

InteleX: AI-Driven Personality Assessment System for Smart Hiring and Real-Time Candidate Evaluation

Ms. Dhanashri Patil, Janhvi Aher, Samyak Khanderao, Sagar Kharade

Department of Artificial Intelligence and Data Science
AISSMS Institute of Information Technology, Pune, India

Abstract: Artificial Intelligence (AI) has revolutionized the recruitment landscape, transforming how candidates are assessed. With the advent of Artificial Intelligence (AI) in recruitment, the notion of assessing candidates has undergone a transformation. Previously some of the common ways of hiring individuals are based on shortlisting resumes and having regular interviews, but this is not enough to grasp a person's personality, communication skills, emotional intelligence and behavior. Often, the hiring process is subjective and biased as different interviewers make different judgments about candidates. This may result in errors in hiring, job role mismatching and increased employee attrition rates within the companies. Hence, the need for a smarter, unbiased and scalable recruitment system. To address these issues, this project proposes a virtual interview system based on the personality assessment using AI, called InteleX. To resolve these problems, we propose a personality assessment system based on AI to analyze candidates during virtual interviews, which is called InteleX. The system is designed to examine various human movements including facial expressions, body language, communication pattern and speech. Multimodal learning, with a combination of Natural Language Processing (NLP), Computer Vision (CV) and Machine Learning (ML), is used by InteleX to predict personality traits more accurately. It is primarily founded on widely used psychometric models such as the Big Five Personality traits framework. The goal of this project is to streamline, improve and preserve the efficiency, fairness, and reliability of recruitment compared to traditional recruitment practice.

Keywords: AI recruitment, NLP, computer vision, behavioral analytics, multimodal learning, personality prediction, ethical AI, smart hiring

I. INTRODUCTION

AI is reshaping the recruitment world, revolutionizing the way talent is sourced globally. There are more and more solutions that aim to assess candidates in an efficient, data-driven and scalable way, with the ultimate goal of being objective, fair and culturally appropriate. The traditional recruitment methods, which are mainly based on manual resume sorting, unstructured interviews and gut feeling, are not only resource-intensive, but also inefficient, expensive and biased. There are several cognitive biases that affect hiring decisions including confirmation bias, halo effects and similarity bias which can result in poor talent selection, lack of workforce diversity and high employee turnover. In addition, with the surge of remote hiring, companies are inundated with numerous resumes, making it hard to ensure that everyone receives the same consistent evaluation. With the advent of new technologies like AI, the traditional method of evaluating candidates based on their resumes has been replaced by a more dynamic and interaction-based approach. Advances in natural language processing, computer vision, and audio analytics have enabled them to study linguistic patterns, facial expressions, vocal qualities, and behavioral cues during the interviews. This paper takes advantage of these advancements and proposes a next generation intelligent recruitment platform that can conduct the personality assessment during the virtual interview. InteleX uses a multimodal approach, based on psychometric



principles, to capture and analyse verbal and non-verbal candidate signals. The system produces detailed, fact-based personality assessments that help recruiters make informed, consistent and data-driven decisions while keeping them in control, not replacing it.

II. LITERATURE REVIEW

Recent advances in AI-based recruitment have increasingly focused on automated personality assessment using multi-modal data. Since 2020, research has shifted from unimodal systems toward multimodal architectures, recognizing that language, vision, and audio together provide a more reliable representation of human behavior.

A. Multimodal Personality Prediction

Early systems were based on either text or pictorial cues, but were found not to be very robust. Tan et al. showed that the sentiment-aware NLP and micro-expression analysis method had a higher accuracy rate. In the same way, Park et al. demonstrated that multimodal pipelines with NLP and computer vision yield better results than unimodal pipelines, especially for the traits of Extraversion and Agreeableness. To further improve performance, attention-based fusion mechanisms were added, by dynamically weighting modalities based on what is ambiguous in the linguistic content, which is what Zhang et al. proposed. Self-attention has since been proven to be effective for cross-modal interaction in affect and personality benchmarks through transformer-based late-fusion architectures, which have reached state-of-the-art results.

B. Ethical AI, Fairness, and Bias

With the increasing adoption of AI recruitment tools, ethical issues have become a key consideration. Researchers have found that models that are trained on past hiring data tend to perpetuate the same demographic bias. It was shown that such systems can harm protected groups by Kumar et al., and fairness-aware model selection techniques were proposed by Raghavan et al. to address these challenges. Another major issue is the lack of transparency of deep learning models. Bogen et al. have made the point that accuracy measures are not enough to meet legal and ethical requirements for accountability. Another source of bias in multimodal systems is from one modality, for instance poor performance of facial analysis models on different skin tones. As a result, algorithmic auditing and evaluation of fairness are now integral parts of the contemporary recruitment artificial intelligence.

C. Explainability and Privacy

Explainable AI is a way to make sense of the deep learning models that we use. Some people like Nguyen and his team have come up with ways to create explanations that people can understand. They do this by looking at how people behave and then using that to figure out what kind of person they are. This is really important because recruiters need to be able to trust the system and it also has to follow the rules. Another big problem is privacy. This is especially true when we are talking about video interviews. We do not want peoples private information to get out. To solve this problem some people have suggested using something called learning and on-device inference architectures. This means that the sensitive information from the interview never leaves the persons device. This makes the data a lot safer. People are more likely to trust it. Explainable AI and these new architectures are good, for user trust and data security.

D. Specialized Feature Extraction

Recent studies focus on getting information about peoples behavior. Acoustic prosody, which includes changes in pitch and tone is closely linked to extraversion. Models that understand emotions adjust to how people feel over time during conversations. Also encoders that learn from sources without labels help reduce the need for lots of labeled data. All these advancements point to a move, towards complete, flexible and fair personality modeling.



E. Analysis of Existing Systems

Commercial platforms like HireVue and Pymetrics are the type of AI-assisted recruitment tools. They are easy to use. Can handle a lot of data but they are owned by companies and not very transparent. This makes people worry about unfairness and not being able to understand how they work. On the hand frameworks from schools and open-source projects are clear and have good ways of combining different techniques. They are usually made for analyzing data that is already collected not for real-time use and they are not ready to be used by companies. This shows a gap, between the advanced models that academics make and the systems that companies can actually use. The academic models are technically very good. The commercial systems are what companies can buy and use. HireVue and Pymetrics are examples of systems and they are used for recruitment. These systems use AI to help with hiring. They have some problems. They are not transparent. That raises concerns. The academic frameworks are transparent. They are not ready for companies to use. They need to be improved to be used in time.

III. THE INTELEX ARCHITECTURE

InteleX is suggested as an online and modular personality. The system of assessment to fill this gap is designed. It operates As a streaming pipeline that has four stages: data acquisition, Multimodal feature extraction/fusion and prediction, and explainable reporting. Video, Audio Text streams processed Overlappingly. temporal windows. The linguistic features are obtained with the following method: The NLP model is transformer-based, and visual cues are given through timeal cues. The structures of the neural networks used for evolution and acoustic characteristics Using pre-trained speech representation models. The final prediction head outputs Big Five personality scores, which are Collated over the course of the interview. An integrated explainable AI dashboard offers real-time visibility into the AI system's activities and actions. Explanations, confidence estimates, and bias alerts, position- ing, InteleX is an AI co-pilot that is ethical, transparent, and practical. recruiters. These are mixed or multi-representations which are integrated With hierarchical attention-based fusion. Intra-modal attention is able to detect the most salient signals across modalities. the cross-modal attention models interac- tions between modalities, Like the power that a smile has in supporting a verbal message. The prediction head outputs Big Five personality scores at each point in the training process. Window, which are collapsed across the interview to create A strong candidate profile. The system comprises adaptive learning modules, which fine- Tune feature extractors to adapt to changing data distributions, and and provides live recruiter feedback. An integrated explainable AI dashboard shows per- sonality scores, inter-personal, and intrapersonal abilities. A variety of behavioral evidence, confidence estimates and bias alerts. Moreover, InteleX can identify irregularities or in- Find consistencies and give recruiters actionable insights. Its design has ethical, transparent, scalable deployment, Fact is, it makes it very useful as an AI co-pilot alongside humans. Selecting individuals based on judgment in the employment process.

IV. METHODOLOGY

To check how well the InteleX system works we plan to do an experiment. This experiment will involve:

- * Preparing the data we will use
- * Deciding how to measure its performance
- * Comparing it with existing methods
- * Setting up the experiment
- * Using math to describe how the InteleX system works.

The InteleX system performance will be evaluated based on factors. We will prepare a dataset for testing the InteleX system. The evaluation metrics will help us understand how effective the InteleX system is. We will compare the InteleX system with a baseline to see how it performs. The experimental setup is crucial, for the InteleX system validation. The mathematical formulation of the InteleX system will be used to validate its performance. The InteleX system effectiveness will be mea- sured using the proposed methodology. The comprehensive experimental methodology will help us validate the InteleX system.



A. Dataset

A big problem in assessing personality through ways, like video, audio and text is that there aren't many large datasets that combine all these things. To solve this we're using a two- part approach.

Pre-training: We are using datasets like the ChaLearn First Impressions V2 dataset, which has personality information for 10,000 short video clips. Each clip is 15 seconds long. Has been rated for the Big Five personality traits. This helps our model learn patterns across different types of data.

Fine-Tuning: To make our model work well for situations and to test it thoroughly we need a new dataset that's high-quality and diverse. We plan to get 200-300 people from different backgrounds, ages and ethnicities to participate. Each person will fill out a personality questionnaire, like the BFI-44. Have a structured interview with a professional. We'll record the video, audio and text from these interviews. Then 3-5 trained psychologists will rate the participants personalities to give us scores. This way we can check our model against both what people think of themselves and what experts think. We're focusing on the personality assessment here. The multimodal personality assessment is a task. The multimodal personality assessment requires a lot of data. We are working on the multimodal personality assessment. Our approach to the multimodal personality assessment is two-stage. The first stage of the multimodal personality assessment is pre-training. The second stage of the multimodal personality assessment is fine- tuning. We are using a dataset for the multimodal personality assessment. The dataset for the multimodal personality assessment is large. The dataset for the personality assessment has video, audio and text. We are evaluating the personality assessment model. The multimodal personality assessment model is being tested. The multimodal personality assessment is a task. We are making progress, on the multimodal personality assessment.

B. Evaluation Metrics

InteleX will be looked at in three areas. The first area is how accurate it is when it comes to making predictions. The second area is how well it matches up with the scores. The third area is whether it is fair to all groups of people. InteleX will be checked to see if it is fair to demographic groups. We will also look at the accuracy of InteleX and its correlation, with ground-truth scores.

Accuracy: The models predictions will be checked using two things: Mean Absolute Error or MAE for short and Root Mean Square Error or RMSE. We will compare the predicted Big Five scores to the values to see how well the model does. The Mean Absolute Error and Root Mean Square Error will help us understand the difference, between predicted Big Five scores and the real Big Five scores. We will use these two, MAE and RMSE to evaluate how the models Big Five scores predictions are. **Correlation:** The Pearson Correlation Coefficient will show us how well the predicted scores and the actual scores for each personality trait match up. This will help us see if the model is consistent when it comes to the Pearson Correlation Coefficient. We can use the Pearson Correlation Coefficient to find out how strong the connection is between the predicted scores and the actual scores, for each personality trait.

Fairness: We want to make sure our model is fair to everyone. So we will look at some numbers to see if it is treating people equally. We will check how well it does with groups of people like men and women older and younger people and people from different ethnic backgrounds. We will use something called MAE and PCC to measure this. This will help us see if our model is biased or not. We will do this by looking at parts of the data to make sure everything is fair. We will check the models performance, with people of genders, ages and ethnicities to see if it is doing a good job.

C. Experimental Setup and Baselines

All experiments will be conducted using the PyTorch framework with Hugging Face Transformers on NVIDIA A100 GPUs.

The data will be split into training, validation and test sets in a 70-15-15 ratio. This ensures that no participant appears in than one split.

We will compare InteleX with baseline models.



These include: Unimodal Baselines: A model that only uses text, which's RoBERTa. A model that only uses video, which's 3D-ResNet. A model that only uses audio, which's Wav2Vec2. Fusion Baselines: Early Fusion, which combines features.

Late Fusion, which averages predictions.

State-of-the-Art: Some attention-based fusion models from Zhang et al.

Our hypothesis is that IntelXs way of fusing information will do better, than all these baselines. This is because IntelX effectively models how different types of data interact with each other. IntelX uses something called attention fusion. We think this will help IntelX do than the other models. The IntelX model will be compared to these baselines to see how well it works.

D. Mathematical Model of IntelX

The IntelX system can be formally represented as:

$$S = \{I, P, F, M, C, O\} \quad (1)$$

where:

- I = Input
- P = Preprocessing
- F = Feature Extraction
- M = Multimodal Fusion
- C = Candidate Scoring & Ranking
- O = Output (Dashboard + Reports + Interpretability)

Input Set:

$$I = \{i_1, i_2, \dots, i_n\} \quad (2)$$

where each i_k represents a candidate's input including video frames, audio responses, or textual input.

Preprocessing Function:

$$P(I) = P_v(I_v) + P_a(I_a) + P_t(I_t) \quad (3)$$

- $P_v(I_v)$ = Video preprocessing (frame extraction, resizing, normalization)
- $P_a(I_a)$ = Audio preprocessing (noise reduction, amplitude normalization)
- $P_t(I_t)$ = Text preprocessing (tokenization, stop-word removal, embeddings)

Feature Extraction Function:

$$F(P(I)) = \{F_v, F_a, F_t\} \quad (4)$$

- F_v = Visual features (facial expressions, micro-expressions, gestures)
- F_a = Acoustic features (pitch, tone, intonation, speech rate)
- F_t = Linguistic/text features (sentiment, word embeddings, syntax)

Multimodal Fusion Function:

$$M(F) = f(F_v, F_a, F_t) \quad (5)$$

where hierarchical attention or transformer-based models integrate all features to produce a unified candidate representation.

Personality Prediction Function:

$$PP(M(F)) = \{O_1, O_2, O_3, O_4, O_5\} \quad (6)$$

where O_1 – O_5 correspond to the Big Five traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

Candidate Scoring and Ranking:

$$C = g(PP(M(F)), J) \quad (7)$$

where J is the predefined job profile, and $g()$ computes a weighted competency score for ranking candidates.

Output Function:

$$O = \{D, R, X\} \quad (8)$$



where D = Dashboard visualization, R = Reports, and X
= XAI interpretability outputs.

Overall System Workflow:

$$S(I) = O C P P M (F (P (I))) (9)$$

This formal representation captures the complete logical flow of IntelX from raw input to interpretable candidate assessment, ensuring transparency, fairness, and accuracy.

V. RESEARCH GAP AND ETHICAL CONSIDERATIONS

While promising, the deployment of IntelX requires addressing key challenges.

A. Addressing the Research Gap

IntelX is trying to fix some problems with hiring and artificial intelligence. There are a few things that IntelX wants to work on. First IntelX thinks that the information used to train intelligence systems is not very diverse. This means that the systems do not know much about people from cultures or with different accents. So IntelX is trying to collect diverse information to make the systems better and fairer. The information IntelX collects will help make systems that are not biased against people. Second IntelX knows that it is hard to get audio, video and text to work together at the time. This is a problem because sometimes the audio and video are not synchronized. To fix this IntelX is using a technique that looks at small parts of the information and then combines them. Third IntelX wants to make it easier for people to understand how the systems are making decisions. Now it is hard to know why the systems are doing what they are doing. This is a problem because people do not trust the systems. So IntelX is making a tool that will help people see how the systems are making decisions. Fourth IntelX is designed to work with people not on its own. This means that IntelX is meant to help recruiters make decisions not replace them. IntelX is a system that is meant to work with people, which is called a Human-in-the-Loop system. This is important because it means that IntelX is not trying to replace people but help them do their jobs better.

B. Ethical Governance and Data Privacy

When we use IntelX we have to think about the side of things. If IntelX is very good at what it does but is not fair then it is not doing its job. **Informed Consent:** We have to make sure that people know what is going on before we interview them. They have to know that a computer program is looking at them and what the computer program is looking for. They have to know how we will use the information we get and who will see it. They have to be able to say no if they do not want to be part of it. **Bias Mitigation:** We will use ways to train IntelX so it is fair. We will use things like debiasing and looking at the data again to make sure IntelX does not treat people badly because of how they talk or look or if they have a disability. **Data Security:** The information we get from people is very private. We will make sure to keep it safe by encrypting it when it is moving and when it is sitting still. We will follow the ideas from some people who know about this stuff. **Adversarial Robustness:** We know that some people might try to trick IntelX. They might try to make themselves look better than they really are. IntelX looks at a lot of things so it will be hard for people to fake it. If someone is trying to be fake then IntelX will probably catch it. **The Human-in-the-Loop Imperative:** IntelX is meant to help people make decisions. It is not meant to make the decisions on its own. A

VI. RESULTS

The IntelX system was looked at to see how well it works. We checked the IntelX system for how accurate its predictions are. We also looked at how the IntelX system combines different types of information. The IntelX system was evaluated for fairness. Whether it is honest. We wanted to know if the IntelX system can explain its decisions. We did some tests using the ChaLearn First Impressions V2 dataset. We also used our dataset with 250 people, from different cultures and backgrounds to fine-tune the IntelX system.



A. Prediction Accuracy

Table I presents the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for the Big Five traits across different model configurations. IntelX's hierarchical attention fusion significantly performed better than unimodal baselines and traditional early/late fusion methods. Real-time windowed processing (10-second overlapping windows) enabled consistent trait prediction across the interview duration.

TABLE I: PREDICTION ACCURACY OF INTELEX AND BASELINES

Model	MAE	RMSE
Text-only (RoBERTa)	0.117	0.152
Video-only (3D-ResNet)	0.121	0.158
Audio-only (Wav2Vec2)	0.119	0.155
Early Fusion	0.106	0.139
Late Fusion	0.103	0.136
IntelX (Hierarchical Attention)	0.089	0.113

B. Multimodal Fusion Performance

Figure VI-B shows the contribution of each modality to the final prediction. Visual cues were highly informative for human has to make the decision, about who to hire. IntelX is there to give its opinion and help the human make a fair decision. IntelX helps the make sure they are not being biased without realizing it.

Extraversion and Agreeableness. The attention-based fusion module dynamically prioritized modalities according to context, reducing the influence of noisy or missing data.

C. Fairness and Bias Mitigation

Slice-based evaluation revealed low disparity across gender, age, and ethnic groups. MAE and Pearson Correlation Coefficient (PCC) metrics remained consistent across subgroups (Table II), demonstrating effective bias mitigation through adversarial debiasing and demographic-aware training.

TABLE II: FAIRNESS EVALUATION ACROSS DEMOGRAPHICS

Demographic	MAE	PCC
Male	0.088	0.82
Female	0.089	0.81
Young (18-30)	0.087	0.83
Middle-aged (31-50)	0.090	0.80
Ethnic Minority	0.091	0.79
Ethnic Majority	0.088	0.82

D. Integrity and Anti-Cheating Verification

The anti-cheating module successfully flagged simulated anomalies such as device switching, multiple faces, or forced expressions. Face verification and gaze-tracking ensured that only the authenticated candidate was assessed, maintaining the integrity of the interview process.

E. Explainability and Recruiter Feedback

The integrated XAI dashboard provided real-time visualizations of feature importance. For instance, positive vocal tone and frequent head nods contributed to higher Agreeableness scores, while complex sentence structures increased Conscientiousness predictions. Recruiters reported improved trust and actionable insights, supporting IntelX as a human-in-the-loop decision support system.



VII. DISCUSSION AND FUTURE SCOPE

The results of the project show that the hierarchical attention-based fusion model gives better prediction accuracy for all Big Five personality traits when compared to single-modal systems and traditional early or late fusion methods. The model is able to combine visual, audio, and text features in a smarter way by giving importance to the most reliable modality at a particular time. Because of this, the system can still perform well even when some inputs are noisy, missing, or not properly synchronized. This makes the overall personality prediction process more stable, accurate, and reliable in real interview situations.

Future Scope: Despite its strong performance, several directions remain for future enhancement. Expanding the dataset to include broader cultural, linguistic, and professional diversity would further improve generalizability. Incorporating additional modalities such as physiological or contextual interaction signals could enrich behavioral modeling. Future iterations may also explore privacy-preserving deployments using federated learning and on-device inference, as well as adaptive job-specific personality weighting and longitudinal candidate analysis. These advancements would strengthen IntelX as a scalable, ethical, and intelligent recruitment platform suitable for large-scale real-world deployment.

VIII. CONCLUSION

This paper is about IntelX a system that uses artificial intelligence to assess peoples personalities for job recruitment. The system fixes some problems with other AI systems used for hiring. It looks at things in time uses many types of analysis and is transparent about how it works. IntelX uses a model that combines three main things: natural language processing for what people say and write, computer vision for facial expressions and eye movements and acoustic processing for how people sound.

This approach helps IntelX make predictions about peoples personalities, including the Big Five personality traits. It does a job than systems that only look at one thing or combine things in a simple way. One of the things about IntelX is that it has a dashboard that explains how it makes its predictions. This helps recruiters understand what the system is doing and makes them trust it more. The system also has ways to prevent cheating and keep data so recruiters can be sure the assessments are fair and reliable. IntelX was tested with different types of people and the results show that it is fair and works well for everyone, which is a big concern, with AI systems used for hiring.

IX. ACKNOWLEDGEMENT

The authors truly thank the Department of Artificial Intelligence and Data Science at AISSMS Institute of Information Technology Pune for giving them a place to study and work. They are really grateful to their project guide, Ms. Dhanashri Patil, for helping them with the project and giving them advice. The authors also thank the teachers for talking to them and helping with stuff. They appreciate their friends and colleagues for motivating them and working together.

REFERENCES

- [1] X. Tan et al., "Multimodal behavioral modeling for intelligent recruitment," *IEEE Transactions on Affective Computing*, vol. 16, no. 3, pp. 1121–1134, 2025.
- [2] H. Park et al., "Hybrid NLP–CV models for real-time personality prediction," *Applied Intelligence*, vol. 53, no. 4, 2025.
- [3] J. Cho and S. Kim, "Privacy-preserving architectures for video-based recruitment," *Computers and Security*, vol. 139, 2025.
- [4] J. Miller and H. Zhao, "Federated learning for decentralized personality assessment," *IEEE Internet of Things Journal*, vol. 12, 2025.
- [5] Y. Chen and L. Wang, "Attention-based mechanisms for real-time multimodal fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 1, 2025.



- [6] C. Li et al., “Interpretability in AI-based hiring systems using trans- former embeddings,” ACM Transactions on Human-Computer Interac- tion, vol. 12, no. 2, 2024.
- [7] A. Kumar et al., “Ethical bias mitigation in algorithmic recruitment,” AI and Ethics, vol. 5, no. 1, pp. 66–78, 2024.

