Digital Object Identifier 10.48165/iitmjit.2026.12.1.11

NeuroHire: Intelligent Virtual Interviewing with AI

Sushma Malik[1], Anamika Rana[2], Priyanshi Kunte[3], Pratham Sachdeva[3]
[1] Assistant Professor, [2] Associate Professor, [3]Research Scholar, Department of Computer Applications
Maharaja Surajmal Institute, New Delhi, India
[1]sushmalik25@gmail.com, [2]anamica.rana@gmail.com, [3]priyanshikunte@gmail.com,
[4]prathamsachdeva1602@gmail.com

Abstract: For many job seekers, interviews are high-pressure situations where success often depends more on performing well under stress than on demonstrating actual skills and knowledge. Despite the critical role interviews play in recruitment, many candidates lack access to realistic and personalized interview practice environments. This paper introduces NeuroHire, an AI-driven interview coaching system designed to address this gap by providing interactive virtual interview experiences.

NeuroHire incorporates the candidate's resume and the target job description to generate tailored interview scenarios. Instead of relying solely on keyword matching, the system applies semantic embeddings to group candidate skills through unsupervised learning techniques. The platform also enables voice-based interaction, allowing users to practice professional communication in a realistic setting. Additionally, Natural Language Processing (NLP) techniques are used to assess the relevance, clarity, and quality of candidate responses.

Upon completion of the interview session, users receive comprehensive feedback highlighting their strengths and areas for improvement. NeuroHire functions as an adaptive interview simulation framework powered by Large Language Models (LLMs). Experimental results show that the system achieves a query classification accuracy of 84.21%**.** Furthermore, evaluation of the generative component using the Mistral model produced a semantic similarity score of 72.84% compared with reference responses. These results demonstrate the potential of NeuroHire as an effective tool for realistic mock interview preparation.

Keyword: Neuro Hire, Virtual Interviewing, Natural Language Processing, Large Language Models

1. Introduction

The integration of Artificial Intelligence (AI) is transforming talent acquisition by shifting recruitment from traditional human-centered processes to more advanced data-driven systems. For many years, recruitment processes have been affected by inefficiencies and vulnerabilities to human biases associated with factors such as gender, personal affinity, and appearance, which often compromise fairness in hiring decisions [1]. Traditional hiring methods are typically slow, labor-intensive, and resource-consuming, as recruiters must manually review large volumes of resumes, resulting in increased costs and extended hiring cycles [2].

Research has also highlighted persistent inequalities in recruitment outcomes. Studies indicate that minority candidates often receive fewer positive responses compared to majority candidates[3]. For example, both male and female evaluators have been shown to rate male applicants as more competent and offer them higher salaries than equally qualified female candidates, particularly for roles historically dominated by men, [3].

AI-powered tools have been introduced to address these challenges by improving efficiency and introducing a higher degree of objectivity into the hiring process [4]. These systems are capable of analyzing large volumes of applications in a significantly shorter time compared to manual screening, thereby reducing recruitment costs and accelerating hiring timelines [5]. As a result, AI has evolved from an experimental technology to a strategic tool that leverages data to enhance human resource acquisition.

Historically, the use of AI in recruitment can be traced back to the introduction of Applicant Tracking Systems (ATS) in the late 1990s. These early systems primarily relied on simple keyword-matching techniques to filter resumes [6]. With technological advancements in the early 2000s, machine learning (ML) and Natural Language Processing (NLP) techniques were integrated into recruitment tools, enabling systems to better understand contextual information within resumes rather than relying solely on keyword matching [5], [6]. This development led to the emergence of AI-driven recruitment solutions capable of identifying passive candidates and using predictive analytics to anticipate future hiring requirements [7].

More recently, virtual interview simulation platforms have emerged that evaluate candidates in real time using multimodal analysis. These systems assess candidates by integrating multiple forms of data. For instance, Natural Language Processing (NLP) can analyze interview transcripts to evaluate the relevance, quality, and tone of verbal responses. Computer vision techniques can analyze visual cues such as facial expressions and body language to infer candidate confidence or nervousness. Additionally, vocal analysis can evaluate tone, pitch, and speech patterns to provide insights into a candidate's emotional state and communication style [8].
Despite their potential benefits, these systems also present several challenges. Candidates may adapt their behavior to align with algorithmic expectations, a phenomenon sometimes referred to as "metric gaming," where superficial compliance may be prioritized over genuine competencies [9]. Furthermore, AI systems may inadvertently reproduce existing biases present in historical data used for training. For instance, an experimental recruitment tool developed by Amazon reportedly learned to penalize resumes containing the term "women's," as it was trained on historical resume data that predominantly represented male candidates [10]. Such incidents highlight the importance of addressing bias in AI-driven hiring systems.
To mitigate these issues, organizations must adopt responsible AI practices. This includes conducting bias audits to ensure equitable outcomes across demographic groups [11]. Additional strategies include improving data quality through careful data preprocessing, integrating fairness-aware algorithms, and maintaining effective human oversight to review and override automated decisions when necessary[12]. Regulatory frameworks are also evolving to address these risks. For example, the European Union's AI Act classifies recruitment-related AI systems as "high-risk," requiring strict standards for transparency, data quality, and human supervision[13], [14]. Similarly, in the United States, the Equal Employment Opportunity Commission (EEOC) enforces anti-discrimination laws that hold organizations accountable for discriminatory outcomes produced by AI systems, even when such systems are provided by third-party vendors[15].
As AI continues to influence talent acquisition, it is more likely to augment rather than replace human recruiters. By automating repetitive and data-intensive tasks, AI allows recruiters to focus on strategic activities such as relationship-building and nuanced decision-making. The future of recruitment should therefore emphasize a balanced integration of technological innovation with fairness, transparency, and human-centered values.
The remainder of this paper is organized as follows. Section II reviews existing mock interview platforms and prior research on video interviewing. Section III describes the proposed platform architecture and methodology. Section IV presents the experimental results and evaluation, and Section V concludes the paper.

2. Literature Review

In today's digital economy, organizations are increasingly adopting Artificial Intelligence (AI) technologies to automate various aspects of talent acquisition and recruitment processes[16]. The early development of AI-based interview simulators was primarily motivated by the need for unbiased, consistent, and scalable solutions for candidate screening. Initial research focused on applying Natural Language Processing (NLP) and computer vision techniques to analyze both verbal and non-verbal cues in candidate responses, enabling automated systems to replicate several elements of traditional human-led interviews.
For example, earlier studies demonstrated that automated interview systems using the Multiple Mini Interview (MMI) methodology achieved strong test–retest reliability, with intraclass correlation coefficients ranging between 0.65 and 0.81. These systems also reported positive user experiences when applied to medical school admissions processes [17]. Subsequent technological advancements have enabled more detailed and real-time assessments of candidate performance during interviews [18].
Modern AI-driven interview platforms increasingly rely on multimodal analysis to evaluate candidates more comprehensively. These platforms integrate multiple data sources and provide structured feedback for both candidates and recruiters [19]. Recent survey studies highlight several emerging AI approaches used in this domain, including multi-agent systems that support collaborative and asynchronous interviews [20], generative AI techniques for dynamic interview question generation, and affective computing methods for emotion recognition during interviews [21].
Performance prediction can also be enhanced through the analysis of facial expressions, vocal tone, and speech patterns. However, the use of such technologies raises several ethical concerns, including privacy issues, algorithmic bias, and inconsistent evaluation outcomes across different demographic groups [22]. As a result, researchers emphasize the importance of designing systems that incorporate interpretable evaluation metrics,

human-in-the-loop supervision, and transparent data auditing mechanisms to ensure fairness and accountability[23].

Despite the promising potential of AI-driven interview systems to improve efficiency and fairness, several limitations remain. These include the risk of algorithmic bias, variations in candidate experience due to differences in technological familiarity, and the ongoing need to validate AI-based assessments against human judgment and real-world job performance outcomes[20], [22]. Overall, existing literature suggests that AI-powered virtual interview simulators can serve as effective and scalable tools for automated candidate evaluation. However, maintaining system transparency, minimizing bias, and ensuring adequate human oversight remain critical challenges for their successful implementation in recruitment processes.

Recent work has also explored the use of Large Language Models (LLMs) to enhance virtual interview systems. For example, Reference [24] proposed an LLM-based platform that supports multilingual simulated interviews, enabling candidates to practice skills across different languages. The system provided personalized interview scenarios and evaluated performance using user-centric metrics such as satisfaction ratings. However, the study identified limitations in the system’s ability to generalize across different cultural and business contexts. The training data and model behaviors were not fully aligned with regional or industry-specific nuances.

Furthermore, additional research is required to examine long-term skill development and the impact of multilingual virtual interview environments on candidate confidence and domain-specific learning. The reported accuracy of descriptive feedback generated by such systems varies considerably, and their applicability to high-stakes interview scenarios—such as medical admissions or executive-level hiring—has not yet been comprehensively validated. As summarized in Table I, many existing systems primarily focus on formative interview practice rather than high-stakes evaluation. The transition from formative training tools to summative assessment systems represents a significant shift in both system objectives and design requirements.

To address these challenges, this paper proposes **NeuroHire**, a comprehensive framework for developing an AI-based virtual interview simulator designed to support realistic interview practice, skill evaluation, and personalized feedback. The proposed system aims to enhance candidate preparation while incorporating scalable AI technologies and explainable evaluation mechanisms.

3. Methodology

The methodology adopted for this study follows a systematic and phased approach to design and implement the proposed AI-based interview simulation system. As illustrated in Figure 1, the architecture of the proposed model aims to enhance mock interview experiences by integrating voice interaction and intelligent analysis. The system begins by analyzing the candidate’s resume and the corresponding job description to create a customized interview scenario tailored to each user. This is followed by a dynamic interview simulation powered by a large language model (LLM), which conducts the interaction in a manner similar to a human interviewer. Finally, the system compiles the results and generates a detailed performance report that provides feedback and recommendations for improvement.

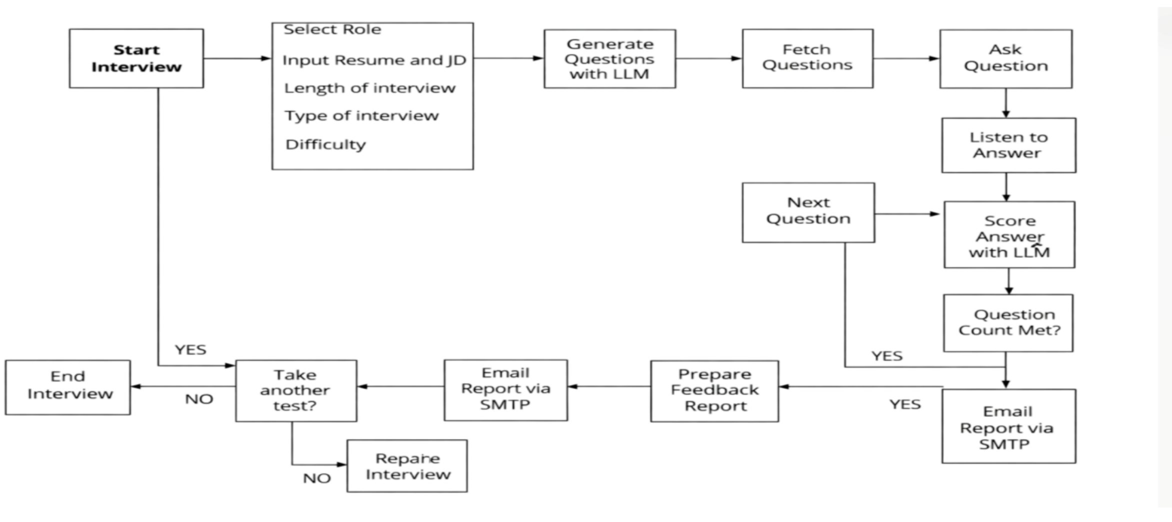

Fig. 11.1: AI interview simulator process flow

3.1 Resume Analysis and Tailored Question Generation

The first phase focuses on generating a personalized mock interview by evaluating the applicant's professional profile extracted from the resume and comparing it with the requirements specified in the target job description.

1. **Resume and Job Description Ingestion and Analysis:**

   The process begins when the user uploads their resume and the job description for the desired role. The Natural Language Processing (NLP) module extracts relevant information using tools such as Optical Character Recognition (OCR), OpenCV, PyMuPDF, and spaCy. These tools identify and categorize key entities including skills, educational qualifications, and professional experience. Subsequently, a similarity analysis is performed to determine the alignment between the candidate's profile and the job requirements.

2. **Skill Cluster Generation:**

   To enable a more comprehensive assessment, the system moves beyond simple keyword matching. Feature extraction techniques are applied to identify relevant skills from the resume, and the K-Means clustering algorithm is used to group related skills into clusters. This approach helps identify both explicit and implicit competencies of the candidate.

3. **Personalized Question Generation:**

   The generated skill clusters serve as the foundation for creating customized interview questions. The system retrieves relevant content from a curated corpus to ensure that the questions remain current and aligned with industry standards. This information is then incorporated into prompt templates used by the question-generation module.

3.2 Interactive Simulation and Intelligent Verbal Analysis

The second phase focuses on conducting the mock interview, capturing candidate responses, and performing real-time analysis of performance.

1. **Core Generative Engine (Ollama):**

   The core generative component of the system is powered by a Large Language Model (LLM) deployed through the Ollama framework, which can be containerized using Docker. The model processes the entire conversation context to generate relevant and coherent responses, simulating the behavior of a human interviewer. This enables the system to conduct natural and context-aware interviews across different professional domains.

2. **Speech-to-Text Transcription:**

   To provide a realistic interview experience, the system incorporates voice-based interaction. Speech-to-text and text-to-speech technologies such as Whisper are used to convert spoken responses into textual form with high accuracy. The transcription process is capable of handling variations in accents and moderate background noise. Real-time processing allows the AI system to immediately analyze candidate responses.

3. **Performance Analysis and Adaptive Learning:**

   Machine learning algorithms analyze the transcribed responses to evaluate candidate performance. The system examines response length, relevance to the question, and the presence of key concepts expected in an ideal answer. Using predictive modeling and historical response patterns, the system identifies strengths and weaknesses in the candidate's responses and adapts the evaluation accordingly.

3.3 Comprehensive Assessment and Recommendation

In the final stage, all performance-related data collected during the interview session is aggregated to generate a comprehensive evaluation report. The report provides structured feedback highlighting the candidate's strengths and areas requiring improvement. Additionally, a recommendation engine suggests learning resources and strategies to address identified skill gaps. By combining realistic conversational interaction with detailed performance analysis, the system provides an effective framework for interview preparation and skill development.

4. System Design and Architecture

The proposed system is designed to ensure reliability, security, scalability, and continuous performance improvement. It is developed to operate efficiently across different devices while accommodating diverse user requirements. The architecture is implemented using a React.js frontend and a Python-based backend**,** as illustrated in Figure 2.

The system incorporates a responsive User Interface (UI) built with React.js to provide a seamless experience across both web and mobile platforms. The design emphasizes an intuitive User Experience (UX)**,** enabling easy navigation and clear interaction pathways for users. Engaging interface elements further enhance usability and accessibility throughout the interview simulation process.

To improve system flexibility and maintainability, the platform follows a microservices architecture, where core components such as user management, the AI analysis engine, and data processing modules operate as independent services. This modular design allows individual components to be developed, tested, and scaled independently.

Furthermore, the system employs an event-driven architecture, enabling asynchronous communication between services. This approach supports real-time data processing and allows the platform to deliver immediate feedback during interview simulations. As a result, the architecture ensures efficient performance, scalability, and responsiveness while maintaining a robust and adaptable framework for AI-driven interview evaluation.

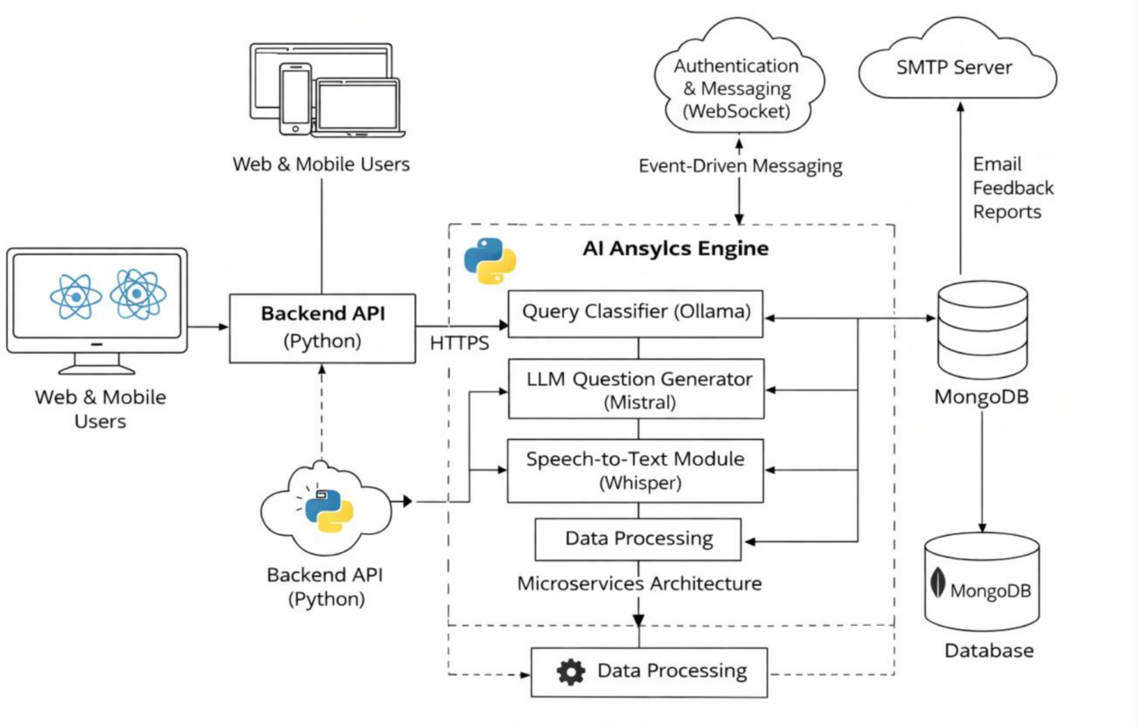

Figure 11.2: End to End Flow of AI Interview Evaluation

5. Results

The performance of the proposed system was evaluated based on two major components: query classification and response generation. The experimental results demonstrate that the system performs effectively in semantic understanding and response generation. The model prioritizes producing fluent and coherent responses, which occasionally results in a slight compromise in factual precision. This design choice ensures that generated answers remain natural, smooth, and contextually connected**,** improving the overall interview simulation experience.

5.1 Query Classification Module Performance

The query classification module is responsible for categorizing interview questions into either "HR" or "Technical" classes. The evaluation results, illustrated in Figure 3 and Figure 4**,** show that the classifier exhibits balanced performance across both categories.

For the HR category, the model achieved a precision score of 0.97 and a recall score of 0.78**.** This indicates that while most HR questions were correctly identified, a small number were incorrectly classified as Technical questions.

For the Technical category, the classifier obtained a recall score of 0.96 with a precision score of 0.71. This suggests that the majority of Technical questions were correctly detected, although a few HR queries were mistakenly labeled as Technical.

Table 11.1: Precision, Recall, and F1-Score for HR and Technical Question

Category	Precision	Recall	F1
HR	0.97	0.78	0.86
Technical	0.71	0.96	0.82

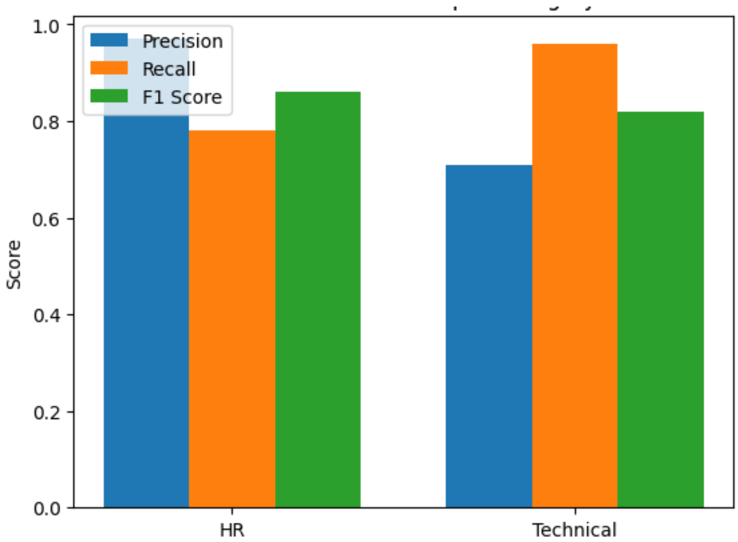

Figure 11.3: Classification Metrics per Category

The confusion matrix in Figure 4 further illustrates the classification performance. Out of 92 HR questions**,** 20 were incorrectly classified as Technical, while 48 out of 50 Technical questions were correctly identified. Overall, the query classification module achieved an accuracy of 84.21% (0.842), demonstrating reliable performance in distinguishing between interview question types.

Table 11.2: Confusion Matrix for HR and Technical Question Classification

Actual / Predicted	HR	Technical
HR	72	20
Technical	2	48

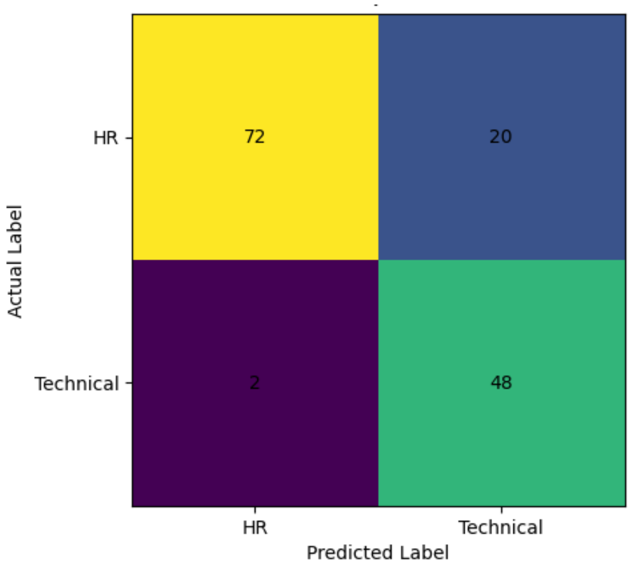

Figure 11.4: Confusion Matrix of Question Classification

Table 11.3: Comparison of Model Accuracies for Virtual Interview Simulation

Model	Accuracy	Reference
KNN	0.61 (average)	[25], [26]
LLM (Gemini AI)	0.82	[27]
LLM (Claude)	0.56	[28]

The comparison indicates that large language models generally outperform traditional machine learning models, although performance varies depending on the architecture and training methodology.

5.2 Generative Model Quality

The response generation capability of the Mistral model was evaluated using two main metrics: semantic similarity and keyword recall. These metrics help measure how well the generated responses align with expected answers in both meaning and key content.

As illustrated in Figure 5, there is a noticeable difference between the two-evaluation metrics. The model achieved a semantic similarity score of 72.84%**,** indicating strong contextual understanding of interview questions. However, the keyword recall score was comparatively lower at 14.65%**.** This difference highlights the model's emphasis on generating meaningful and coherent responses rather than strictly reproducing specific keywords**.**

Table 11.4:.Average Scores for Semantic Similarity and Keyword Recall in Response Generation

Metric	Score
Semantic Similarity	72.84%
Keyword Recall	14.65%

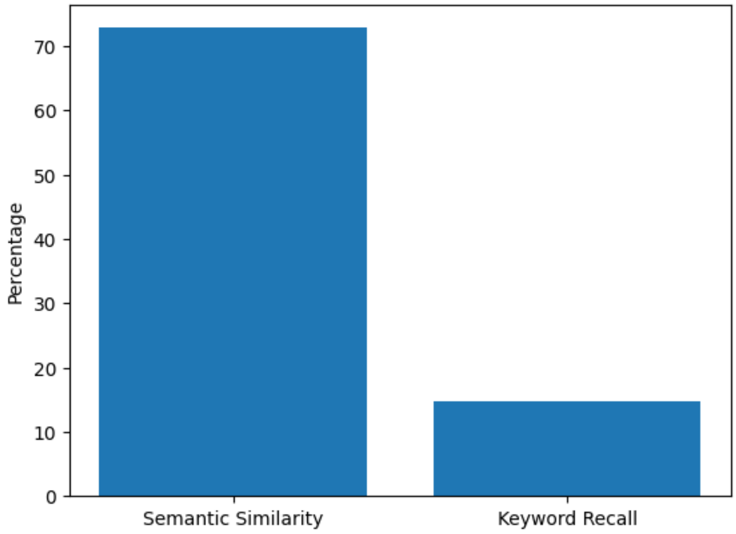

Fig. 11.5: Average Score for Semantic Similarity and Keywords Recall

The distribution of semantic similarity scores shown in Figure 6 follows a near-normal distribution, demonstrating consistent performance across different queries. The highest concentration of scores appears between 0.55 and 0.65**,** indicating that most responses fall within the moderate-to-high semantic quality range. A smaller number of responses achieved higher similarity values, which slightly increased the overall average. The final average semantic similarity score was approximately 73%, reflecting the model's ability to generate contextually appropriate interview responses.

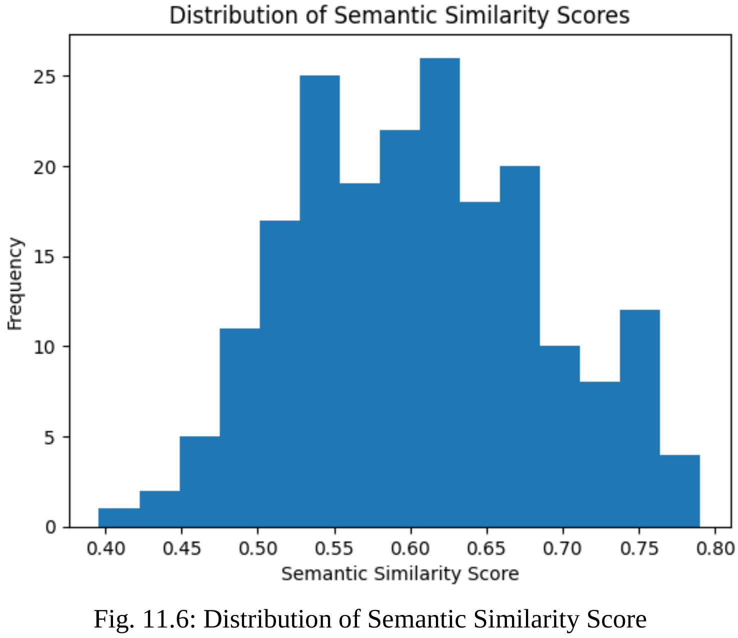

Fig. 11.6: Distribution of Semantic Similarity Score

6. Conclusion

NeuroHire demonstrates the practical applicability of Artificial Intelligence in modern recruitment systems**.** The proposed model helps bridge the gap between theoretical knowledge and real-time interview performance by simulating a realistic interview environment. By leveraging Large Language Models (LLMs) and Natural Language Processing (NLP) techniques, the system dynamically generates interview questions tailored to each candidate. This personalization is achieved through the analysis of both the candidate's resume and the job description, enabling more relevant and context-aware interview scenarios.

Additionally, NeuroHire provides personalized feedback to help candidates improve their interview performance. The system emphasizes semantic similarity in responses, encouraging candidates to produce fluent, coherent, and contextually meaningful answers rather than relying solely on keyword matching.
Future work will focus on enhancing the user interface through gamification techniques to increase user engagement and motivation. Further improvements will also include extensive evaluation and testing to ensure greater fairness, reliability, and overall system efficiency.

References

1. M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: Evaluating claims and practices," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020, pp. 469–481.
2. L. Lippens, A. Dalle, F. D'hondt, P.-P. Verhaeghe, and S. Baert, "Understanding ethnic hiring discrimination: A contextual analysis of experimental evidence," *Labour Econ.*, vol. 85, p. 102453, 2023.
3. A. Neschen and S. Hügelschäfer, "Gender bias in performance evaluations: The impact of gender quotas," *J. Econ. Psychol.*, vol. 85, p. 102383, 2021.
4. S. Kassir, L. Baker, J. Dolphin, and F. Polli, "AI for hiring in context: A perspective on overcoming the unique challenges of employment research to mitigate disparate impact," *AI and Ethics*, vol. 3, no. 3, pp. 845–868, 2023.
5. P. R. Chavan, Y. Chandurkar, A. Tidake, G. Lavankar, S. Gaikwad, and R. Chavan, "Enhancing recruitment efficiency: An advanced applicant tracking system (ATS)," *Industrial Management Advances*, vol. 2, no. 1, p. 6373, 2024.
6. V. R. Uma, I. Velchamy, and D. Upadhyay, "Recruitment analytics: Hiring in the era of artificial intelligence," 2023.
7. R. Sumathi and L. Kumar, "AI-Driven Resume Screening Integrated with Predictive Analytics for Data-Driven Recruitment," in *2025 International Conference on Sustainable Communication Networks and Application (ICSCN)*, IEEE, 2025, pp. 913–918.
8. A. Nayak and I. Satpathy, "The Impact of Predictive Analytics on Recruitment: Going Beyond Traditional Methods," in *Data-Informed Leadership in Higher Education: An Executive Playbook for Institutional Excellence*, Bentham Science Publishers, 2025, pp. 103–122.
9. A. Wang, S. Kapoor, S. Barocas, and A. Narayanan, "Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy," *ACM Journal on Responsible Computing*, vol. 1, no. 1, pp. 1–45, 2024.
10. Z. Chen, "Ethics and discrimination in artificial intelligence-enabled recruitment practices," *Humanit. Soc. Sci. Commun.*, vol. 10, no. 1, p. 567, 2023.
11. N. A. Khan, "Ensuring Ethical and Responsible Use of Artificial Intelligence," *Journal of Computer Science and Technology Studies*, vol. 7, no. 5, pp. 376–385, 2025.
12. S. A. Joseph, T. M. Kolade, O. O. Val, O. O. Adebiyi, O. S. Ogungbemi, and O. O. Olaniyi, "AI-powered information governance: Balancing automation and human oversight for optimal organization productivity," *Asian Journal of Research in Computer Science*, vol. 17, no. 10, pp. 110–131, 2024.
13. A. P. Nayak and V. B. Sindhu, "Ethical Considerations in AI-Enabled HR Practices: A Comprehensive Review," *Harnessing AI to Transform Human Resources in Future Workplace Practices*, pp. 65–92, 2025.
14. N. A. Khan, "Ensuring Ethical and Responsible Use of Artificial Intelligence," *Journal of Computer Science and Technology Studies*, vol. 7, no. 5, pp. 376–385, 2025.
15. E. Dave, A. Rajurkar, and V. Nagila, "Employment Law in the Age of Automation: Legal Protections Against AI-Driven Workplace Discrimination," *Authorea Preprints*, 2025.
16. R. Pillai and B. Sivathanu, "Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations," *Benchmarking: an international journal*, vol. 27, no. 9, pp. 2599–2629, 2020.
17. A. Callwood *et al.*, "Feasibility of an automated interview grounded in multiple mini interview (MMI) methodology for selection into the health professions: an international multimethod evaluation," *BMJ Open*, vol. 12, no. 2, p. e050394, 2022.
18. S. Khapekar, S. Bothara, T. Babar, and R. Kine, "AI-driven smart interview simulator with real-time speech and emotion analysis," *Int. J. Adv. Eng. Res. Sci.*, vol. 12, no. 3, 2025.
19. X. Wu *et al.*, "Candidate Evaluation with Multimodal Data-Driven for Recruitment," in *International Conference on Pattern Recognition*, Springer, 2024, pp. 81–96.
20. M. Xie and B. Liu, "EvalNet: Sentiment Analysis and Multimodal Data Fusion for Recruitment Interview Processing," in *2025 7th International Conference on Artificial Intelligence Technologies and Applications (ICAITA)*, IEEE, 2025, pp. 444–448.

21. P. Sinha, Khushi, and A. Dagur, “Improved framework model to train and evaluate difficulty of interview question using generative AI,” in *International Conference on Artificial Intelligence and its Applications in the Age of Digital Transformation*, Springer, 2024, pp. 175–188.
22. A. K. Kamboj, L. E. Raffals, J. A. Martin, and V. Chandrasekhara, “Virtual interviews during the COVID-19 pandemic: a survey of advanced endoscopy fellowship applicants and programs,” *Tech. Innov. Gastrointest. Endosc.*, vol. 23, no. 2, pp. 159–168, 2021.
23. L. McCormack and M. Bendechache, “A comprehensive survey and classification of evaluation criteria for trustworthy artificial intelligence,” *AI and Ethics*, vol. 5, no. 3, pp. 1973–1994, 2025.
24. Y. Ma, Y. Zeng, T. Liu, R. Sun, M. Xiao, and J. Wang, “Integrating large language models in mental health practice: a qualitative descriptive study based on expert interviews,” *Front. Public Health*, vol. 12, p. 1475867, 2024.
25. S. Chopra and S. Urolagin, “Interview data analysis using machine learning techniques to predict personality traits,” in *2020 Seventh International Conference on Information Technology Trends (ITT)*, IEEE, 2020, pp. 48–53.
26. L. Hickman, R. Saef, V. Ng, S. E. Woo, L. Tay, and N. Bosch, “Developing and evaluating language-based machine learning algorithms for inferring applicant personality in video interviews,” *Human Resource Management Journal*, vol. 34, no. 2, pp. 255–274, 2024.
27. J. Geathers *et al.*, “Benchmarking generative AI for scoring medical student interviews in objective structured clinical examinations (OSCEs),” in *International Conference on Artificial Intelligence in Education*, Springer, 2025, pp. 231–245.
28. Y. S. Kruthika and N. S. Nandini, “AI-Enhanced HR Interview Simulation for Realistic Candidate Assessment,” *INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH*, vol. 13, no. 4, pp. 914–917, 2025.