# AI - Based Resume Scanning System

Shaikh Saniya Jameel[1]; Padma Manoj Sharma[2]; Katkure Dhanshree Yuvraj[3]

[1,2,3]Student of B. Tech in Computer Engineering and Technology Department M.S. Bidve College of Engineering, Latur, Maharashtra, India

**Abstract:** This project focuses on the development of an AI-Powered Resume Screening System designed to overhaul traditional, inefficient talent acquisition processes. The system leverages a multi-model AI architecture combining Rule-Based logic, advanced Deep Learning (LSTM and fine-tuned DistilBERT) models, and the Gemini Pro Large Language Model (LLM) to provide objective, data-driven matching scores between candidate resumes and specific job descriptions. Key functionalities include: high-speed screening, real time analytics for HR teams (e.g., market trends, skill gaps), and personalized, constructive feedback generated by the LLM for applicants.

**How to Cite:** Shaikh Saniya Jameel; Padma Manoj Sharma; Katkure Dhanshree Yuvraj (2025) AI - Based Resume Scanning System. *International Journal of Innovative Science and Research Technology,* 10(12), 781-786. https://doi.org/10.38124/ijisrt/25dec380

## I. INTRODUCTION

➤ *Problem Statement*

Traditional resume screening is inherently inefficient, slow, and prone to inconsistency and human bias due as recruiters must manually process the high volume of digital applications generated by modern job portals. This inefficiency not only delays the hiring cycle but often leads to the loss of potentially high-value candidates who are overlooked. This project is motivated by the critical need to create a fast, objective, and scalable solution for talent acquisition using cutting-edge Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques.

➤ *System Objectives*

- Efficiency and Speed: Streamline the initial screening process by an estimated 80% compared to manual review.
- Accuracy and Robustness: Provide highly accurate and robust match scores by utilizing a multi-model validation approach.
- Data-Driven Decisions: Present HR teams with a comprehensive, real-time analytics dashboard to guide hiring strategy, allowing them to visualize skill demands and market trends.
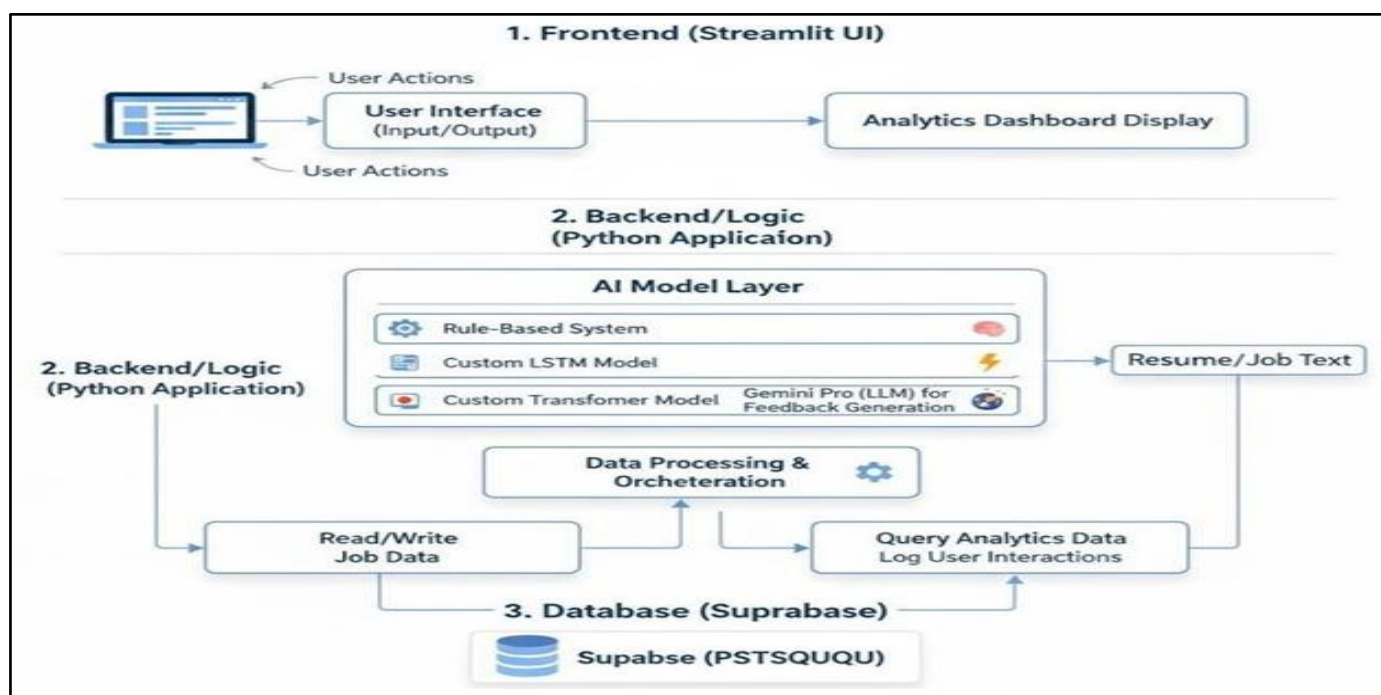


Fig 1 Architecture of the System

## II. SYSTEM ARCHITECTURE AND METHODOLOGY

The system follows a three-tier architecture, ensuring separation of concerns between the user interface, application logic, and data storage.

➢ Frontend: The Streamlit UI handles all user interactions (file uploads and viewing results).
➢ Backend/Logic: The Python application layer manages the data processing, model orchestration, and API calls.
➢ Database: Supabase (PostgreSQL) provides persistent data storage and retrieval.
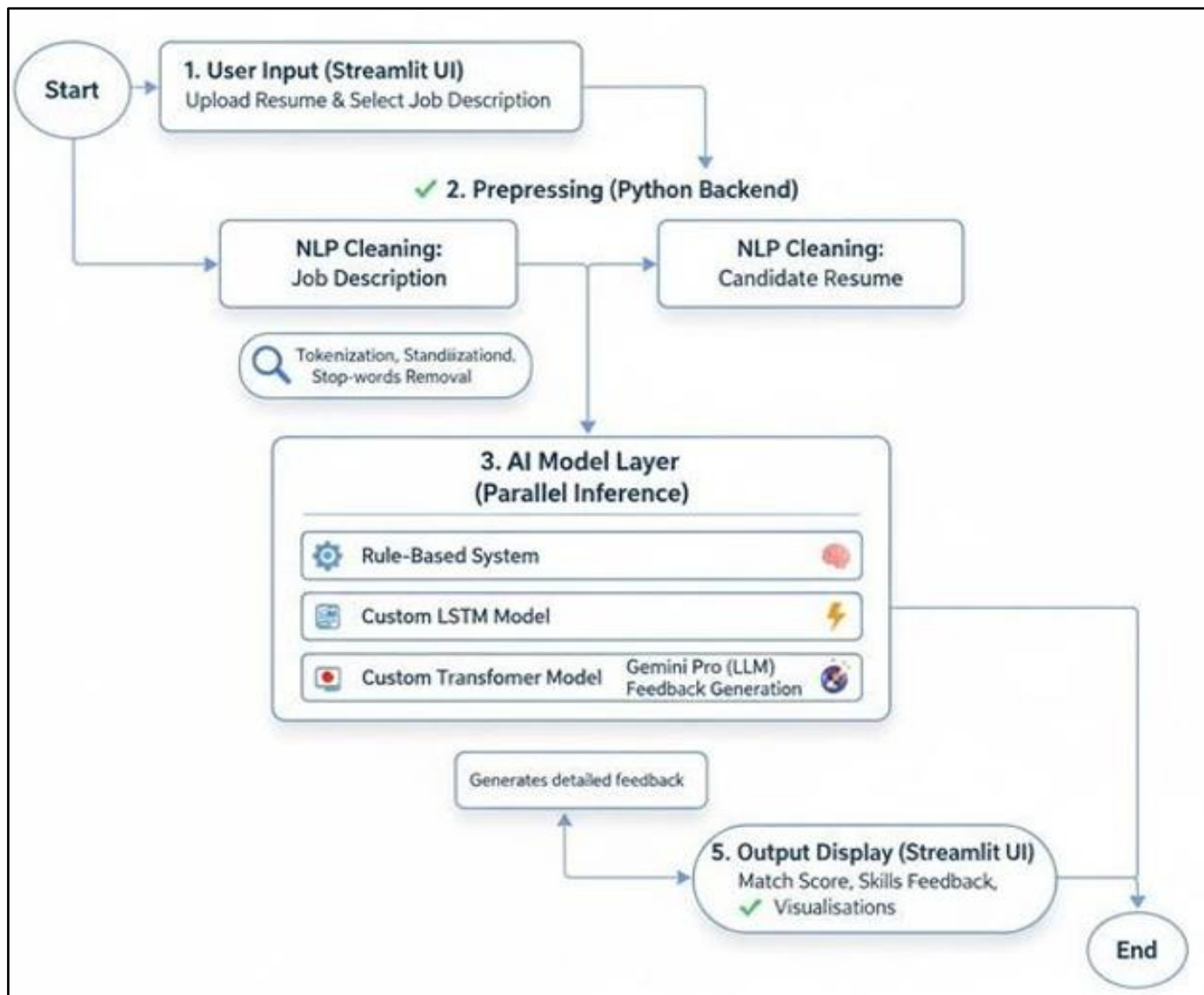


Fig 2 Data Flow

## III. FEATURES AND USER INTERFACE

➢ *Features of HR Professionals*
This section focuses on the administrative control panel for HR users. The core functionality centers around the Job Management Interface, which provides CRUD (Create, Read, Update, Delete) operations for job descriptions. These job descriptions are critical, as they serve as the ground truth and primary feature set against which all applicant resumes are measured.

• Job Listing Management: HR can quickly input new job details (title, requirements, required skills, and experience level) directly into the Supabase database. Existing listings can be easily edited or archived.

• Prediction Data Export: A high-value feature is the one-click export functionality. This allows HR to download all processed candidate data in a standardized spreadsheet format (e.g., CSV or Excel). The exported data typically includes:

✓ Candidate Identifier
✓ Composite Match Score
✓ Score Breakdown by Model (LSTM, Transformer, Rule-Based)
✓ Raw LLM Feedback
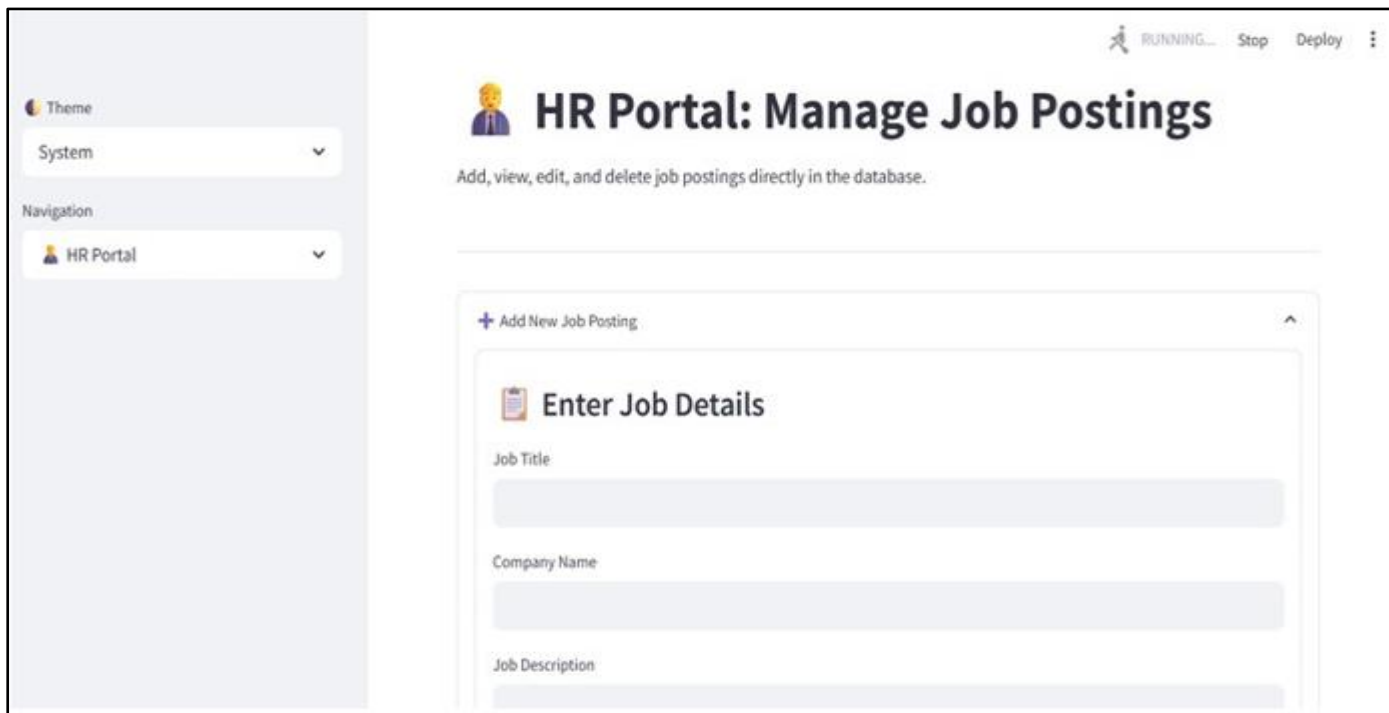✓ Processing Timestamp Ultrasound and receiving the wave that the object reflects.

Fig 3 HR Portal

> *Features of Job Applicants*

The applicant-facing interface prioritizes ease of use and transparency in the matching process.

- Secure Resume Upload: The system accepts common document formats (e.g., PDF, DOCX) and immediately processes the text securely in the backend, ensuring data privacy and integrity.
- Targeted Matching: Applicants select a specific active job from the listing pool. The system then runs their resume against the job description to generate a tailored assessment.

- Composite Match Score: The system returns a prompt, easily digestible score. This score is a weighted aggregation of results from the different AI models, providing a more robust, validated, and reliable assessment than reliance on any single algorithm.
- Constructive Feedback: Detailed, personalized feedback (generated by Gemini Pro) is provided, offering actionable suggestions on missing keywords, relevant skill additions, and formatting improvements.
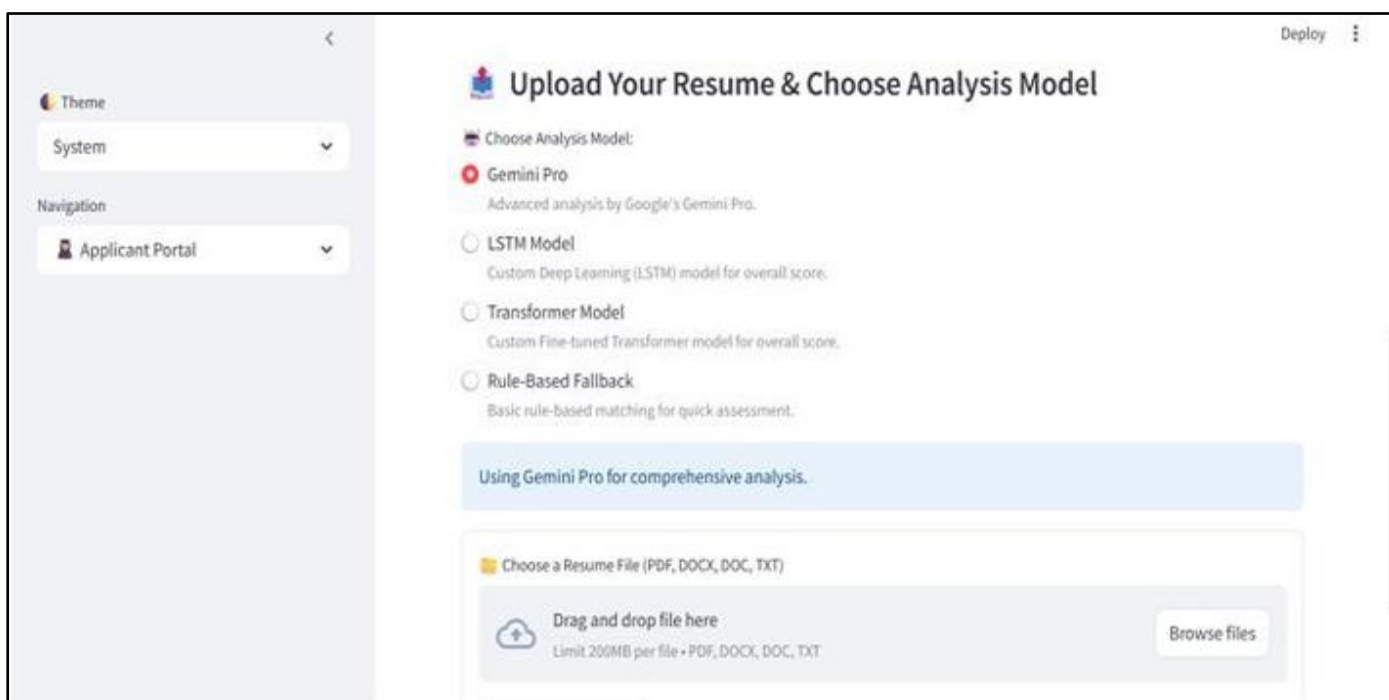


Fig 4 Applicant Portal

## IV. ARTIFICIAL INTELLIGENCE AND NLP MODELS

➢ *Rule-Based Scoring System*
This serves as the fastest, most transparent baseline using deterministic logic and predefined weightings.

- Logic: Scores based on exact keyword matching for non-textual criteria (e.g., years of experience, specific certifications).
- Weighting: Criteria are assigned specific weights; the final score is an accumulation of satisfied weighted requirements.
- Utility: Provides a quick, auditable score for mandatory requirements, ensuring immediate filtering of unqualified candidates.

➢ *Deep Learning Models*
These models are the core for semantic similarity scoring, moving beyond simple keywords.

- Custom LSTM Model: A Recurrent Neural Network effective for processing sequences of text. It is trained to map vector representations of the resume and job to an initial match probability.
- Custom Transformer Model (Fine-tuned DistilBERT): Uses attention mechanisms for superior contextual understanding. It is fine-tuned for the task of semantic similarity, outputting a highly accurate match score by recognizing skills despite variations in phrasing.

➢ *Large Language Model (Gemini Pro)*
The Gemini Pro LLM is leveraged for its generative and reasoning capabilities, not primary scoring.

- Purpose: To provide nuanced, human-readable, and personalized feedback.
- Contextual Input: The LLM receives the raw text and the numerical scores from other models via prompt engineering.
- Outputs: Generates an explanation of the composite score, specific actionable suggestions, and identifies semantically related missing skills.
- Mitigation: Adds a crucial layer of human-like context to offset the "cold" nature of numerical filtering.

➢ *Model Training*
This section ensures the reliability and effectiveness of the deep learning models.

- Dataset: Training requires a robust, labeled dataset of diverse resume job pairs to teach the models the concept of "fit."
- Training Strategy: Models are trained using cross-validation and early stopping to ensure generalization and prevent overfitting.

## V. RESULT, CONCLUSION, FUTURE SCOPE
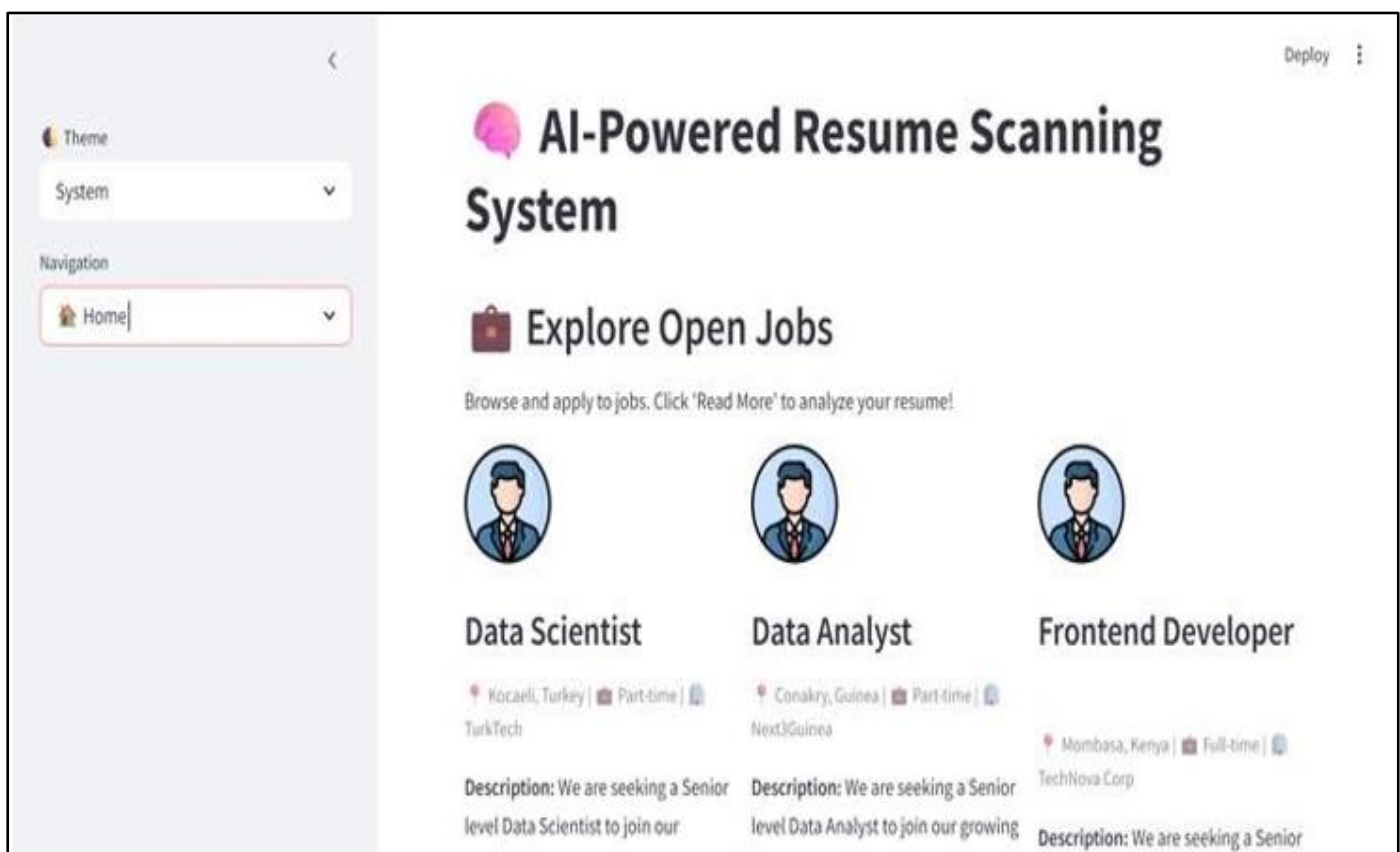
➢ *Result*

- *Home Page*
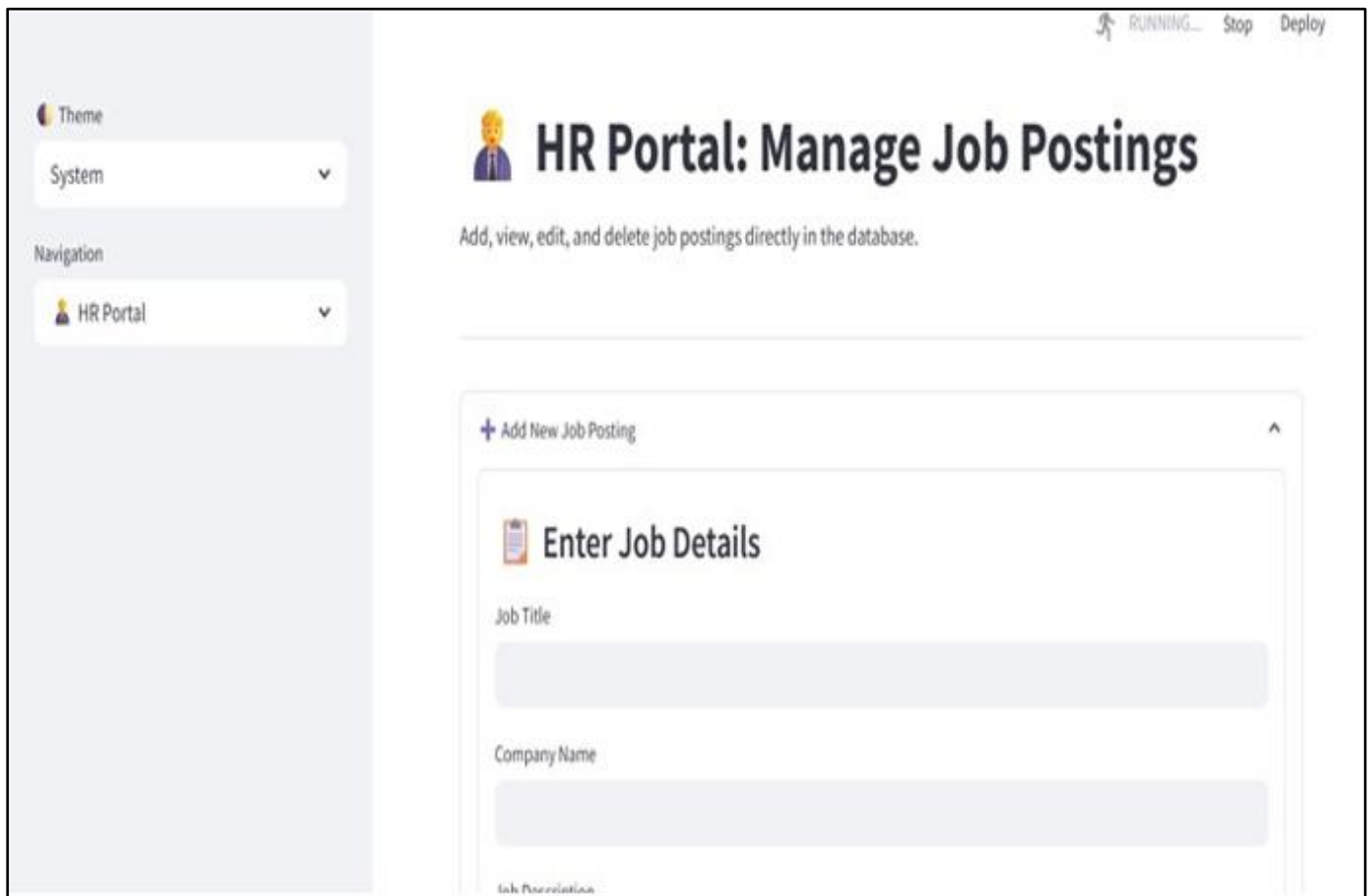


Fig 5 Home Page

- *HR Portal*



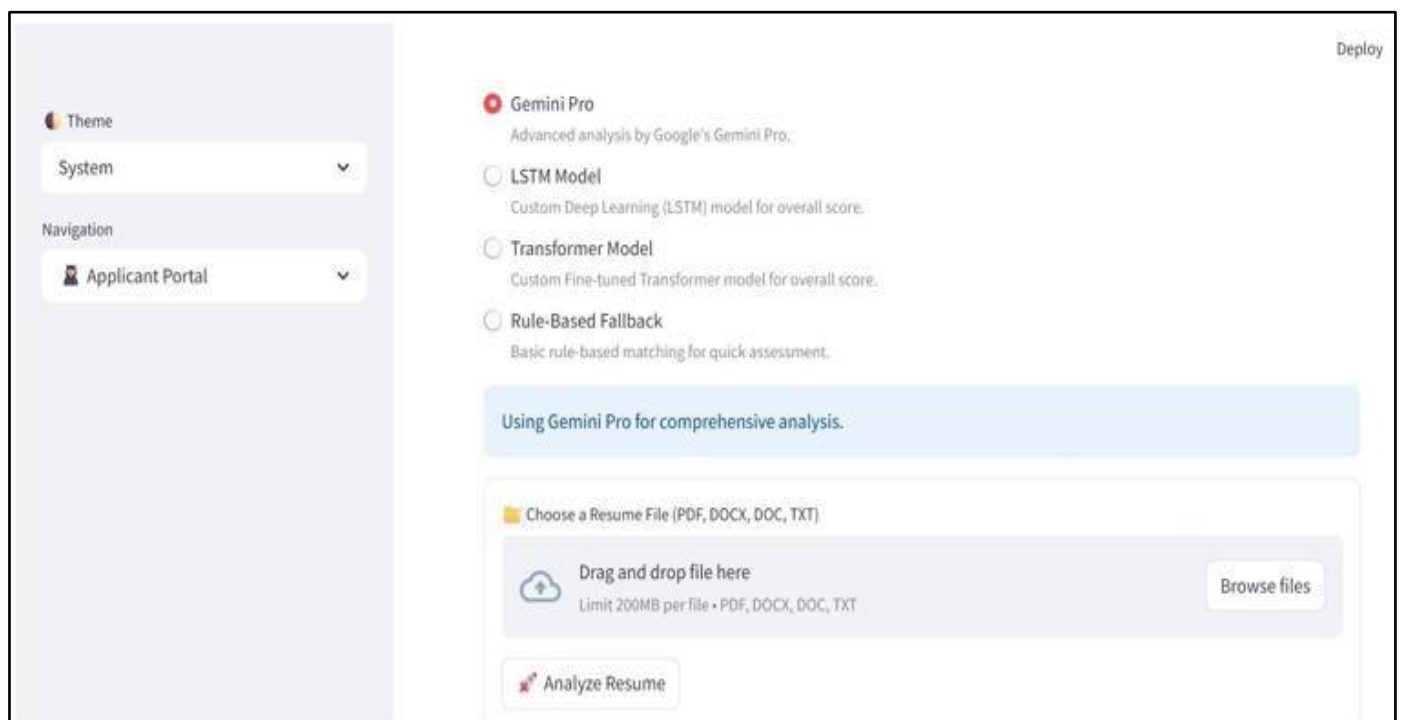Fig 6 HR Job Postings Page

- *Applicant Portal*



Fig 7 Applicant Portal Page
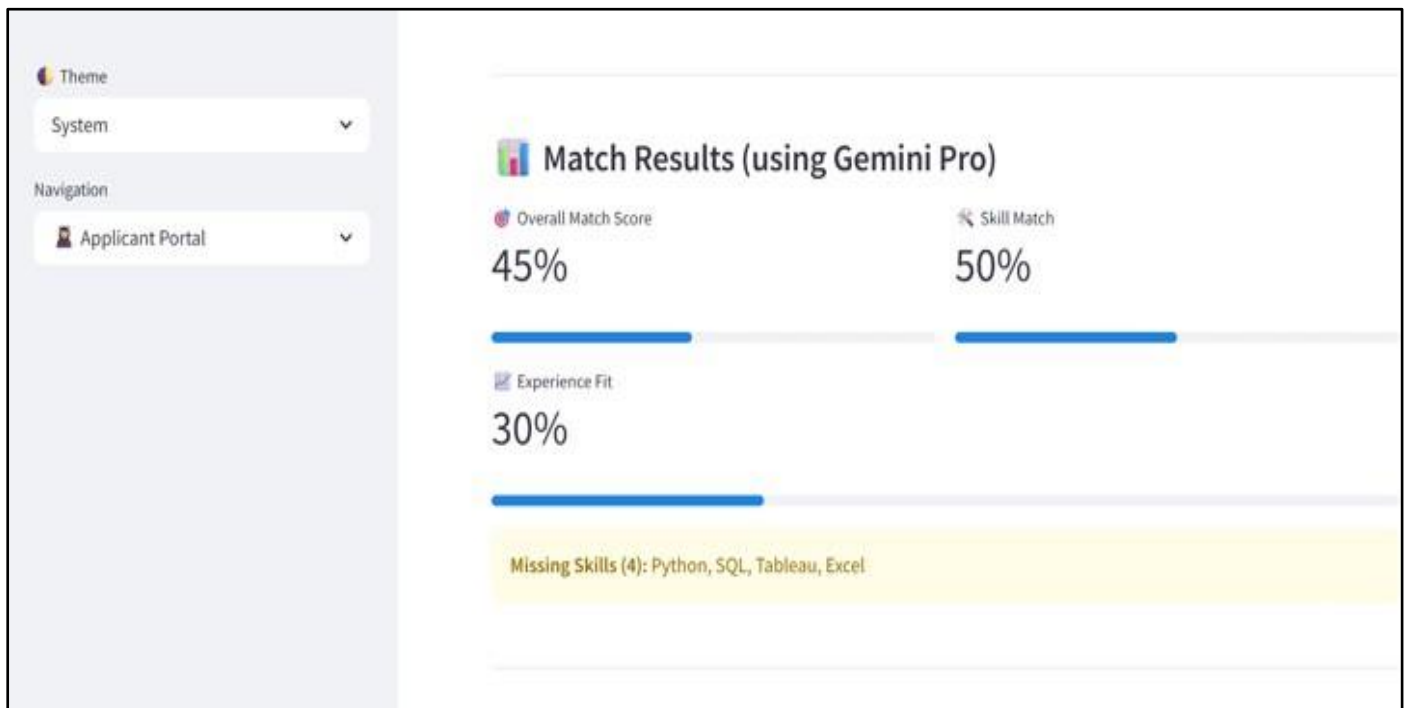
- *Resume Analysis*



Fig 8 Resume Analysis Results

> *Conclusion*

The AI-Powered Resume Screening System successfully met its primary objective of providing an automated, objective, and efficient solution for candidate screening. By integrating a sophisticated multi-model strategy—combining the speed of rule- based logic, the contextual accuracy of the Transformer/DistilBERT model, and the reasoning power of the Gemini Pro LLM the system delivers a composite score that is both reliable and actionable.

The project demonstrates a successful application of contemporary MLOps principles, deploying a complex AI pipeline through a user-friendly Streamlit interface and securing data via Supabase. The outcome is a powerful tool capable of drastically reducing the time spent on manual resume review while offering personalized, constructive feedback that benefits both HR professionals and job applicants. The project is positioned for commercialization through the proposed future enhancements.

> *Future Scope*

- Full ATS Integration: Develop seamless API integration (via Webhooks) with major Applicant Tracking Systems (e.g., Greenhouse, Lever) to automatically push the generated match scores and LLM feedback directly into the existing HR candidate workflow.
- Authenticated User Management: Implement a robust Role-Based Access Control (RBAC) system to manage separate, secure accounts for HR administrators and applicants, enhancing security and compliance.
- Explainable AI (XAI) and Bias Auditing: Integrate visual XAI tools (e.g., attention heatmaps) to show which

sections of the resume contributed most to the score. Implement continuous bias detection metrics to audit model fairness.
- Multi-Modal Input Analysis: Extend the system to process non-text inputs, such as analyzing the context and activity from a candidate's LinkedIn profile or GitHub repository (via external APIs) to gather richer, real-time professional context.

## REFERENCES

[1]. Vaswani, A., et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems (NIPS), 30. (Source for the foundational Transformer architecture.)
[2]. Sanh, V., et al. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108. (Source for the specific deep learning model used.)
[3]. Google. (2024). Gemini API Documentation: Generative AI. Retrieved from https://ai.google.dev/docs
[4]. Streamlit. (2024). Streamlit Documentation. Retrieved from https**://**docs**.**streamlit**.**io