Smartprep AI: Advanced Interview Simulation Using AI, Speech Recognition and Real Time Feedback Systems

Tanay A. Jadhav Department of Computer Science and Engineering
K.K Wagh Institute Of Engineering Education and Research
tanayjadhav007@gmail.com

Ritesh S. Shevare Department of Computer Science and Engineering
K.K Wagh Institute Of Engineering Education and Research
riteshshevare@gmail.com

Om M. Taskar
Department of Computer Science and Engineering
K.K Wagh Institute Of Engineering Education and Research
ommaheshtaskar@gmail.com

Pratik D. Kotkar
Department of Computer Science and Engineering
K.K Wagh Institute Of Engineering Education and Research
Pratikkotkar7575@gmail.com

Dr. Y.D. Bhise
Professor Of Department of Computer Science and Engineering
K.K Wagh Institute Of Engineering Education and Research
tanayjadhav007@gmail.com

Abstract - Smart Prep AI is an advanced interview preparation platform designed to address the limitations of traditional methods by providing real-time, personalized feedback. In today's competitive job market, candidates need more than just qualifications—they must excel in communication, confidence, and emotional intelligence. Smart Prep AI analyzes user resumes to generate domain-specific interview questions, creating a tailored simulation experience. Responses are captured in audio, video, and text formats, enabling a comprehensive evaluation of accuracy, grammar, confidence, and emotional tone. With detailed feedback and progress tracking, users can identify areas for improvement and enhance their interview performance. By offering a structured, data-driven approach, Smart Prep AI helps candidates become more confident and well-prepared, significantly increasing their chances of securing their desired job.

Keywords - Interview, Bot, Machine Learning, ChatBot

I. INTRODUCTION

Smart Prep AI is an advanced, AI-powered platform designed to simulate professional interview environments, offering a comprehensive prepare on tool for job seekers. User will upload their Job Role and Qualities & Experience , this three thing will be recorded and analyzed by the system to generate tailored, domain specific interview questions relevant to the user's qualifications and job application. This intelligent ques on genera on ensures that the interview process aligns with the user's field, providing a personalized and focused experience.

During the simulated interview, Smart Prep AI presents questions in real time, allowing users to respond verbally. These responses are captured in audio, video, and text formats, mimicking an actual interview setting. The system's advanced analysis capabilities then evaluate these responses across multiple dimensions, including fact-checking, sentiment analysis, grammar, confidence levels, and emotion detection. This comprehensive assessment provides users with a detailed scorecard, offering valuable feedback on their performance and highlighting areas for improvement.

II. LITERATURE SURVEY

The paper *"QAI: An AI-powered Mock Interview Bot for Enhancing the Performance of Aspiring Professionals"* presents a system designed to help job seekers improve their interview skills through AI-

driven mock interviews. This system generates personalized interview questions based on a user's resume and analyzes their responses for sentiment factual accuracy, language proficiency, confidence, and emotional expression. The platform provides detailed feedback and a scorecard to help users track their performance over time, thereby

offering a structured approach to refining their interview skills.

Data Preparation: The uploaded resume is converted to text, and key features are extracted to generate relevant interview questions.

Training Phase: Transformer models like BERT or ChatGPT analyze the resume, generate questions, and assess responses based on various parameters such as confidence and accuracy.

Testing Phase: The user's responses, captured in audio, video, and text, are processed using models for sentiment analysis, emotion detection, and language proficiency evaluation.

Final Output: A detailed scorecard is generated, providing personalized feedback on each aspect of the interview, helping users refine their performance.

The mathematical framework includes sentiment analysis and emotion detection using pre-trained models like Gensim Word2Vec for sentiment evaluation and CNN-based models for facial emotion recognition. The system applies various machine learning classifiers, such as SVM and Naïve Bayes, to categorize sentiment and confidence, ensuring that the feedback is both accurate and comprehensive. [1]

Significance of Sentiment Analysis Approaches Using Machine Learning (ML) Techniques

The document *"Significance of Sentiment Analysis Approaches Using Machine Learning (ML) Techniques"* presents an in-depth study of how machine learning is applied to sentiment analysis, focusing on identifying and classifying emotions or opinions from various data forms, including text, audio, and video. It explores different techniques, challenges, and methodologies in sentiment analysis, covering both text-based and multimedia (audio/video) approaches. Additionally, it highlights the growing importance of this technology in business, social media monitoring, customer feedback analysis, and market research.

Speech-to-Text Conversion: Before applying sentiment analysis to audio or video data, speech- to-text conversion is performed using technologies like Google Cloud Speech-to-Text, which transcribes spoken content into text.

Multimodal Sentiment Analysis: This technique combines audio and video data, leveraging both textual content and emotional cues from speech and facial expressions for a more comprehensive sentiment analysis.

Emotion Recognition: Machine learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are trained to recognize emotions like happiness, sadness, and anger based on audio or visual data.

The mathematical framework involves models like BERT for contextual sentiment understanding, along with algorithms such as SVMs, Naïve Bayes, and RNNs for emotion classification. For audio sentiment analysis, features like Mel-frequency cepstral coefficients (MFCCs) are extracted and analysed to identify emotional cues in speech. These models process data, including pitch and rhythm, to infer sentiment from both text and multimedia sources. The integration of deep learning models enhances the ability to recognize complex emotional patterns, making sentiment analysis more robust and accurate. [2]

Conversational Etiquette: The system tracks metrics such as the speak-to-listen ratio, silence ratio, and speaking pace to evaluate how well the interviewer engages in two-way communication.

Behaviour Monitoring: Metrics such as the interviewer's video presence, background noise, and punctuality are tracked to assess professional behaviour during interviews.

Interview Structure: SmartView detects and classifies interview questions into categories (e.g., functional, situational) to evaluate how structured and relevant the interview is.

This work involves using models like BERT for question classification and rVAD for voice activity detection. For question detection and classification, BERT is fine-tuned on labeled interview data to categorize questions into predefined types. In interviewer behavior analysis, OpenFace is used to process video frames and detect visual presence. This

combination of NLP and computer vision techniques enables SmartView to generate detailed and accurate feedback for interviewers based on real-world data. [3]

III. PROPOSED METHODOLOGY

The users of **Smart Prep AI** can sign up or log into the desktop application. From the home page, users will enter their desired job role, specific skills, and experience details. Based on this information, Smart Prep AI generates domain- specific interview questions relevant to the job application. A real-time environment is utilized to ask each generated question, mimicking an actual interview setting. The user's responses are recorded in audio, video, and text formats and sent for further processing.

An advanced analysis system evaluates responses based on factual accuracy, sentiment, grammar, confidence levels, and emotion detection. Based on this assessment, users receive scores and feedback, which they can track over time to monitor their progress. The system consists of four major modules: **User authentication module, question generation module, user interaction module and analysis module.**

The architecture of the proposed methodology is given in Figure 1.

A. User Authentication Module

Registered users can log in using their username and password, which are validated against stored credentials in the database. New users can sign up by entering their details, such as name, date of birth, email, username, and password. Once verified, users are directed to the home page of the application.

B. Question Generation Module

Users provide details such as job role, specific skills, and prior experience. The system utilizes transformer models like BERT, RoBERTa, or ChatGPT to analyse this input and generate relevant interview questions. These AI models ensure high-quality, contextually appropriate questions that align with the chosen job role and skill set. The generated questions are stored for use in the interview session.

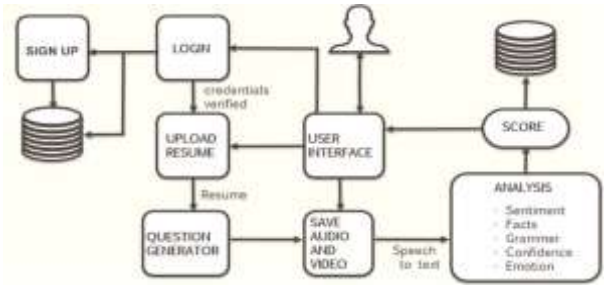

Fig 1: Architecture of the Proposed Methodology

User Interaction Module

The system presents questions to the user one by one, simulating a real interview. The user responds verbally, and the responses are captured via the camera and microphone. The audio is converted into text format, and the responses are forwarded for analysis. The system ensures a realistic experience by activating the microphone within seconds after displaying a question. The user has options to submit their response or quit the interview.

C. Analysis Module

1) Language Proficiency

Python's language tool library is used to detect grammatical mistakes in user responses. The text is parsed into words, phrases, and sentences, and linguistic rules are applied iteratively to check for grammar violations and stylistic inconsistencies. Errors are categorized by grammar correctness, sentence structure, and clarity. A proficiency score (1-5) is assigned based on the severity of detected errors.

2) Factual Accuracy

User responses are summarized and sent to an AI model, such as ChatGPT, for fact-checking against the context of the question. A summarization model (e.g., Hugging Face Transformers) refines the text before AI evaluation. The AI assigns a factual accuracy score (1-5) based on content correctness and relevance.

3) Emotion Detection

Video responses are processed frame-by-frame using convolutional neural networks (CNNs) to detect facial expressions and classify emotions such as happiness, sadness, and anger. The DeepFace library simplifies this process by analyzing facial features and identifying the dominant emotion throughout the interview.

4) Sentiment Analysis

Text-based responses undergo data preprocessing, including stop word removal, tokenization, stemming, and lemmatization. Pre-trained word embeddings from Gensim's Word2Vec model

transform responses into numerical vectors. A classifier such as SVM (Support Vector Machine) determines whether the sentiment is positive or negative. The final sentiment score represents the ratio of positive to negative sentiment during the interview.

IV. RESULTS AND DISCUSSIONS

The different phases of SmartPrep AI are evaluated in this section. The results obtained from different modules are summarized.

A. Question Generation Module

SmartPrep AI introduces an advanced role- based analysis system that transforms candidate information into meaningful insights. By employing cutting-edge transformer models like BERT, RoBERTa, and GEMINI, the system generates precise and context-aware interview questions tailored to the selected role. Through API-driven analysis, it interprets key professional attributes, ensuring the questions align with industry expectations and skill requirements. This AI- powered approach enhances the relevance of interview preparation, offering a structured and intelligent way to assess a candidate's expertise. The generated questions are highly refined, closely mirroring real-world interview expectations, making the system an invaluable tool for professional growth and readiness.

B. Language Proficiency

Language proficiency is evaluated using natural language processing techniques, where grammar and spelling are systematically analyzed. Sentences are assessed for proper noun-phrase construction, and errors are identified and quantified. Each detected error is assigned a specific weight, contributing to an overall language proficiency score. This structured approach ensures a precise assessment of communication skills, helping candidates refine their linguistic accuracy and clarity.

C. Factual Accuracy

Factual accuracy is determined by measuring the similarity between the given and expected responses using cosine similarity techniques. Advanced transformer models, trained on diverse datasets, are employed to evaluate the correctness of answers. This AI-driven approach ensures reliable accuracy verification, enabling a more precise assessment of a candidate's knowledge and response credibility.

D. Emotion Detection

Upon receiving a video input, the system processes it using a robust facial emotion recognition module based on DeepFace, which identifies the most likely emotion in each frame. The system aggregates this information to determine the emotion that predominates over time, quantifying it as a percentage. This streamlined method delivers precise emotion analysis, offering valuable insights into the dominant emotional expression throughout the video, and enhancing the system's capacity for detailed emotion tracking and interpretation.

E. Sentiment Analysis

The system processes user text responses through a detailed pre-processing pipeline, which includes removing unwanted characters, converting text to lowercase, eliminating punctuation and stopwords, and applying tokenization, stemming, and lemmatization. A pre- trained 300-dimensional word embedding model from Gensim, based on Word2Vec, is then used to capture semantic relationships within the text. The transformed data is converted into numerical vectors for further analysis. Machine learning classifiers are employed to determine whether the sentiment expressed is positive or negative. Key evaluation metrics such as precision, recall, and F1-score are used to assess accuracy. The system calculates the ratio of positive to negative sentiment throughout the interaction, offering a quantitative measure of the overall sentiment conveyed. This structured approach ensures a deep and precise understanding of sentiment trends in textual responses.

V. CONCLUSION

In today's competitive job landscape, mastering interview skills is essential for career advancement. SmartPrep AI emerges as a cutting- edge solution, redefining interview preparation through advanced artificial intelligence and an intuitive interface. Designed to create an immersive and structured practice environment, this system tailors questions based on the user's selected role, ensuring relevance and depth. By simulating real-world interview scenarios, it enables users to enhance their communication, problem-solving, and presentation abilities. Future advancements could include AI-driven interviewer avatars for a more interactive experience, along with seamless integration into university placement portals to expand accessibility for aspiring professionals.

REFERENCES

[1]. Mandal, Rubi, Pranav Lohar, Dhiraj Patil, Apurva Patil, and Suvarna Wagh. "AI-Based mock interview evaluator: An emotion and confidence classifier model." In 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS),
pp. 521-526. IEEE, 2023.

[2]. Pandey, Raj, Divya Chaudhari, Sahil Bhawani, Omkar Pawar, and Sunita Barve. "Interview Bot with Automatic Question Generation and Answer Evaluation." In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 1279- 1286. IEEE, 2023.

[3]. Zhang, Zhiming, Chao Sun, and Beibei Guo. "Transfer-learning guided Bayesian model updating for damage identification considering modeling uncertainty." Mechanical Systems and Signal Processing 166 (2022): 108426.

[4]. Lin, Wan-Yu, Wei J. Chen, Chih-Min Liu, Hai-Gwo Hwu, Steven A. McCarroll, Stephen J. Glatt, and Ming T. Tsuang. "Adaptive combination of Bayes factors as a powerful method for the joint analysis of rare and common variants." Scientific reports 7, no. 1 (2017):
13858.

[5]. Berg, Thomas, and Peter N. Belhumeur. "Tom-vs- Pete Classifiers and Identity-Preserving Alignment for Face Verification." In Bmvc, vol. 2, p. 7. 2012.

[6]. Mao, Tianqi, Lingran Ren, Faqiang Yuan, Chenxiao Li, Li Zhang, Min Zhang, and Yu Chen. "Defect recognition method based on HOG and SVM for drone inspection images of power transmission line." In 2019 international conference on high performance big data and intelligent systems (HPBD&IS), pp. 254-257. IEEE, 2019.

[7]. Ahmad, Munir, Shabib Aftab, and Iftikhar Ali. "Sentiment analysis of tweets using svm." Int. J. Comput. Appl 177, no. 5 (2017): 25-29. [12]. Tyagi, Abhilasha, and Naresh Sharma. "Sentiment analysis using logistic regression and effective word score heuristic." International Journal of Engineering and Technology (UAE) 7, no. 2 (2018): 20-23.

[8]. Venkatesan, Ramachandran, Sundarsingh Shirly, Mariappan Selvarathi, and Theena Jemima Jebaseeli. "Human Emotion Detection Using DeepFace and Artificial Intelligence." Engineering Proceedings 59, no. 1 (2023): 37.