InceptoAI: AI Powered Mock Interview Platform

Anjali Deshpande, Abhilash Shinde, Shantanu Hazra, Ram Wasankar, Dr. Vikas Nandgaonkar

Computer Engineering / Indira College of Engineering and Management / Savitribai Phule Pune University, Pune, India

Corresponding Author Email: vikas.nandgaonkar@indiraicem.ac.in

anjalide464@gmail.com | ORCID: https://orcid.org/0009-0003-7355-1786

abhilashshinde4@gmail.com | ORCID: https://orcid.org/0009-0007-6757-530X

hazrashantanu051@gmail.com | ORCID: https://orcid.org/0009-0000-7972-050X

ramwasankar07@outlook.com | ORCID: https://orcid.org/0009-0006-8982-1172

ABSTRACT

InceptoAI is an intelligent, AI-powered Mock Interview platform designed to simulate and evaluate real-world interview and communication scenarios through multimodal inputs. By integrating advanced machine learning and deep learning algorithms, the system captures and processes both audio and video streams using the user's webcam and microphone. The platform evaluates parameters such as confidence, speech clarity, emotional stability, and engagement using facial expression analysis and voice sentiment recognition models. Powered by Gemini and LangChain APIs, it generates dynamic, domain-specific interview questions and evaluates the correctness and contextual accuracy of responses. Additionally, the system provides a comprehensive feedback report that breaks down strengths, weaknesses, and behavioral metrics into quantifiable scores. InceptoAI's modular architecture comprising input capture, processing, AI-based analysis, and automated report generation enables seamless performance across interview training, education, psychotherapy, customer service quality analysis, and corporate skill assessment. Through its AI-driven, real-time analysis, InceptoAI bridges the gap between human communication and computational understanding, empowering individuals to enhance their interpersonal and professional presentation skills.

Keywords: Mock Interview Platform, Multimodal AI, Facial Expression Recognition, Speech Sentiment Analysis, Behavioral Assessment, AI-Based Feedback System.

I. INTRODUCTION

InceptoAI is an artificial intelligence-driven mock interview platform designed to evaluate human communication behavior in simulated interview environments using multimodal analysis. By integrating computer vision and speech processing techniques, the system captures real-time audio and video data through a user's webcam and microphone to assess both verbal and non-verbal cues. Facial emotion recognition and voice-based sentiment analysis are employed to evaluate critical behavioral attributes such as confidence, clarity of speech, emotional stability, engagement, and response accuracy.

Unlike conventional mock interview tools that focus primarily on question-answer practice, InceptoAI combines behavioral observation with domain-specific interview question generation powered by large language models, enabling a more realistic and adaptive interview experience. The system follows modular architecture encompassing input-capture, AI-based processing, automated evaluation, and feedback visualization, culminating in a comprehensive performance report that converts subjective behavioral traits into quantifiable metrics.

The growing reliance on digital recruitment and remote assessments has exposed limitations in traditional interview training methods, particularly their dependence on human evaluators and lack of objective, data-driven feedback on emotional and behavioral aspects.

This highlights a research gap in the availability of intelligent platforms capable of simultaneously analyzing speech content, facial expressions, and emotional tone to deliver holistic interview assessments. The objective of this research is to design and implement a self-sufficient, unbiased, and scalable AI-based mock interview system that provides actionable feedback and tracks user progress over time, thereby supporting continuous improvement in communication and professional presentation skills while extending applicability to domains such as education, recruitment, counselling and corporate training.

II. LITERATURE REVIEW

Artificial intelligence has increasingly influenced recruitment, selection, and interview-related processes by enabling scalable, technology-supported evaluation mechanisms that complement traditional human judgment. Nikolaou [1] reviews the role of digital technologies in recruitment and selection, highlighting how algorithmic tools and data-driven systems improve efficiency and standardization across hiring stages, while also raising concerns regarding transparency and candidate perceptions. Although this work does not focus specifically on mock interview platforms, it establishes the broader context in which AI-assisted interview preparation systems have emerged.

Recent advances in natural language processing have significantly improved automated language understanding and generation capabilities used in interview-related applications. Transformer-based architectures such as BERT [2] enable deep contextual representation of textual responses, while large language models such as GPT-3 [3] demonstrate strong few-shot learning abilities that support adaptive and context-aware question generation. These models provide a foundation for dynamic, role-specific questioning, improving upon static or rule-based interview templates.

Despite these advancements, prior research cautions that algorithmic decision-support systems in hiring may reproduce or amplify existing biases. Raghavan et al. [4] critically examines claims surrounding bias mitigation in algorithmic hiring tools, demonstrating that many systems lack transparency and may encode demographic or occupational biases. Complementing this perspective, Mitchell et al. [5] introduce Model Cards as a framework for improving transparency, accountability, and responsible reporting of machine learning models.

Research in human resources and organizational psychology further indicates that AI-based interviewing tools can enhance recruitment efficiency when used to support, rather than replace human judgment. Langer et al. [6] review the application of AI in HR functions and identify both operational benefits and ethical challenges associated with automated assessments. Similarly, Chamorro-Premuzic et al. [7] caution against overreliance on opaque algorithmic scoring in high-stakes decision-making.

Interview performance is influenced not only by verbal responses but also by non-verbal and behavioral communication factors. Baltrušaitis et al. [8] introduce OpenFace, an open-source facial behavior analysis toolkit that enables extraction of facial action units and head pose features. Zadeh et al. [9] propose the Tensor Fusion Network, which models interactions across language, visual, and acoustic modalities. In the interview domain, Abouelenien, Burzo, and Mihalcea [10] demonstrate that combining verbal and non-verbal cues enables effective prediction of interview performance (e.g., multimodal behavior analysis).

III. METHODOLOGY

This research adopts a system-oriented experimental design focused on the development and evaluation of an AI-powered mock interview platform capable of analyzing multimodal behavioral data. The study follows a modular engineering methodology integrating machine learning, deep learning, and natural language processing techniques to assess verbal and non-verbal communication skills in simulated interview scenarios.

Data collection is performed through real-time audio and video capture using a user's webcam and microphone within a browser-based environment. The WebRTC library is used to acquire synchronized audio-video streams, which are processed locally or transmitted to a backend analytics pipeline depending on user consent. Video frames are extracted at fixed intervals and analyzed using automatic speech recognition and sentiment analysis models (Whisper and transformer-based NLP models) to evaluate speech clarity, fluency, tone and emotional stability.

Natural language understanding and interview question generation are handled using Gemini and LangChain APIs, enabling domain-adaptive and context-aware interview flows. User responses are evaluated

semantically for relevance and correctness and behaviorally using multimodal cues. Extracted features from facial expressions, vocal prosody, and textual content are aggregated and normalized to generate quantifiable performance metrics, collectively referred to as Incepto Score.

Analysis is performed programmatically using Python-based frameworks, with results compiled into automated feedback reports that highlight strengths, weaknesses, and improvement trends across sessions. No centralized database is used, instead, anonymized session metrics are stored locally using IndexedDB to ensure privacy. Ethical considerations include explicit user consent for media access, local-first data processing, absence of cloud storage by default, and complete user control over data retention, ensuring compliance with responsible AI and data privacy principles.

Fig. 1. Implementation overview of InceptoAI: Mock Interview Platform

IV. RESULTS AND DISCUSSION

The InceptoAI implementation demonstrates a practical AI-powered mock interview platform capable of real-time multimodal behavioral analysis. It integrates browser-based audio and video capture with machine learning models for facial emotion recognition, speech fluency assessment, and semantic response evaluation. Prototype testing confirmed stable real-time processing without external installations, supporting cross-browser use.

The facial analysis module effectively identified dominant emotional states, neutral, confident and stressed, providing insights beyond text-based evaluation. Speech analysis captured features like speaking rate, paused, and filler words, assessing clarity and engagement. NLP modules generated adaptive, domain-specific questions and evaluated responses in context.

Automated feedback reports combined verbal, vocal and visual cues into structured summaries, helping users recognize strengths and areas for improvement. Unlike traditional tools that focus mainly on answer correctness, InceptoAI offers a holistic assessment, incorporating non-verbal and paralinguistic factors, validating its architecture as a foundation for intelligent interview training and communication skill development.
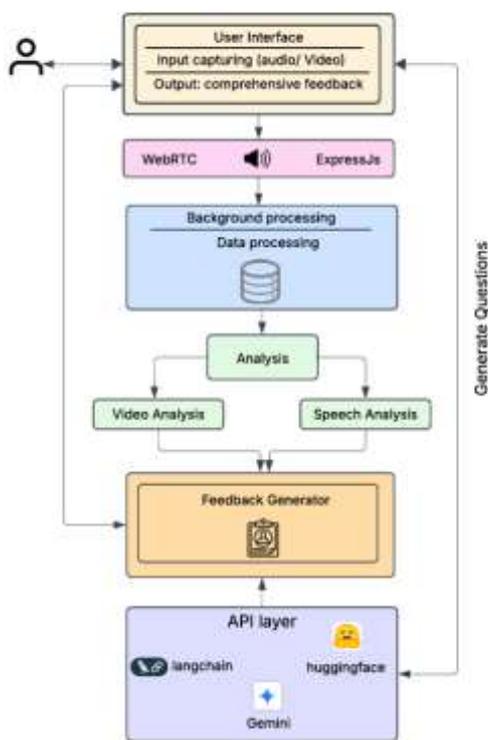
V. CONCLUSION

InceptoAI demonstrates the practical value of combining multimodal AI analysis with real-time feedback to enhance communication and interview skills. Its contribution lies in transforming subjective evaluation into objective, actionable insights, enabling users to understand and improve non-verbal, vocal, and behavioral aspects of performance. By providing a holistic, learner-focused platform, it offers a novel approach to skill development that goes beyond conventional mock interviews.

Future work can expand its impact through large-scale user studies, integration of multiple languages, advanced bias mitigation techniques, and longitudinal tracking of performance improvements. Additionally, adapting the system for specialized domains such as psychotherapy, customer service, or corporate training could further validate its versatility and practical relevance.

REFERENCES

[1] A. Nikolaou, "What is the role of technology in recruitment and selection?" International Journal of Selection and Assessment, vol. 29, no. 3–4, pp. 261–273, 2021.

[2] J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL-HLT, 2019.

[3] T. Brown et al., "Language models are few-shot learners," NeurIPS, 2020.

[4] M. Raghavan et al., "Mitigating bias in algorithmic hiring," FAccT, 2020.

[5] M. Mitchell et al., "Model Cards for model reporting," FAccT, 2019.

[6] LaMihai Burzo, Mohamed Abouelenien, Veronica Perez-Rosas, and Rada Mihalcea. 2018. Multimodal deception detection. The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition - Volume 2. Association for Computing Machinery and Morgan & Claypool, 419–453.

[7] T. Chamorro-Premuzic et al., "New talent signals," Industrial and Organizational Psychology, 2016.

[8] T. Baltrušaitis et al., "OpenFace 2.0," IEEE FG, 2018.

[9] A. Zadeh et al., "Tensor Fusion Network," EMNLP, 2017.

[10] Mihai Burzo, Mohamed Abouelenien, Veronica Perez-Rosas, and Rada Mihalcea. 2018. Multimodal deception detection. The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition - Volume 2. Association for Computing Machinery and Morgan & Claypool, 419–453.