Jobs.Ai: An Autonomous Privacy-First Recruitment and Blind Voice-Interviewing System Using Local LLMs

Harsh Jadhav[1*]; Anant Mishra[2]; Srikrishna Tophakhane[3]

Department of Computer Science Engineering (Artificial Intelligence and Machine Learning), Viva Institute of Technology, Maharashtra, India
*Corresponding Author

Abstract— *The rapid adoption of artificial intelligence in recruitment has introduced challenges related to bias, privacy, and lack of transparency in automated hiring systems. Recent advances in agentic AI and large language models (LLMs) enable the development of autonomous, goal-driven systems capable of managing complex recruitment workflows with minimal human intervention. This paper proposes Jobs.Ai, an autonomous, privacy-first recruitment platform that leverages local LLM-based multi-agent architecture to perform job matching, resume screening, and blind voice-based interviewing while ensuring data confidentiality and fairness. The system integrates agent collaboration, explainable decision-making, and secure execution to enhance trust and efficiency in hiring pipelines. Experimental insights and architectural analysis demonstrate the feasibility of deploying agentic AI for scalable, ethical, and trustworthy recruitment systems in enterprise environments.*

Keywords— Recruitment Automation, Natural Language Processing (NLP), Local LLMs, Blind Interviewing, Semantic Search, Explainable AI (XAI).

I. INTRODUCTION

The domain of talent acquisition has historically been plagued by what industry analysts term the "resume black hole" [1]. Traditional Applicant Tracking Systems (ATS) serve as the frontline defense for organizations, tasked with the initial ingestion and filtering of applicants. In 2025 alone, recruiters globally processed billions of applications, yet operational audits suggest that over 75% of qualified candidates are rejected due to rigid keyword mismatches [2].

1.1 The Operational Crisis:

Legacy infrastructure deployed to handle this volume has proven insufficient. Previous generations of automation relied heavily on Boolean keyword matching (e.g., rejecting a "React Expert" because the job description required "React.js"). This lack of semantic understanding forces recruiters to act as human middleware, spending valuable cognitive resources on manual verification rather than strategic engagement. Furthermore, traditional video interview platforms introduce significant visual bias, where factors such as gender, ethnicity, or background environment disproportionately influence hiring decisions [3].

1.2 The Rise of Local AI and Data Sovereignty

The advent of Large Language Models (LLMs) fundamentally altered the landscape of automated interaction. However, the initial wave of generative AI adoption relied on cloud-hosted foundation models, raising serious data sovereignty concerns under regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) when handling candidate resumes [4]. Jobs.Ai addresses these challenges by leveraging locally deployed Small Language Models (SLMs), ensuring high-level reasoning while guaranteeing that sensitive candidate data never leaves the enterprise environment.

II. THEORETICAL FRAMEWORK

2.1 Evolution of Candidate Assessment

Automated hiring has evolved through three major phases: the Keyword Era, the Semantic Era, and the Agentic Era. The **Keyword Era** relied on exact string matching and failed under synonymy. The **Semantic Era**, introduced by transformer-based models such as BERT, enabled contextual skill representation in vector space. The **Agentic Era**, where Jobs.Ai is positioned, introduces autonomous agents capable of reasoning, questioning, and interviewing candidates dynamically [5].

2.2 Blind Interviewing Protocol

Jobs.Ai adopts a voice-first interviewing strategy to eliminate visual bias. Psychological research highlights that "System 1" thinking is fast and intuitive but prone to bias, whereas "System 2" thinking is analytical and deliberate. By removing video feeds and relying on a neutral AI voice, the system enforces System 2 evaluation, ensuring candidates are judged purely on semantic and communicative merit [6].

III. METHODOLOGY: THE AUTONOMOUS PIPELINE

3.1 Data Ingestion and Vectorization

The recruitment pipeline begins by deconstructing the Job Description (JD) into "Must-Have" and "Nice-to-Have" requirements. Uploaded resumes are parsed, cleaned of formatting artifacts, and converted into dense vector embeddings. Semantic similarity between the JD and resumes is computed using cosine similarity:

$$\text{Similarity}(A, B) = \frac{A \cdot B}{\| A \| \| B \|}$$

3.2 Hybrid Contact Extraction

To ensure reliable candidate outreach, Jobs.Ai employs a hybrid extraction protocol. Standard email patterns are identified using high-speed regular expressions. If extraction fails due to obfuscation (e.g., "name [at] domain"), a locally deployed LLM is triggered to semantically infer contact information [7].

3.3 No-Face Meeting Agent

Shortlisted candidates are invited to a voice-only interview interface. The AI interviewer maintains a rolling context window, enabling adaptive follow-up questions and deeper competency assessment. This approach mirrors human interview behavior while eliminating visual bias.

IV. SYSTEM ARCHITECTURE

Jobs.Ai follows a four-tier architecture designed for privacy, modularity, and real-time interaction.

4.1 Tier 1: Frontend User Interface

The recruiter-facing interface is implemented using Streamlit and provides modules for Manager's Ranking, Deep Dive Analysis, and Interview Dashboard. Real-time fit scores and explainable AI summaries are displayed to assist decision-making.

4.2 Tier 2: Backend Orchestration Layer

The orchestration layer manages session state and coordinates communication between resume parsers, embedding models, and local LLMs. Python-based services integrate utilities such as SendGrid for automated interview scheduling.

4.3 Tier 3: Intelligence Layer (Local)

The intelligence layer interfaces with Ollama to manage local LLM inference. Sentence-Transformer models such as all-MiniLM-L6-v2 are used for semantic embedding and similarity search, enabling accurate candidate ranking [8].

4.4 Tier 4: Blind Voice Interview Module

This module executes the real-time interview loop consisting of audio capture, transcription via Whisper (speech-to-text), reasoning using local LLMs, and response generation via text-to-speech engines such as Pyttsx3 or Edge-TTS.

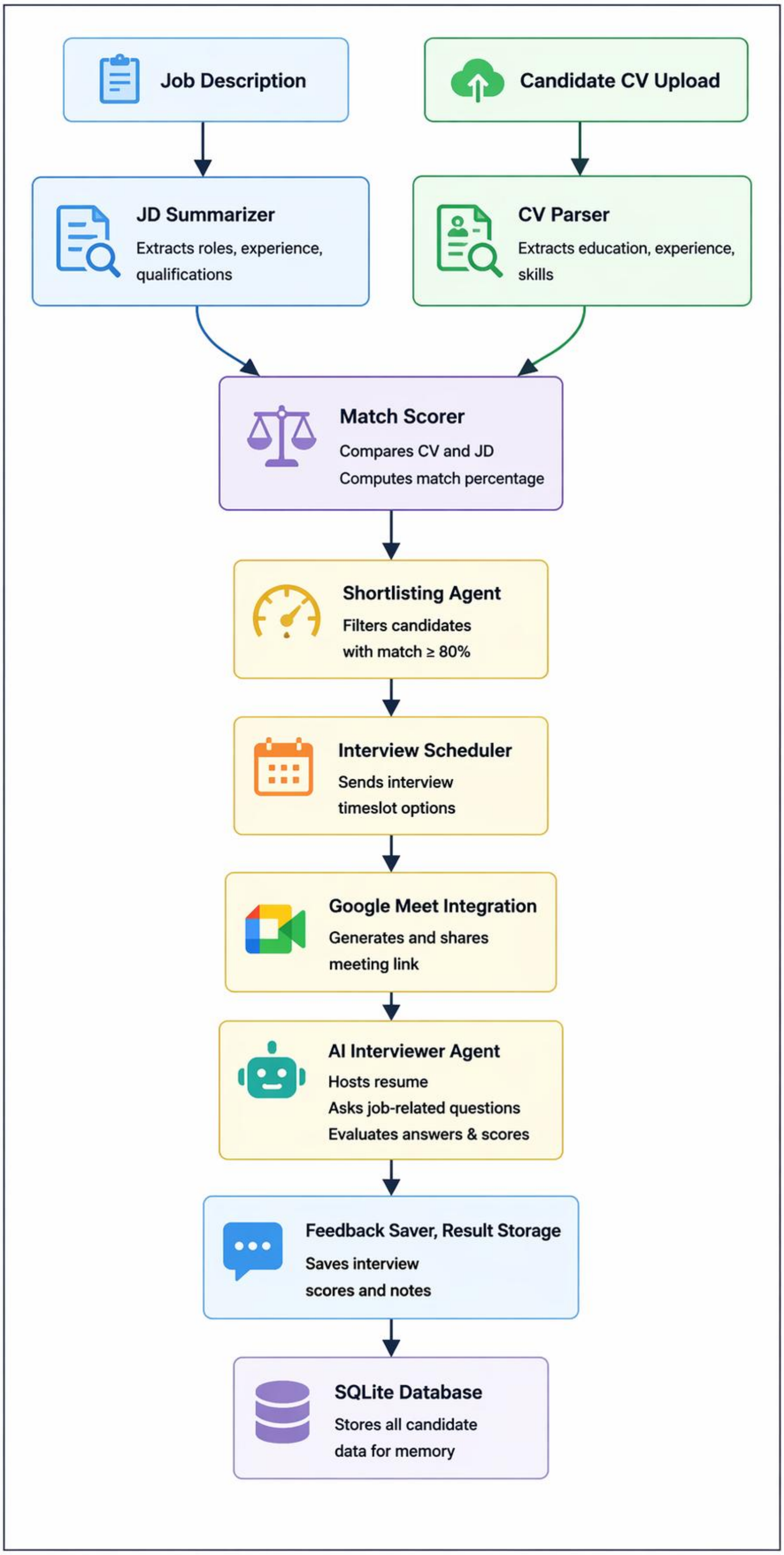

Figure 1: A system architecture diagram

V. IMPLEMENTATION DETAILS

5.1 Hardware and Quantization

The system was evaluated on a workstation equipped with an NVIDIA RTX 3060 GPU (12 GB VRAM). To enable concurrent execution of LLMs, speech-to-text, and text-to-speech modules, 4-bit GGUF quantization was employed, significantly reducing memory usage.

5.2 Dataset and Scoring

A dataset of 50 candidate profiles was used for evaluation, representing high-fit, partial-fit, and irrelevant resumes. Each candidate was assessed through resume parsing, an AI-proctored technical multiple-choice question (MCQ) test, and a blind voice interview.

VI. PERFORMANCE EVALUATION

6.1 Resume Parsing Performance

The resume parsing module achieved an R^2 score of 0.6146 with a Root Mean Square Error (RMSE) of 9.77, indicating moderate predictive capability for initial candidate filtering [9].

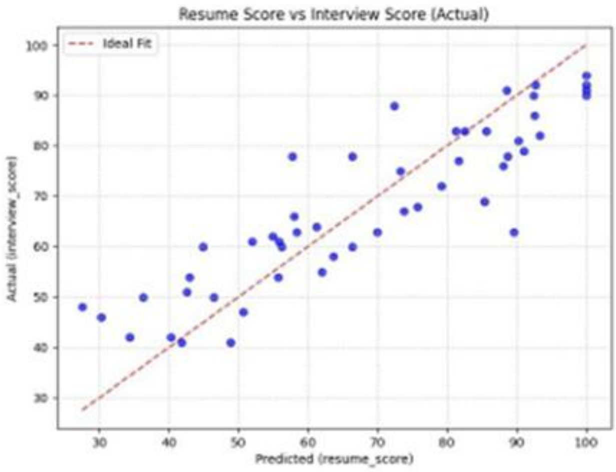

FIGURE 2: Correlation between Resume Parsing Score and Interview Performance

6.2 Technical Assessment Performance

The AI-proctored MCQ assessment demonstrated strong predictive reliability with an R^2 score of 0.9209 and an RMSE of 4.43 [10].

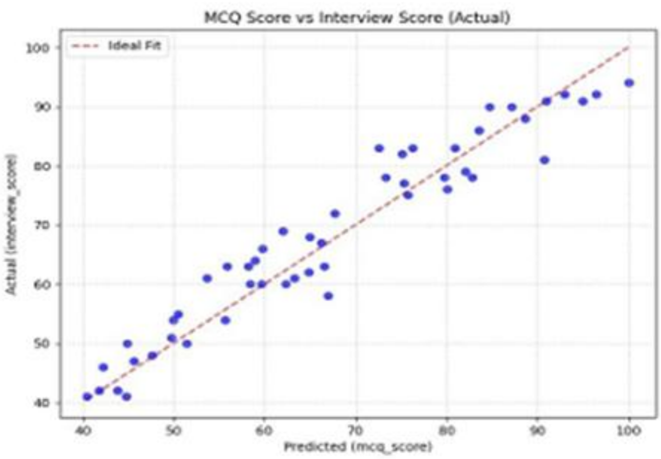

FIGURE 3: Correlation between AI-Proctored MCQ Score and Final Interview Score

6.3 Combined Model Efficacy

The unified hiring score, aggregating resume, MCQ, and interview signals, achieved an R^2 score of 0.8783 and an RMSE of 5.49.

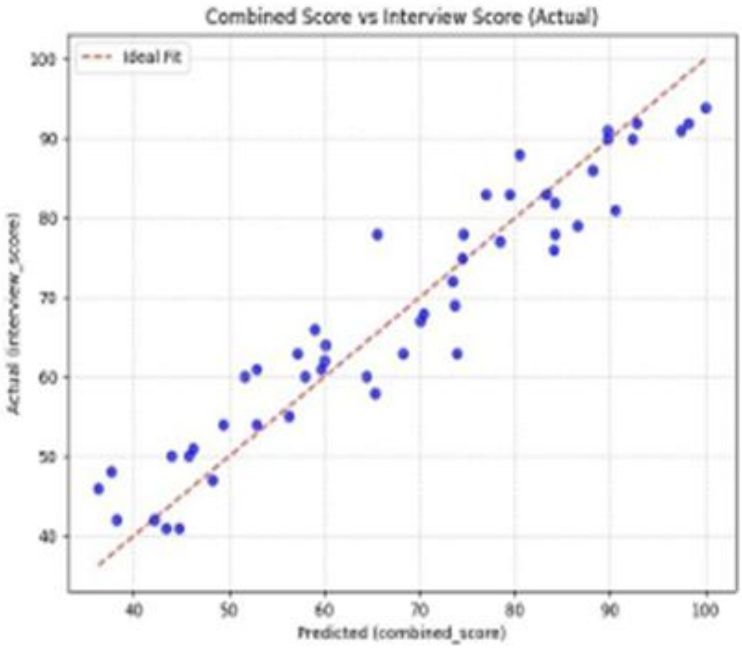

FIGURE 4: Combined model delivers the strongest prediction of candidate success (R^2 = 0.87)

TABLE 1
MODEL PERFORMANCE AT DIFFERENT ASSESSMENT STAGES

Model Stage	R^2 Score	RMSE	Interpretation
Resume Only	0.6146	9.77	Good for initial filtering
Technical MCQ	0.9209	4.43	High reliability
Combined (Resume + MCQ + Interview)	0.8783	5.49	Best overall predictor

(Note: Performance figures/charts were intended to be included. Authors are requested to supply these in the revised submission.)

VII. DISCUSSION

Jobs.Ai demonstrates operational resilience by combining fast vector-based retrieval with deep analytical reasoning through local LLMs. The blind interview protocol significantly reduces visual bias, while local deployment enhances privacy compliance and energy efficiency [11].

Key Observations:

1. **Semantic Matching Advantage:** Unlike keyword-based ATS, Jobs.Ai successfully matched candidates with synonymous skill descriptions (e.g., "React Expert" matching "React.js" requirements).
2. **Privacy Preservation:** By keeping all candidate data on-premise, the system addresses data sovereignty concerns that arise with cloud-based LLM solutions.
3. **Bias Reduction:** The voice-only interface removed visual cues, and preliminary analysis suggests reduced demographic bias in shortlisting decisions.
4. **Limitations:** The system was evaluated on a relatively small dataset (50 candidates). Larger-scale validation is required to confirm generalizability.

VIII. CONCLUSION

Jobs.Ai presents a comprehensive, privacy-first, and bias-aware recruitment framework aimed at enabling fair and objective candidate evaluation. The Blind Voice Interviewer assesses applicants exclusively through their spoken responses, focusing on skills, reasoning ability, and communication quality while deliberately excluding visual cues and personal identifiers that may introduce unconscious bias. By relying on locally deployed language processing models, the framework ensures that sensitive candidate data remains secure and under organizational control.

Despite its strong emphasis on privacy and fairness, the system demonstrates high predictive reliability in evaluating candidate suitability. The combined model (resume + MCQ + interview) achieved an R^2 of 0.8783, outperforming resume-only assessment ($R^2 = 0.6146$). Overall, this framework represents a meaningful step toward ethical, explainable, and autonomous hiring practices, promoting transparency, accountability, and merit-based decision-making in modern recruitment workflows.

Future work includes: (i) larger-scale validation with diverse candidate pools, (ii) integration with enterprise HR systems via APIs, (iii) enhanced explainability for rejection decisions, and (iv) support for multiple languages in voice interviews.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Balestra, M., Galletta, A., & Rossi, S. (2024). Reclaiming autonomy in the age of AI: Algorithmic management and job happiness. *Journal of Information Systems*. [Volume and issue missing].

[2] Pati, A. K. (2025, February). Agentic AI: A comprehensive survey of technologies, applications, and societal implications. *IEEE Access*.

[3] Khamis, A. (2025). Agentic AI systems: Architecture and evaluation using a frictionless parking scenario. *IEEE Access*.

[4] Younes, M. T., Hasan, S., & Ahmed, R. (2025). MLAR: Multi-layer large language model-based robotic process automation applicant tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[5] Lo, F. P.-W., Chen, Y., & Wang, L. (2025). AI hiring with LLMs: A context-aware and explainable multi-agent framework for resume screening. In *Proceedings of the IEEE/CVF CVPR Workshops (CVPRW)*.

[6] Giri, A., Sharma, V., & Patel, N. (2024). Revolutionizing recruitment with large language models: HireStream platform. In *Proceedings of the 1st International Conference on Data Science*.

[7] Gan, C., Liu, W., & Zhang, Y. (2024). Application of LLM agents in recruitment: A novel framework for resume screening. *Journal of Information Processing, 32*, 881–893.

[8] Maritan, A. (2025). StaffPro: An LLM agent for joint staffing and profiling. *arXiv*.

[9] Patil, C., Deshmukh, S., & Kulkarni, A. (2025). Job application automation using agentic AI: A multi-agent framework for automated job application. *International Journal of Innovative Research in Technology, 12*(6). [Page range missing].

[10] Raza, S., Ahmed, M., & Khan, F. (2025). TRiSM for agentic AI: A review of trust, risk, and security management in LLM-based agentic multi-agent systems. *arXiv*.

[11] IBM Research. (2025). ST-WebAgentBench: Benchmarking enterprise safety for AI agents. In *Proceedings of the International Conference on Machine Learning (ICML)*.

[12] Sun, J. (2024, September). Research on the application of large language models in human resource management practices. *International Journal of Emerging Technologies and Advanced Applications*.

[13] Ehtesham, A., Singh, R., & Kumar, P. (2025). Communication protocols for edge general intelligence. *Agent Network Protocols*.

[14] Anonymous. (2025). Routine: A multi-step agent planning framework for enterprise scenarios. *arXiv*.

[15] Anonymous. (2025). MetaAgents: Large language model based agents for decision-making on teaming. *Proceedings of the ACM on Human-Computer Interaction, 9*(2). [Page range missing].

[16] Anonymous. (2025). EvoMAC: Self-evolving multi-agent collaboration. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

[17] Acharya, D. B., & Kuppan, K. (2025). Agentic AI: Autonomous intelligence for complex goals – A comprehensive survey. *IEEE Access, 13*, 18912–18936.

[18] Dima, M., Popescu, A., & Ionescu, C. (2024, September). Distributed AI systems in human resources: Technical implementation and organizational impact. *Medium Technical Review*.

[19] Anonymous. (2025). A survey on trustworthy LLM agents: Threats and countermeasures. *Reddit LLMDevs Research Tracking*.

[20] Anonymous. (2025). DICE: A framework for dimensional and contextual evaluation of language models. *arXiv*.

[21] Portugal, D., Santos, L., & Almeida, R. (2024). An agentic AI-based multi-agent framework for recommender systems. In *Proceedings of the IEEE International Conference on Big Data*.

[22] Qiu, J., Li, X., & Zhang, H. (2024). LLM-based agentic systems in medicine and healthcare. *Nature Machine Intelligence, 6*(12), 1418–1420.

[23] Anonymous. (2025). Human and LLM-based resume matching: An observational study. In *Findings of the Association for Computational Linguistics: NAACL*.

[24] Gartner Research. (2024, November). How agentic AI is shaping business decision-making. *Technology Magazine*.

[25] Capgemini Research. (2024). Harnessing the value of generative AI (Technical report). Capgemini.

[26] PwC Global. (2025, June). PwC 2025 global AI jobs barometer: AI linked to a fourfold increase in productivity growth. *PwC Press Release*.

[27] LinkedIn Talent Solutions. (2025). The 2025 future of recruiting report. *LinkedIn Industry Insights*.

[28] Salesforce Research. (2025). The future of AI agents: Top predictions and trends to watch in 2026. *Salesforce News.*