International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Autonomous interview intelligence engine

P. Mahaboob Chand[1], S.MD. Mushraf Hussain[2], S.MD. Jabiullah[3], S. Fayaz[4], S.MD. Thahir[5], K. Ravi Teja[6]

[1]Assistant Professor, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[2]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[3]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[4]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[5]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
[6]UG Student, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
Corresponding Author E-mail: ravitejakanaguduru@gmail.com

Abstract

The growing need to recruit large quantities of candidates without biases has led to the rapid use of intelligent automation in screening candidates. In this paper, I will introduce an Autonomous Intelligence Interview Engine that is a real-time interview analysis and evaluation system that will help in standardising initial recruiting. The system allows registering of candidates, which is administratively approved and authenticated using two factors so as to control access. Interviews are held in built schedules whereby structured questions are asked one by one with time restrictions. This is done by recording candidates with the automated voice recorders and later transcribing them with speech recognition technology. Natural Language Processing (NLP) techniques are applied to the textual responses to measure their relevance to the keywords, semantic correspondence with the anticipated responses, fluency, and sentimental attributes. These parameters are combined into a weighted scoring system producing quantitative scores on performance as well as qualitative feedback. Administrative dashboard allows to track the status of examinations, the participation of the candidates and the management of the feedback. The proposed system is implemented with Java-based technologies with embedded speech-to-text and NLP modules, which provide the system with real-time processing, security enforcement, and scalability. The site is convenient in conducting academic mock interviews, preparation of campus recruits, and screening of possible hires at the initial stage, which offers a well-organized and unbiased assessment system.

Keywords: Interview, Natural Language Processing(NLP), Candidates, Exam

1. Introduction

The high pace of digitalization of the recruitment procedure has hastened the implementation of smart systems to conduct a preliminary screening and assessment of candidates. Using automated instruments to enhance efficiency, reduce the amount of manual labor, and lessen subjectivity in assessing interviews have been adopted by organizations. Face-to-face interviews although effective in assessing interpersonal skills are usually deficient in consistency in questioning patterns, rated by evaluators, time and cost and reachability. Such

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

difficulties have prompted the creation of AI-based interview systems capable of standardizing the process of evaluations and preserving the structured evaluation criteria.

Recent studies have examined adaptive interview mechanism which can adjust interaction strategies depending on the behavior of the user. Nagasawa et al. [1] revealed that multimodal nonverbal cues can be used to recognize the speaking willingness of a candidate, which can be applied to improve questioning strategies and increase engagement. In a similar vein, VR based-interview systems suggested by Artiran et al. [2], [3] have examined the gaze behavior and head orientation to examine how social modulation is affected during simulated interviews. Even though these systems help in understanding behavioral dynamics, they mainly involve training and behavioral analysis and not secure and scalable automated evaluation.

Automated assessment based on language has also received interest as transformer architecture has developed. In their study, Stoev et al. [4] used the BERT-based models to classify transcripts of structured interviews, revealing that the linguistic patterns can be effectively captured by using the contextual embeddings. Simultaneously, speech processing and large language models were used to establish simulators of interviews with avatars and interviews with gamers proposed by Mongkoljaturong et al. [5] to provide participants with a sense of immersion. Although these methods make the experiences more realistic and more prepared the candidates, it does not meet the criteria of standardized scoring, guaranteed scheduling and centralized administrative control envisaged in real-world recruitment systems.

There has been also a concern on fairness and bias due to increasing use of automated interview assessment. Kim et al. [6] and Putra et al. [7] have suggested multimodal learning models based on fairness to address demographic differences in predicting interviews through videos. Though such approaches play an important role in the ethical use of AI, they rely on the use of multimodal inputs that are based on video, which adds to the complexity of their computations and infrastructure needs. In addition, the privacy issues relating to visual data processing are also one of the important questions.

Although there has been significant improvement in adaptive interaction, multimodal assessment and immersive training environments, there is still the need to have a lightweight, secure and structured interview examination platform which incorporates scheduled assessment, controlled access, automated voice recording, speech-to-text transcription, natural language analysis, and objective weighted scoring into a single framework. To fill this gap, the design proposed the Autonomous Intelligence Interview Engine, which is a voice-assisted assessment system with secure authentication and administrative monitoring. Focusing on reducing bias and the evaluation of responses with the help of NLP and structured questioning, the system is intended to offer scalable, objective and real-time candidate evaluation that would be applicable to academic mock interviews, campus recruitment preparation and initial screening applications.

2. Literature Review

2.1 Existing Systems

The recent progress of intelligent interview systems is a multidisciplinary contribution involving artificial intelligence, social signal processing, virtual reality, and natural language understanding to improve the process of interview evaluation and training. Nagasawa et al. [1] designed an adaptive interview robot to estimate willingness of a speaker based on multitask nonverbal cue, and dynamically selects questions. They make the findings that adaptive strategies may boost high-willingness responses, yet recognition accuracy is mediocre and relies on behavioral sensing reliability. Artiran et al. [2] compared gaze modulation in virtual reality based mock interviews, where there were observed differences between neurotypical and neurodivergent participants that were quantifiable. Artiran et al. [3] also studied gaze and head orientation patterns in triadic VR interviews in a follow-up study and revealed how conversational roles affect the distribution of attention. Although such systems can give useful behavioral feedback and training features, they are more concerned with skills training as opposed to systematic automated feedback.

Interpretation Interpretation of interviews has also been studied using language-model-based methods. Stoev et al. [4] studied the automated Adult Attachment Interview classification with BERT based embeddings and conventional linguistic characteristics. They found that transformer representations can be used to attain competitive classification; the study is domain-specific and does not set the context of real-time recruitment screening. In the same way, Mongkoljaturong et al. [5] developed an AI-based MetaHuman interviewer with speech processing and large language models to develop immersive serious-game interview training. Even though the system is more realistic and engaging to the user, the system encourages experiential learning more than standardized scoring systems.

The issues of impartiality and prejudice in automated interview systems are also considered in the area of multimodal learning. Kim et al. [6] recommended fairness-sensitive learning models to video-based interview evaluation with regularization approaches to balance forecasting and demographic fairness. This direction was further pursued by Putra et al. [7], who incorporated adversarial reweighting techniques into multimodal transformer structures to enhance fairness without sensitive attribute considerations. Although these methods have a major impact on the application of ethical AI in the recruitment process, they are based on the fact that multimodal inputs (videos) play a key role, which complicates computational costs and system structure.

According to Hosseini et al. [8], the researchers examined the avatar-based feedback in the interview training setting and found that there were enhancements in confidence levels of the participants and a decrease in anxiety levels. Despite being a useful system in terms of coaching, the system does not introduce an autonomous assessment channel with secure scheduling and administrative oversight. In these papers, the existing solutions target either adaptive behavioral interaction, immersive simulation, fairness-aware multimodal prediction or psychological classification. Nonetheless, few of them feature a unified system of secure user authentication, managed interview scheduling, automated voice-recorded interviews, speech-to-text conversion, structured natural language assessment, weighted feedback and centralized monitoring in a scalable architecture. In general, although the studies above show that there is great advancement in the automation and evaluation of interviews conducted by voice, the lack of a secure voice-driven autonomous interview examination system that does not depend on intensive use of expensive multimodal sensing networks has been found.

Current automated interviewing systems include adaptive robot systems, virtual reality platforms, language-model-based transcription analysis systems and multimodal video evaluation systems. Interpretative interview robots use nonverbal behavioral cues to estimate the internal state of the candidate, e.g. the willingness to speak, and dynamically control the choice of questions to help maintain interest in the interaction. Although this enhances flow of interaction, the accuracy of behavioral inference is moderate and such systems are very reliant on specialized sensing devices and sophisticated signal processing which means that they do not scale up well when dealing with large numbers of applicants. Interview platforms that are made with virtual reality are more concerned with training and less with evaluation. These systems examine gaze patterns, head direction, and attention changes in simulated interviews to promote the skill development in communication. Despite delivering quantifiable information on the patterns of social interaction, and being handy in confidence building, they need immersive hardware implementations and are not optimized on secure and standardized scoring of candidates. They usually aim at development rather than organized recruitment screening.

Interview transcript classification has also been done using transformer based language models. These methods use contextual embeddings to classify interview responses to predefined classes, usually in the field of psychological or behavioral assessment. Although they show high linguistic modeling, they are typically used in the context of post-interview transcript data as opposed to live-time implementation of interviews, complete authentication, schedule control, and administrative audits. Multimodal automated video interview evaluation systems are systems that strive to forecast hiring advice based on textual, vocal, and facial attributes. These models have been equipped with fairness-conscious learning methods and adversarial training approaches to deal with the bias issue. Though they help to enhance metrics of equity and predictive performance, they raise the demands of the computational load and bring up considerations of privacy, as they process visual data. It is also not very easy to deploy in resource constrained settings due to the reliance on high dimensional multimodal inputs.

Serious-game interview simulators based on avatars are more realistic, as they incorporate speech recognition and text to speech synthesis and the use of large language models to produce interactive digital interviewers. These systems also have personalized feedback and immersive experiences and enhance user preparedness. Nevertheless, they are more training instruments and do not generally impose any sort of organized arrangement, safeguarded test administration, or standardized weighted scoring systems demanded in legitimate recruitment screening. In general, the available systems focus on the adaptive behavioral interaction, immersive training, multimodal predictive modeling, or fairness-aware assessment. Not many offer secure user authentication, scheduled interviews, voice capture, real-time speech-to-text conversion, structured natural language assessment, weighted scoring and central administrative control in a single and scalable system. This loophole would necessitate the development of a simplistic, voice-activated autonomous interview assessment system that can provide the objective and standardized results, which do not require the multimodal systems, which are complex to implement.

2.2 Proposed System

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The proposed system is an Autonomous Intelligence Interview Engine that is aimed at computerizing and automating the interview examination and evaluation process with the help of secure voice-driven architecture. The system is designed to have three major modules namely the User Module (Candidate), the Exam Server and the Admin Module which are integrated to provide controlled access, structured assessment and objective assessment. Within the User Module, the candidates will start with registration and logging on. The system is protected by the two-step verification where only authorized users can take part in scheduled interviews. After authentication, the candidate will be presented with the exam interface, in which questions about the interview will be presented in a series one after another based on the set schedule by the administrator. The system includes time limits both on an individual question and on the entire interview, which guarantees the same conditions of assessment to the candidates.

The Exam Server is the main control unit, which is in charge of exam scheduling, session validation, and inter-module communication. The voice recording module starts recording the verbal answers of the candidate when the interview starts, and it records responses to each question. The audio is then run into a speech-to-text engine that in real-time translates spoken words into texts. Such transcription is sent to the Natural Language Processing (NLP) analysis module where various parameters of evaluation are calculated. The ones are the relevance of a keyword, semantic relatedness with the anticipated responses, fluency markers, and sentimental features. These parameters are smushed together to produce a set of objective performance scores on each of the questions and also an overall interview score using a weighted scoring mechanism. Result and Feedback Generation module is the module, which assembles analytical outputs into detailed performance reports. Quantitative scores and qualitative feedback are included in these reports and they represent the strengths and improvement areas. The system makes sure that all candidates are graded through the same means by using a set of pre-determined scoring weight and the set of uniform analysis rules.

The Admin Module gives a user a centralized dashboard to monitor and manage on the administrative side. Administrators are also able to manage the process of scheduling interviews, accept their applications, and monitor their status, such as active interviews, passed tests, and unsatisfied ones. The system produces performance reports which can be accessed in a dashboard where the administrators can check and make corrections to the feedback. According to the assessed results, the administrator is able to make the ultimate decision to accept or deny a candidate. Altogether, the suggested system incorporates an alternative level of authentication, a timed examination control, voice capture and subsequent transcription, NLP-based assessment, localized scoring and centralized observation into a single structure. The platform offers a scalable, secure, and unbiased platform based on voice analysis with structured input; it does not rely on hardware-intensive multimodal input nor is it limited to academic mock interviews, training of campus recruitment, and preliminary candidate screening.

Methodology and Modules

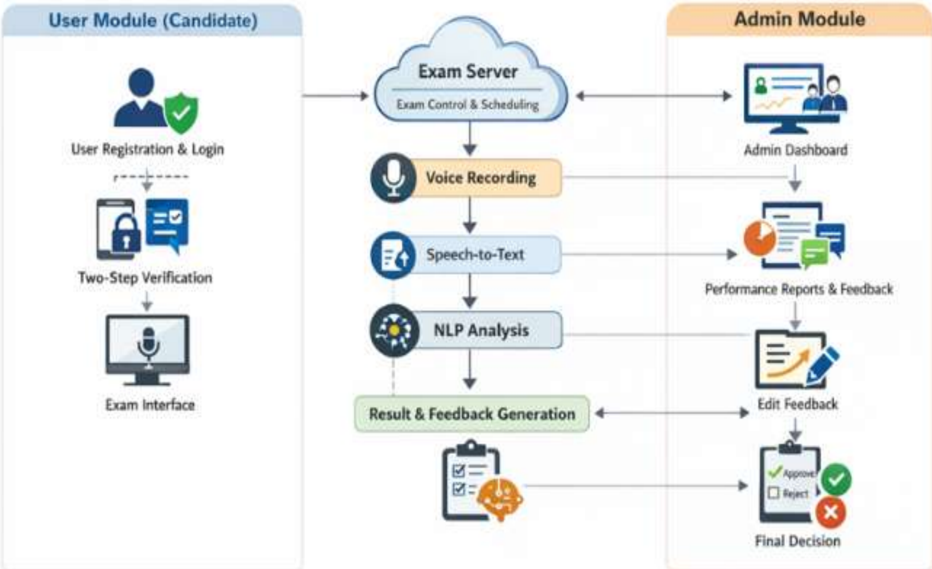
The system has a systematic approach to a methodology that integrates secure access control, automated voice processing, natural language assessment, and administrative control to foster fairness, scale and consistency during the assessment of the interview. The methodology starts

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

with the controlled candidate authentication where the users are verified before taking part. After the authentication the system initiates a timed interview workflow which is centrally handled by exam server. The audio recordings are taken and converted into a document through speech recognition technology. The responses provided in form of text are then subjected to Natural Language Processing capabilities to assess semantic relevancy, contextual accuracy, fluency, and conformity of the keywords to previously known answer models. A grading system assigns grades based on a weighted algorithm to determine the performance on questions and on the level of performance. Analytical results are used to produce automated feedback, although the administrator can continue to make final decisions. This approach will guarantee standardization, minimized human bias during initial assessment, and a scoring transparency approach.

The User Module handles the candidate side activities such as registration processes, login processes and the two step verification of the candidate to ensure security. It gives the interview interface that the applicants are asked questions, recorder their answers, and end the session within set time limits. This module makes it usable and implements a set of rules of examination.

The Exam Server Module is the focal processing and coordination place. It manages the scheduling of exams, validation of session and real time monitoring and inter-component communication in the system. It activates voice recording, controls the flow of data to the



speech-to-text engine, and activates the interaction of NLP analysis and scoring.

Voice Processing Module is one that captures the audio responses and translates them into machine readable text by means of speech to text processing. It guarantees clarity improvement, noise management and correct transcription prior to transmitting data to be analyzed.

The NLP Analysis Module measures the responses that have been transcribed on the basis of semantic similarity analysis, key-word matching, contextual analysis, and linguistic quality analysis. It uses prearranged weights of scoring and creates structured evaluation measures of every response.

Result and Feedback Module is used to combine the outputs of analytical processes into elaborate performance reports. It produces both quantitative scores as well as qualitative recommendations automatically according to identified strengths and weaknesses.

The Admin Module gives a dashboard interface to a centralized control. Administrators are able to arrange interviews, see the running sessions, look at the generated reports, amend feedback when necessary and finalize on approvals or rejections. The module is provided to guarantee the supervision but maintain the automated integrity of the assessment process. Combined, these modules and methodologies form a system of automated assessments of interviews which are scalable, integrative and secure and which balances automation with administrative control.

4. Results and Discussion

Application of the suggested system of automated interview evaluation shows that automated voice-based assessment will save substantial amount of manual work and will preserve consistency in the evaluation. System testing facilitated the candidates to undertake the interview process without any hassles using the secure login and two-step verification and this proved that the authentication mechanism is effective in avoiding unauthorized access. The speech-to-text and voice recording features were effective in converting the audio responses into text with sufficient accuracy in transcription and at the standard audio conditions. Cases of heavy accents or background noise showed a slight decrease, which demonstrates that the environmental factors have a direct impact on the reliability of transcription. The NLP-based evaluation module came up with a stable scoring results on the basis of the preset semantic similarity scales, keyword matching, and contextual relevance analysis. In comparison to the conventional manual evaluation, the automated scoring process has resulted in a significant reduction of evaluation time and removal of variability due to subjective evaluation by a human. Nonetheless, findings also point to a trade-off: even though the system works well in the evaluation of structured and domain-specific responses, it can undervalue the responses that are highly creative or unconventional and do not follow answer patterns as anticipated but are otherwise logically correct. This shows that it is dependent on the quality and coverage of predefined answer models. The feedback generation and result production component was found to be effective in generating structured performance reports that would unambiguously determine areas areas of strength and improvement. Administrators have also noted increased efficiency because of centralized dashboards which enable the ability to monitor, review and make final decisions without processing raw interview data. The human validation that is there by being able to edit feedback prior to final approval is a needed layer that the system can be used to aid in decision-making and not to remove the entire process.

Scalability wise, the architecture can handle multiple simultaneous interviewing sessions run by the exam server proving useful with campus recruitment processes, training evaluations as well as initial screening tests. Centralized processing model also makes it easy to update systems and monitor their performance. Scalability though requires a large server capacity and efficiency of real-time processing particularly when it comes to voice transcription and NLP analysis simultaneously. Altogether, the findings suggest that the proposed system attains its main goals, namely automation, standardization, less bias, and enhanced operational efficiency. Among the areas of improvement that are also revealed in the discussion is the need to improve the resilience of speech recognition, extend the NLP semantic coverage, and refine the scoring algorithms that would enable it to capture more subtle responses. It is not a fully automated system that substitutes qualified human assessment but can be effectively used as the first level screening and structured assessment tool that helps to decrease workload and preserve fairness and consistency.

5. Conclusion

The suggested automated interview evaluation platform reveals that a voice-based, structured assessment platform can greatly enhance results in terms of consistency, scalability and operational effectiveness in the case of the interview based evaluation. The system uses secure authentication, centralized control of exams, speech-to-text translation, semantic analysis of the text with the help of NLP, and automated feedback to decrease reliance on the manual evaluation and the subjective bias in the process of screening out the candidates at the initial stage. The architecture provides standardized assessment conditions of all participants but administrative oversight to final decision-making. The findings suggest that the system is effective in structured and domain-concentrated interviews, it saves on the time in evaluation, and generates reports efficiently using a centralized dashboard. Nevertheless, speech clarity, environmental noise, and depth of knowledge on pre-established answer models affect performance, which is why data preparation and constant model update should be performed with great care. In general, the system is not a fully-fledged substitute of human judgment as it is a solid first-level screening and structured evaluation tool.

The way forward of the system is in the advancement of intelligence, flexibility and strength. Using the latest deep learning-based acoustic models and accent adaptation mechanisms, speech recognition can be enhanced. Transformer-based language models can be used to further extend the NLP module to understand the context better and detect creativity, and score an open-ended response beyond simple keyword matching. Emotion and sentiment analysis through voice tone should be integrated to give more behavioral information during interviews. This system can also develop into a multimodal platform that includes facial expression analysis and behavioral tracking under the condition that the issue of privacy and ethical considerations will be considered. Also, it is possible to introduce adaptive questioning mechanisms, in which other questions dynamically change depending on prior answers, and this can make the interview simulation more realistic. It would be further supported to scale to large-scale institutional and corporate usage by deployment on cloud-based scalable infrastructure with improved security protocols. On-going dataset growth, bias checking, and fairness reviewing will become essential aspects to have responsible and transparent AI-motivated assessment in the future applications.

References:

Nagasawa, F., Okada, S., Ishihara, T., & Nitta, K. (2024). interview based on interviewees' speaking willingness recognition for interview robots. IEEE Transactions on Affective Computing, 15(3), 942–957. https://doi.org/10.1109/TAFFC.2023.3309640

Artiran, S., Ravisankar, R., Luo, S., Chukoskie, L., & Cosman, P. (2022). Measuring social modulation of gaze in autism spectrum condition with virtual reality interviews. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30, 2373–2384. https://doi.org/10.1109/TNSRE.2022.3198933

S., Bedmutha, P. S., & Cosman, P. (2024).orientation and joint attention in autism with triadic VR interviews. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 759–769. https://doi.org/10.1109/TNSRE.2024.3363728

Mongkoljaturong, K., Manitsakulwong, M., & Moodleah, S. (2025). AI-powered MetaHuman interviewer: game for student job interview skills. IEEE Access, 13, 205505–205520. https://doi.org/10.1109/ACCESS.2025.3640017

Salehi, P., et al. (2024).empirical analysis of 2D and 3D virtual environments in training for child interview skills. IEEE Access, 12, 131842–131864. https://doi.org/10.1109/ACCESS.2024.3442297

L'Yi, S., Wang, Q., & Gehlenborg, N. (2023). The role of visualization in data analysis workflows: The interviews. In 2023 IEEE Visualization and Visual Analytics (VIS) (pp. 101–105). https://doi.org/10.1109/VIS54172.2023.00029

Peftitsis, D., & Mavroudi, A. (2022). perceptions on teaching design using student-response systems: Thematic content analysis of interviews. IEEE Access, 10, 64059–64071. https://doi.org/10.1109/ACCESS.2022.3183584

Muiruri, D., Lwakatare, L. E., Nurminen, J. K., & Mikkonen, T. (2022). Practices and infrastructures systems: An interview study in Finnish organizations. Computer, 55(6), 18–29. https://doi.org/10.1109/MC.2022.3161161

Oliva, R., Beacco, A., Gallego, J., Abellan, R. G., & Slater, M. (2023). The making of a newspaper interview in virtual reality .IEEE Computer Graphics and Applications, 43(6), 117–125. https://doi.org/10.1109/MCG.2023.3315761

Kim, C., Choi, J., Yoon, J., Yoo, D., & Lee, W. (2023). Fairness-aware multimodal learning in automatic video interview assessment. IEEE Access, 11, 122677–122693. https://doi.org/10.1109/ACCESS.2023.3325891

Putra, B., et al. (2024). MAG-BERT-ARL for fair automated video interview assessment. IEEE Access, 12, 145188–145205. https://doi.org/10.1109/ACCESS.2024.3473314

Aranda, O., Dieste, O., Panach, J. I., & Juristo, N. (2026). Visibility of domain elements in the elicitation process interviews: A family of empirical studies. IEEE Transactions on Software Engineering. https://doi.org/10.1109/TSE.2026.3662599