Real Time Mock Interview Using Deep Learning

Keerthi Kumar S[1]

Department of electronics and Communication,

Jawaharlal Nehru New College of Engineering,

Karnataka, 577201

keerthikeerthikumar712@gmail.com

Darshan A R[2]

Department of electronics and Communication,

Jawaharlal Nehru New College of Engineering,

Karnataka, 577201

darshan110904@gmail.com

Ankith G S[3]

Department of electronics and Communication,

Jawaharlal Nehru New College of Engineering,

Karnataka, 577201

ankithgs01@gmail.com

Chandan N[4]

Department of electronics and Communication,

Jawaharlal Nehru New College of Engineering,

Karnataka, 577201

chandangowda1289@gmail.com

Smitha S M[5]

Department of electronics and Communication, Jawaharlal Nehru New College of Engineering,

Karnataka, 577201 smithasm@jnnce.ac.in

Abstract: Real Time Mock Interview Using Deep Learning system is a web application helpful for users to practice for interviews. Nowadays many companies are conducting interviews virtually through online mode. So, this is the need of the day to develop a system where users can practice for these online interviews. This system will help candidates to practice for mock interviews by facing mock interviews. It also provides feedback including facial preference, head nodding, reaction time, speaking rate and volume to let users know their own performance within the mock interview. The system provides speech-to-text conversion for checking grammar in the candidates reply and suggests required corrections. Results are given in a graphical format by using these two or more interviews can be compared to track the progress of the candidates and corrective action will be taken in order to give better performance in the next interviews. This AI-powered tool aims to democratize interview preparation by offering 24/7 access to realistic practice sessions, reducing dependency on human interviewers. The project also incorporates a feedback dashboard that visualizes performance trends and offers improvement tips. This work has potential applications in career counseling, edtech platforms, and recruitment training modules, contributing to more effective and equitable job preparation. Preparing for interviews can be stressful, especially without proper feedback or practice opportunities. This project proposes a Real-Time Mock Interview System using Deep Learning to help users improve their interview performance through simulated, intelligent practice sessions. The system uses deep learning models to analyze spoken responses, facial expressions, and voice tone to mimic the behavior of a real interviewer. It asks domain-specific questions and evaluates the candidate's answers based on accuracy, clarity, and confidence. The system includes modules for speech-to-text conversion, emotional analysis through facial recognition, and tone detection using audio signals. Based on the user's performance, it provides real-time feedback and improvement suggestions. A performance report is generated at the end of each session, highlighting strengths and areas for growth. This tool is especially useful for students, job seekers, and professionals who want to practice interviews anytime without needing a human interviewer.

I. INTRODUCTION

Deep learning is a specialized branch of machine learning that utilizes artificial neural networks with multiple layers to model complex patterns in data. Unlike traditional machine learning methods that rely heavily on handcrafted features, deep learning algorithms automatically extract relevant features from raw data through layered hierarchical structures. These models are particularly effective when dealing with large and unstructured datasets such as text, images, and audio. Their ability to generalize from vast examples has led to breakthroughs in a variety of domains including computer vision, speech recognition, natural language processing, and autonomous systems. The strength of deep learning lies in its ability to mimic the learning process of the human brain, making it suitable for tasks that require perception and reasoning. Applications such as virtual assistants, self-driving cars, recommendation systems, and facial recognition are built using deep learning techniques. Due to its powerful capabilities, deep learning forms the foundation for building intelligent systems that can simulate real-world scenarios and respond adaptively—one such application being mock interview systems for training and evaluation. To develop an intelligent mock interview system, several deep learning methodologies are combined to interpret user input in real time. One of the most important models used in visual data analysis is the Convolutional Neural Network (CNN). CNNs are designed to detect spatial features from images and are commonly used for tasks like facial recognition, emotion detection, and gesture analysis. These capabilities make CNNs suitable for analyzing candidate facial expressions and non-verbal behavior during an interview setting—if the system is extended to support video input. Another critical component of the system is Natural Language Processing (NLP). NLP allows machines to understand and process human language in both written and spoken form. Using techniques such as sentiment analysis, semantic similarity, intent recognition, and part-of-speech tagging, the system can evaluate the quality of candidate responses. It can detect nervousness through tone or word choice, assess grammar and fluency, and provide suggestions for improvement. Real-time inspection and feedback mechanisms are also essential in such a system. These involve continuously monitoring the user's input during the interaction and generating instant evaluations. This feature helps create a dynamic, interactive environment that closely mimics real interview conditions. Whether analyzing spoken answers, typed text, or facial cues, the goal is to provide immediate, targeted feedback that helps users recognize their strengths and areas for development. 1 The main objective of this project is to design and develop a real-time mock interview system using deep learning techniques. The system serves as a training platform that simulates an interview environment and evaluates candidate performance using AI models.In the proposed system, NLP techniques are used to evaluate the content of the candidate's responses. This includes checking the relevance of answers to common interview questions, analyzing sentence structure, and identifying emotional tone using sentiment analysis. If facial or gesture input is later included in the system, CNN-based models could be trained to detect non- verbal indicators such as eye contact, facial tension, or hand movements, which often contribute significantly to interview outcomes. The system architecture is modular, allowing future integration of additional features like real-time speech recognition, visual emotion detection, or even AI- based interviewer bots. Currently, the project emphasizes robust textual analysis and adaptive feedback mechanisms, making it a powerful self-assessment tool for interview preparation. The output is presented to users in a detailed report format, highlighting their strengths and offering constructive tips for improvement. The real-time mock interview system has broad and practical applications in today's competitive job market. One of its primary use cases is in student placement training, where it can be deployed in universities and coaching centers to help students practice and prepare for campus recruitment drives. The system provides a cost-effective and scalable solution, eliminating the need for human interviewers while still delivering personalized feedback. Overall, the project combines the strengths of deep learning, NLP, and real-time analysis to deliver a smart, interactive training tool that adapts to the needs of modern interview preparation. As the system evolves, it holds the potential to become an integral part of AI-powered skill development and recruitment ecosystems.

II. LITERATURE REVIEW

This study proposes an approach to dialog state tracking and action selection supported deep learning methods. First, the interview corpus during this study is collected from 12 participants, and is annotated with dialog states and actions. Next, a long-short term memory and a man-made neural network are employed to predict dialog states and therefore the Deep RL is adopted to find out the relation between dialog states and actions. Finally, the chosen action is employed to get the interview question for interview practice. to gauge the proposed method in action selection, an interview coaching system is made. Experimental results show the effectiveness of the proposed method for dialog state tracking and action selection. In this study, an interview coaching system is proposed and constructed for dialog state tracking and action selection. LSTM and ANN are employed to predict dialog states and the deep RL is used to learn the relation between dialog states and actions. Finally, the anticipated dialog state and action pair are wont to generate an interview question. For performance evaluation on the proposed method in dialog state tracking and action selection, an interview coaching system was constructed, AND AN encouraging result was obtained for dialog state tracking, action selection and interview question generation [1]

It is documented that syntactic constraints, when applied to speech recognition, greatly improve accuracy. However, until recently, constructing an efficient grammar specification to be used by a connected word speech recognizer was performed by hand and has been a tedious, time-consuming task susceptible to error. For this reason, very large grammars haven't appeared. We describe a compiler for constructing optimized syntactic di-graphs from easily written grammar specifications. These are written during a language called grammar specification language (GSL). The compiler features a pre-processing (macro expansion) phase, a parse phase, graph code generation and compilation phases, and three optimization phases. Digraphs also can be linked together by a graph linker to make larger digraphs. Language complexity is analysed during a statistics phase. Heretofore, computer generated digraphs were often crammed with redundancies. Larger graphs were constructed and optimized by hand so as to realize the specified efficiency. We demonstrate that the optimization phase yields graphs with even greater efficiency than previously achieved by hand. We also discuss some preliminary speech recognition results of applying these techniques to intermediate and enormous graphs. With the introduction of those tools it is now possible to supply a speech recognition user with the power to define new task grammars within the field. GSL has been employed by several untutored users with good success. Experience with GSL indicates that it's a viable medium for quickly and accurately defining grammars to be used in connected speech recognition systems [2].
With the development of artificial intelligence (AI), the automatic analysis of video interviews to recognize individual personality traits has become an active area of research and has applications in personality computing, human-computer

interaction, and psychological assessment. Advances in computer vision and pattern recognition based on deep learning (DL) techniques have led to the establishment of convolutional neural network (CNN) models that can successfully recognize human nonverbal cues and attribute their personality traits with the utilization of a camera. during this study, an end-to-end AI interviewing system was developed using asynchronous video interview (AVI) processing and a Tensor Flow AI engine to perform automatic personality recognition (APR) supported the features extracted from the AVIs and therefore the true personality scores from the facial expressions and self-reported questionnaires of 120 real job applicants. The experimental results show that our AI-based interview agent can successfully recognize the "big five" traits of an interviewee at an accuracy between 90.9% and 97.4%. Our experiment also indicates that although the machine learning was conducted without large-scale data, the semi supervised DL approach performed surprisingly well with reference to automatic personality recognition despite the lack of labor-intensive manual annotation and labeling. The AI-based interview agent can supplement or replace existing self-reported personality inventory methods that job applicants may distort to realize socially desirable effects [3].To avoid the complex process of explicit feature extraction in traditional countenance recognition, a face recognition method supported a convolutional neural network (CNN) and a picture edge detection is proposed. Firstly, the countenance image is normalized, and therefore the fringe of each layer of the image is extracted within the convolution process. The extracted edge information is superimposed on each feature image to preserve the sting structure information of the feel image. Then, the dimensionality reduction of the extracted implicit features is processed by the utmost pooling method. Finally, the expression of the test sample image is assessed and recognized by employing a Soft max classifier. To verify the robustness of this method for countenance recognition under a posh background, a simulation experiment is meant by scientifically mixing the Fer-2013 countenance database with the LFW data set. The experimental results show that the proposed algorithm are able to do a mean recognition rate of 88.56% with fewer iterations, and therefore the training speed on the training set is about 1.5 times faster than that on the contrast algorithm [4].Emotion detection and recognition from text could also be a recent essential research area in tongue Processing (NLP) which may reveal some valuable input to a selection of purposes. Nowadays, writings take many sorts of social media posts, micro- blogs, news articles, customer review, etc., and thus the content of those short-texts are often a useful resource for text mining to urge an unhide various aspects, including emotions. The previously presented models mainly adopted word embedding vectors that represent rich semantic/syntactic information and other people models cannot capture the emotional relationship between words. Recently, some emotional word embeddings are proposed but it requires semantic and syntactic information the other way around. to affect this issue, we proposed a completely unique neural specification, called SENN (Semantic-Emotion Neural Network) which can utilize both semantic/syntactic and emotional information by adopting pre-trained word representations. SENN model has mainly two sub-networks, the first sub-network uses bidirectional Long-Short Term Memory (BiLSTM) to capture contextual information and focuses on semantic relationship, the second sub-network uses the convolutional neural network (CNN) to extract emotional features and focuses on the emotional relationship between words from the text. We conducted a comprehensive performance evaluation for the proposed model using standard real-world datasets. We adopted the notion of Ekman's six basic emotions. The experimental results show that the proposed model achieves a significantly superior quality of emotion recognition with various state-of-the-art approaches and further are often improved by other emotional word embedding [5].

III. OBJECTIVES AND METHODOLOGY

1. To develop an AI-powered mock interview system that analyses facial expressions using the inspection method for real-time feedback.

2. To implement a speech analysis module using a spectrogram to assess tone, clarity, and confidence.

3. To utilize NLP to evaluate vocabulary and grammar, providing insights for better communication.

4. To create a user-friendly web-based platform for seamless interaction.

METHODOLOGIES:

1. Data Acquisition:

• Video and Audio Capture: The system begins by capturing real-time video and audio streams from the user's webcam and microphone during the mock interview session.

• Preprocessing: Video frames are extracted at a consistent rate and resized to the input size required by the VGG16 model. Audio signals are processed to enhance clarity and reduce noise before transcription.

2. Facial Expression Recognition:

• Feature Extraction with VGG16: Thepre-trained VGG16convolutional neural network is employed to extract facial features from the video frames.

• Expression Classification: These features are passed through a classifier that categorizes the facial expressions into predefined classes such as happy, sad, angry, surprised, and neutral.

• Real-Time Emotion Feedback: The detected emotions are logged and used to provide instant feedback about the user's non-verbal cues like confidence or ner vousness.

3. Speech-to-Text Conversion:

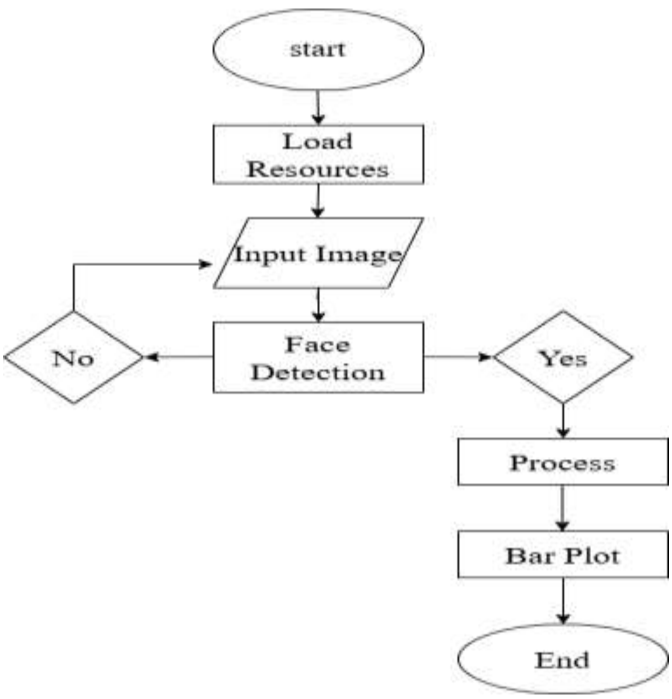
• Audio Transcription: Captured audio is processed through a speech recognition engine (e.g., Google Speech-to-Text API or OpenAI Whisper) to convert spoken responses into text.

• Text Cleaning: The raw transcript undergoes cleaning to remove fillers, hesita tions, and background noise artifacts for clearer analysis.

4. Natural Language Processing and Deep Learning Analy sis:

• Sentiment and Tone Analysis: Deep learning NLP models analyze the tran scribed text to detect sentiment, tone, and relevance of the responses.
• Fluency and Clarity Assessment: The system identifies filler words, pauses, and repetitive phrases to evaluate the fluency and clarity of speech.
5. Grammar Specification Language (GSL) Processing:
• Grammar Parsing: Using GSL, the system parses the transcribed text to check grammatical structure, syntax, and semantic correctness.
• Error Detection and Suggestions: Grammar mistakes or incoherent sentences are flagged, and suggestions for improvement are generated based on GSL rules.
6. Feedback Generation and Reporting:
• Multi-Modal Data Aggregation: Data from facial expression recognition, speech analysis, and grammar checking modules are combined.
• Real-Time Feedback: The system delivers immediate insights on emotional ex pression, verbal clarity, and grammatical accuracy during the interview.
• Comprehensive Post-Interview Report: After completion, a detailed perfor mance report is generated highlighting strengths, areas for improvement, and per sonalized recommendations.
7. User Review and Iteration:
• Session History Storage: Interview data and feedback are stored securely for user review.
• Progress Tracking: Users can analyze past sessions to monitor their improvement over time and focus on weak areas in subsequent interviews.

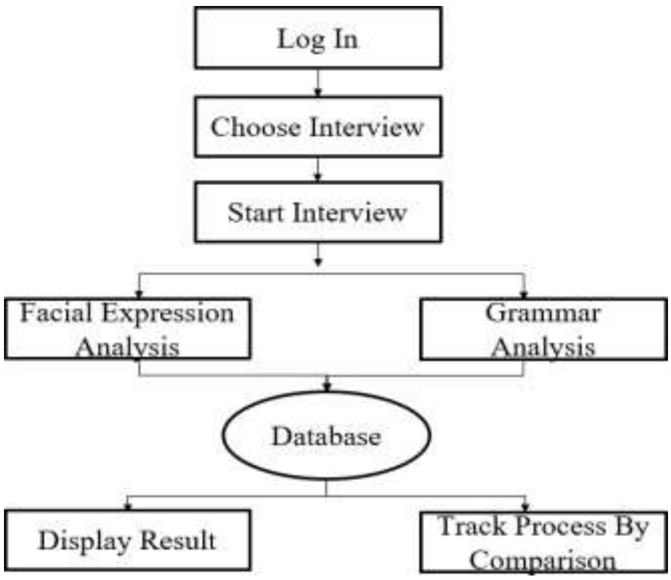
IV. BLOCK DIAGRAM



The given flowchart illustrates the **process of face detection and analysis** using an input image. It begins with the *Start* node, followed by the step to *Load Resources*, where necessary libraries, models, and data are initialized. Next, the user is prompted to *Input Image*, which serves as the main data for processing. The image is then passed to the *Face Detection* block, where the system attempts to locate faces within the image. A decision node checks whether any face is detected — if *No*, the process loops back, allowing the user to input a new image.

If a face is successfully detected (*Yes*), the process moves to the *Processing* stage, where detected face data is analysed or features are extracted. The results are then visualized in the form of a *Bar Plot*, providing a graphical representation of the analysis. Finally, the flow reaches the *End* node, signifying the completion of the task. Overall, the diagram effectively outlines a sequential workflow for automated face detection and result visualization in an image processing system.

IMPLEMENTATION



Log In – First the user has to sign up in the system. After that user will get a username and password. Using the credentials, user can login in the system.

Choose Interview – After successfully logging in, user will have to choose interview of his choice based upon the interviews present in the system. User can choose interviews to track his performance and progress which will be saved in the database while comparing the results.

Start Interview – Once the user chooses the interview of his preference, after that the interview will start, while the interview proceeds users progress and performance will be saved in the database of the system. User's progress and performance saved in the database will be further used to compare different parameters at the end while displaying the result.
Facial Expression Analysis – When the user is giving the interview with the help of the webcam user's facial expression will be analysed. When the user chooses to start interview the OpenCV runs in background and starts to record the video of

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

the interview. This facial expression will be analysed based on the dataset imported in the system and this data will be stored in the database. This proposed system will help users who are nervous or anxious while giving the interviews. The interviewers keep a keen attention on the expressions of the candidates they are interviewing because many candidates are rejected because they are not confident while giving the interviews. Therefore, facial expression analysis is required or essential so that the interviewee can improve his performance in front of the panel interviewing the user by taking mock interviews which will help the interviewee while giving actual company recruitment interviews.

Grammar Analysis – When the user will be giving the interview, whatever user speaks will be converted into text. This conversion is essential to keep an eye on user's grammatical mistakes. This analysis will help the user improve his vocabulary for actual interviews. Along with confidence of the interviewee, interviewers also keep an eye on the vocabulary of the candidate. Therefore, this analysis will be saved in the database which will be used while displaying result to make user aware of his weaknesses in vocabulary. Hence the user can work on the factors which are affecting his performance in actual interviews.

Database – Database is an important factor of this proposed system because user's progress and performance are saved in the database. This saved data of user will help to improve performance every time he logs in the system to practice for actual interviews. Also, user's data like user's preferences, personal information will be also saved in the database.

Display Result – After user completes the interviews based on performance and progress result will be displayed. This result will help user to improve on the factors in which the result is bad. Result will be displayed based on data analysis visualization. User's facial expression analysis and grammar analysis will be displayed in the result. As the facial expressions and grammatical errors will be converted to some dataset, the data will be then analysed and predicted. Both factors will be considered and result will be displayed in graphical format using data visualization tools. The result of two or more interviews can be compared to track the progress of the candidates. The result of users performance in particular interview will also depend on his facial preference, head nodding, reaction time and speaking rate.

V. RESULTS

FIG-1: Admin login

The Admin Login page of the Mock Interview Portal provides a secure interface where administrators can enter their username and password to access the admin panel. It features a clean and minimal layout with navigation options such as Home, Register, Student Login, and Admin Login at the top for easy access.

FIG-2: Admin dashboard

The Admin Dashboard provides a centralized interface for managing content within the Mock Interview Portal. It allows administrators to add materials by entering details such as the title, subject, type of content, and uploading corresponding files. The page also includes navigation options like Home, Add Interview Question, and Logout, enabling smooth access to administrative functions.

FIG-3: Question

The "Add New Question" page allows administrators to create and add new interview questions to the portal by specifying the subject, entering the question, and providing four answer options. It also includes a dropdown to select the correct answer and a button to submit the question, all within a clean and organized interface for easy content management.

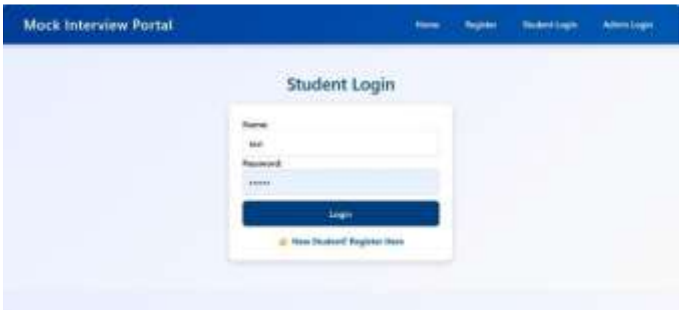


FIG-4: Student login

The Student Login page of the Mock Interview Portal provides a simple interface where users can enter their name and password to access the system. It also includes navigation links such as Home, Register, Student Login, and Admin Login, along with an option for new students to register.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

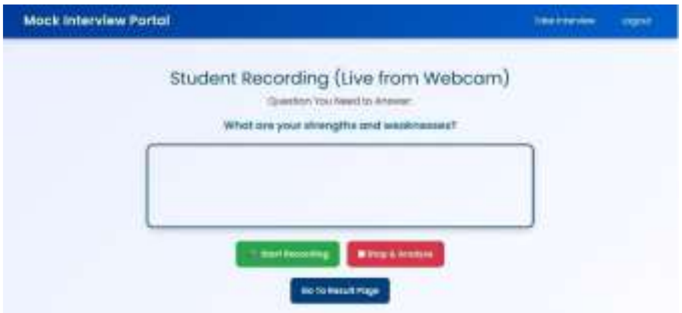


FIG-5: User Interface

The Mock Interview Portal provides students with a login interface to access personalized interview practice. It features a webcam-based recording module where users respond to questions like “What are your strengths and weaknesses?” for self-analysis.

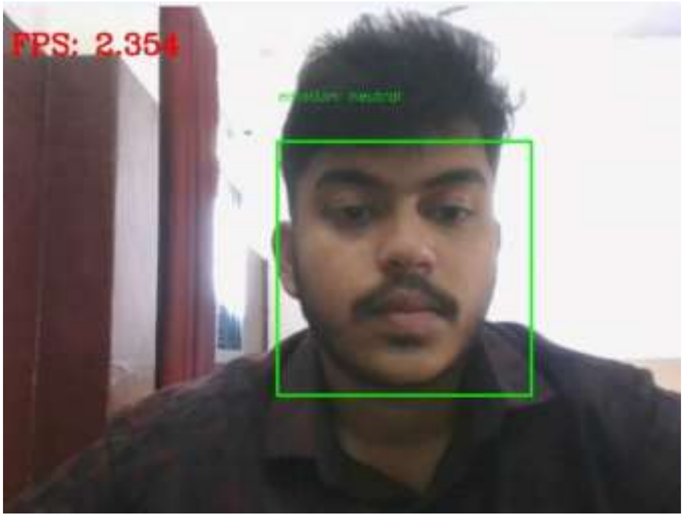


FIG-6: Facial Emotion detection

The system successfully detected the user’s face and analyzed the facial expression in real time. Based on the processed video frame, the model identified the user’s emotion as Neutral. This indicates that the user maintained a calm and steady facial expression during the interview session. The system operated with an FPS of approximately 2.35, showing that real-time facial expression recognition was achieved. The output demon strates that the facial expression recognition module is functioning correctly, accurately classifying emotions and providing consistent feedback.

FIG-7: Speech-to-Text Engine

After the user responds to the interview question, the system successfully captures the spoken audio and processes it through the Speech-to-Text engine. The backend accurately converts the audio input into text, demonstrating the functionality of the transcription module. This confirms that the audio recording, processing pipeline, and speech recognition model are working correctly. The module effectively transforms the spoken response into machine-readable text, enabling further analysis such as grammar checking, total error score and total interview score.

FIG-8: Result

user’s facial expression, speech input, and grammatical accuracy. The final output displays multiple performance parameters, confirming that all modules—visual analysis, emotion detection, speech confidence scoring, grammar checking, and error evalua tion—are functioning correctly. The system identified the candidate’s emotion as Positive, with a Visual Confidence Score of 0.3548 and a high Speech Confidence Score of 0.9779, indicating strong verbal clarity. Grammar analysis detected minor text errors, such as the correction for “problem solving,” resulting in a Grammatical Score of 80 and a Total Error Score of 2. Based on the overall evaluation, the system computed a Total Interview Score of 81.33, which meets the passing criteria. As a result, the system displays a confirmation message: “Congrats! You Have Been Selected for Next Round.” This output verifies that the system is able to assess multiple communication pa rameters and provide an automated final decision regarding the candidate’s interview performance

FIG-9:Student dashboard

This page serves as the student dashboard in the Mock Interview Portal, where the logged-in student is greeted along with the selected subject, such as Java. It provides access to the interview paper, along with a clearly visible option to view notes, while also including navigation links for taking an interview or logging out.

VI. CONCLUSION

In this paper, we presented an approach that enhances a virtual agent by the ability to interpret and respond to social cues of users participating in a simulated job interview. In order to achieve seamless credible interaction, our system automatically recognizes the user’s social cues in real time. Based on these, the virtual recruiter reacts and adapts to the user’s behaviour. Furthermore, the interaction with the virtual agent can be recorded and presented to the user to enhance the learning effect, for example, by identifying crit ical incidents during the simulated interview. The scenario manager was used to model the virtual recruiter's interactive behaviour allowing the character to react to various social users recognized by the social cue recognition module. More precisely, we mod elled mirroring and turn taking behaviour. Despite several reported problems, such as the realism of the character's appearance, all participants' reactions were mainly positive saying they would use such a system to train for real job interviews.

VI.REFERENCE

[1] Seamus Ryan, Camille Nadal, And Gavin Doherty "Integrating Fairness in the Software Design Process": An Interview Study With HCI and ML Experts Received 17 February 2023, accepted 7 March 2023, date of publication 22 March 2023, date of current version 28 March 2023.

[2] Pei-hsuan Lin And Shih-yeh Chen "Design and Evaluation of a Deep Learning Recommendation Based Augmented Reality System for Teaching Programming and Computational Thinking" Received January 14, 2020, accepted February 24, 2020, date of publication March 2, 2020, date of current version March 16, 2020.

[3] Liu Qianqian, Wan Qian, Xie Boya, Lin Churan, Xie Zhenyou,pan Shu, And Peng Guoqing "Research on Behavior Analysis of Real-Time Online Teaching for College Students Based on Head Gesture Recognition" Received 20 June 2022, accepted 6 July 2022, date of publication 22 July 2022, date of current version 8 August 2022.

[4] Mahsa Raeiszadeh, Amin Ebrahimzadeh, Roch H. Glitho (Life Senior Member, IEEE), Johan eker, and Raquel a. f. mini "Asynchronous Real-Time Federated Learning for Anomaly Detection in Microservice Cloud Applications". [5] Hung-yue Suen, Kuo-en Hung, And Chien-liang Lin "TensorFlow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews" Received February 1, 2019, accepted February 20, 2019, date of publication March 5, 2019, date of current version May 22,2019.

[6] D. Baviskar, S. Ahirrao, V. Potdar, and K. Kotecha, "Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions," Received March 25, 2021, accepted April 7, 2021, date of publication April 13, 2021, date of current version May 24, 2021.

[7] M. Palviainen, T. Harviainen, M. Bordallo López, and J. Mäntyjärvi, "Boosting Business With Machine Learning-Based Automated Visual Data Processing: Results of Finnish Company Interviews," Received March 23, 2020, accepted April 22, 2020, date of publication May 11, 2020, date of current version June 8, 2020.

[8] S. Kusal, S. Patil, J. Choudrie, K. Kotecha, S. Mishra, and A. Abraham, "AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions," Received 6 August 2022, accepted 18 August 2022, date of publication 23 August 2022, date of current version 8 September 2022.