

AI Resume Scanner for HRM Integration

s3926243 - Nguyen Hoang Khanh Duc
s3926553 - Nguyen Thi Phuong Anh
s3978141 - Than Phung Dung
s3975154 - Nguyen Ba Lam Quang Thai
s3979170 - Nguyen Chi Nghia

Industry Supervisors:
Mr. Le Vu Hoang – Chief AI Officer
Mr. Do Van Vu – Project Manager
VMO Holdings

Academic Supervisors:
Dr. Sreenivas Tirumala
Dr. Anh Hoang



Scan for more information

BACKGROUND & MOTIVATION

Modern recruitment faces growing pressure from globalization, online job portals, and a surge in applicants. Traditional processes are slow, repetitive and error-prone, while current HRM systems rely on limited keyword searches. Advanced AI tools exist but are often too expensive or opaque for SMEs. To address this, we developed an AI Resume Scanner that converts unstructured resumes into structured, searchable profiles. It delivers:

- Accurate parsing of text- and image-based resumes
- Context-aware candidate-job matching with semantic embeddings
- Scalable, explainable AI integration

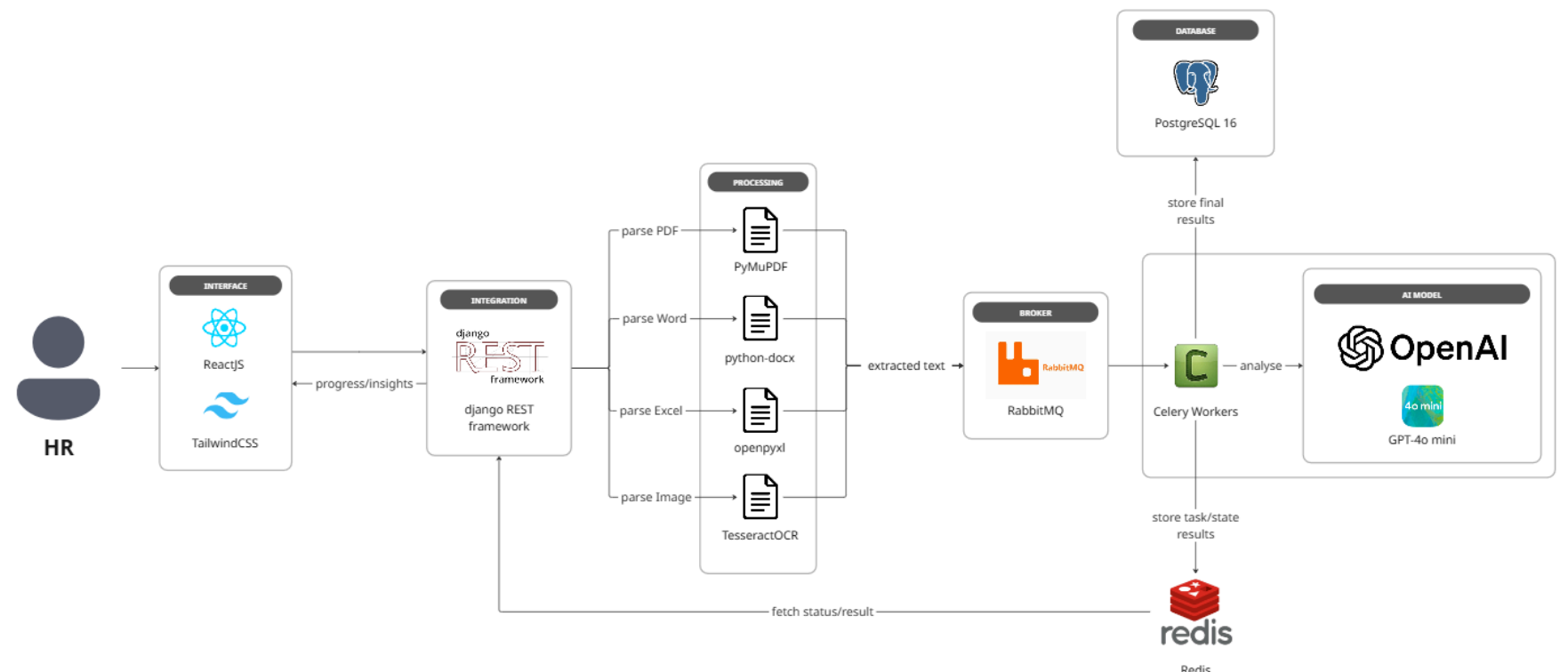
In collaboration with our industry partner, VMO Holdings, we adapted our scope to align with real-world HR practices. Their feedback guided us to prioritize transparent filtering, advanced search tools, and usability, instead of opaque AI scoring systems. This partnership ensures that our solution is not only technically feasible but also industry-relevant, equipping HR professionals with a practical tool to reduce inefficiencies and shorten the recruitment cycle.

OBJECTIVE

- **Automate Resume Parsing:** Extract structured data from PDF, Words Document, Excel spreadsheet and image-based CVs.
- **Error Reduction:** Maintain <1% error rate, surpassing manual entry.
- **Scalability:** Handle 10,000+ resumes with sub-second processing.
- **Practical Usability:** Provide an intuitive HR dashboard with advanced search and filters.
- **Broader Impact:** Extendable AI framework for industries like legal, insurance, and customer service.

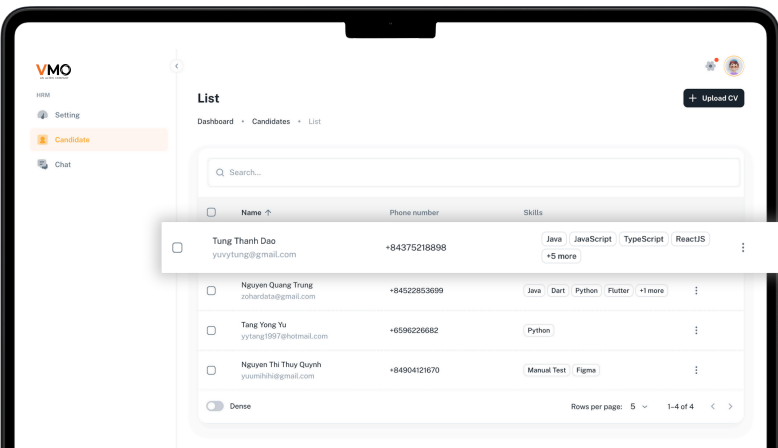
METHODOLOGY

- **User Interface:** HR staff use a ReactJS and TailwindCSS dashboard to upload resumes, track progress, and view insights through a responsive, intuitive UI.
- **Integration Layer:** Django REST Framework handles secure communication between frontend and backend, managing data flow and workflow execution.
- **Processing Layer:** Resumes are parsed using:
 - PyMuPDF (PDFs)
 - python-docx (Word)
 - openpyxl (Excel)
 - TesseractOCR (scanned/image-based)
- **Broker & Task Management:** Extracted text is sent to RabbitMQ for queuing; Celery Workers process tasks asynchronously, forwarding data to the AI model.
- **AI Model:** GPT-4o mini performs entity recognition, context analysis, and semantic matching for candidate-job fit.
- **Database & Caching:** Parsed data is stored in PostgreSQL 16; Redis caches task states and results for fast, real-time updates.



System Architecture of AI Resume Scanner

EXAMPLE GENERATION



CONCLUSION & FINDINGS

The AI Resume Scanner shows how automation can transform recruitment by improving speed, accuracy and fairness while reducing inefficiencies in workflow. Beyond hiring, it offers a flexible framework for processing unstructured documents, with applications in legal, insurance and customer service. This project demonstrates the power of AI and LLMs to modernize enterprise workflows for smarter and more efficient processes.

EXPERIMENTS & RESULTS

We evaluated our AI Resume Scanner across four areas: parsing accuracy, semantic matching, scalability and integration. Using 4,500 resumes in diverse formats (PDF, Word, Excel, images), we applied regex, GPT-4o mini prompts and OCR for information extraction. Job-resume matching was tested against curated benchmarks to measure precision and recall.

Scalability was assessed with Locust load tests, processing up to 10,000 resumes in batch mode while tracking latency, throughput and processing time. Finally, integration was validated in VMO's HRM sandbox, confirming end-to-end automation, concurrent user handling and the robustness of our microservice model.

Our experiments showed strong results across all areas. Parsing reached 100% accuracy on basic fields and over 95% on context-dependent fields, with OCR boosting performance on scanned and Canva-style resumes. The semantic engine achieved 85% precision and 80% recall, ensuring solid JD-CV alignment.

The system processed 100+ resumes per minute with 300–500 ms latency, handling 10,000 resumes in under two hours. Integration with the VMO HRM sandbox confirmed seamless automation, while usability testing scored ≥4/5. Microservice deployment remained stable under concurrent requests, validating enterprise readiness and pointing toward future improvements in throughput.

4,500
CVs in different formats

10,000
CVs processed in 2 hours

