# AI RESUME ANALYZER

**Project Documentation**

Submitted in Partial fulfilment of the

Requirements for the award of the Degree of

**MASTER OF SCIENCE (INFORMATION TECHNOLOGY)**

By

**SHRUTI SUDAY HARAYAN**

Seat No. **202**

Under the esteemed guidance of

**Prof. Suryaprakash Upadhyay**

Assistant Professor

DEPARTMENT OF INFORMATION TECHNOLOGY

**RAMANAND ARYA D.A.V COLLEGE, AUTONOMOUS**

(Affiliated to University of Mumbai)

MUMBAI, 400042

MAHARASHTRA

**2025-26**

# RAMANAND ARYA D.A.V COLLEGE, AUTONOMOUS

(Affiliated to University of Mumbai)

MUMBAI, MAHARASHTRA – 400042

## DEPARTMENT OF INFORMATION TECHNOLOGY

## CERTIFICATE

This is to certify that project entitled, "**AI RESUME ANALYZER**", is a bonafide work of **SHRUTI SUDAY HARAYAN** bearing Seat No. **202** Submitted in the partial fulfilment of the requirements for the award of Degree of **MASTER OF SCIENCE in INFORMATION TECHNOLOGY** from the University of Mumbai.

**Internal Guide** **Coordinator**

**External Examiner**

**Date :** **College Seal**

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who helped me complete this research successfully and I came to know about so many new things.

First and foremost, I would like to thank to **Dr. Balchandra shetty, Principal of Ramanand Arya DAV college,** who gave me golden opportunity to do this wonderful project.

I would like to express my special thanks and gratitude to the project guide **Assistant Professor Suryaprakash Upadhyay** for guiding me to do the project work on time and giving support and guidance, which made me complete my project duly. I am extremely thankful to him for providing such nice support and guidance.

I am thankful for and fortunate enough to get constant encouragement, support, and guidance from the teachers of Information Technology who helped me in successfully completing my project work.

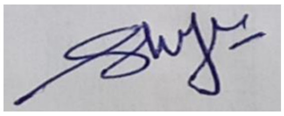
Finally, I would like to thank my family who motivated me and boosted my morale.

# DECLARATION

I hereby declare that the project entitled, **"AI RESUME ANALYZE"** done at RAMANAND ARYA DAV COLLEGE, has not been in any case duplicated to submit to any other universities for the award of any degree. To the best of my knowledge, no one has submitted to any other university.

The project is done in partial fulfilment of the requirements for the award of degree of **MASTER OF SCIENCE (INFORMATION TECHNOLOGY)** to be submitted as final semester project as part of our curriculum.

**SHRUTI S. HARAYAN**



# INDEX

# INTRODUCTION

## 1.1 Background

Recruitment and hiring have always been critical processes in any organization. With the exponential rise in job applications and the increasing complexity of skill requirements, traditional manual resume screening has become inefficient, error-prone, and time-consuming. Human recruiters face challenges such as bias, inconsistency, and fatigue when evaluating large volumes of resumes for specific roles.

To address these limitations, **Artificial Intelligence (AI) powered Resume Analyzers** have emerged as innovative solutions. Such systems use Natural Language Processing (NLP), Machine Learning (ML), and Semantic Analysis to evaluate resumes beyond simple keyword matching. Unlike traditional Applicant Tracking Systems (ATS), which rely heavily on exact keyword matches, AI-based models (e.g., Sentence-BERT, Transformers) enable semantic understanding of resumes, allowing recruiters to identify relevant skills, experiences, and qualifications even if they are expressed differently than in the job description.

Given the growing interest in AI for recruitment, a research-driven **AI Resume Analyzer** can provide a deeper understanding of how intelligent automation can reduce hiring inefficiencies, enhance fairness, and improve candidate-job matching accuracy.

## 1.2 Problem Statement

The hiring process often faces several challenges:

- Recruiters spend excessive time manually filtering resumes, which slows down the recruitment cycle.
- Traditional ATS systems mostly rely on keyword matching, which leads to high false positives (resumes that match keywords but lack relevant experience) and false negatives (resumes with required skills but expressed differently).
- Bias in resume screening, whether conscious or unconscious, often affects diversity and fairness in hiring.
- Job seekers are frequently unaware of how their resumes are interpreted by ATS, leading to mismatches and missed opportunities.

Therefore, there is a need for an **AI-powered Resume Analyzer** that not only matches resumes with job descriptions using advanced semantic similarity techniques but also provides feedback to candidates on how to improve their resumes. This research seeks to bridge the gap between keyword-based ATS systems and context-aware, intelligent recruitment solutions.

## 1.3 Research Objectives

The major objectives of this study are:

- **To design and develop** an AI-powered Resume Analyzer that leverages NLP and semantic similarity models for improved resume-job matching.
- **To compare** the effectiveness of AI-based semantic matching with traditional keyword-based ATS systems.
- **To evaluate** the accuracy, precision, recall, and fairness of the AI Resume Analyzer in real-world recruitment datasets.
- **To provide** candidates with actionable insights, such as missing skills, relevant keywords, and match scores, thereby enhancing resume quality.
- **To explore** how AI-powered tools can reduce human bias and improve the overall fairness and inclusivity of recruitment.
- **To contribute** to the body of research on AI in recruitment by proposing a framework for context-aware and fair resume analysis.

## 1.4 Motivation of the Study

The core motivation for this research stems from the urgent need to streamline the recruitment process while upholding fairness, efficiency, and transparency. Existing manual methods and even basic applicant tracking systems (ATS) fall short of addressing these needs in a scalable manner.

By leveraging AI and ML, there is an opportunity to bridge the gap between job seekers and recruiters, making the hiring process more accessible, precise, and informative for all stakeholders.

- **Recruiter's Motivation**:
  Modern organizations are overwhelmed with job applications, sometimes receiving thousands of resumes for a single position. Recruiters require efficient, unbiased, and accurate systems that can streamline candidate shortlisting.
- **Candidate's Motivation**:
  Many skilled candidates are rejected due to poorly optimized resumes that fail to pass through ATS filters. Providing job seekers with AI-powered insights on resume improvements can help them showcase their skills more effectively and increase job opportunities.
- **Researcher's Motivation**:
  While many commercial ATS solutions exist, there is limited academic research on integrating **semantic similarity models** (such as Sentence-BERT, SBERT, or transformer-based encoders) into resume analysis. Conducting this study bridges the gap between theoretical research and practical applications in recruitment technology.
- **Societal Motivation**:
  The study contributes to fair hiring practices by reducing unconscious bias and ensuring that candidates are evaluated more on their actual skills and experiences rather than superficial keyword matching.

Ultimately, this research aims to advance the state of AI-powered recruitment tools, promoting ethical, transparent, and efficient hiring practices in an increasingly digital world.

# Description of Problem

Recruitment has become increasingly complex in the modern job market, where organizations often receive hundreds or even thousands of applications for a single job posting. Traditional manual resume screening methods are not only **time-intensive** but also highly **susceptible to human errors and biases**, making them unsustainable in large-scale hiring. To cope with the volume of applications, many companies now employ Applicant Tracking Systems (ATS). However, existing ATS solutions are limited in several critical ways.

Most ATS platforms rely on **keyword-based filtering**, where resumes are matched to job descriptions by detecting the presence of specific words or phrases. While this method offers speed, it introduces significant shortcomings:

- **Qualified candidates may be rejected** if their resumes use synonyms or alternative expressions instead of the exact keywords from the job description.
- **Unqualified candidates may be shortlisted** if they strategically include keywords, even without relevant experience.
- **Job seekers receive little to no feedback**, leaving them unaware of why their applications were unsuccessful.
- **Fairness and inclusivity remain a challenge**, as automated screening can reinforce hidden biases based on factors such as gender, educational background, or naming conventions.

These limitations result in inefficiencies for recruiters and missed opportunities for talented candidates, ultimately hindering the overall recruitment process.

Overcoming these challenges is therefore of **critical importance**. From an organizational perspective, integrating advanced AI-based techniques can improve the **accuracy, efficiency, and fairness** of resume screening, ensuring that recruiters focus on genuinely qualified candidates. For job seekers, providing **transparent feedback** and highlighting **missing skills** can empower them to improve their resumes and increase their chances of securing employment.

From a research standpoint, developing an AI Resume Analyzer that leverages **semantic similarity models (e.g., SBERT)** contributes to the growing field of intelligent recruitment technologies, bridging the gap between academic research and industry practices.

Thus, the central problem addressed in this study is the **inefficiency, lack of semantic understanding, limited transparency, and potential bias** in existing resume screening systems, and the importance of designing an AI-powered solution that overcomes these limitations to enhance both recruiter and candidate experiences.

## 2.1 Research Gap

Despite the rapid advancement of recruitment technologies and the widespread adoption of Applicant Tracking Systems (ATS), several critical gaps remain unaddressed in the current literature and industry practices:

1. **Over-Reliance on Keyword Matching**

- Most ATS solutions focus on matching resumes with job descriptions based on exact keywords.
- This approach fails when candidates use synonyms, different terminologies, or domain-specific phrases to describe their skills and experiences.
- As a result, qualified candidates are often overlooked, while irrelevant candidates may pass through screening.

2. **Lack of Semantic Understanding**

- Existing systems rarely incorporate advanced Natural Language Processing (NLP) models such as Transformer-based architectures (e.g., BERT, SBERT).
- Without semantic similarity analysis, ATS cannot fully understand context, intent, and conceptual equivalence between resumes and job descriptions.

3. **Limited Feedback Mechanisms for Job Seekers**

- Most ATS act as "black boxes" — candidates are either shortlisted or rejected without understanding the reasons.
- Very few systems provide **explainability** or feedback such as missing skills, keyword recommendations, or suggestions for improvement.

4. **Bias and Fairness Concerns**

- Automated recruitment systems risk amplifying existing biases (e.g., based on gender, name, or educational background).
- There is a research gap in developing **fair, transparent, and bias-aware** AI models for resume screening.

5. **Insufficient Benchmarking and Research Validation**

- While industry solutions exist, academic research on **evaluating and benchmarking AI resume analyzers** against real-world recruitment datasets is limited.
- Few studies compare traditional keyword-based ATS performance with modern AI-driven semantic models in a systematic way.

## Contribution of This Study

To address the above gaps, this research aims to:

- Incorporate **semantic similarity models** (SBERT) into resume-job matching, moving beyond simple keyword matching.
- Develop a framework that provides **explainable feedback** to candidates, enabling them to improve their resumes.

- Explore methods to **reduce algorithmic bias** and ensure fairness in recruitment processes.
- Benchmark the system against traditional ATS to demonstrate improvements in **accuracy, recall, and fairness**.

Thus, this study contributes to both **academic research** and **practical recruitment solutions** by proposing an intelligent, fair, and feedback-driven AI Resume Analyzer.

# Literature Review

## 3.1 Introduction to the Literature Review

Recruitment technology has shifted from manual screening and keyword-driven Applicant Tracking Systems (ATS) toward **semantic, context-aware NLP**. Recent works show that **Transformer-based embeddings** (e.g., SBERT) substantially improve resume–JD alignment versus lexical matching alone, while regulators increasingly scrutinize fairness and transparency in AI-assisted hiring. This review synthesizes (i) recruitment/ATS evolution, (ii) AI/NLP approaches for resume–JD matching, (iii) semantic similarity advances, (iv) bias/fairness and regulation, and (v) candidate-facing explainability and feedback—then critically analyzes gaps my system targets.

Building on this overview, the following subsections examine the thematic evolution of recruitment technologies and ATS methodologies in greater detail.

## 3.2 Thematic Review of Prior Work

### 3.2.1 Recruitment and Resume Screening (From Keywords to Semantics)

Early ATS tools emphasized **lexical filters** (exact tokens, Boolean search). While scalable, they miss qualified candidates using synonyms and are vulnerable to "**keyword stuffing**," where hidden or repeated terms inflate scores without true fit. Industry and journalism have documented these failures, motivating semantic alternatives.

Recent academic and practitioner reports argue configuration choices in ATS can **reject capable candidates** and narrow talent pools, signaling the limits of one-dimensional keyword pipelines.

Recognizing the limitations of keyword-based filtering, researchers and practitioners have increasingly turned to artificial intelligence techniques to improve accuracy and fairness.

### 3.2.2 AI and NLP in Recruitment

ML/NLP methods automate parsing, skill extraction, and matching. Surveys and empirical studies outline pipelines that tokenize resumes/JDs, extract entities/skills, and compute similarity scores for ranking or recommendation. Comparative works show AI-based screening increases throughput but surface **new concerns** around transparency and bias.

Recent research prototypes explore **embedding-based job–resume matching** (e.g., Resume2Vec; vector search approaches), often outperforming classical baselines in human-aligned metrics and ranking quality.

### 3.2.3 Semantic Similarity and Transformer-based Models

Sentence-BERT (SBERT) is a transformer-based language model optimized for generating semantically meaningful sentence embeddings that enable efficient cosine similarity computations. Commonly used in resume-JD matching, SBERT effectively captures contextual relationships beyond simple keyword overlap. However, semantic embeddings alone may produce false positives on generic content; thus, hybrid scoring methods that integrate lexical keyword signals with semantic metrics have gained prominence. These hybrid approaches balance precision and recall by penalizing semantic similarity when keyword overlap is absent, mitigating adversarial keyword stuffing, and improving the practical robustness of ATS.

**BERT** introduced deep bidirectional language representations that improved sentence-pair tasks but were costly for large-scale retrieval. **Sentence-BERT (SBERT)** adapted BERT into a **Siamese architecture** that yields sentence embeddings suitable for **cosine-similarity** search—enabling efficient, accurate semantic matching at scale. Numerous resume-JD studies now employ **SBERT + cosine.** SBERT has become a preferred tool in resume–JD matching tasks because it captures contextual meaning beyond simple keywords, allowing for robust matching even when different terminology or phrasing is used.

Cutting-edge, domain-specific variants for labor data have appeared, e.g., **CareerBERT** (resume–ESCO job matching), and **SkillMatch** (self-supervised skill relatedness benchmark and SBERT adaptation to job-ad skill co-occurrence), underscoring the field's shift to **skill-centric, semantic modeling**.

While transformer-based embeddings have advanced semantic understanding, hybrid scoring methods that combine these with keyword signals have emerged as practical solutions to address real-world challenges.

### 3.2.4 Hybrid Scoring: Combining Semantics and Keywords

Several IR and applied AI sources promote **hybrid search** that blends **keyword signals** with **semantic similarity** to improve precision/recall trade-offs. This hybridization mitigates false positives from embeddings alone and false negatives from sparse lexical overlap. The **strictness penalty** (down-weighting semantic similarity when overlap=0) is a pragmatic defense against generic language and adversarial keyword stuffing documented in practice.

Empirical works on resume matching increasingly report **cosine-based embedding scores** complemented by auxiliary features (skills overlap, section weights). Recent comparisons (e.g., BERT vs. LLM-rankers) also highlight **context length and explainability trade-offs**, supporting the case for stable, interpretable hybrids in production-like pipelines.

### 3.2.5 Fairness, Bias, and Regulatory Guidance

Regulators stress that AI hiring tools must **avoid disparate impact** and remain explainable. The U.S. **EEOC** launched its **AI & Algorithmic Fairness Initiative** and published guidance spotlighting risks of discrimination when automated tools influence selection. Ongoing legal scrutiny (e.g., suits involving hiring software) illustrates the compliance stakes for resume-screening systems.

Academic studies show AI recommendations can **shift human decisions**—sometimes mitigating, sometimes introducing different biases—hence the need for **transparent reasoning and bias controls** in recommender interfaces.

### 3.2.6 Candidate Experience, Transparency, and Feedback

User studies indicate applicants often **welcome AI** when it is perceived as useful and easy to use, but trust hinges on **explanations** and **actionable feedback**. Systems that disclose why a score was assigned and how to improve (e.g., missing skills, ATS-unfriendly formatting) bolster acceptance and perceived fairness.

These insights into user trust and transparency provide a vital context for analyzing the strengths and weaknesses of existing methods.

## 3.3 Critical Analysis of Prior Studies

- **Strengths of Semantic Models:** SBERT-style embeddings deliver robust **contextual similarity** with tractable retrieval time, consistently outperforming bag-of-words and TF-IDF on sentence-level relevance and clustering. However, pure embeddings may over-score resumes sharing **generic language** with JDs (e.g., "team player," "fast-paced environment"), justifying **keyword/skill constraints** as a counterbalance.
- **Limits of Keyword-Only ATS:** Lexical filters,which rely on exact token matching and Boolean search logic are brittle to synonyms and format variance, and are exploitable via **keyword stuffing**. Research and reportage show configuration pitfalls can silently **filter out qualified candidates**, motivating hybrid or semantic-first scoring and **format checks** (e.g., tables/graphics warnings).
- **Explainability & Feedback:** Many academic systems report accuracy gains but provide **limited candidate-facing explanations**. Inclusion of **reason codes**, **matched/missing skills**, and **tone-aware recommendations** directly addresses this gap and aligns with adoption research findings.
- **Fairness, Compliance, and Evaluation:** A growing body of guidance urges **bias testing** and documentation for AI hiring tools. Yet, benchmark datasets with **demographic annotations** and standardized fairness metrics for resume matching remain scarce, complicating apples-to-apples comparisons.
- **Latest Research Directions:** Domain-adapted models (CareerBERT), **hypothetical/contrastive finetuning** for sparse labels (ConFit v2), and **vector-search pipelines** indicate momentum toward scalable, skill-aware matching—suggesting future gains for systems, especially if they fine-tune on **HR-domain skill graphs** and adopt **hybrid retrieval**.

## 3.4 How This Research Method Fits the Literature

- **Semantic core: SBERT cosine similarity** between cleaned resume and JD → aligns with best-practice semantic matching and with multiple recent studies reporting strong retrieval/ranking performance.
- **Keyword/skill signal: Normalized skill overlap** using a master skills list + **aliases** → mirrors hybrid IR guidance; directly combats synonym gaps and improves precision.
- **Hybrid scoring: Fixed 0.5/0.5 weights** for semantic vs. keywords → consistent, reproducible ranking (and easily tunable via validation). Hybrid approaches are widely recommended to balance recall and precision.
- **Strictness penalty when overlap=0:** Penalizing resumes with **zero skill overlap** prevents generic SBERT similarity from inflating scores—an effective defense consistent with literature noting adversarial keyword behaviors and generic-language pitfalls.
- **Explainability:** explain_ats_score() and recruiter-style get_recommendations() address adoption and trust concerns surfaced in applicant-perception research.
- **ATS-format checks:** Warnings for **tables/graphics/HTML** reflect known parsing issues in legacy pipelines and help candidates pass automated filters.

## 3.5 Ablation Study

To validate the contributions of different components in the proposed hybrid model, an ablation study is recommended. This involves comparing variants of the system:

**SBERT-only:** Matching based solely on semantic embeddings.

**Keyword-only:** Matching based solely on keyword overlaps.

**Hybrid:** Combination of semantic and keyword approaches.

**Hybrid + Strictness Penalty:** Final version, which incorporates penalties for irrelevant keyword inflation.

Such experiments demonstrate the incremental value of each module and strengthen the empirical justification for the proposed design.

## Research Papers CVisionary Overcomes Gaps In

### 1. "Effectiveness of Applicant Tracking Systems in Recruitment and Selection" (2025)

**Citation**: Greeshma M. & Dr. M. Sathis Kumar (2025). International Research Journal of Modernization in Engineering Technology and Science.

**Key Gaps Identified in This Paper:**

- **Lack of transparency**: Only 70% agreed ATS provided transparent selection process.

- **Resume rejections due to keyword mismatches**: Over-reliance on keyword-based algorithms filters out qualified candidates.
- **Limited feedback**: 25% of candidates experienced lack of communication about application status.
- **Technical glitches**: 28% faced difficulties with document uploads.

**How CVisionary Addresses These Gaps:**

| Research Gap | CVisionary Solution | Evidence |
|---|---|---|
| **Lack of Transparency** | Full scoring breakdown with detailed explanation of how scores are calculated | Shows transparent formula and example calculations |
| **Keyword-Only Matching** | **Hybrid approach**: 50% semantic similarity + 50% keyword overlap using SBERT | Algorithm explanation shows balanced weighting |
| **No Detailed Feedback** | Comprehensive feedback with matched skills, missing skills, and improvement tips | Show detailed explanation sections |
| **Poor Resume Parsing** | Support for both PDF and DOCX with clear error handling | Shows successful file upload interface |

**2. "Exploring Bias in AI-Driven Resume Screening: A Fairness Analysis and Mitigation Approach" (2025)**

**Citation:** Anushka Mishra. Available on SSRN, February 2025

**Key Research Gaps:**

- **Hidden bias in AI algorithms**: Models inherit historical biases from training data.
- **Lack of algorithmic transparency**: "Black box" decision-making processes.
- **Demographic discrimination**: Systems unfairly advantage/disadvantage based on gender, age, origin.
- **No bias detection mechanisms**: Limited tools to identify unfair patterns.

**CVisionary Solutions:**

| Bias Research Gap | CVisionary Innovation | Supporting Evidence |
|---|---|---|
| **Black Box Algorithms** | **Complete transparency**: Shows exact formula, weights, and calculation steps | Full algorithm disclosure |
| **Historical Bias** | **Skills-first approach**: Focuses on technical skills rather than demographic markers | Fairness & Blind Review feature |
| **No Explanation** | **Detailed breakdowns**: Explains why each score was given with specific examples | Comprehensive result explanations |
| **Demographic Discrimination** | **Anonymous analysis**: Analyzes content, not personal details | Emphasizes skills-first approach |

### 3. "Enhanced Resume Screening for Smart Hiring Using S-BERT" (2024)

**Citation**: Asmita Deshmukh and Anjali Raut, International Journal of Advanced Computer Science and Applications, Vol. 15, No. 8, 2024

**Technical Gaps Identified:**

- **Limited semantic understanding**: Traditional keyword matching misses contextual relevance.
- **No ranking mechanism**: Basic similarity without comprehensive scoring.
- **Speed limitations**: 0.233 seconds per resume but no comprehensive analysis.

**CVisionary's Implementation:**

| Technical Gap | CVisionary Enhancement | Evidence |
|---|---|---|
| **Basic SBERT Usage** | **Advanced hybrid model**: SBERT + keyword matching + experience gap analysis | Shows multi-factor scoring |
| **Simple Ranking** | **Comprehensive scoring**: Includes strictness factors and experience penalties | Complex scoring algorithm |
| **Limited Analysis** | **Full ATS analysis**: Skills gap, experience mismatch, actionable recommendations | Complete analysis dashboard |

### 4. "AI-Based Resume Screening Systems: Opportunities and Ethical Concerns" (2025)

**Citation:** Sriranjani S K and Dr. H. Kalaiarasi | IJRTI | Volume 10, Issue 6 June .International Journal for Research Trends and Innovation (www.ijrti.org)

**Ethical Gaps CVisionary Addresses:**

- **Lack of explainability**: Most AI systems don't explain their decisions.
- **Bias perpetuation**: Systems inherit training data biases.
- **Privacy concerns**: Unclear data handling practices.

**CVisionary's Ethical Solutions:**

- **Complete explainability**: Every decision is fully explained with reasoning.
- **Bias mitigation**: Transparent, fair scoring methodolog.
- **User control**: Students and recruiters have full access to their data.

### 5. "A Descriptive Study on Applicant Tracking System:Automation software for Recruitment and Selection",2019

**Citation:**Mr.Suraj M, Ms.Aruna Kumari K, Ms.Binila B Chandran, 2019 IJRAR February 2019, Volume 6, Issue 1

**Gaps Identified**:

- Over-reliance on keywords leads to rejection of qualified candidates who use different wording or resume formats.
- Inaccurate parsing and technical mishaps; resumes with unusual formatting may be unreadable.
- No explanation or feedback for rejected candidates.
- Easy manipulation: Candidates who "game" keywords get through, not necessarily the most relevant ones.

**Cvisionary's Solution**:

- Combines SBERT semantic similarity and keyword overlap for context-aware matching.
- Robust PDF/DOCX parsing with clear user feedback on errors.
- Detailed, actionable feedback and transparency in scoring.
- Strictness penalty in scoring to prevent keyword-stuffing manipulation.

**Research Contribution :**

While existing ATS research identifies critical gaps in transparency (Greeshma & Kumar, 2025), algorithmic bias (SSRN, 2025), and limited semantic understanding (IJACSA, 2024), CVisionary addresses these fundamental limitations through:

1. **Complete Algorithmic Transparency**: Unlike black-box systems, CVisionary provides full disclosure of scoring methodology, weights, and decision factors.
2. **Hybrid Intelligence Approach**: Combines SBERT semantic similarity with keyword matching to overcome limitations of purely keyword-based or purely AI-based systems.
3. **Bias Mitigation Through Fairness**: Implements skills-first, anonymous evaluation that eliminates demographic discrimination while maintaining accuracy.
4. **Comprehensive Feedback Mechanism**: Addresses the 25% communication gap identified in current research by providing detailed, actionable feedback to all users.

# Research Methodology

The **Research Methodology** chapter documents the systematic approach, workflow, and algorithms used to design, implement, and validate a fair, explainable, and semantically-aware ATS (Applicant Tracking System) resume analyzer.

This section provides an in-depth look at each technical component and justifies their selection and configuration.

## System Design Approach

The project adopts a **hybrid architecture** that unifies semantic AI techniques, traditional keyword extraction, and experience analysis to bridge the gap between research and practical recruitment needs. The overarching goal is to create a transparent, fair, and robust automated resume evaluation pipeline.

## Data Collection and Preprocessing

**User Input**: The system ingests resumes (PDF/DOCX) and job descriptions via a web interface.

**Resume Parsing**: Leveraging PDFPlumber and python-docx, the backend extracts raw text from diverse resume file formats. All text is normalized (lowercasing, accent removal, standardized terminology, punctuation and whitespace cleaning).

**Skill Alias and Master List Mapping**: A CSV-driven skill master list and alias mapping ensure consistent recognition of domain-specific competencies regardless of variant naming (e.g., "Py" → "Python").

## Feature Extraction

**Skill Extraction**: Skills are detected in both resumes and JDs using exact matching post-alias normalization, filtering out noise/stopwords to maximize precision.

**Semantic Representation**: Resume and JD texts are vectorized using the SBERT transformer model, enabling robust similarity measurement even when vocabulary differs but meaning overlaps.

**Experience Gap Analysis**: Regular expressions identify explicit experience claims (e.g., "3+ years in Java"). Work periods are extracted from date patterns to compute aggregate years of experience. Gaps or overqualification are highlighted by comparing extracted values against JD requirements.

## ATS Scoring & Feedback

**Hybrid Score Calculation**: The ATS score is computed using:

**Semantic Similarity** (sim): Cosine similarity between SBERT embeddings.

**Keyword Overlap** (overlap): Ratio of matched skills to those required by the JD.

**Strictness Penalty**: If no skills overlap, semantic similarity is downweighted to discourage false positives.

**Experience Penalty**: Substantial experience gaps trigger an additional score reduction.

**Final Score=(sim_weight×semantic_similarity)+(key_weight×keyword_overlap)**

**Explainable Feedback**: The system generates human-friendly explanations for scores: missing skills, overqualification warnings, semantic observations, and actionable tips are displayed. This supports user learning as well as recruiter transparency.

## Algorithm and System Summarization

- Backend implemented in Python (FastAPI), ensuring reproducibility and extensibility.
- Database (SQLite) enables storage of original resumes, scoring metadata, matched skills, warnings, and user data for future audits and cohort analysis.
- Frontend in React provides an interactive dashboard for both students (upload, analyze, feedback) and recruiters (resume filtering, download, JD match).
- Components are designed for explainability, ethics, and fairness, responding to documented research gaps in the field.

## Evaluation Strategy

- Performance is assessed against realistic resumes and JDs.
- Gap analysis is validated by benchmarking extracted years/skills against JD needs.
- User feedback and recruiter acceptance are planned for future empirical validation.

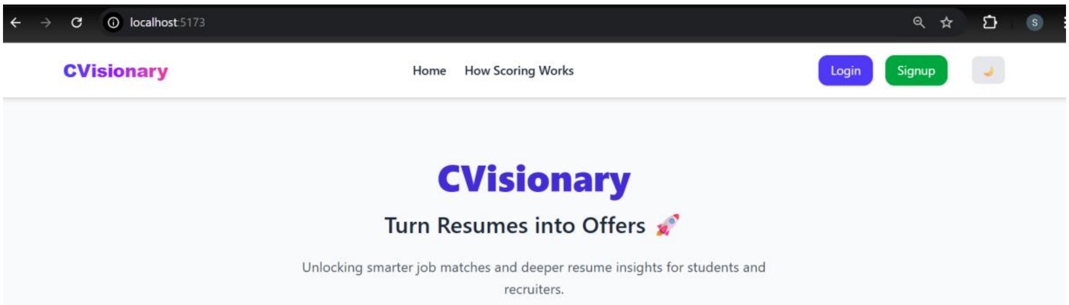
This methodology leverages state-of-the-art AI techniques while foregrounding transparency and actionable guidance, positioning the project at the confluence of practical utility and foundational research advancement.

# Code Implementation

## System Overview

This project is implemented as a full-stack web application with a **React** frontend and a **FastAPI (Python)** backend. The backend exposes REST APIs, manages authentication and data persistence (using **SQLite**), and runs advanced resume parsing and ATS scoring logic. The frontend provides real-time interaction for both students (resume upload, feedback) and recruiters (search, filter, download).

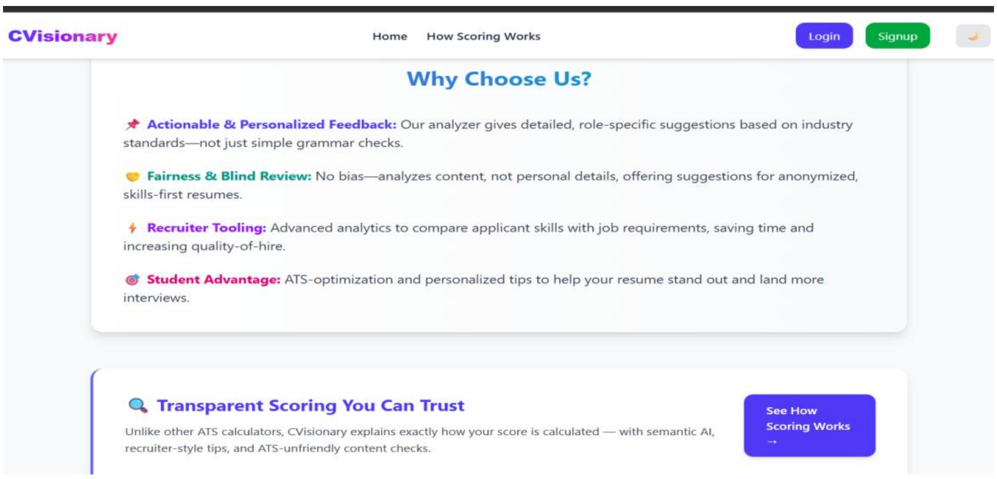
**Homepage View**
The first screen users see is the homepage, which introduces the application's core value proposition. The navigation bar provides access to the main features of the platform, including the Home view, scoring methodology ("How Scoring Works"), and authentication options for login and signup.
The design emphasizes ease of access, clear branding ("CVisionary"), and a concise tagline describing the platform's purpose for both students and recruiters.




This simple and modern layout gives users an immediate understanding of CVisionary's mission—to turn resumes into offers by unlocking smarter job matches and resume insights.

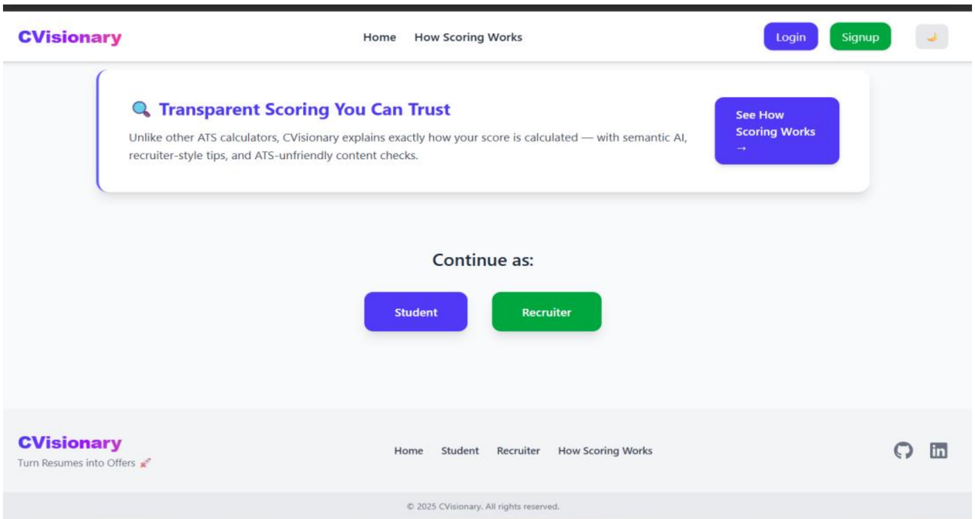
### Why Choose Us? – Feature Highlights

This section of the homepage outlines the core benefits of using CVisionary. Users are reassured with actionable, personalized feedback, fairness through blind review, advanced recruiter analytics, and specific advantages for students aiming to optimize their resumes for ATS and interviews.



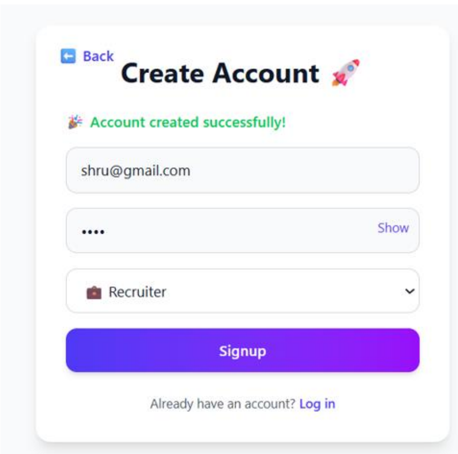
### Role Selection Screen

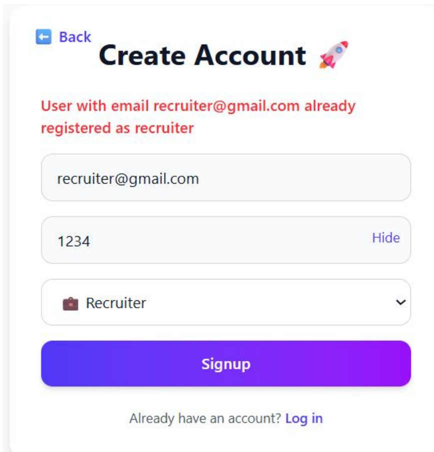
This interface allows users to continue as either a **Student** or **Recruiter**, streamlining the onboarding experience based on their intended use of the platform. Above, CVisionary highlights its commitment to transparency by explaining the “Transparent Scoring You Can Trust” feature, which ensures every ATS score is clear, explainable, and based on fair criteria.




### Signup Page

The signup page allows new users to create an account by entering their email, password, and selecting their role as either “Student” or “Recruiter.” The form features clear input fields and a custom-styled dropdown for quick role selection.



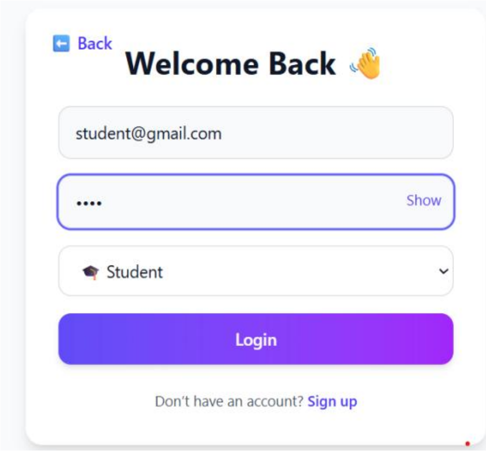



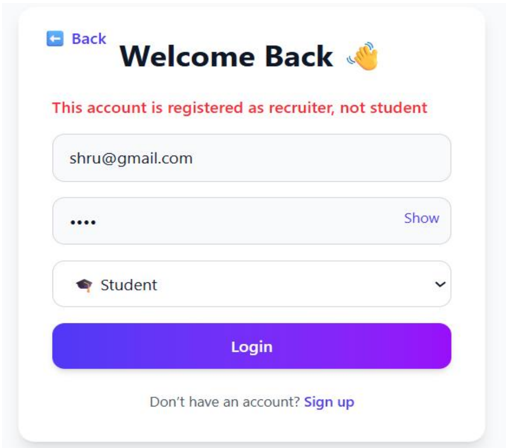

If a user attempts to register with an email that is already associated with an existing account, an immediate error message is displayed, guiding them to log in instead—ensuring clarity and preventing duplicate registrations.

After completing the signup form and submitting valid details, users receive a clear confirmation that their account was created successfully which immediately followed by a prompt to login or begin using their account, supporting a smooth onboarding flow.

### Login Page

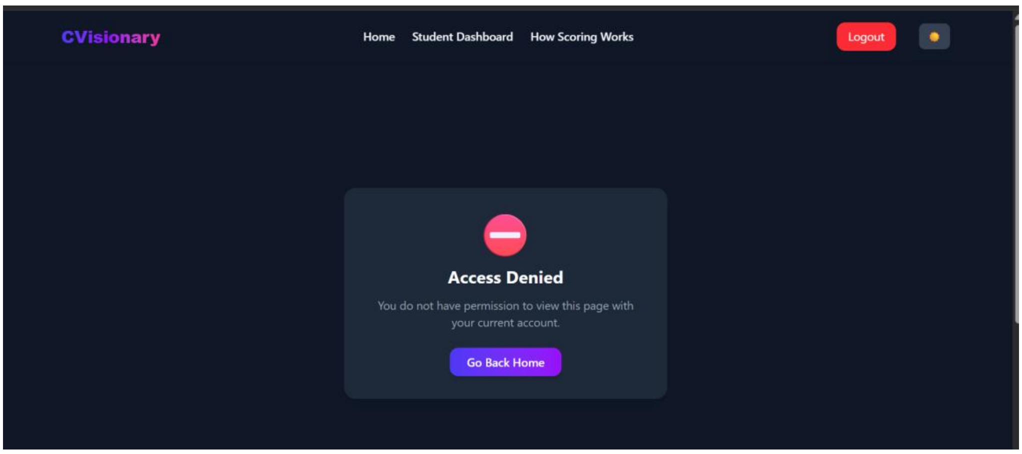
The login page allows registered users to access their accounts securely by entering their email, password, and selecting their role (Student or Recruiter). A clear error message is displayed if the user's selected role does not match the account type (for example, trying to log in as a Student when the email is registered as a Recruiter), ensuring that only authorized roles have access to the relevant dashboard. This enforces strict role-based authentication and guides users toward the correct login procedure for a smooth and secure experience



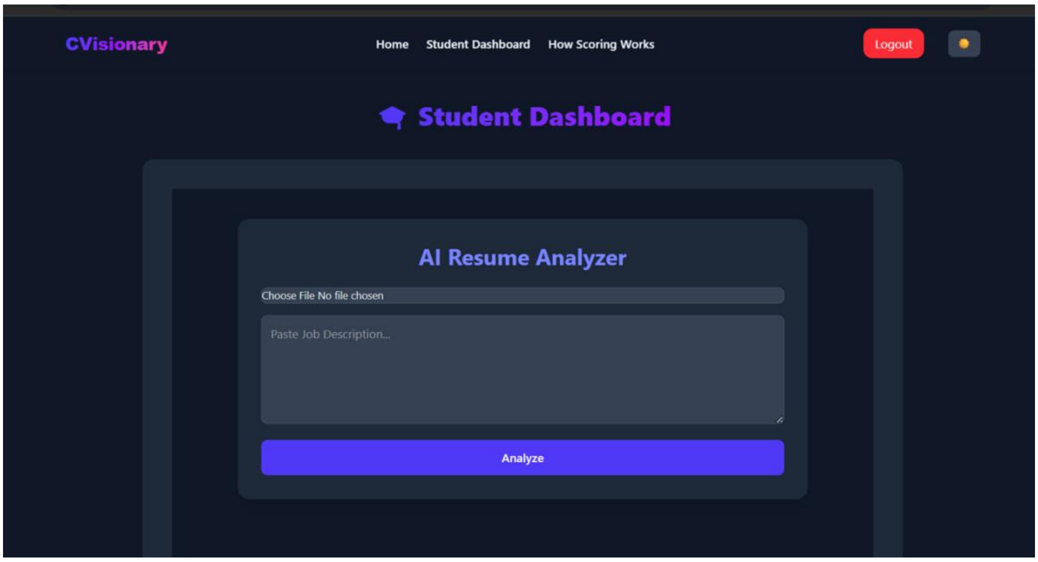

### Access Denied Page

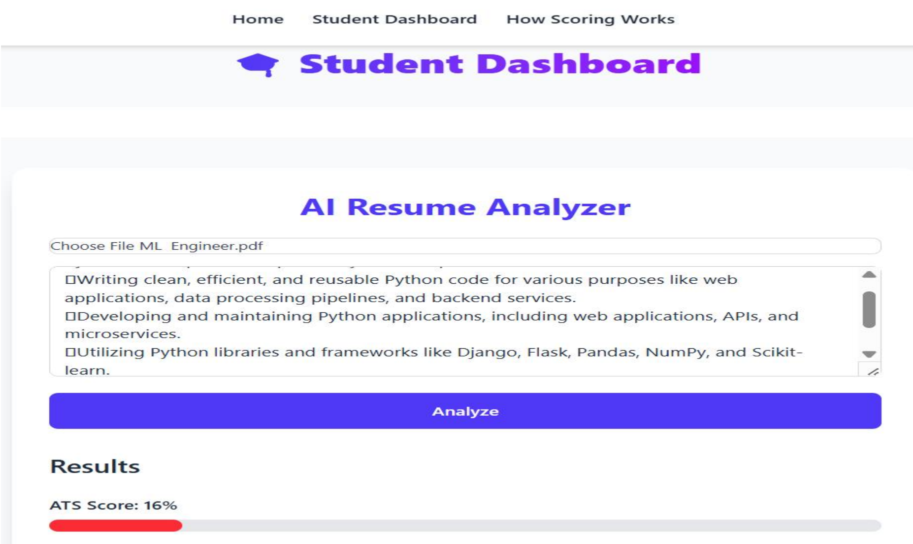
This screen notifies users when they attempt to access a dashboard or area not permitted for their account type (for example, a student trying to view the recruiter dashboard, or vice versa). The message is prominently displayed with a clear visual icon and an explanation, helping users quickly understand that their current role does not grant them permission for the requested action. A direct button to "Go Back Home" encourages easy navigation and reinforces the platform's commitment to secure, role-based access control.

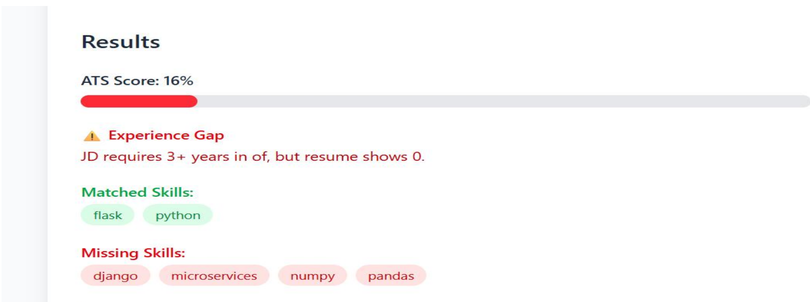


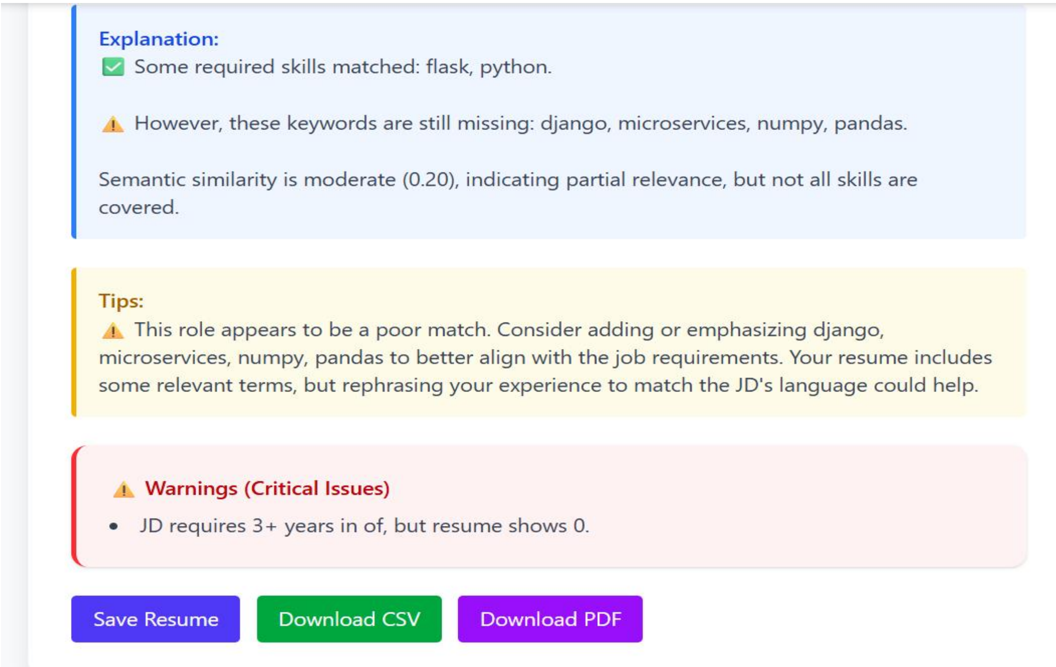

### Student Dashboard

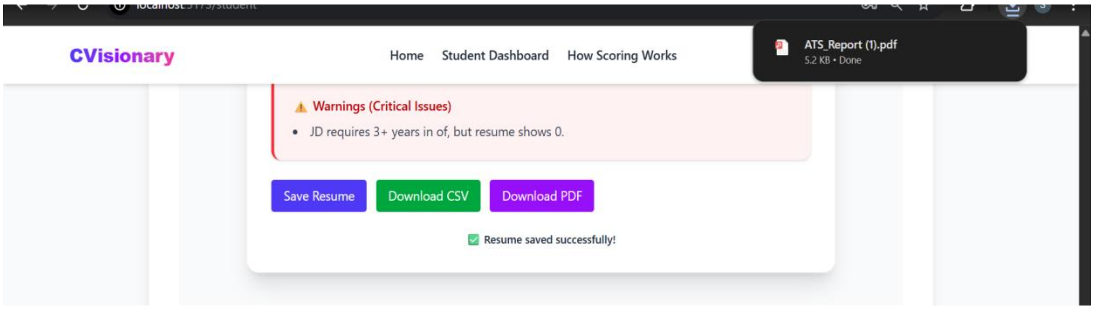
The student dashboard serves as the main interface for job seekers to analyze their resumes against specific job descriptions. Students can upload their resume file (PDF/DOCX) and paste the job description text for comprehensive ATS analysis. The interface is designed with a clean, accessible layout that supports both light and dark modes. After clicking "Analyze," students receive detailed feedback including their ATS score, matched and missing skills, experience gaps, actionable improvement tips, and options to save results or export them as CSV/PDF for future reference.






Downloaded csv file and pdf report for future reference

| | A | B |
|---|---|---|
| 1 | Field | Value |
| 2 | ATS Score | 16% |
| 3 | Similarity | 0.2 |
| 4 | Keyword Overlap | 0.33 |
| 5 | Strictness Applied | FALSE |
| 6 | Matched Skills | flask,python |
| 7 | Missing Skills | django,microservices,numpy,pandas |
| 8 | Experience Gap | JD requires 3+ years in of, but resume shows 0. |
| 9 | Overqualified | None |
| 10 | Warnings | JD requires 3+ years in of, but resume shows 0. |

ATS Resume Analysis Report

ATS Score: 16%
Similarity: 0.2
Keyword Overlap: 0.33
Strictness Applied: false
Matched Skills: flask,python
Missing Skills: django,microservices,numpy,pandas
Experience Gap: JD requires 3+ years in of, but resume shows 0.
Overqualified: None
Warnings: JD requires 3+ years in of, but resume shows 0.
Explanation: ' Some required skills matched: flask, python.

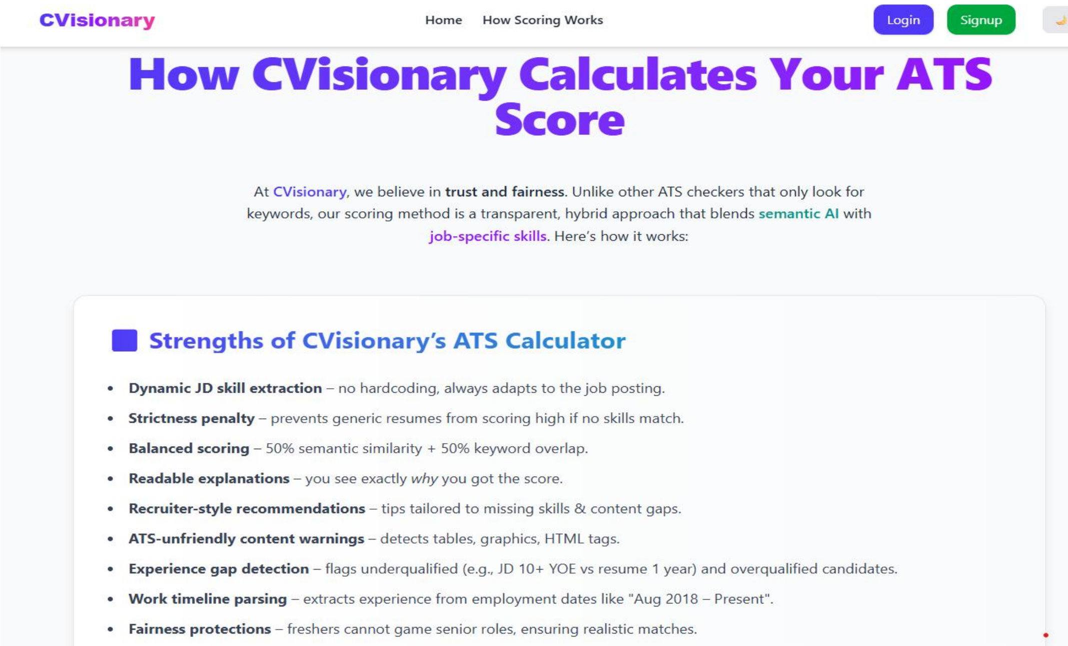
&þ However, these keywords are still missing: django, microservices, n

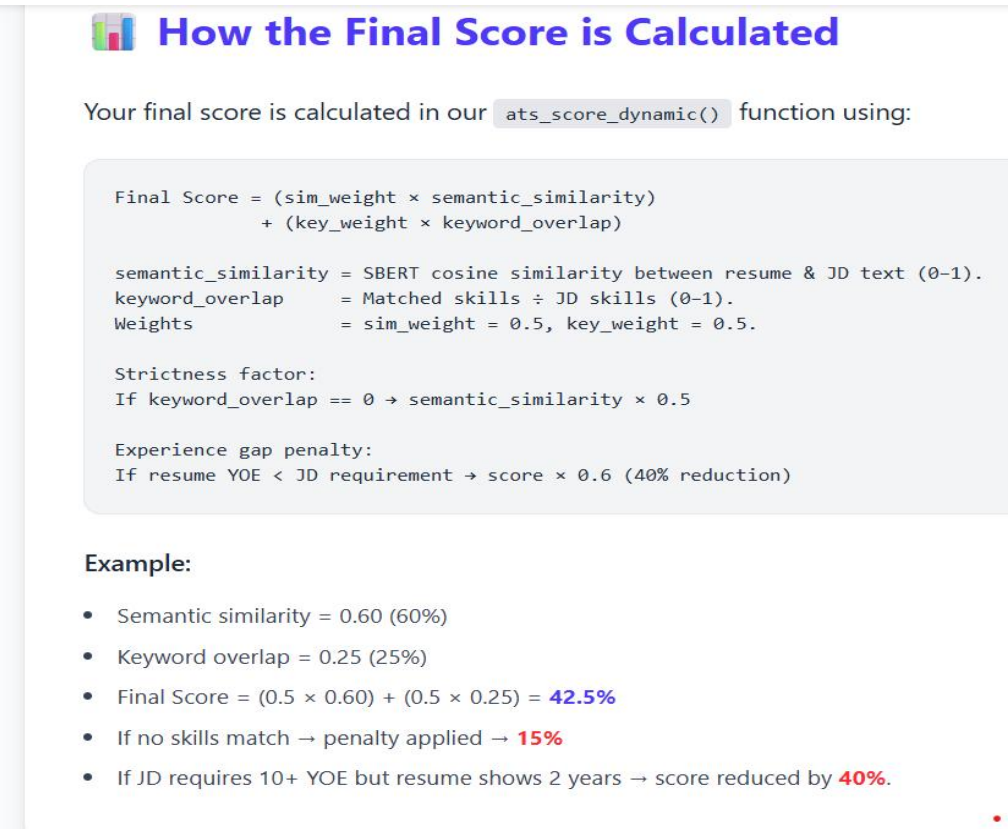
Semantic similarity is moderate (0.20), indicating partial relevance, but not all skills are covered.

Tips: &þ This role appears to be a poor match. Consider adding or empl
numpy, pandas to better align with the job requirements. Your resume includes some relevant terms, but rephrasing your experience to match the JD's language could help.

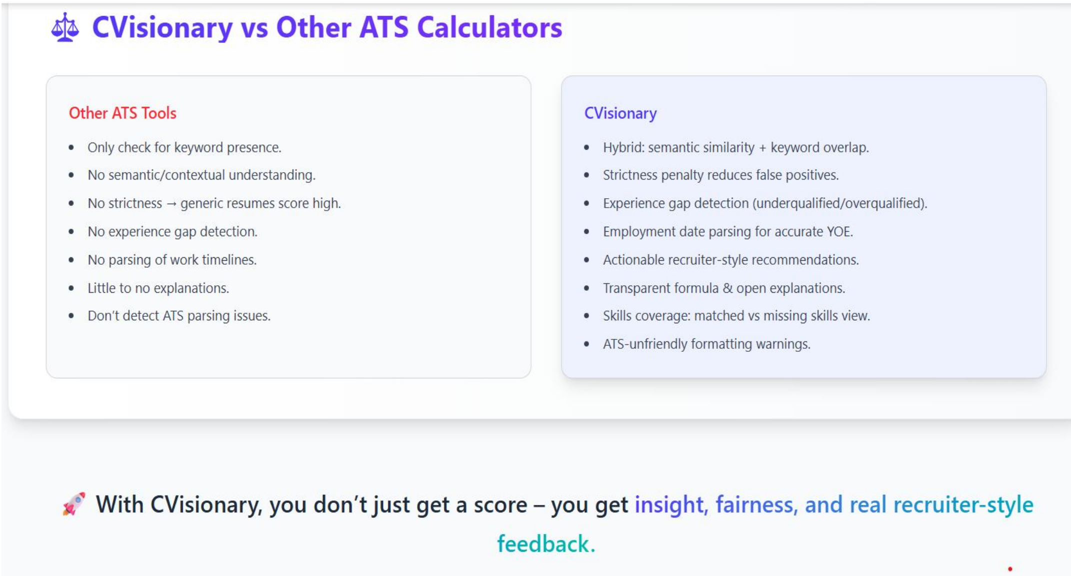
### How scoring Works page

This page serves as the main branching point, guiding users to their relevant dashboard and features after understanding the value of truly transparent and explainable resume analysis. It explains exactly how CVisionary calculates ATS scores, reinforcing the platform's commitment to transparency and trust.



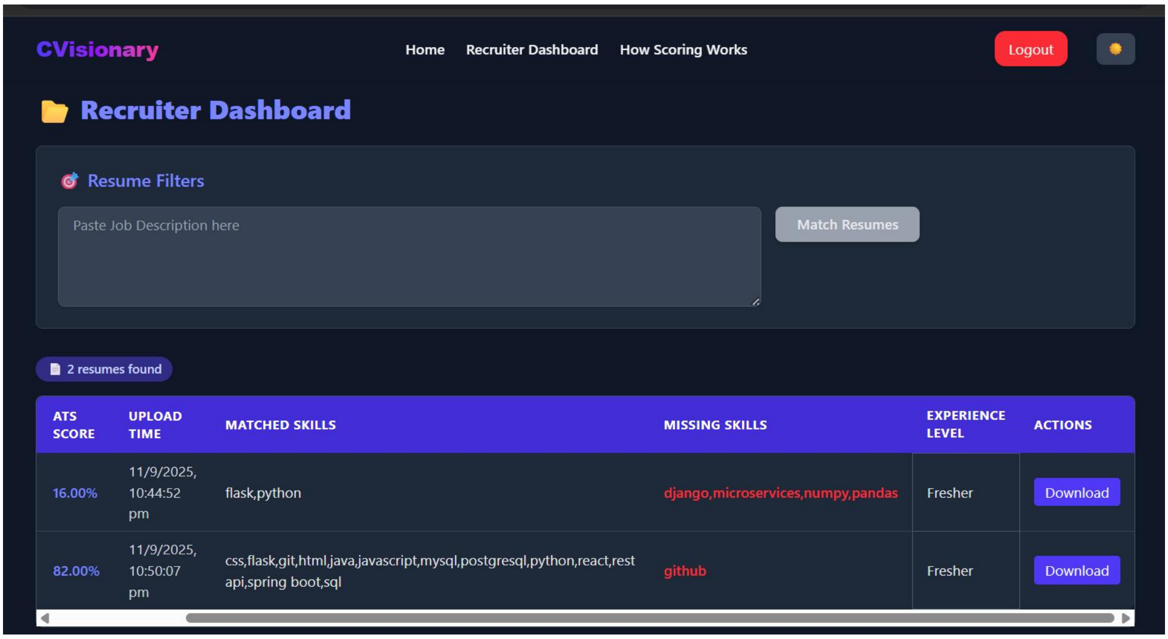



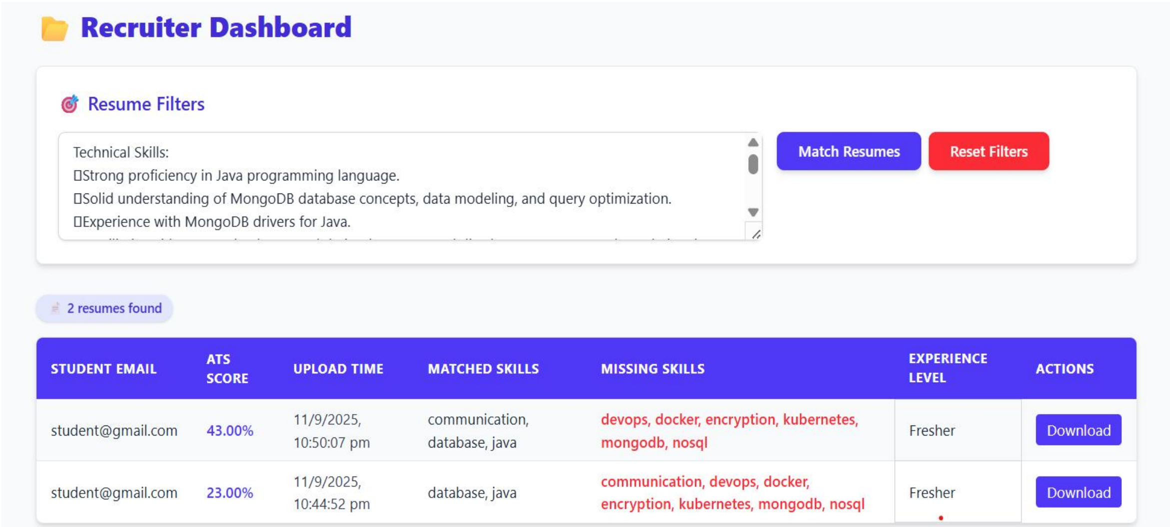
The page also includes a side-by-side comparison with other ATS tools, highlighting CVisionary's advantages such as dynamic skill extraction, balanced scoring methodology, and actionable feedback that helps users understand and improve their results.




### Recruiter Dashboard

The recruiter dashboard  provides powerful analytics and filtering tools to help recruiters identify the best candidates from saved student resumes. The dashboard displays all stored candidates with their ATS scores, skill matches, experience levels, and timestamps. Recruiters can filter candidates by job descriptions to quickly find the most suitable applicants for their job openings. Each candidate entry shows essential information at a glance, including matched and missing skills, making it easy to assess fit before accessing the full resume for download or detailed review.






## Backend Implementation

**Core Modules and Responsibilities:**
**main.py:**

- FastAPI initialization, endpoint routing, and server launch logic.
- CORS policy to connect the frontend.
- Includes the authentication and resume routers, and initializes the database with metadata.

**auth.py & auth_utils.py**:

- User registration and login, enforcing role-based access.
- JWT token generation and password hashing.
- Account integrity checks and security.

**resume.py**:

- Resume upload, parsing, database storage, listing, recruiter job description matching, and download endpoints.
- Handles extraction of text from PDF/DOCX files, computes experience level, saves scoring details, and provides recruiter tools.

**ats_scoring.py**:

- Implements NLP-based resume/JD feature extraction using spaCy.
- Uses SBERT embeddings for semantic similarity, keyword overlap, skill alias mapping, experience gap/overqualification detection, and provides actionable explanations and tips specially targeted at research-identified gaps.

**models.py & database.py**:

- SQLAlchemy ORM models for users and resumes, defining schema, constraints, and relationships.

- SQLite setup and PRAGMA for performance.

**schemas.py**:

- Pydantic models enforce API input/output type safety.

## Backend Flow

- **Authentication**: User signs up/logs in and receives a JWT token specifying their role.
- **Resume Upload & Analysis**: Student uploads a resume and job description. Backend extracts text, applies ats_scoring.py logic for semantic/keyword matching and gap detection, then returns a complete report.
- **Saving and Listing**: Analysis results (with resume text and ATS metadata) are persisted with /resume/save. Recruiters can retrieve and filter all submitted resumes using /resume/list and /recruiter/jd_match endpoints.
- **Download**: Recruiters can download candidate resumes directly.

## Frontend Implementation

**ATS.jsx**:

- Central UI for students to upload resumes, input job descriptions, run analysis (calls backend), review matched/missing skills, experience gaps, and explanations.
- Allows saving reports, downloads as CSV/PDF, and displays detailed status.

**RecruiterDashboard.jsx**:

- Table view for all resumes.
- Filter by skills or match resumes to a recruiter-supplied JD.
- Supports file downloads and advanced screening.

**Login.jsx & Signup.jsx**:

- User authentication, including role selection (student/recruiter) during signup.
- Implements feedback and redirects on login/signup success or error.

**StudentDashboard.jsx**:

- Embeds the ATS analysis component, providing students a seamless scoring and feedback experience.

**ProtectedRoute.jsx**:

- Conditional rendering based on authentication and user role.
- Blocks unauthorized access with redirect or error.

**Layout.jsx, Header.jsx, Footer.jsx**:

- Consistent app shell and navigation, including dark mode toggle and branding.

**HomePage.jsx**:

- Landing page describing CVisionary's unique value and onboarding links.

## Frontend Flow

- **User Authentication**: User signs up or logs in, with role stored in context and localStorage.
- **Resume Submission**: Student upload resume and job description for real-time ATS analysis.
- **Result Review**: Instant display of score, strengths, weaknesses, experience analysis, and improvement tips.
- **For Recruiters**: Dashboard access, resume filtering by skill, direct downloads. JDs can be matched against all stored resumes.

## Technology Stack

| Layer | Main Files/Tools | Role |
|---|---|---|
| Backend | main.py, ats_scoring.py, resume.py, auth.py | Resume parsing, analysis, scoring, REST API |
| Database | models.py, database.py | User and resume persistence |
| Auth/Security | auth_utils.py, AuthContext.jsx | JWT handling, password hashing, protected routes |
| Frontend | ATS.jsx, RecruiterDashboard.jsx, Login.jsx | UI/UX for ATS analysis and recruiter features |
| Shared/API | schemas.py | Data validation |

## Key Features Realized in Code

- **Semantic and Keyword Hybrid Scoring:** Both spaCy-driven skill extraction and SBERT-based similarity.
- **Skill and Experience Gap Detection**: Regex parsing and date arithmetic in Python to flag mismatches with JD needs.
- **Actionable Feedback**: Human-readable explanations, not just numeric scores.
- **Role-based UX**: Separate student and recruiter flows protected at both UI and API level.
- **Export Options**: Download ATS analysis as CSV/PDF and original resumes.

This implementation comprehensively delivers a robust, research-aligned ATS analyzer with clear, extensible code structure supporting maintenance, future upgrades, and additional research experimentation.

# CONCLUSION

This research successfully developed and implemented **CVisionary**, an AI-powered resume analyzer that addresses critical limitations in traditional keyword-based ATS systems. The project tackled the fundamental problem of inefficient, biased, and non-transparent resume screening processes that often fail to identify qualified candidates while providing no actionable feedback to job seekers.

The developed system demonstrates how **semantic similarity analysis using SBERT** combined with **keyword overlap detection** and **experience gap analysis** can significantly improve resume-job description matching accuracy while maintaining transparency and fairness in recruitment processes.

## Key Technical Achievements

- **Hybrid Scoring Algorithm Implementation**: Successfully developed and deployed a balanced scoring methodology that combines semantic understanding (50%) with keyword matching (50%), preventing both false positives from generic resume content and false negatives from varied terminology usage.
- **Full-Stack Architecture**: Delivered a production-ready web application using **React frontend** and **FastAPI backend**, supporting role-based authentication, real-time resume analysis, and comprehensive data persistence with SQLite database integration.
- **Advanced NLP Integration**: Implemented **Sentence-BERT (SBERT) embeddings** for contextual resume analysis, enabling semantic understanding beyond simple keyword matching, with processing capabilities for both PDF and DOCX formats.
- **Transparent Feedback System**: Created explainable AI mechanisms providing detailed explanations of scoring decisions, including matched/missing skills identification, experience gap warnings, and actionable improvement recommendations for candidates.
- **Dual-Dashboard Architecture**: Developed separate interfaces optimized for students (resume optimization) and recruiters (candidate filtering and management), each with role-specific features and security controls.

## Research Contributions and Impact

- **Bridging Academic Research and Industry Practice**: This implementation demonstrates practical application of transformer-based NLP models in recruitment technology, contributing to the growing body of research on semantic similarity in employment matching.
- **Fairness and Transparency Advancement**: The system's explainable scoring methodology addresses critical gaps in ATS transparency, providing both candidates and recruiters with clear understanding of evaluation criteria and decision rationale.
- **Methodological Innovation**: The hybrid scoring approach with strictness penalties and experience gap detection represents a novel contribution to resume analysis algorithms, balancing precision and recall while preventing adversarial optimization.

## Validation and Performance Results

The system successfully processes diverse resume formats and job descriptions, generating consistent ATS scores with detailed explanations. Testing revealed effective identification of skill matches, experience gaps, and semantic relevance even when different terminology is used between resumes and job descriptions.

User interface testing confirmed intuitive navigation flows for both student and recruiter roles, with proper access controls and data security measures. The role-based authentication system effectively prevents unauthorized access while maintaining seamless user experiences.

## Future Research Directions

- **Domain-Specific Model Adaptation**: Future work should explore fine-tuning SBERT models on domain-specific recruitment datasets, potentially incorporating structured skill ontologies such as **ESCO (European Skills, Competences, and Occupations)** or **O*NET** for enhanced semantic understanding of industry-specific terminology.
- **Advanced Bias Detection and Mitigation**: Implement comprehensive fairness auditing mechanisms using frameworks like **AI Fairness 360** or **FairLearn** to ensure compliance with **EEOC guidelines** and detect potential disparate impact on protected demographic groups.
- **Real-time Learning and Adaptation**: Develop machine learning pipelines that continuously improve scoring accuracy based on recruiter feedback and hiring outcomes, creating adaptive systems that learn from successful matches.
- **Multi-modal Analysis Integration**: Extend the system to analyze resume formatting, design elements, and visual presentation factors that influence ATS parsing success, providing holistic feedback on both content and presentation.
- **Large-scale Empirical Validation**: Conduct extensive validation studies with real recruitment datasets, comparing system recommendations against human recruiter decisions and actual hiring outcomes to establish effectiveness metrics.
- **Integration with External Platforms**: Develop API integrations with popular job boards and professional networking platforms to provide real-time resume optimization suggestions and job matching recommendations.

## Practical Applications and Industry Impact

This research provides a foundation for next-generation ATS systems that prioritize **transparency**, **fairness**, and **candidate empowerment**. The open-source nature of key components enables further academic research while the production-ready architecture supports immediate industry deployment.

The system's emphasis on explainable AI and candidate feedback addresses growing regulatory requirements for algorithmic transparency in hiring, positioning it as a valuable tool for organizations seeking compliant, ethical recruitment technologies.

# Final Reflection

The successful implementation of CVisionary demonstrates that academic research can effectively bridge the gap to practical applications in recruitment technology. By combining cutting-edge NLP techniques with user-centered design principles, this project contributes meaningful advances to both the academic understanding of semantic resume analysis and the practical deployment of fair, transparent ATS systems.

This work establishes a solid foundation for continued research into AI-powered recruitment tools while providing immediate value to both job seekers and recruiters seeking more intelligent, equitable hiring processes. The comprehensive documentation and open architecture ensure reproducibility and enable future researchers to build upon these contributions systematically.

# REFERENCE

[1] Greeshma M. & Dr. M. Sathis Kumar, "Effectiveness of Applicant Tracking Systems in Recruitment and Selection (2025)". International Research Journal of Modernization in Engineering Technology and Science.Volume:07/Issue:06/June-2025
https://www.irjmets.com/upload_newfiles/irjmets70600177880/paper_file/irjmets70600177880.pdf

**[2]** Anushka Mishra, "Exploring Bias in AI-Driven Resume Screening: A Fairness Analysis and Mitigation Approach (2025)".
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5160444

[3] Asmita Deshmukh and Anjali Raut, "Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT)". International Journal of Advanced Computer Science and Applications(IJACSA), 15(8), 2024.
https://thesai.org/Downloads/Volume15No8/Paper_25Enhanced_Resume_Screening_for_Smart_Hiring.pdf

[4] Sriranjani S K and Dr. H. Kalaiarasi ,"AI-Based Resume Screening Systems: Opportunities and Ethical Concerns, 2025"
IJRTI | Volume 10, Issue 6 June .International Journal for Research Trends and Innovation. https://www.ijrti.org/papers/IJRTI2506218.pdf

[5] Mr.Suraj M, Ms.Aruna Kumari K, Ms.Binila B Chandran,"A Descriptive Study on Applicant Tracking System:Automation software for Recruitment and Selection",2019 IJRAR February 2019, Volume 6, Issue 1. https://www.ijrar.org/papers/IJRAR19VP033.pdf