K. Rajkumar, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

Development Of A Smart Interview Assistant For Real-Time Candidate Evaluation

[1]k Rajkumar,[2]Donthuri Pranitha

[1]Assistant Professor, [2]M.Tech Student, Department of CSE,

Megha Institute Of Engineering And Technology For Womens, Edulabad (Village),

Ghatkesar (Mandal), Medchal District, Telangana

Abstract—Automated intelligent software agents may mimic human conversational behaviours, allowing for more natural and interesting interactions with people, thanks to the fast development of Conversational AI. By allowing for the replacement of human interviewers with intelligent autonomous software agents, these developments pave the way for the automation of the candidate interview process. Agents outfitted with Conversational AI can mimic human interviewers in every way: asking questions, comprehending and evaluating responses, and starting up dynamic discussions. More effective and equitable recruiting processes are the result of this automation, which streamlines the interview process as a whole and guarantees consistent and impartial judgement. The purpose of this research article is to provide a thorough analysis of the design and implementation of an AI-driven interview system for real-time applicant assessments. As part of the system, various AI agents carry out various tasks, such as choosing from a set of predefined questions, assessing candidates' responses, analysing speech for sentiment and emotion, and finally, combining these analyses to generate separate scores for answers and emotions in the performance evaluation.

Keywords—AI-Driven Interviews, Automated Candidate Assessment, Multimodal Emotion Detection, Natural Language Processing (NLP), Generative Pre-trained Transformers (GPT)

I. INTRODUCTION

Since Artificial Intelligence (AI) was widely used, talent acquisition has become much better [1]. Assessing candidates quickly and accurately throughout the hiring process is more crucial than ever in today's competitive employment market. When it comes to evaluating candidates, the conventional interview process has its limitations due to human subjectivity and time limits. However, automating the interview process using AI-driven solutions allows for a more thorough examination. To provide a more impartial, thorough, and scalable alternative for candidate evaluation, this study introduces a novel framework that makes use of cutting-edge technology.

An army of AI software agents, each trained to handle a certain facet of the interview, work together in this system. A question management agent is assigned to the interview process and its main function is to choose predetermined questions that are relevant and Suitable based on the circumstances. To provide a personalised and adaptable interview experience, the agent makes use of GPT-4 and specialised Natural Language Processing (NLP) models to choose questions according on the candidate's past responses. A sentiment analysis agent records and examines the emotional components of the candidate's communication after a response management agent records, processes, and stores the candidate's answer.

Using a combination of textual analysis and audio input, the system is capable of multimodal emotion detection. Using this method, we can pick up on nonverbal signs of emotion, such tone, pitch, and speech tempo, that could otherwise go unnoticed when analysing text alone. Assessments of "soft skills," like EQ and IC, benefit greatly from this kind of investigation. These agents coordinate their efforts to provide a complete picture of the candidate by

recording not just the substance but also the tone of their answer. A holistic assessment of the applicant is produced by a comprehensive evaluation agent, which integrates the data and insights produced by other agents. Both quantitative and qualitative measures, including ratings and sentiment analysis, are part of an all-encompassing review. Alignment with job-specific competences, language correctness, and coherence are all evaluated by this AI using NLP approaches.

A comprehensive knowledge of each applicant is used to make the final judgement, which helps to minimise biases and maximise assessment objectivity. In order to give a thorough data-driven evaluation of the candidates' performance, the interview system can integrate with natural language processing (NLP) [2], convolutional neural networks (CNN) [3], recurrent neural networks (RNN) [4], and generative pre-trained transformers (GPT) [5]. This allows it to ask questions, score answers, and analyse speech for sentiment and emotion. The use of an AI-driven interview technology might drastically alter the hiring procedure. Human resource managers are free to concentrate on strategic decision-making rather than administrative duties thanks to the system's automation of key interview processes, which improves overall efficiency and decreases expenses.

II. RELATED WORK

In order to measure respondent behaviour, Priya et al. [6] suggested an automated method that would evaluate audio and visual signals. We used Support Vector Machine (SVM) to categorise the facial expressions and aural inputs. Using natural language processing (NLP) and deep learning (DL), Senarathne et al. [7] suggested an intelligent interviewing application that automates the interview process. In order to determine a candidate's worth, the system verifies and predicts their ratings based on their responses. An automated computational framework was suggested for the purpose of recognising verbal and non-verbal behaviours that occur during employment interviews. Based on prosodic characteristics, language, and facial expressions, the system forecasts interview ratings [8]. To solve the problem of unfairness in AI interview systems, Kim et al. [9] suggested a technique that makes use of multimodal data. To strike a compromise between accuracy and fairness, the method included a regularisation term. Reducing the Wasserstein gap between sensitive groups is another way the technique minimises prejudice. In order to categorise interview responses, Romadon et al. [10] contrasts TF-IDF with word embeddings that use Artificial Neural Networks (ANN). When it comes to evaluating job interviews, TF-IDF is the way to go since it minimises dimensionality, bias, and human mistakes while outperforming word embeddings. Using natural language processing (NLP) capabilities, an interview chatbot was developed that could automatically produce questions and replies according to the talents shown on the résumé [11].

In their study, Pickard et al. looked at how various interview formats affected the amount of personal information that was divulged during in-person interviews. Research found that participants were more forthcoming with personal information while using the faceless Audio-only Computer Assisted Self Interview (ACASI) mode as opposed to the human-like Embodied Conversational Agent (ECA) and the human interviewer mode with visible faces [12]. Machine learning techniques for evaluating candidates' cultural fit were investigated by Yusuf et al. [13] via the analysis of interview transcripts. In the research, SVM, Naive Bayes, and K-Nearest Neighbours (KNN) were evaluated as classifiers. As compared to the other algorithms, SVM proved to be the most successful solution for this particular job on several occasions.

To determine symptoms and their effects on quality of life, Fang et al. [14] investigated several natural language processing techniques for categorising qualitative, unstructured text data derived from patient interviews. Researchers found that the BERT model, which stands for Bidirectional Encoder Representations from Transformers, performed the best when it came to categorising the effects and symptoms that patients

described. Researchers Jiang et al. [15] looked into how televideo interview data (including voice, facial, and linguistic expressions as well as cardiovascular modulation) may be used to distinguish between people with mental problems. Automated mental health evaluations that are scalable, accessible remotely, and economical may be possible using a multimodal strategy, according to the results. The significance of scientific methods in the classroom and evaluation is highlighted in the research [16]. It goes on to say that ML methods might be very useful for evaluating science education programmes.

III. PROPOSED METHODOLOGY

The suggested method for creating an AI-powered interview system incorporates state-of-the-art tools including GPT, RNN, CNN, and natural language processing. Through real-time monitoring of the applicant's emotional state reactions, the system intends to automate and improve candidate assessment by dynamically altering the questioning method. A. Gathering Information The SAVEE[17], TESS[18], and CREMA-D [19] datasets were used to train the audio emotion identification model. As part of the Surrey Audio-Visual Expressed Emotion (SAVEE) multimodal dataset, four male actors were recorded acting out fifteen phrases representing seven distinct emotional states. The 200 phrases recorded by two women for the Toronto Emotional Speech Set (TESS) aim to portray seven distinct emotions.

Audiovisual recordings of actors exhibiting various emotions are included in the Crowdsourced Emotional Multimodal Actors Dataset (CREMA-D). The International Survey on Emotion Antecedents and Reactions (ISEAR) dataset was used to train the model for emotion recognition from text [20]. ISEAR aims to investigate emotional reactions in various cultural and environmental settings. This research includes data from more than 3,000 individuals from 37 different nations who described their emotional experiences in reaction to different stimuli. Multiple-Choice Test for Medical Purposes Initiating conversation between the interviewer and the applicant was done using MedMCQA [21] for the question and answer dataset. With over 194,000 questions covering a wide range of medical issues, along with detailed explanations for each answer, MedMCQA is an extensive and comprehensive assessment tool.

Questions in the AIIMS and NEET PG admission exams are based on actual medical practice, and this data is intended to answer such questions. Part B: Interview System Design Figure 1 is an architectural schematic of an AI-driven interview system that shows how the interviewer, applicant, and backend components interact with one another. The interviewee interacts with the system in real-time, responding to questions given by the interviewer. After the interviewer retrieves the questions from the database, the system begins to transmit the candidate's audio replies so that they may be analysed. The system's essential parts are: The Conversational Server acts as the nerve centre for all back-end AI model interactions and data flow involving the interviewee, the interview, and the applicant. Transcribing the candidate's voice replies into text is the job of speech-to-text (STT) technology.

Text-to-Speech (TTS): This technology translates written content, such questions and comments, into spoken language so that you may converse with the applicant orally. Maintains control of the interview's environment and status via session management. Handles the streaming of audio data in real-time for emotion identification APIs, or application programming interfaces, allow for the integration of various software agents. The interview process is overseen by a group of software agents known as QMA (III-C), RMA (III-D), MESAA (III-E), and CESA (III-F).

The central analytical component that integrates different AI capabilities to handle and analyse data is the AI engine. To find the most important points made by the candidate, use natural language processing models to examine the transcript of their speech. AI Models: A few examples of AI models include those that analyse text and audio. • GPT: Help choose from a

list of sample interview questions, revise the language as needed, and check the candidates' answers for clarity and relevance to the situation.

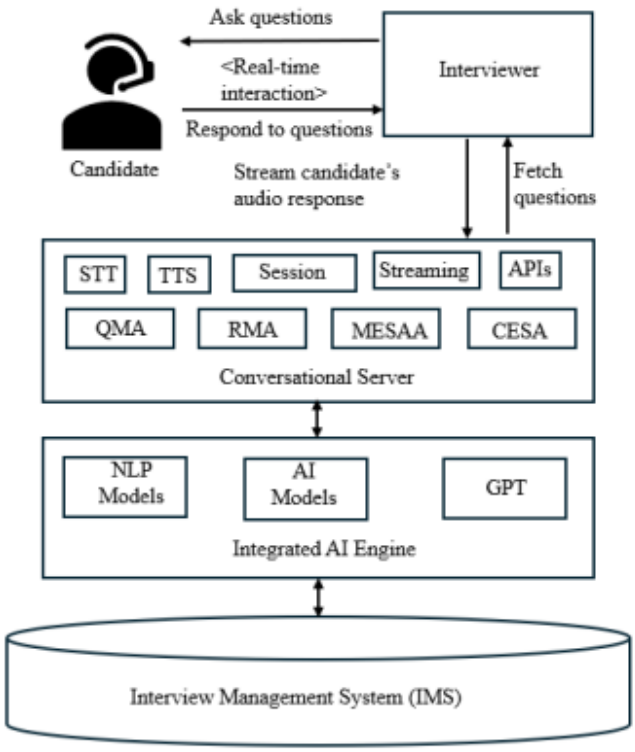

Fig. 1. Architecture of the Interview system IMS:

All information pertaining to the interview process is stored and managed by this backend system. For effective index-based search, the document database keeps track of queries and answers. • Relational database: Stores analysed and aggregated data to provide detailed reports on candidate sentiment, performance, and emotions. • Vector database: Allows for retrieval with context by use of interview material that has been semantically embedded.

Part C: The Question Management Agent Based on the applicant's history, the preset criteria, and their real-time responses, this agent is in charge of dynamically sending questions to the candidate. To make sure queries are appropriate to the context, the bot employs GPT and bespoke NLP models to dynamically change them. Additionally, this agent controls the interview's progression by determining, in real-time, using analysis of responses and established criteria, when to ask the next question.

The MedMCQ questions and their corresponding answersElasticsearch stores datasets in a document-based database that also serves as a distributed search and analytics engine. It allows for effective and fast data retrieval and is scalable. Figure 2 displays the fields that are pertinent to the document database. The interview system may make advantage of this database's full-text search capabilities to choose and retrieve questions according to keywords, relevancy, or contextual similarities. To keep the interviewer-candidate interaction flowing smoothly, it's important to be prepared to manage the following typical events and actions (see Table). I.

TABLE 1: Candidate's Response Scenarios And Implementation Strategies

#	Scenario	Action	Implementation strategies
1	Providing correct answers	Acknowledgment, next question, more in-depth questions	Custom NLP: Managed with contextually relevant keyword
2	Providing partially correct answers	Encouragement, guidance to complete the answers	GPT: Adapt system response based on candidate's response or choose additional questions
3	Incorrect answers	Correction, clarification, next question	GPT: Generate detailed and contextually relevant responses
4	Completely different answers	Rephrase the question, reiterate the question	Custom NLP: Measure semantic similarity between the response and the question
5	Correct answers but for a different question	Clarification, acknowledge and redirect to original question	Custom NLP: Detect mismatches using semantic similarity analysis, topic modeling, and keyword matching
6	Asks for clarification or repeats the question	Rephrase the question, offer additional context to the question	Custom NLP: Refine question wording with additional context
7	Unable to answer or cannot provide a response	Encouragement to respond, move on to another question or topic	Custom NLP: Supportive feedback and transitioning to next topic or question
8	Provides vague or general answers	Clarification, reiterate the question	GPT: Interpret vague response, provide contextually appropriate feedback and prompt for more details

For different cases, we employed a mix of GPT-4 and our own bespoke NLP models. Easy jobs like keyword extraction, rule-based processing, and predetermined answer creation were handled by custom NLP models. Using GPT's language creation skills is unnecessary for

these jobs. Concerns like data protection, scalability, speed, and the amount of customisation all play a role in deciding between GPT and bespoke NLP models.

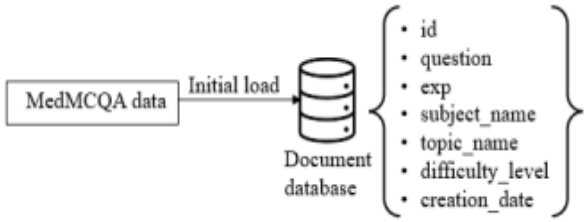

Fig. 2. Data Ingestion and Storage Schema of MedMCQA Dataset

Part D. RMA, or Response Management Agent Capturing, processing, and storing the candidate's replies are the functions of the response management agent. To make sure the data is ready for analysis, this agent checks that the system correctly records each answer and associates it with the relevant question. The information is stored in an RDBMS like PostgreSQL, and the pertinent fields are shown in Figure 3.

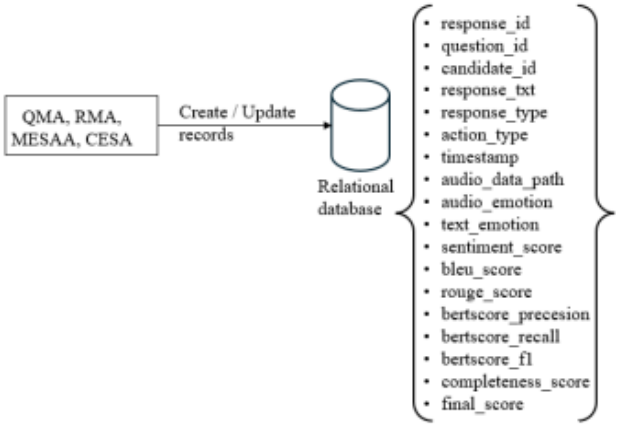

Fig. 3. Data Management and Record Maintenance in the Interview System

An in-depth study of a candidate's emotional and sentimental expressions throughout the interview process may be provided by the Multimodal Emotion and Sentiment Analyser Agent (MESAA). Tone, pitch, loudness, and speech rate are just few of the vocal characteristics that MESSA analyses using CNN and other deep learning methods. The candidate's emotional state and degree of involvement may be better understood with the use of this analysis, which helps detect emotions such as happiness, sadness, anger, and worry. In order to identify certain feelings and thoughts expressed in text, MESAA employs a Bidirectional Long Short-Term Memory (BiLSTM) network. Part F. C. ESA While conducting interviews, CESA collects and analyses data from a variety of sources.

A thorough evaluation and score report are the principal outputs of CESA, which aims to provide an accurate, fair, and all-encompassing review of each applicant. CESA assesses the quality of answers using a number of measures. There are a handful of important measures that may be used to assess the relevancy and quality of a candidate's replies. The following metrics provide light on various parts of the reaction in their own special way. • Assessment That is Bilingual The BLEU understudy assesses how similar a candidate's response is to a reference answer in terms of words and n-grams. When comparing a candidate's response to a reference, a higher BLEU score implies more lexical similarity, which means the candidate used comparable phrases and word selections.

The ROUGE metrics place an emphasis on the candidate's recollection and the inclusion of important substance in their answer. Two ROUGE variations, ROUGE-1 and ROUGE-L, evaluate the response's coherence and structure by measuring the overlap of unigrams and the longest common subsequence, respectively. This measure is great for summarising, and in the interview system, ROUGE scores show whether applicants covered important topics.

BERTScore: This contextual embedding method compares potential candidates' responses to reference replies based on how close they are in meaning. In contrast to BLEU and ROUGE, BERTScore evaluates the overall meaning of the answers rather than just whether they match word for word. Using one text as the "ground truth" and the other as the "candidate,"

the Completeness Score assesses the similarity of texts and runs from 0 to 1. Alg. 1 is used to determine the completeness score. By combining them, we may assess applicant replies in a thorough and comprehensive manner.

Algorithm 1 Calculating Completeness Score

```
Split the candidate answer into list of individual sentences.
Split the actual answer into list of individual sentences.
Initialize coverage to 0.
for each sentence in candidate answer list do
    Initialize sentence coverage as an empty list.
    for each answer in actual answer list do
        Calculate SIMILARITY for answer & sentence.
        Append SIMILARITY to sentence coverage.
    end for
    Add maximum value in sentence coverage to coverage value.
end for
Divide coverage by length of candidate sentence list & save as completeness score.
```

IV. EXPERIMENTAL RESULTS

On Google Colab, the BiLSTM and CNN models were trained using NVIDIA T4 Tensor Core GPUs. This research recorded a full suite of training and assessment metrics. At the conclusion of 14 epochs, the model attained an overall accuracy of 70% on the training dataset for emotion and sentiment analysis in text. This was accomplished using BiLSTM. On the other hand, a 60% accuracy rate on the validation set suggests a passable capacity to identify emotions in written content.

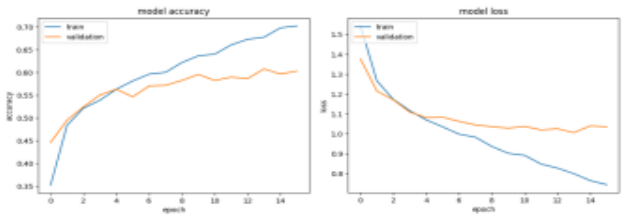

Fig. 4. Text Emotion Recognition Model Training and Validation Accuracy

F1-Score, Precision, and Recall: With the best recall (0.76) and F1-score (0.70) for Class 3 (Joy), the model clearly does a good job of detecting occurrences of this class. It is challenging to accurately identify all occurrences of Class 4 (neutral) due to its poorer recall (0.47) and F1-score (0.52). Class 0, 1, 2, 5, and 6 stand for astonishment, sorrow, wrath, fear, and disgust, in that order.

	precision	recall	f1-score	support
0	0.57	0.66	0.61	228
1	0.58	0.50	0.54	208
2	0.65	0.59	0.62	235
3	0.65	0.76	0.70	224
4	0.59	0.47	0.52	190
5	0.62	0.54	0.58	208
6	0.59	0.71	0.64	202
accuracy			0.61	1495
macro avg	0.61	0.60	0.60	1495
weighted avg	0.61	0.61	0.60	1495

Fig. 5. Precision, Recall and F1-Score for Emotion Detection from Text

Convolutional Neural Networks for Detecting Sound Emotions Precision: As a whole, the model was 95% accurate on the training dataset and 67% accurate on the validation dataset.

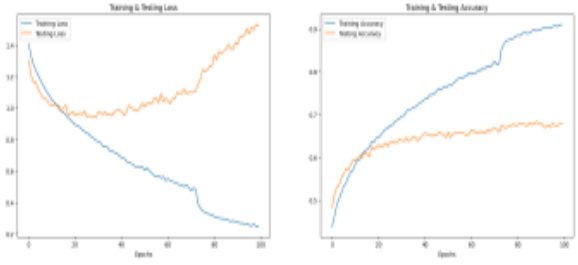

Fig. 6. Audio Emotion Recognition Model Training and Validation Accuracy

With an F1-score of 0.86, the "surprise" class offers the best mix of recall and accuracy. While the "disgust" emotion class has a reasonable recall of 0.65, its poor accuracy of 0.57 and F1-score of 0.61 indicate that the model has difficulty accurately predicting disgust.

	precision	recall	f1-score	support
angry	0.79	0.75	0.77	1099
calm	0.74	0.82	0.78	130
disgust	0.57	0.65	0.61	1098
fear	0.68	0.61	0.64	1082
happy	0.64	0.64	0.64	1122
neutral	0.67	0.62	0.64	1001
sad	0.65	0.68	0.67	1138
surprise	0.83	0.90	0.86	495
accuracy			0.68	7165
macro avg	0.70	0.71	0.70	7165
weighted avg	0.68	0.68	0.68	7165

Fig. 7. Precision, Recall and F1-Score for Emotion Detection from Audio

Part C: Metrics for evaluating quality Emotional intelligence and topic knowledge are two distinct criteria that are evaluated independently. The calculation of emotion scores involves classifying feelings as either anticipated or unexpected, according to the criteria laid forth in Table II. Every feeling is given a score between zero and one. Positive and anticipated emotions get higher ratings, whereas negative and unwelcome feelings get lower marks. The maximum possible score is for happiness, while the lowest possible value is for rage. Each question's emotion score is determined by capturing emotions from both voice and text and then using Equations 1 and 2.

$$\text{Emotion Score} = \frac{\text{Voice Emotion Score} + \text{Text Emotion Score}}{2} \quad (1)$$

$$\text{Average Emotion Score} = \frac{\sum_{i=1}^{n} \text{Emotion Score}_i}{n} \quad (2)$$

"n" is the number of enquiries. • The i-th question's emotion score is denoted as Emotion Scorei. The average emotion score is determined by adding together all of the scores and then dividing by the total number of questions using the given equation. A final knowledge score is computed for response assessment by adding together all of the individual scores in a weighted manner, as shown in Eq. 3.

$$\begin{aligned}\text{Final Score} = {} & 0.1 \times \text{BLEU} \\ & + 0.2 \times \text{Completeness} \\ & + 0.3 \times \text{ROUGE-L F1} \\ & + 0.4 \times \text{BERTScore}\end{aligned} \quad (3)$$

In this experiment, we generated five potential replies by randomly selecting one from the dataset that had four lines. prospective respondent's response 1 supplied a response that was a paraphrase of the original, comprising only the first sentence. The second candidate's response consisted of two lines of paraphrasing.

TABLE 2: Classification And Scoring Of Expected Vs. Unexpected Emotions

Interview response	Emotion	Score
Expected emotion	Fear	0.5
	Surprise	0.6
	Neutral	0.8
	Happy	1
Unexpected emotion	Anger	0.1
	Disgust	0.2
	Sad	0.4

for the first two choices, and similarly for the next two. The response given by the fifth contender was a carbon duplicate of the first. A final score was generated by adding the individual scores of each candidate's answers. As seen in Figure 8, the findings were displayed on a graph. As seen in the graph, an answer is deemed accurate if the knowledge score is 0.45 or above; answers with values ranging from 0.2 to 0.45 are deemed incomplete and need more explanation.

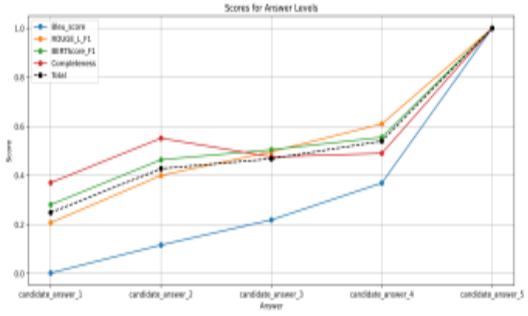

Fig. 8. Metrics of Scores at Various Answer Levels

V. CONCLUSION

An excellent applicant assessment process is guaranteed by the automated interview system that makes use of artificial intelligence and machine learning technologies. These technologies are able to identify emotions from audio and text in several ways, analyse vocal replies, and reply appropriately to talks. By combining key components including QMA, RHA, MESAA, and CESA, the system offers a comprehensive assessment of applicants. Eventually,

we'll make it so the machine can read candidates' body language, facial expressions, and other nonverbal clues in real time. We will also be testing this in the real world with a variety of applicant pools and recruiting procedures in the near future. Finally, the automated interview system that is driven by AI is a huge step forward in the assessment and hiring of candidates.

REFERENCES

1. J. Attupuram, P. Sequeira, and A. H. Sequeira, "Talent Acquisition Process in a Multinational Company: A Case Study," Management of Innovation e-Journal, CMBO, Dec. 24, 2015.
2. T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing [Review Article]," IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55-75, Aug. 2018.
3. O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional Neural Networks for Speech Recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 10, pp. 1533-1545, Oct. 2014.
4. G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," IEEE Access, vol. 7, pp. 51522- 51532, 2019.
5. T. Wu et al., "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development," IEEE/CAA Journal of Automatica Sinica, vol. 10, no. 5, pp. 1122-1136, May 2023.
6. K. Priya, S. M. Mansoor Roomi, P. Shanmugavadivu, M. G. Sethuraman, and P. Kalaivani, "An Automated System for the Assessment of Interview Performance through Audio & Emotion Cues," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 1049-1054.
7. P. Senarathne, M. Silva, A. Methmini, D. Kavinda, and S. Thelijjagoda, "Automate Traditional Interviewing Process Using Natural Language Processing and Machine Learning," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-6.
8. I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated Analysis and Prediction of Job Interview Performance," IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 191-204, April-June 2018.
9. C. Kim, J. Choi, J. Yoon, D. Yoo, and W. Lee, "Fairness-Aware Multimodal Learning in Automatic Video Interview Assessment," in IEEE Access.
10. A. W. Romadon, K. M. Lhaksmana, I. Kurniawan, and D. Richasdy, "Analyzing TF-IDF and Word Embedding for Implementing Automation in Job Interview Grading," 2020 8th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, 2020, pp. 1-4.
11. R. Pandey, D. Chaudhari, S. Bhawani, O. Pawar, and S. Barve, "Interview Bot with Automatic Question Generation and Answer Evaluation," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1279- 1286.
12. M. D. Pickard and C. A. Roster, "Using computer automated systems to conduct personal interviews: Does the mere presence of a human face inhibit disclosure?," Computers in Human Behavior, vol. 105, 2020, Art. no. 106197.

13. M. Yusuf and K. M. Lhaksmana, "An Automated Interview Grading System in Talent Recruitment using SVM," 2020 3rd International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2020, pp. 34-38.
14. C. Fang, N. Markuzon, N. Patel, and J.-D. Rueda, "Natural Language Processing for Automated Classification of Qualitative Data From Interviews of Patients With Cancer," Value in Health, vol. 25, no. 12, pp. 1995-2002, 2022.
15. Z. Jiang et al., "Multimodal Mental Health Digital Biomarker Analysis From Remote Interviews Using Facial, Vocal, Linguistic, and Cardiovascular Patterns," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1680-1691, Mar. 2024.
16. E. P. Beggrow, M. Ha, and R. H. Nehm, "Assessing Scientific Practices Using Machine-Learning Methods: How Closely Do They Match Clinical Interview Performance?," J. Sci. Educ. Technol., vol. 23, no. 2, pp. 160-182, 2014.
17. S. Haq, P. J. Jackson, and J. Edge, "Speaker-dependent audio-visual emotion recognition," in Proc. AVSP, vol. 2009, pp. 53-58, 2009.
18. N. Neubauer and K. Dupuis, "Toronto Emotional Speech Set (TESS)," Aging and Communication Lab, Univ. of Toronto, 2011, Scholars Portal Dataverse.
19. H. Cao, D. G. Cooper, M. K. Keutmann, R. C. Gur, A. Nenkova, and R. Verma, "CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset," IEEE Transactions on Affective Computing, vol. 5, no. 4, pp. 377-390, Oct.-Dec. 2014.
20. K. R. Scherer and H. G. Wallbott, "International Survey on Emotion Antecedents and Reactions (ISEAR) [Data set]," Swiss Center for Affective Sciences, Univ. of Geneva, 1997.