IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 6, November-December 2024

Impact Factor: 7.394

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106016

A Web Application-Based Mock Interview Using NLP Techniques

Ms. B. Padmaja[1], Ch. Sudheer Kumar[2], V. Preethi[3], B. Sai Someswar[4], Akankhya Gantayat[5]

Assistant Professor, Department of IT, GMR Institute of Technology, Vizianagaram, Andhra Pradesh, India[1]

UG Student, Department of IT, GMR Institute of Technology, Vizianagaram, Andhra Pradesh, India[2,3,4,5]

ABSTRACT: Mock interviews have long been a critical tool in preparing candidates for real-world interviews. However, traditional mock interview processes can be time-consuming, resource-intensive, and often lack personalization. To address these challenges, this system introduces a web-based mock interview platform integrated with machine learning and Natural Language Processing (NLP) technologies, designed to provide a scalable, personalized, and interactive interview experience to enhance interview preparation for job candidates. The platform aims to simulate a realistic interview environment by leveraging key AI-driven components, including a Resume Parser, Question Generator Bot, Speech-to-Text, and Text-to-Speech technologies. To evaluate the candidate's performance, the application incorporates Machine Learning techniques, specifically Cosine Similarity and BERT score. These methods assess the semantic relevance and coherence of the candidate's spoken answers in comparison to the expected or ideal responses. Additionally, the ROUGE score is employed to measure the overlap between the candidate's responses and reference answers, ensuring that the content is thoroughly evaluated for both meaning and accuracy. This application provides a highly interactive and intuitive interface for candidates to practice their interviewing skills, offering immediate feedback based on AI-driven metrics. By integrating these technologies, the system not only personalizes the interview experience but also provides a robust mechanism for evaluating performance through objective, data-driven techniques. This approach holds significant potential for revolutionizing the interview preparation process, making it more effective, accessible, and reflective of real-world interview scenarios.

KEYWORDS: BERT score, Cosine Similarity, Natural Language Processing (NLP), Question Generator Bot, Resume Parser, ROUGE score, Speech – to – Text, Text – to – Speech.

I. INTRODUCTION

Natural Language Processing (NLP) is a dynamic and essential subfield of artificial intelligence (AI) that focuses on equipping machines with the ability to comprehend, interpret, generate, and respond to human language in ways that are meaningful and contextually relevant. By integrating the principles of linguistics, computer science, and machine learning, NLP facilitates the processing and analysis of vast volumes of text and speech data, enabling machines to bridge the gap between computational systems and human communication. NLP comprises several core linguistic components. Syntax examines the rules and structures governing the arrangement of words and phrases to form grammatically correct sentences. Semantics delves into the meanings of individual words and combinations, ensuring that language processing aligns with its intended message.

Pragmatics adds another layer by considering the context, intent, and subtleties that influence how language is understood in real-world situations. For spoken language, NLP addresses additional complexities through phonetics (the study of speech sounds) and phonology (the systematic organization of these sounds in language). Together, these elements form a comprehensive foundation for interpreting both written and spoken language. The applications of NLP are far-reaching and transformative. In business and technology, NLP powers tools such as sentiment analysis to gauge opinions and emotions from text, machine translation to break language barriers, and chatbots to automate customer interactions. In media and communication, NLP enables text summarization for condensing large documents, speech recognition for converting spoken words into text, and search engines for providing relevant results to user queries. These applications are pivotal in improving user experiences and automating labour-intense tasks across industries.

The evolution of NLP techniques has been significant. Early systems relied on rule-based approaches, where explicit linguistic rules were encoded manually, and statistical models that analysed text based on probabilities. With advancements in machine learning, more sophisticated methods emerged, particularly with neural network architectures such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory networks (LSTMs), which excelled at

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106016

handling sequential data. The introduction of Transformer-based models, like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), marked a turning point. These models leverage vast datasets and advanced training techniques to understand and generate text with remarkable accuracy, enabling applications such as conversational agents, code generation, and creative writing. Despite its progress, NLP faces numerous challenges.

In today's job market, where competition is fierce and employers have high expectations, the importance of thorough interview preparation cannot be overstated. Candidates are often judged not only on their skills and qualifications but also on their ability to perform under pressure during interviews. For many, the key to improving interview performance lies in practice—particularly through mock interviews. However, access to quality mock interviews can be challenging. Professional coaching or mock interview services are often expensive and scheduling them can be difficult due to limited availability. Additionally, the time investment required can be a barrier for many people, especially those balancing job searches with existing responsibilities. Many current mock interview platforms fail to fully address these challenges. Some rely on a set of predefined questions, offering little flexibility or relevance to individual users. Others use basic automated systems that may assess answers in a generic manner, providing superficial feedback that doesn't help candidates identify specific areas for improvement. These platforms often overlook the need for personalized, adaptive feedback that reflects the nuanced nature of real interview experiences. Additionally, ethical concerns, such as the presence of bias in models stemming from biased training data, pose significant obstacles to the development of fair and inclusive systems. Nonetheless, the potential of NLP remains vast. By continuously advancing its methodologies and addressing its challenges, NLP is shaping the future of human-computer interaction. From virtual assistants like Siri and Alexa to automated translation systems that connect people across linguistic boundaries, and text analytics tools that extract insights from unstructured data, NLP is revolutionizing how we communicate and interact with technology.

As research and innovation in the field persist, NLP continues to unlock new possibilities for enhancing understanding between humans and machines. In today's highly competitive job market, preparing for interviews is critical for candidates seeking to improve their chances of success. However, access to professional mock interviews can be limited by factors such as cost, availability, and time constraints. Many existing mock interview platforms either rely on predefined questions and manual evaluations, or use basic automated systems that fall short in providing meaningful, personalized feedback. To address these limitations, we are building a web-based mock interview application that leverages Natural Language Processing (NLP), speech-to-text technology, and advanced AI models like BERT. This system aims to provide users with a dynamic, immersive, and adaptive interview preparation tool that closely simulates real-life interviews, while offering real-time, data-driven feedback. In today's competitive job market, effective interview preparation is critical for candidates aiming to secure employment.

mock interviews can be expensive, time-consuming, and limited by availability, making them inaccessible for many job seekers. Furthermore, existing mock interview platforms often rely on static questions and basic automation, which lack the depth and personalization needed to provide meaningful feedback. To address these challenges, we are developing a web-based mock interview application that leverages Natural Language Processing (NLP), speech-to-text technology, and advanced AI models like BERT. This innovative platform will simulate real interview dynamics, generating tailored questions based on the user's target role and industry, analyzing responses for relevance, clarity, and tone, and providing real-time, objective feedback. Unlike conventional methods, our application offers accessibility and flexibility, allowing candidates to practice at their convenience, while receiving data-driven, unbiased evaluations.

To bridge this gap, we are developing a web-based mock interview platform that leverages advanced technologies such as Natural Language Processing (NLP), speech-to-text capabilities, and AI models like BERT (Bidirectional Encoder Representations from Transformers). By integrating these cutting-edge tools, the platform can simulate dynamic and realistic interview scenarios, adapting questions based on the user's responses and providing real-time, in-depth feedback.

The use of NLP allows the system to understand and analyse spoken language in a sophisticated way, while speech-to-text technology ensures that spoken answers are accurately transcribed for evaluation. By employing BERT and similar AI models, the platform can assess the content of a user's responses in a more contextual and meaningful manner, identifying areas where the candidate might improve, such as communication style, clarity of thought, or the depth of their answers.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106016

What sets this system apart from others is its ability to offer personalized feedback. Instead of generic responses, the AI-driven evaluation can pinpoint specific strengths and weaknesses in a candidate's interview performance, helping them understand how to adjust their answers to fit the expectations of different job roles or industries.Furthermore, the platform's adaptability means that users can engage in a wide variety of interview types, whether it's a technical interview for a software engineering position, a behavioural interview for a management role, or any other professional scenario.

Ultimately, this application aims to democratize access to high-quality mock interviews by making them more affordable, flexible, and tailored to individual needs. It will serve as a powerful tool for job seekers to refine their interview skills and boost their confidence, giving them a competitive edge in the hiring process.

II. LITERATURE SURVEY

The framework developed to predict interview performance and social traits uses a dataset of 138 mock interviews with MIT students, where judges rated traits like engagement, friendliness, and stress. It integrates multiple data modalities and uses techniques such as regression models, feature extraction, and dimensionality reduction to offer insights into social behaviours. This system can automate feedback on interview performance, accurately predicting traits such as engagement and excitement (correlation > 0.70), with an overall interview performance correlation of 0.62. Positive behaviours identified include reducing filler words, smiling more, and maintaining eye contact. However, the model has limitations. It depends on subjective ratings, which may introduce bias, and its dataset is homogenous, limiting generalizability. The mock setup lacks real-world stress and job relevance, and cultural differences in interpreting social cues, along with Mechanical Turk ratings, reduce robustness. Although promising for general feedback, the system needs refinement to be applicable to diverse and job-specific contexts. [1]

This study examines gaze and head orientation in individuals with autism, especially focusing on joint attention in social contexts like job interviews. A VR-based job interview simulation was developed, incorporating machine learning to analyse gaze, head movements, speech levels, and attention on interviewers. Though IRB restrictions prevent public access to the dataset, the system captures comprehensive behavioural data through VR, gaze and head tracking, machine learning, and voice activity detection. The VR system offers a controlled environment for safe practice, enabling detailed behavioural analysis and self-practice customization. It objectively tracks both neurotypical and autistic behaviours, providing insights into interview-related challenges. However, limitations include a small sample size, VR usability issues, and restricted data access. Additionally, participants don't receive real-time feedback, and the set interview questions limit realism. The findings reveal differences between autistic and non-autistic participants: non-autistic individuals show more consistent gaze and head orientation, while autistic individuals display lower joint attention and less behavioural mirroring. Despite these challenges, the VR system shows potential as a tool for both neurotypical and autistic individuals to improve interview skills. [2]

This study explores stress measurement in autistic individuals during simulated job interviews, using physiological data to provide insights into stress management for both interviewees and interviewers. It introduces a unique physiological dataset on stress in autistic adults and tests various supervised learning models, with Elastic Net Regression achieving 84.8% accuracy in individual stress predictions, and Support Vector Regression (SVR) performing well in group predictions with 75.4% accuracy. The approach offers non-invasive, real-time stress detection, capturing involuntary responses specific to autistic individuals and featuring adaptive learning and visualization. However, challenges include the need for personalized models due to individual variability, sensor sensitivity, and a limited feature set, as well as technical issues like data complexity and potential overfitting. The study is limited by its small sample size of 15 autistic adults, which affects generalizability, and some participants experienced discomfort with sensors. Additionally, focusing solely on physiological data restricts stress detection; future studies should include multimodal data like gaze and speech. Individual model training is time-consuming, and sensor noise requires significant preprocessing. Despite these limitations, the study shows that Elastic Net Regression is effective for individual stress prediction, while SVR works well across groups. Visualization tools highlighted stress peaks during open-ended questions, suggesting areas for focused stress management. [3]

The "Real Time Mock Interview Using Deep Learning" system is a web-based platform for virtual job interview practice. It analyses facial expressions, head movements, reaction time, speaking rate, and grammar to provide immediate, actionable feedback, helping users improve their interview skills and confidence through repeated practice. The system uses a dataset of 1,500 emotion-eliciting video clips, including 28 standardized clips representing six emotions and a neutral expression, aiding in emotion recognition during interviews. User performance metrics and progress data are stored to track improvements over time. The system employs models like SVM, KNN, Attention-LSTM, and SENN for accurate emotion and facial expression recognition. It provides benefits such as high accuracy, real-time feedback, a user-friendly interface, and customizability. However, challenges include limited realism in virtual interviews, a constrained emotional range, dependence on technology, and data privacy concerns. Limitations include potential inaccuracy in facial and grammar analysis, user engagement issues, and data privacy risks. Despite these challenges, users reported positive experiences, finding the system helpful for tracking and improving performance in areas like facial expressions, speaking rate, and grammar. This suggests it holds promise as an effective tool for interview preparation. [4]

The AI-based mock interview evaluator system is designed to assess emotions and confidence in real-time, aiming to improve recruitment practices and promote inclusivity using machine learning (ML) and AI. It employs convolutional neural networks (CNNs) to classify emotions like happiness, sadness, and anger, while recurrent neural networks (RNNs) analyse speech patterns and body language. Additionally, speech recognition transcribes responses, sentiment analysis detects emotional cues in language, and facial expression and gesture recognition capture nonverbal cues. Using a diverse dataset of labelled mock interviews, the system analyses emotions and confidence through speech and body language. It provides real-time feedback to enhance emotional intelligence assessment and candidate confidence. While it offers a user-friendly interface for remote interviews, improving objectivity and efficiency, it faces challenges such as potential misinterpretation of emotions, dataset limitations, AI bias, and reduced human interaction. This system supports candidates from diverse backgrounds, helping them prepare for interviews by building confidence and skills. Despite limitations, it represents an innovative step toward using AI to create a fairer, more inclusive recruitment process. [5]

The paper *"AI-based Behavioural Analyser for Interviews/Viva"* by Dissanayake et al. (2021) introduces an AI-driven system for evaluating candidate behaviour during interviews by analysing nonverbal cues like facial expressions, eye movements, and head gestures. It utilizes datasets such as SMILEs and SPOS for smile analysis, Eye-Dataset and Eye Aspect Ratio for gaze and blink detection, and Chon Kanade for emotion recognition, enabling real-time assessment of attention, emotions, and smile genuineness. The system also incorporates personality evaluation based on the Big Five model (openness, conscientiousness, extraversion, agreeableness, neuroticism) inferred from nonverbal behaviours. Technical implementations, including CNNs and ResNet-18, enhance feature extraction and classification. Tested on real-time interview recordings, the system provides a structured, data-driven approach to candidate evaluation, with additional capabilities in emotion detection and verbal analysis through NLP for fluency and coherence assessment.[6]

Jadhav et al. (2024) review advancements in AI-driven simulations and machine learning models for virtual interview preparation, addressing the limitations of traditional mock interviews in scalability and accessibility. They highlight the integration of pose estimation, natural language processing (NLP), and emotion recognition to enhance the realism and effectiveness of virtual interviews. Machine learning algorithms, including CNNs for visual analysis and RNNs for text generation, underpin responsive interview bots capable of assessing verbal and nonverbal cues. Technologies like BERT and GPT enable semantic understanding and nuanced questioning, while feedback mechanisms using cosine similarity and deep learning models evaluate user responses. Pose-based systems track micro-expressions and postural shifts to assess confidence and stress, enhancing user engagement and providing deeper insights into interview dynamics. The paper discusses challenges like ensuring real-time response accuracy and handling multimodal data efficiently, concluding with the development of an AI-powered mock interview simulator that integrates pose analysis, NLP, and machine learning for improved virtual interview training.[7]

The paper *"Computer Vision-based Online Job Interview Proctoring for Campus Placement"* by Srivastava, Tripathi, and Pant (2024) proposes a computer vision system to ensure secure, fair, and efficient online job interviews, particularly for campus placements. Leveraging machine learning and image processing, the system monitors candidates' behaviour, detecting cheating, impersonation, and suspicious activities in real time. It addresses the limitations of traditional surveillance methods, such as human error and system overload, by tracking gestures, facial expressions, and eye movements. This solution enhances the interview experience by ensuring integrity and providing real-time feedback. The paper highlights its scalability for large recruitment drives, ensuring authenticity and unbiased candidate evaluation.[8]

IJARETY | ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106016

The paper *"Automated Analysis and Behavioural Prediction of Interview Performance using Computer Vision"* by Priya, Malavika, and Lijiya (2022) presents a method to evaluate and predict interview performance using computer vision techniques. It addresses the limitations of subjective traditional evaluations by automating the analysis of non-verbal cues like facial expressions, body language, and gestures, which reveal emotional state, confidence, and demeanour. The paper reviews methods like facial expression recognition, posture detection, and gesture analysis, emphasizing the value of combining verbal and non-verbal cues for a comprehensive assessment. Machine learning models, particularly CNNs, enhance accuracy in recognizing subtle behaviours. Challenges such as privacy concerns, cultural diversity, and model biases are acknowledged, with suggestions for future improvements, including the integration of multimodal data, refined predictive models, and ethical considerations for automated recruitment.[9]

Su et al.'s (2016) paper *"Dialog State Tracking and Action Selection using Deep Learning Mechanism for Interview Coaching"* examines the application of advanced dialog systems to improve interview coaching. Traditional methods are critiqued for their lack of personalization and adaptability. The study highlights dialog state tracking (DST) as essential for maintaining context and understanding user intent. While earlier rule-based and statistical approaches to DST faced challenges in scalability and managing complex dialogues, deep learning methods, particularly RNNs and LSTMs, have advanced DST and action selection by enabling nuanced, real-time conversational flows and enhanced user engagement. The authors review prior applications of dialog systems in language learning and coaching, identifying a gap in leveraging these technologies for interview preparation. They propose a deep learning-driven system to enhance DST and adaptive action selection, addressing the limitations of traditional systems such as inflexibility, reliance on predefined rules, and difficulty handling complex dialogues.[10]

III. METHODOLOGY

This study aimed to evaluate bias and fairness in an Automated Video Interview (AVI) system by leveraging a comprehensive dataset of mock interview videos. The methodology focused on balancing the dataset for gender representation, extracting multimodal features, and assessing fairness through machine learning models. By integrating verbal, paraverbal, and visual modalities, the research provided insights into how gender influences model performance and fairness metrics.

Dataset and Participants

The dataset consisted of 4,255 mock interview videos from 733 participants, primarily upper-level undergraduate students recruited through universities and Prolific. These participants were likely to be actively seeking employment, making them an ideal sample for evaluating hireability in managerial or team lead roles. Participants were compensated differently based on recruitment sources: university participants received a $10 Amazon gift card, while Prolific participants were paid $7.20 directly. Each participant underwent a structured asynchronous interview through an online platform, answering six questions designed to assess hireability. Videos were included in the analysis only if features could be extracted from at least four responses.

To address potential gender bias, the dataset was balanced across genders. While 262 participants identified as men and 465 as women, six non-binary individuals were excluded from the analysis due to their small representation. A gender-matching algorithm was applied to down-sample the majority class (women) to create a dataset with equal representation (262 men and 262 women). Statistical tests such as the Fligner-Killeen and Kolmogorov-Smirnov tests confirmed that the gender-balanced dataset had equivalent variances and distributions, ensuring fair comparison across genders.

Annotation and Hireability Scoring

A team of trained research assistants annotated the videos, with at least three raters per video to ensure consistency. The raters underwent frame-of-reference training, which involved defining evaluation constructs, familiarizing themselves with rating scales, completing practice ratings, and resolving disagreements collaboratively. Hireability was assessed using two Likert-scale items: (1) “I would recommend this person be hired,” and (2) “If hired, I believe this person would perform well on the job.” Each participant’s hireability score was the average of these ratings. Inter-rater reliability, computed as ICC(1,k)=0.67, indicated moderate agreement among raters. The construct validity of these ratings was supported by positive correlations between hireability scores and general mental and verbal ability.

Three machine learning models were trained using Random Forest regressors to evaluate bias and fairness: (1) a baseline model using all features, (2) a gender-normed model with features normalized separately for men and women, and (3) a reduced-features model optimized to minimize gender predictability. Feature reduction was achieved

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106016

iteratively by eliminating the most gender-predictive features until gender prediction accuracy was close to random chance (AUROC ≈ 0.5). Hyperparameter tuning was conducted for each modality using stratified cross-validation, and final model performance was assessed using predefined train/test splits to ensure reproducibility. Metrics such as Spearman rank correlation (accuracy), Adverse Impact Ratio (fairness), and AUROC (gender bias) were evaluated across models to measure robustness and mitigate gender bias effectively.

Feature Extraction Across Modalities

To capture a comprehensive set of features, the study analyzed three modalities: verbal, paraverbal, and visual.

- Verbal Features: These included n-gram frequencies and categories from the Linguistic Inquiry and Word Count (LIWC) tool. Features were processed using Term Frequency-Inverse Document Frequency (TF-IDF) weighting and filtered based on high Pointwise Mutual Information (PMI) values. This yielded approximately 5,653 verbal features per participant.
- Paraverbal Features: These comprised acoustic parameters such as loudness, jitter, and shimmer, extracted using the Geneva Minimalistic Acoustic Parameter Set (GeMAPS). A total of 125 paraverbal features were derived.
- Visual Features: These included facial expressions and body motion data, analyzed with tools such as Emotient's FACET and Motion Tracker. Approximately 250 visual features were extracted. Statistical functionals, such as medians and standard deviations, summarized feature values across each participant's six videos.

Machine Learning Models for Bias and Fairness

Three Random Forest regression models were developed to evaluate bias and fairness:

1. **Baseline Model**: This model utilized all extracted features across the three modalities without modifications. It served as the control for comparison with other models.
2. **Gender-Normed Model**: In this approach, features were normalized separately for men and women to reduce the influence of gender on predictions. The intention was to account for inherent differences in feature distributions between genders.
3. **Reduced-Features Model**: This model aimed to minimize gender predictability by iteratively removing the most gender-informative features. The process continued until gender prediction accuracy reached random chance (AUROC ≈ 0.5).

Hyperparameter tuning for each modality was conducted through stratified cross-validation to ensure optimal performance. Final model evaluation was performed using predefined train/test splits to enhance reproducibility.

Resume Parser

In today's data-driven landscape, resumes are valuable sources of information that can transform recruitment and talent management when processed effectively. This potential is harnessed through resume parsing, which converts unstructured resume text into structured data. The resume parsing pipeline starts with a resume parser that uses natural language processing (NLP) and machine learning to extract key information. Next, normalization standardizes the text, followed by tokenization, which breaks it into smaller units. Stop words are removed to focus on essential content, while lemmatization and stemming reduce words to their root forms to group related terms. Part-of-speech (POS) tagging and frequency analysis then help identify important keywords, which are ultimately selected based on job relevance through rule-based and machine learning techniques.

Fine Turing

BERT, or Bidirectional Encoder Representations from Transformers, is a pre-trained model that can be fine-tuned for various tasks. In this case, it's being trained with the SQuAD (Stanford Question Answering Dataset) and an additional dataset containing 88,000 entries, with KeyBERT used to generate questions from context by extracting keywords.BERT comes in two main variants: BERT-base with 12 encoder layers, 768 hidden units, 12 attention heads, and 110 million parameters, and BERT-large with 24 encoder layers, 1024 hidden units, 16 attention heads, and 340 million parameters.

Text and Speech Recognition

Speech-to-Text (STT) and Text-to-Speech (TTS) are two natural language processing technologies that facilitate communication between humans and machines. STT converts spoken language into text, making it valuable for voice assistants, transcription, and accessibility features. TTS transforms written text into spoken language, useful for

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106016

applications like virtual assistants and helping visually impaired individuals by reading content aloud. Together, they enhance device interaction by bridging audio and text. In mock interview platforms, TTS (using Speech Synthesis Utterance) converts interview questions into spoken audio, with customizable options for voice, speed, pitch, and volume to enhance realism. The audio allows users to focus on formulating responses, simulating real interview dynamics. STT, via speech recognition, captures and transcribes users' spoken answers, helping the platform analyse responses. It handles variations in accents and speech patterns, adapting to diverse user inputs and detecting pauses to signify response completion.

IV. RESULT AND DISCUSSION

Procedural Fairness (Gender Predictability):

Figure 3 highlights the gender prediction AUROC distributions across modalities for the baseline models. The optimal AUROC score of 0.5 represents the model's inability to predict gender, ensuring gender blindness. In this study, the balanced dataset contained equal numbers of men and women, with a prior probability of 0.5 for guessing gender correctly, serving as a baseline for assessing gender predictability. Results showed that all modalities contributed to gender predictability in the baseline models, though the extent varied. Verbal and visual baseline models exhibited AUROC scores of 0.73 and 0.79, respectively, indicating limited ability to predict gender. However, paraverbal and combined modality baseline models nearly achieved perfect gender prediction due to features like vocal pitch, which inherently carry gender information.

Figure 4 evaluates how gender bias mitigation strategies influenced gender predictability using the combined modality. Gender predictability in both baseline and gender-normed models remained above the ideal AUROC of 0.5, suggesting that z-scoring predictors did not substantially diminish gender-related information, especially in multimodal settings. For unimodal gender-normed models, visual and paraverbal modalities showed significant reductions in AUROC (from 0.79 to 0.58 and 0.99 to 0.69, respectively). However, gender-norming was ineffective for verbal data due to inherent linguistic differences between genders. Reduced-features models achieved AUROC scores close to 0.5 across modalities, successfully mitigating gender predictability.

Validity (Accuracy)

Figure 5 illustrates the distribution of Spearman's rank correlation coefficients (ρ) across 100 trials for baseline models using unimodal and combined features. The small standard deviations (0.02–0.03) highlight the robustness of results despite random variations in data partitions and training processes. Among unimodal models, the verbal modality performed best, achieving a moderate accuracy of $\rho = 0.45$, indicating strong alignment between predictions and ground truth. Paraverbal and visual modalities underperformed compared to verbal data. The combined feature set only slightly outperformed the verbal modality alone, with $\rho = 0.46$.

Figure 6 focuses on combined modality model variants. The baseline and gender-normed models showed nearly identical performance, while the reduced-features model exhibited a notable drop in ρ (from 0.46 to 0.38). This decline, although statistically significant ($p < 0.01$), reflects the removal of features critical to hireability predictions but also correlated with gender. Despite the reduction, the combined reduced-features model retained sufficient predictive power to make meaningful assessments, albeit with slightly lower accuracy compared to other variants.

Bias (Differential Correlational Accuracy)

Bias was assessed through gender differences in Spearman's ρ (women minus men). Figure 7 shows the distribution of these differences for baseline models across modalities. Ideally, ρ differences should be zero, indicating equal accuracy for men and women. However, results revealed positive differences for most modalities, indicating models were generally more accurate for women. For the visual modality, 100% of trials showed positive differences, with a substantial bias magnitude (mean = 0.15). Paraverbal and verbal modalities exhibited smaller biases (means = 0.07 and 0.04, respectively).

Figure 8 highlights bias variations across combined modality model variants. While baseline and gender-normed models showed consistent bias favoring women, the reduced-features model achieved a mean ρ difference of zero, effectively mitigating gender bias. Similar trends were observed in unimodal settings. For the paraverbal modality, bias reduced from 0.07 to 0.05 under the reduced-features model. The verbal modality demonstrated near-complete bias elimination, with differences dropping from 0.04 to 0.01. Conversely, visual modality bias remained unaffected, potentially due to low baseline accuracy ($\rho = 0.16$).

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106016

Distributive Fairness (Adverse Impact)

Adverse impact was quantified using the AI ratio under the assumption that the top 10% of participants were selected. Figure 9 displays AI ratio distributions for each baseline modality. The ideal AI ratio is 1.0, reflecting equal selection rates for men and women. Legally, an AI ratio above 0.8 is acceptable under the U.S. four-fifths rule. For verbal and paraverbal modalities, the baseline AI ratios (mean = 0.87 for both) were within the acceptable range in 70% and 72% of trials, respectively. However, the visual modality had a mean AI ratio of 0.52, falling below the legal threshold in 99% of trials, signaling clear evidence of discrimination. The combined baseline model also fell short, with a mean AI ratio of 0.76 and 73% of trials below the 0.8 threshold.

Figure 10 shows how mitigation strategies impacted fairness in the combined modality. Gender-normed models slightly improved fairness, raising the mean AI ratio from 0.76 to 0.81, while reduced-features models significantly enhanced fairness (mean AI ratio = 0.87). Unimodal results in Table 1 indicated that gender-norming improved fairness for visual predictors (AI ratio = 0.60) and reduced features further improved it to 0.80. The verbal and paraverbal modalities, already within acceptable fairness ranges, showed minimal changes across models. Overall, the reduced-features approach demonstrated the greatest effectiveness in promoting distributive fairness across modalities.

V. CONCLUSION

The development of a web-based mock interview platform using advanced NLP and machine learning revolutionizes interview preparation by addressing the limitations of traditional methods, such as high costs, limited access, and static feedback. This AI-driven platform incorporates resume parsing, dynamic question generation, speech-to-text processing, and semantic evaluation via BERT and cosine similarity, providing tailored, real-time adaptive interview simulations. Tools like ROUGE scores enable precise evaluation of content relevance, coherence, and accuracy, ensuring personalized, actionable feedback. Key features include accessibility for users in underserved regions, bias-free automated evaluations, and adaptability across industries and job roles. Dynamic question generation and real-time feedback on relevance, clarity, and body language empower users to prepare for various interview types effectively. Challenges such as refining AI models, mitigating training data biases, and integrating multimodal feedback must be addressed, alongside ensuring robust data privacy and security. This project demonstrates the transformative potential of combining AI, NLP, and machine learning, offering a scalable, interactive solution that enhances interview preparation, democratizes access, and sets new standards in recruitment technology.

REFERENCES

1. Naim, I., Tanveer, M. I., Gildea, D., & Hoque, M. E. (2021). Automated analysis and prediction of job interview performance. IEEE Transactions on Affective Computing, 9(2), 191-204
2. Artiran, S., Bedmutha, P. S., & Cosman, P. (2024). Analysis of Gaze, Head Orientation and Joint Attention in Autism with Triadic VR Interviews. IEEE Transactions on Neural Systems and Rehabilitation Engineering.
3. Migovich, M., Adiani, D., Breen, M., Swanson, A., Vogus, T. J., & Sarkar, N. (2024). Stress Detection of Autistic Adults during Simulated Job Interviews Using a Novel Physiological Dataset and Machine Learning. ACM Transactions on Accessible Computing, 17(1), 1 25.
4. Patil, R., Butte, A., Temgire, S., Nanekar, V., & Gavhane, S. (2021). Real Time Mock Interview using Deep Learning. IJERT, 10(5).
5. Amrutha, M. B., Kumar, T. N., Sathwika, M. N., & Pavansai, P. (2024). Enhancing Interview Evaluation: AI-Based Emotion And Confidence Analysis In Mock Interview. Journal of Nonlinear Analysis and Optimization, 15(1).
6. Migovich, M., Adiani, D., Breen, M., Swanson, A., Vogus, T. J., & Sarkar, N. (2024). Stress Detection of Autistic Adults during Simulated Job Interviews Using a Novel Physiological Dataset and Machine Learning. ACM Transactions on Accessible Computing, 17(1), 1-25.
7. Jadhav, B., Sawant, A., Shah, A., Vemula, P., Waikar, A., & Yadav, S. (2024, March). A Comprehensive Study and Implementation of the Mock Interview Simulator with AI and Pose-Based Interaction. In 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU) (pp. 01-05). IEEE.
8. Srivastava, A. K., Tripathi, V., & Pant, B. (2024, February). Computer Vision based Online Job Interview Proctoring for Campus Placement. In 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 663-668). IEEE.

9. Priya, D. H., Malavika, P., &Lijiya, A. (2022, November). Automated Analysis and Behavioural Prediction of Interview Performance using Computer Vision. In *2022 IEEE 19th India Council International Conference (INDICON)* (pp. 1-6). IEEE.
10. Su, M. H., Huang, K. Y., Yang, T. H., Lai, K. J., & Wu, C. H. (2016, November). Dialog State Tracking and action selection using deep learning mechanism for interview coaching. In 2016 International Conference on Asian Language Processing (IALP) (pp. 6-9). IEEE.
11. CJ, J. M., Sabi, M., Benson, M., Baburaj, G., & Saritha, S. (2024, April). Q&AI: An AI Powered Mock Interview Bot for Enhancing the Performance of Aspiring Professionals. In 2024 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI) (pp. 1-5). IEEE.
12. Booth, B. M., Hickman, L. Subburaj, S. K., Tay, L., Woo, S. E., & D'Mello, S. K. (2021, October). Bias and fairness in multimodal machine learning: A case study of automated video interviews. In Proceedings of the 2021 international conference on multi-model interaction (pp. 268-277).
13. Varalakshmi, P., & Bugatha, N. M. K. (2024, March). AI-Powered Resume Based QA Tailoring for Success in Interviews. In 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1-6). IEEE.

IJARETY

INNO SPACE
SJIF Scientific Journal Impact Factor

doi
crossref

निस्केयर
NISCAIR

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com