice sessions and provide real-time feedback along with predicted performance scores. By integrating Natural Language Processing (NLP), speech recognition, and Explainable Artificial Intelligence (XAI), the system not only evaluates responses accurately but also provides transparent insights into the factors influencing the evaluation results.

The proposed interview assistant system is implemented using machine learning and NLP techniques combined with speech recognition technologies. The dataset used in this study consists of interview questions and candidate responses collected from interview preparation platforms and simulated interview sessions. The responses are categorized based on evaluation parameters such as relevance, completeness, communication clarity, and confidence level. Each response is labeled according to its quality and effectiveness to enable supervised learning.The collected responses are organized into structured datasets and divided into training, validation, and testing subsets to ensure reliable and unbiased model evaluation. Prior to model training, several preprocessing operations are performed to prepare the data for analysis. For textual responses, preprocessing steps include converting text into lowercase, removing punctuation symbols, eliminating stop words, and performing tokenization. These operations standardize the textual inputs and reduce noise within the dataset. In addition, lemmatization is applied to transform words into their root forms, allowing the model to recognize similar semantic structures across different responses. In the case of speech-based interview responses, audio recordings are converted into textual transcripts using speech recognition technology such as Whisper. The converted transcripts are then processed using the same NLP preprocessing pipeline applied to textual responses. This ensures that both speech and text responses can be analyzed uniformly within the machine learning framework.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155140

To capture the contextual meaning of candidate responses, the processed text is transformed into numerical feature vectors using semantic embedding models such as BERT. These embeddings encode linguistic and contextual relationships between words and sentences, enabling the model to evaluate the semantic quality of answers more effectively. For response classification and score prediction, several machine learning models are implemented and evaluated, including Logistic Regression, Decision Tree, and Random Forest classifiers. Among these models, the **Random Forest classifier demonstrates the highest prediction accuracy and robustness. Random Forest operates as an ensemble learning method that constructs multiple decision trees during training and combines their outputs to produce a final prediction. This approach improves generalization capability and reduces the risk of overfitting. During the training phase, feature vectors extracted from candidate responses are provided as input to the machine learning models. The models learn patterns associated with high-quality and low-quality responses by analyzing labeled training data. Model optimization is achieved through hyperparameter tuning and cross-validation techniques. The trained model is then stored and integrated into the interview assistant system for real-time prediction. Although machine learning models can provide accurate evaluation results, their internal decision-making mechanisms often lack interpretability. To address this challenge, the proposed system integrates Explainable Artificial Intelligence (XAI) techniques to provide transparent feedback to users. The system employs Local Interpretable Model-Agnostic Explanations (LIME) to identify the most influential words or phrases in a candidate's response that contribute to the predicted score.

LIME works by generating multiple perturbed variations of the candidate's response and observing how these variations affect the model's prediction probabilities. Based on these observations, a locally weighted linear surrogate model is constructed to approximate the behavior of the original prediction model. The surrogate model determines the contribution of individual words or phrases to the final evaluation result. The explanation results are visualized by highlighting important keywords or phrases within the candidate's response. These highlighted segments indicate which parts of the answer influenced the system's evaluation positively or negatively. This approach helps candidates understand the strengths and weaknesses of their responses and provides actionable feedback for improvement. The final system output includes several components: the predicted response quality classification, the overall performance score, highlighted keywords influencing the evaluation, and detailed textual feedback. Visualization is performed using standard data visualization and text processing libraries to ensure clear and interpretable presentation of results. Experimental evaluation demonstrates that the proposed system achieves a classification accuracy ranging between 88% and 92%, depending on the size and diversity of the dataset used for training. Speech recognition accuracy for voice inputs is approximately 90%, and the average response evaluation time is less than three seconds, making the system suitable for real-time interview practice environments. By integrating machine learning techniques with explainable AI mechanisms, the proposed implementation not only provides accurate automated interview evaluation but also improves transparency and user trust. The system serves as an intelligent interview preparation tool that helps candidates enhance their communication skills, refine answer structures, and build confidence before participating in real recruitment interviews.

CONCLUSION

This research presented the design and implementation of an AI-Powered Smart Interview Assistant aimed at improving interview preparation through automated evaluation and intelligent feedback mechanisms. The system integrates Natural

DOI: 10.17148/IJARCCE.2026.155140

Language Processing, speech recognition, machine learning algorithms, and Explainable Artificial Intelligence techniques to analyze candidate responses and provide meaningful performance assessments. By combining textual analysis with speech-to-text conversion, the proposed system can evaluate both written and spoken responses, making it adaptable for various interview practice scenarios.

The experimental results demonstrate that the implemented models are capable of accurately predicting response quality and generating reliable performance scores. Among the evaluated machine learning models, the Random Forest classifier achieved the best performance due to its ensemble learning capability and ability to generalize effectively across diverse response patterns. The integration of semantic embeddings enabled the system to capture contextual meaning within candidate responses, thereby improving the accuracy of response evaluation. In addition to prediction accuracy, interpretability plays a crucial role in automated decision-making systems. To address this requirement, the system incorporates Explainable Artificial Intelligence using the LIME technique, which identifies influential words and phrases that contribute to the predicted evaluation outcome. This transparency allows users to understand how their responses are assessed and provides actionable insights for improving communication, relevance, and clarity in future answers. Overall, the proposed AI-Powered Smart Interview Assistant demonstrates the potential of combining machine learning with explainable AI to create an intelligent and user-centric interview preparation platform. The system not only automates response evaluation but also enhances learning by providing clear, interpretable feedback. In the future, the system can be further improved by incorporating advanced deep learning models, emotion detection from facial expressions, and personalized feedback mechanisms to create a more comprehensive and interactive interview training environment

REFERENCES

[1]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the NAACL-HLT Conference*, Minneapolis, USA, 2019, pp. 4171–4186.

[2]. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016, pp. 1135–1144.

[3]. Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4]. Scott Lundberg and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, Long Beach, USA, 2017.

[5]. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[6]. Christopher M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[7]. Jeffrey Pennington, Richard Socher, and Christopher D. Manning, "GloVe: Global Vectors for Word Representation," *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[8]. Matthew E. Peters, Mark Neumann, Mohit Iyyer, et al., "Deep Contextualized Word Representations," *Proceedings of NAACL*, 2018.

[9]. Alec Radford, Jong Wook Kim, Chris Hallacy, et al., "Robust Speech Recognition via Large-Scale Weak Supervision," 2022. (Speech recognition model related to OpenAI Whisper).

[10]. Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.

[11]. W. Samek, T. Wiegand, and K. Müller, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models," *arXiv preprint*, 2017.

[12]. A. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, 2014.

[13]. J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding Neural Networks Through Deep Visualization," *Proceedings of the Neural Information Processing Systems (NeurIPS)*, 2015.

[14]. S. Ahmad, P. Katari, A. Kumar, et al., "Explainable AI: Interpreting Deep Learning Models for Decision Support," *Advances in Deep Learning Techniques*, vol. 4, no. 1, pp. 80–107, 2024.

[15]. L. Chen, Y. Song, and A. Schmidt, "Model Interpretability in Deep Learning: A Comprehensive Review," *IEEE Access*, vol. 8, pp. 86524–86537, 2020.

