IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, October 2024

InterviewX: AI/ML Powered Interview Simulator Using NLP And CNN

Kute Manjusha[1], Bhor Akshada[2], Mate Prachi[3], Mundhe Bhalchandra[4]
Department of AI DS Engineering[1,2,3,4]
JCEI's Jaihind College of Engineering, Kuran, Maharashtra, India
kutemanjusha4@gmail.com, akshadabhor210@gmail.com
mateprachi7@gmail.com, mundheraj.mundhe@gmail.com

Abstract: *Interview-X is an advanced AI and ML-powered simulator designed to mimic real-world interview scenarios, enabling users to effectively prepare for interviews. The platform evaluates both verbal and non-verbal responses, offering comprehensive feedback that helps users refine their performance. By integrating Natural Language Processing (NLP) and Convolutional Neural Networks (CNN), Interview-X delivers precise insights into the user's communication skills and confidence levels. For speech evaluation, the system uses NLP to transcribe and analyze spoken answers. By leveraging the Google Speech-to-Text API, it converts audio inputs into text and evaluates the accuracy and relevance of the responses based on the interview questions. A predefined accuracy threshold determines whether the answer meets the desired standard, offering actionable feedback for improvement. On the non-verbal front, Interview-X uses CNNs for facial expression analysis. With technologies like MTCNN for facial detection and FaceNet /VGGFace for feature extraction, the system assesses the user's confidence by comparing facial landmarks and expressions with predefined confident face templates. This analysis ensures that both verbal content and non-verbal cues are aligned for effective interview performance. The platform aggregates results from both NLP and CNN evaluations, providing users with detailed feedback on their performance. This holistic approach helps individuals enhance both their answers and their presentation, boosting their interview confidence and readiness. Interview-X is a valuable tool for anyone looking to improve their interview skills in a realistic, data-driven environment.*

Keywords: AI, ML, CNN, NLP, Embedding system, Recognition

I. INTRODUCTION

Interview-X is an innovative AI and ML-based platform designed to simulate realistic interview environments, helping users enhance their interview performance through comprehensive feedback. As the job market becomes more competitive, mastering both technical skills and interpersonal communication is essential for candidates aiming to succeed in interviews. Interview-X addresses this need by combining advanced Natural Language Processing (NLP) and Computer Vision (CV) techniques to evaluate users' verbal and non-verbal responses in real time. Interviews not only test a candidate's technical knowledge and problem-solving skills but also assess their communication, confidence, and ability to manage stress. Traditionally, interview preparation has involved mock interviews with human interviewers or static question-answer sessions, which often lack dynamic feedback on key elements such as verbal communication effectiveness, body language, and confidence. As the demand for more personalized and efficient preparation methods grows, leveraging Artificial Intelligence (AI) and Machine Learning (ML) offers a promising solution. The platform's goal is to provide an immersive interview experience that goes beyond traditional preparation methods. While mock interviews often focus solely on the content of a candidate's answers, Interview-X analyzes both what users say and how they say it. By integrating live audio and video data, it evaluates speech accuracy, facial expressions, and confidence, offering users actionable insights to improve their performance. This system is built using modern web technologies and AI/ML models. The front end is developed with React, while the back end uses Node.js and Firebase for real-time communication and data storage. The platform captures live audioand video streams,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, October 2024

processes the data using NLP and CNN pipelines, and generates feedback based on a combination of speech accuracy and facial confidence analysis.

Interview-X marks a significant step forward in interview preparation, offering a powerful tool for users to refine their skills and boost their confidence in professional settings. Through this platform, candidates can better understand their strengths and weaknesses, ensuring they are well-prepared for their next interview opportunity.

This paper presents interview-X, an innovative AI-based interview simulator that creates a realistic interview environment, providing comprehensive feedback on both verbal and non-verbal communication. The platform aims to bridge the gap between traditional preparation techniques and the real demands of professional interviews by analyzing the candidate's responses in real-time. It evaluates both speech content and facial expressions to give a holistic assessment of performance, empowering users to refine their skills before stepping into actual interviews.

II. PROBLEM STATEMENT

Develop an AI/ML-powered interview simulator that uses NLP for analyzing candidate responses and CNN techniques for evaluating speech and facial expressions, providing personalized feedback for interview preparation

II. PROBLEM STATEMENT

Develop an AI/ML-powered interview simulator that uses NLP for analyzing candidate responses and CNN techniques for evaluating speech and facial expressions, providing personalized feedback for interview preparation.

III. GOALS AND OBJECTIVES

- To Enhanced user experience with AI-Power feedback.
- To Boost user Confidence with real time analysis.
- To Ensure accessibility for all experience levels.
- To continuously update with new modules and features.

IV. MOTIVATION AND BACKGROUND

Growing need for effective, scalable interview preparation tools in a competitive job market.

- Limitations of Traditional Methods: Lack of real-time feedback and ability to assess both verbal and non-verbal communication, especially in virtual settings.
- AIAdvancements: Utilizes NLP (Google Speech-to-Text) and computer vision (FaceNet, MTCNN) for real-time analysis of speech and facial expressions.
- Real-Time Feedback: Provides detailed feedback to help users improve both verbal and non-verbal interview per formance.
- Modern Interview Requirements: Bridges the gap between traditional preparation methods and the demands of virtual, AI-driven interview processes.

V. METHODOLOGY

A. Algorithm

Speech Recognition Algorithm (Google Speech-to-Text)

Input: audio data (from user) Output: transcribed text
Step 1: Initialize Google Speech-to-Text API
Step 2: Send audio data to the API
Step 3: Receive the transcribed text from API response
Step 4: Return transcribed text

Text Evaluation Algorithm (Gemini Pro):

Input: transcribed text, interview question Output: accuracy score
Step 1: Initialize Gemini Pro NLP API

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, October 2024

Step 2: Send transcribed text and interview question to the Gemini Pro API
Step 3: Receive accuracy score from the API response (0 to 100
Step 4: If accuracy score ¿= 80, mark answer as correct
Step 5: Return accuracy score.

Facial Detection Algorithm (MTCNN):
Input: video frame Output: face coordinates, facial landmarks
Step 1: Initialize MTCNN for face detection
Step 2: Pass video frame to MTCNN for processing
Step 3: MTCNN detects face coordinates in the frame
Step 4: MTCNN extracts facial landmarks from the face
Step 5: Return face coordinates and facial landmarks.

Feature Extraction Algorithm (FaceNet/VGGFace):
Input: face coordinates, facial landmarks Output: face embedding
Step 1: Initialize FaceNet or VGGFace model
Step 2: Crop the face from the video frame using face coordinates
Step 3: Pass the cropped face and facial landmarks to the model
Step 4: The model generates face embedding (numerical representation
Step 5: Return face embedding.

Graph Comparison Algorithm (Euclidean Distance):
Input: face embedding, stored confident embeddings, stored unconfident embeddings Output: confidence score
Step 1: Initialize comparison metric (Cosine Similarity or Euclidean Distance)
Step 2: Compare face embedding with stored confident embeddings
Step 3: Compare face embedding with stored unconfident embeddings
Step 4: Calculate similarity scores for both comparisons
Step 5: Calculate confidence score based on higher similarity with confident embeddings
Step 6: Return confidence score.

Decision Logic Algorithm:
Input: accuracy score, confidence score Output: pass or fail (boolean)
Step 1: Set pass threshold = 80
Step 2: If accuracy score ¿= pass threshold AND confidence score ¿= pass threshold: pass or fail = PASS Else: pass or fail = FAIL
Step 3: Return pass or fail.

Data Storage and Notification:
Input: user id, accuracy score, confidence score, pass or fail Output: success message (for notification)
Step 1: Initialize connection to Firebase Firestore
Step 2: Store user id, accuracy score, confidence score, and pass or fail in Firestore
Step 3: Send success message to frontend using Socket.IO
Step 4: Notify user of their result (pass or fail)
Step 5: Return success message.

B. System Architecture :

- The frontend captures live audio and video streams using React, Vite, and WebRTC.
- Real-time data is sent to the backend via Socket.IO, where the audio is processed using an NLP pipeline,and video is analyzed by a CNN-based facial detection system

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, October 2024

- Results are aggregated and sent back to the frontend for real-time display.

NLP Pipeline for Speech Transcription:

- Audio Preprocessing: Use librosapydub for audio normalization, denoising, and segmentation.
- Speech-to-Text: Leverage Google Speech-to-Text API for transcription.
- Accuracy Evaluation: Measure accuracy with Gemini Pro, using Word Error Rate (WER).

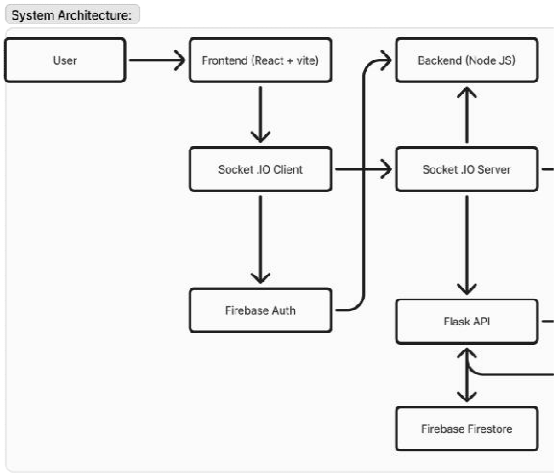

Fig. 1. Architecture of AI ML powered interview simulator using CNN and NLP.

CNN Pipeline for Facial Detection Recognition:

Facial Detection: MTCNN (Multi-task Cascaded Convolutional Neural Networks): This is a deep learning model specifically designed for detecting faces in images. It works by generating bounding boxes around detected faces and identifying key facial landmarks, such as eyes, nose, and mouth. The model uses a series of cascaded CNNs to process images at multiple stages, improving accuracy by refining the face location and landmark points at each stage. MTCNN is highly effective for detecting faces in various lighting conditions, angles, and backgrounds, making it suitable for real-world applications.

Facial Recognition:

FaceNet: Once faces are detected, the next step is facial recognition. FaceNet is a deep neural network that generates facial embeddings — unique numerical representations of faces by analyzing facial features. These embeddings capture essential details and distinguish one face from another. In this process, each face is converted into a 128-dimensional vector. Comparing Embeddings: To recognize a face, FaceNet compares the embeddings of the detected face with those in a database of known faces. Euclidean distance is typically used to measure the similarity between embeddings. A smaller distance indicates a closer match, meaning the faces are more similar, while a larger distance indicates dissimilarity.

Confidence Scoring: Similarity Threshold: During recognition, a confidence score is assigned to each match based on the similarity of the embeddings. This score represents how confident the system is that the detected face matches a known face in the database. The confidence is calculated by comparing the Euclidean distance between embeddings against a predefined threshold. If the distance is below the threshold, it indicates a match with high confidence; otherwise, it is likely not a match. Assigning Scores: The closer the distance to zero, the higher the similarity, and thus the higher the confidence score. This scoring system helps determine if the detected face is an accurate match to a known face or if it is unfamiliar.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, October 2024

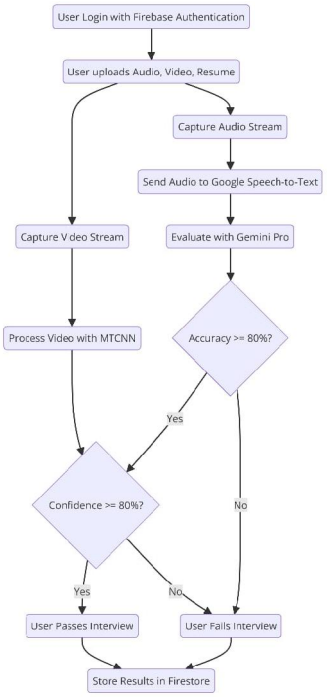

Fig. 2.flowchart of AI ML powered interview simulator using CNN and NLP.

VI. COMPARATIVE RESULT

Predefined Benchmarks or Average Scores: Show how the candidate's performance compares to an average successful candidate's performance (e.g., confidence, correctness of answers).

• Metrics: Percentiles, accuracy scores, and confidence levels.

Past Performance (if available): If the candidate has used the system before, show improvements or declines over time in areas such as facial expressions (confidence), accuracy of responses, and overall scoring. • Metrics: Graphs or trend lines indicating past interview results against current results.

Ideal Responses: Display a side-by-side comparison of the candidate's responses and a model or ideal response, with notes on what could have been improved.

• Metrics: Highlight areas such as relevance, completeness, and appropriateness of answers

Other Candidates' Performance (Anonymized): Show how the candidate compares to anonymized scores of other candidates in the same field or position.

• Metrics: Percentiles, rank positioning, or radar charts showing strengths and weaknesses.

Category	Candidate Score	Average Successful Candidate
Confidence Level	• 82%	• 75%
Answer Accuracy	• 88%	• 85%
Eye Contact	• 79%	• 80%
Body Language	• 76%	• 78%

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, October 2024

VII. CONCLUSION

In this paper the InterviewX project creates a realistic interview environment using AI to help users improve their interview skills. By analyzing answers and facial expressions, it gives feedback on both verbal and non-verbal communication. This project combines modern technologies like React, Firebase, and AI algorithms to provide a useful tool for job seekers to practice and for companies to streamline their hiring process. Overall, InterviewX shows how AI can make interview preparation and assessment more efficient and effective.

REFERENCES

[1] Xinyi Luo1, Yuyang Wang, Lik-Hang Lee, Zihan Xing, Shan Jin, Boya Dong, Yuanyi Hu, Zeming Chen, Jing Yan, Pan Hui, "Using a virtual reality interview simulator to explore factors infuencing people's behavior",28 February 2024

[2] Dr.SurendraMahajan, Prof NileshSonawane, RohitBagal,SaketKulkarni, AyushSuryawanshi, YadnyeshMhaisne, "Generative AI-BASED Interview Simulation and Performance Analysis",International Research Journal of Modernization in Engineering Technology and Science Volume:06/Issue:05/May-2024

[3] PankajRambhauPatil, ShindeRushikeshRajendra, GosaviVishakhaMahendra, BhamareBhagyashriJijabrao, PatilPareshDilip, "Elevating Performance Through AI-Driven Mock Interviews",International Journal for Research in Applied Science Engineering Technology (IJRASET)ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 12 Issue VI June 2024-Available at www.ijraset.com

[4] AmithaCaldera,SathiraHettiarachchi,Bandara H.M.R.M, Abeywickrama Y.S, "Automating Recruitment Processusing Natural ", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395- 0056 Volume: 10 Issue: 10 — Oct 2023 www.irjet.net

[5] PelinVardarlıe "A System That Allows Users to Have a Job Interview Experience ",Sustainability 2023, 15, 16031. https://doi.org/10.3390/ su152216031

[6] Aditi S. More, Samiksha S. Mobarkar, Siddhita S. Salunke, Reshma R. Chaudhari, "SMART INTERVIEWS USING AI", ISSN (PRINT): 2393- 8374, (ONLINE): 2394- 0697, VOLU- ME-9, ISSUE-1, 2022 Experience. Sustainability 2023, 15, 16031. https://doi.org/10.3390/su152216031

[7] RohanPatil , Akash Butte , SahilTemgire , VarunNanekar , Asst. Prof. ShivgangaGavhane "Real Time Mock Interview using Deep Learning",International Journal of Engineering Research Technology (IJERT) ISSN: 2278-0181 IJERTV10IS050213 Vol. 10 Issue 05, May2021