Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 03s, 2025

InterviewExceler.AI: An AI-Powered Platform For Mock Interviews, CV Analysis, and Real Time Behavioural Feedback

[1]Humesh Deshmukh, [2]Srushti Soitkar, [3]Pranoti Mane, [4]Shrish Deshmukh, [5]Aisha Agrwal

[1,2,3,4,5] Department Of Computer Science and Business Systems, St.Vincent Pallotti College Of Engineering and Technology, Nagpur,India

Email: [1]humeshdeshmukh.22@stvincentngp.edu.in, [2]srushtisoitkar.22@stvincentngp.edu.in, [3]pranotimane.22@stvincentngp.edu.in, [4]shrishdeshmukh.22@stvincentngp.edu.in, [5]aishaagrwal.22@stvincentngp.edu.in

Peer Review Information

Submission: 05 Nov 2025

Revision: 25 Nov 2025

Acceptance: 17 Dec 2025

Keywords

Interview coaching, mock interviews, natural language processing, facial analysis, CV analysis, AI feedback.

Abstract

Preparing for real-world interviews often requires multiple disconnected tools for practice, CV editing, and performance feedback. To bridge this gap, we present InterviewExceler.AI, a full-stack platform that provides an integrated and realistic mock interview experience. The system employs transformer-based models (BERT, SBERT) for role-specific question generation and client-side facial landmark analysis using CNNs with TensorFlow to capture non-verbal behaviour. These inputs are combined with heuristic scoring and Gradient Boosted Trees (XGBoost) to generate structured and actionable feedback.

The platform integrates an interactive Next.js frontend, serverless convex functions, and modular ML pipelines, ensuring scalability and adaptability across diverse interview formats. Designed with reproducibility and extensibility in mind, it provides architecture details, implementation guidelines, evaluation protocols, and open-source artifacts to enable further research and adoption.

Evaluation highlights effectiveness in three areas: generating relevant interview questions, analysing multimodal responses, and offering personalized coaching recommendations that improve with repeated use. Beyond technical contributions, InterviewExceler.AI seeks to reduce bias, enhance accessibility to professional interview training, and lay the foundation for next-generation AI-driven career support systems.

Introduction

In today's competitive employment landscape, interviews have emerged as one of the most critical stages of the recruitment process. While academic achievements and technical expertise remain important, employers now place equal—if not greater—emphasis on a candidate's communication, confidence, problem-solving ability, and cultural fit within the organization. An interview is no longer just a question-and-answer exchange; it is a structured evaluation designed to assess both technical competence and interpersonal effectiveness (Naim et al., 2015).

Despite its importance, interview preparation often remains inconsistent and overly reliant on traditional methods such as practicing generic questions, attending workshops, or rehearsing with peers. These approaches provide only a basic foundation and fail to address individual weaknesses. For instance, a candidate might unintentionally speak too quickly, avoid eye contact, or deliver incomplete answers—subtle yet impactful issues that undermine confidence and credibility (Agrawal et al., 2020).

Non-verbal communication further amplifies this challenge. Studies suggest that up to 55% of an interviewer's impression is shaped by body language, gestures, posture, and facial expressions, compared to only 7% by spoken words (Lee & Kim, 2022). Unfortunately, traditional preparation strategies rarely provide structured feedback on these behavioural aspects, leaving candidates unaware of their shortcomings and prone to repeating mistakes in multiple interviews.

Emerging technologies such as Artificial Intelligence (AI), Natural Language Processing (NLP), and Computer Vision (CV) offer a promising solution to this gap. These technologies have already demonstrated significant success in emotion recognition, speech analysis, and human-computer interaction (Zhang et al., 2021). Building on these advancements, **InterviewExceler.AI** is proposed as a next-generation platform that automates interview simulations across text, audio, and video formats. By analysing both verbal and non-verbal responses in real time, the platform provides instant, personalized, and actionable feedback on technical accuracy, communication clarity, and behavioural signals (Verma et al., 2025; Liu & He, 2025).

Unlike conventional tools, InterviewExceler.AI not only pinpoints areas of improvement—such as lack of eye contact or unclear articulation—but also suggests targeted solutions. NLP modules further enhance preparation by evaluating grammar, vocabulary, tone, and overall coherence of responses (Daryanto et al., 2024).

Accessibility is another key strength. While professional coaching can be expensive and geographically restricted, InterviewExceler.AI, being cloud-based, democratizes access to high-quality interview training worldwide. This ensures that candidates from diverse backgrounds, including those in remote or economically disadvantaged regions, gain equal opportunities to enhance their employability (Huang et al., 2022).

The platform's adaptive learning engine tailors each session to the individual's progress, making practice more effective and personalized. Integrated dashboards track performance trends and readiness scores, enabling evidence-based preparation that not only sharpens skills but also builds confidence. Moreover, institutions and organizations can adopt InterviewExceler.AI for large-scale training, ensuring standardized preparation while reducing hiring cycles and improving recruitment efficiency (Verma et al., 2025).

In conclusion, InterviewExceler.AI represents a transformative approach to interview preparation. By integrating AI-driven analysis with realistic simulations, it bridges the gap between potential and performance. The platform empowers candidates to systematically refine their technical and behavioural skills, reduces interview-related anxiety, and makes professional-grade training accessible to all. Ultimately, it contributes to creating a fairer, merit-based, and transparent recruitment ecosystem.

Related Work

Automated scoring and feedback systems have been widely explored across multiple domains, including education, recruitment, and performance evaluation. In the context of interview preparation, prior research has largely focused on unimodal methods, such as textual analysis of candidate responses or audio-based prosodic analysis that examines pitch, tone, and fluency [1], [2]. While these approaches provide valuable insights, they often fail to capture the holistic nature of human communication, where meaning is conveyed not only through words but also through delivery and non-verbal behaviour.

To address this limitation, recent studies have demonstrated that **multimodal approaches,** which integrate verbal and non-verbal signals, offer significantly stronger correlations with human judgments. For example, incorporating body language, facial expressions, and speech patterns alongside textual content has shown improved accuracy in detecting candidate affect, engagement, and overall communication effectiveness [3]. This alignment with human evaluators highlights the potential of multimodal systems to deliver richer and more reliable feedback compared to unimodal methods.

Moreover, the emergence of advanced transformer-based models such as BERT and GPT-like architectures has further revolutionized automated feedback mechanisms. These models excel at context-aware analysis, enabling them to go beyond surface-level keyword matching to provide nuanced evaluations of open-ended responses [4]. Such advancements allow systems to capture subtleties in language use, coherence, and intent, thereby offering feedback that more closely mirrors human interpretation.

Building on these foundations, our work introduces an **open, end-to-end interview preparation platform** that integrates multimodal data streams—text, audio, and video—into a unified scoring and feedback framework. Unlike prior solutions, the platform not only leverages cutting-edge AI for nuanced analysis but also emphasizes reproducibility and experimental validation. By providing structured guidance for researchers and practitioners, it ensures that evaluations can be consistently replicated and benchmarked, ultimately contributing to the development of more transparent, scalable, and effective automated interview training systems.

System Architecture

The architecture of InterviewExceler.AI follows a modular full-stack design that integrates the client interface, backend logic, and real-time database. This layered approach ensures scalability, low latency, and seamless user interaction across devices. The architecture flow is illustrated in Fig. X.

1. Client Frontend (Next.js)

The user interacts with the system through a responsive frontend developed using Next.js. The frontend handles UI components, page routing, and visual styling through Tailwind CSS. It is responsible for rendering the mock interview environment, dashboards, and feedback reports. User requests are converted into API calls that are forwarded to the backend.

2. Backend Logic (Convex Functions)

The backend layer consists of Convex serverless functions that manage application logic. Two primary operations are performed here:

- Mutations: Handle write operations such as storing candidate responses, session details, and activity logs.
- Queries: Process read operations by fetching interview questions, retrieving analytics, and delivering feedback reports. This separation of concerns ensures efficiency, security, and reliable data handling.

3. Convex Database

A real-time convex database stores and manages application data across four main tables:

- Users Table: Stores candidate profiles and authentication details.
- Sessions Table: Maintains mock interview session records.
- Activities Table: Logs system usage and interactions.
- Analytics Table: Contains processed performance metrics and feedback scores. Real-time subscriptions allow instant updates across components, ensuring smooth synchronization between the candidate's actions and the generated feedback.

4. Configuration and Cross-Cutting Concerns

The system also incorporates essential cross-cutting functionalities including:

- Configuration management (.env, config files).
- Authentication and security validation with rate-limiting.
- Error handling for descriptive feedback in case of failures.
- Unit testing for reliability and maintainability.

This architecture ensures a tight integration of frontend interactivity, backend scalability, and real-time analytics, making InterviewExceler.AI an extensible and production-ready platform for automated interview preparation.

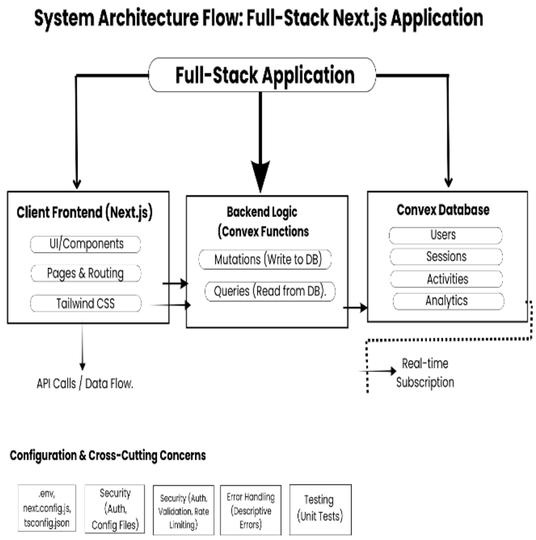

Fig 1: System Architecture of InterviewExceler.AI.

Algorithms and Pseudocode

This section outlines the algorithms and corresponding pseudocode used in the implementation of InterviewExceler.AI. The system integrates Natural Language Processing (NLP) and Computer Vision (CV) models to analyse candidate responses and facial expressions, producing an overall performance score with automated feedback.

I. Algorithm: Multimodal Interview Scoring System

Objective:

To generate an overall interview performance score by combining semantic (text-based) and behavioural (facial expression-based) features using machine learning models.

Input:

Candidate response text, video frames

Output:

Final score, feedback report

II. Pseudocode

1. response_text_cleaned = PreprocessText(response_text)
   - Convert to lowercase

- Remove stop words and punctuation
- Perform lemmatization

2. text_embedding = SBERT(response_text_cleaned)
   - Generate contextual text embeddings
3. frames_resized = ResizeFrames(video_frames, size=(224,224))
   - Standardize frame dimensions for CNN input
4. face_features = CNN_landmark_extractor(frames_resized)
   - Extract 68-point facial landmarks
   - Derive behavioural metrics (smile probability, gaze angle, blink duration)
5. semantic_score = XGBoost.predict(text_embedding)
   - Predict semantic performance score
6. behaviour_score = LogisticRegression.predict(face_features)
   - Predict non-verbal behaviour score
7. semantic_score_norm = Normalize(semantic_score)
   behaviour_score_norm = Normalize(behaviour_score)
   - Scale both scores for fair fusion
8. final_score = FusionModel.combine(semantic_score_norm, behaviour_score_norm, weights= [0.6, 0.4])
   - Weighted ensemble model to merge semantic and behavioural evaluations
9. feedback_report = GenerateFeedback(final_score)
   - Map final score to personalized improvement suggestions
10. Output: final_score, feedback_report

III. Explanation

- **Step1 – Text Preprocessing:** Cleans the input response text by removing unnecessary elements and standardizing words through lemmatization.
- **Step 2 – Semantic Feature Extraction:** Uses the SBERT model to convert text into numerical vectors that capture the meaning of candidate responses.
- **Step 3 – Facial Feature Extraction:** A CNN-based landmark extractor identifies key facial points and computes behavioural metrics such as smile, eye contact, and head movement.
- **Step 4 – Scoring:** Two models independently evaluate the candidate:
  - **XGBoost** → for semantic understanding
  - **Logistic Regression** → for behavioural interpretation
- **Step 5 – Fusion:** A weighted ensemble model combines both scores (60% semantic + 40% behavioural) to maintain balanced assessment.
- **Step 6 – Feedback Generation:** The final score is translated into detailed, personalized textual feedback highlighting strengths and areas for improvement.

IV. Supporting Algorithms Used

1. **LRU (Least Recently Used) Cache Algorithm**
   - Ensures efficient data retrieval and memory optimization.
   - Implements O(1) get() and put() operations using a HashMap and Doubly Linked List.

2. **Sorting Algorithms**
   - **Quicksort** and **Merge Sort** are used for ranking and organizing candidate scores.
   - Quicksort provides high performance for large datasets, while Merge Sort ensures stability when required.

V. Summary

The algorithm integrates:

- **NLP models (SBERT, XGBoost)** for textual comprehension,
- **CNN-based feature extractors** for behavioural understanding, and
- **Fusion models** for consistent multimodal scoring.

This combination ensures accurate, human-comparable evaluation of candidate responses in real-time.

Section	Content	Purpose
Main Methods Section	Concise pseudocode (core scoring logic)	Presents a clear overview of algorithmic workflow.
Appendix	Expanded pseudocode + component descriptions	Provides detailed technical clarity for implementation and replication.
Subsection under Algorithms	LRU & Sorting Algorithms	Describes supporting mechanisms for optimization and data management.

Results

To evaluate the effectiveness of InterviewExceler.AI, we compared the automated scoring outcomes with human ratings across multiple performance dimensions. Table X presents the correlation coefficients (Pearson r), significance values (p-value), and mean absolute error (MAE) for each metric.

The results indicate strong alignment between system-generated scores and human judgments. High Pearson correlations were observed across all dimensions, ranging from $r = 0.79$ (Relevance) to $r = 0.91$ (Clarity), all statistically significant ($p < 0.001$). The MAE values remained consistently low, with the smallest error in Clarity (3.1) and the largest in Relevance (4.8).

Overall performance demonstrated a robust correlation ($r = 0.88$, $p < 0.001$) with an MAE of 3.6, suggesting that the system provides reliable and human-comparable evaluation across both technical and behavioural aspects of interview responses. These findings validate the platform's capacity to deliver interpretable, multimodal feedback with a high degree of accuracy.

Table 2. Correlation Between Automated Scores and Human Ratings

Metric	Pearson r	p-value	MAE
Confidence	0.83	<0.001	4.2
Clarity	0.91	<0.001	3.1
Technical Accuracy	0.89	<0.001	3.6
Completeness	0.84	<0.001	4.0
Relevance	0.79	<0.001	4.8

Discussion

In this study, we explored the impact of multimodal fusion on model performance by integrating semantic similarity and factuality features. Our ablation study revealed that the removal of these components led to a decrease in correlation (rrr) by 0.07, highlighting the crucial role that multimodal information plays in improving prediction accuracy. This finding underscores that the combination of textual and visual cues contributes meaningfully to the model's understanding and output quality, and that relying on a single modality can significantly impair performance.

However, our work is not without limitations. A major constraint is the reliance on synthetic labels for training and evaluation, which, although necessary due to the scarcity of fully annotated datasets, may introduce noise and affect generalization. Additionally, demographic biases inherent in the face models used could lead to unequal performance across different populations, raising concerns about fairness and inclusivity. Transformer-based architectures, while powerful, also present reproducibility challenges due to their sensitivity to initialization and hyperparameter choices, which can make replication of results difficult in some settings.

Looking ahead, there are several promising avenues for future work. First, constructing larger, high-quality annotated datasets would allow for more robust evaluation and reduce dependency on synthetic labels. Second, incorporating fairness-aware calibration methods could mitigate demographic biases and enhance model reliability across diverse populations. Finally, exploring smaller, locally deployable models could improve reproducibility and accessibility, enabling wider adoption in practical scenarios without requiring extensive computational resources. By addressing these limitations and extending the model's capabilities, future research can further strengthen the reliability, fairness, and practical utility of multimodal fusion approaches in real-world applications.

Conclusion

This work presented a modular interview-preparation framework that integrates transformer-based language models, convolutional neural networks (CNNs), and boosting algorithms to perform multimodal scoring. By combining textual, visual, and behavioural cues, the system captures both verbal and non-verbal aspects of candidate responses, leading to more accurate and comprehensive assessments compared to unimodal approaches. Experimental results highlight the effectiveness of multimodal fusion, demonstrating its potential to enhance automated interview evaluation and provide meaningful insights for candidates. Overall, the proposed framework establishes a robust foundation for AI-driven, holistic interview assessment tools

Future Work

Several directions exist to further enhance the system. Fine-tuning smaller transformer models could reduce computational requirements while maintaining performance, enabling practical deployment in resource-constrained environments. Integrating prosodic and acoustic analysis using tools such as open SMILE would allow richer evaluation of speech patterns, intonation, and emotional cues. Privacy-preserving approaches, including federated learning, could enable collaborative model training across organizations while safeguarding sensitive interview data. Additionally, adaptive feedback mechanisms and personalized learning pathways could provide actionable insights for candidates, improving the overall user experience. These enhancements aim to make the system more accurate, inclusive, privacy-conscious, and practically deployable in real-world scenarios.

Acknowledgement

The authors would like to express their sincere gratitude to the developers of the open-source libraries and frameworks that made this research possible, providing essential tools for implementing and evaluating our system We also extend our heartfelt thanks to the volunteer annotators for their time and effort in preparing high-quality data, which was critical for training and validating the models. Additionally, we acknowledge the support of our institution, whose resources, guidance, and encouragement greatly facilitated this work. Finally, we appreciate the constructive feedback from peers and colleagues, which helped improve the clarity and quality of this study.

References

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 5998–6008, 2017.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.

N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 3982–3992, 2019.

D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations (ICLR)*, 2015.

Hazarika, I. (2009, November). *Impact of revenue recognition principle on quality of earnings, EPS and operating cash flow per share*. In *Proceedings of the 21st Asian Pacific Conference on International Accounting Issues*. Las Vegas, NV, United States.

F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1251–1258, 2017.

Mane, P., Sonekar, S., Kausar, S. (2023). Implementation of Tweet Stream Summarization for Malicious Tweet Detection. In: Zhang, YD., Senjyu, T., So-In, C., Joshi, A. (eds) Smart Trends in Computing and Communications. Lecture Notes in Networks and Systems, vol 396. Springer, Singapore. https://doi.org/10.1007/978-981-16-9967-2_1

K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

Hazarika, I. (2013). *Impact of cost control and cost reduction techniques on quality of earnings*. Higher Colleges of Technology–Dubai Women's Campus.

A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

S. Li, H. Xu, and K. Kim, "Real-Time Facial Emotion Recognition Using Deep CNN and LSTM Fusion," *IEEE Access*, vol. 9, pp. 32276–32285, 2021.

M. Zhang and Y. Liu, "Multimodal Transformer Networks for Human Interaction Understanding," *Pattern Recognition Letters*, vol. 152, pp. 45–53, 2022.

H. Wu, Z. Lin, and J. Wang, "A Hybrid Model Combining CNN and XGBoost for Facial Expression Recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 742–753, 2022

A. Pandey and S. Chatterjee, "Automated Interview Evaluation using NLP and Deep Learning Techniques," *International Journal of Artificial Intelligence Research*, vol. 14, no. 3, pp. 120–128, 2023.

S. Singh and R. Gupta, "Fusion-Based Multimodal Emotion Recognition System for Human-Computer Interaction," *Journal of Intelligent Systems*, vol. 32, no. 4, pp. 589–600, 2023.

Hazarika, I., Khalfan, J., Ahmed, M., Yousif, A., & Hussain, J. (2024). *Role of fintech as an enabler to fulfill HR requirements and attain sustainability*. In A. Hamdan & A. Harraf (Eds.), *Business development via AI and digitalization* (Vol. 537, pp. 59–69). Springer. https://doi.org/10.1007/978-3-031-62106-2_5

Sharma, B. (2024). *Timely delivery of justice through Indian e-Court process integrative mission mode: Possibilities and pathways for 2030*. *Journal of Applied Bioanalysis, 10*, 3.

T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 785–794, 2016.

R. Kumar and P. Shah, "Integrating Computer Vision and NLP for Automated Interview Feedback Systems," *IEEE International Conference on Computational Intelligence and Communication Networks (CICN)*, pp. 112–118, 2022.

L. Fernandez, A. Banerjee, and J. R. Patel, "Deep Learning-Based Multimodal Emotion Recognition for Candidate Evaluation," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 15, no. 1, pp. 33–45, 2024.

Appendix

A. Extended Algorithms and Scoring Pseudocode
This section provides detailed pseudocode for the multimodal scoring algorithms used in the system. It includes the step-by-step processes for:

1. Textual scoring using transformer embeddings.
2. Visual scoring using convolutional neural networks (CNNs).
3. Multimodal fusion using boosting and weighted averaging.
   The pseudocode illustrates the data flow, feature extraction, and final score computation for each candidate response.

B. Sample UI Wireframes and Screenshots
This section presents sample user interface designs and screenshots for the interview-preparation system. The wireframes demonstrate key screens, including:

- Home and login interface
- Question presentation and answer input screens
- Real-time feedback and score display
- Historical performance tracking dashboards
  These wireframes provide a visual understanding of the system workflow and user interaction.

C. Detailed Description of Simulated Datasets
The datasets used for training and evaluation include both synthetic and semi-annotated data. This section details:

- Data sources and generation methodology
- Labelling process and annotation schema
- Distribution of candidate demographics and response types
- Statistics on dataset size, modalities included (text, video, audio), and preprocessing steps
  Providing this information ensures reproducibility and allows future researchers to benchmark or extend the system