AI Skill Gap Analyzer

Project Report  
By: Srihitha Kalagarla  
ID : 16

# 1. Abstract

The AI Skill Gap Analyzer Pro is an intelligent system that identifies and analyzes the gap between   
a candidate’s resume and a given job description using Natural Language Processing (NLP) and   
Machine Learning. The application automates skill extraction, similarity matching, and visualization   
to provide a comprehensive skill comparison report. Built using Streamlit, spaCy, and Sentence-BERT,   
it helps both recruiters and candidates understand strengths, weaknesses, and areas for upskilling.

# 2. Introduction

Manual comparison of resumes and job descriptions is time-consuming, inconsistent, and subjective.   
This project aims to automate the process by using AI models to read resumes and JDs, extract   
skills, compute similarity scores, and visualize the results. The goal is to highlight the most   
important missing skills for a target job role.

# 3. Objectives

• Automate the process of resume–JD comparison.  
  
• Extract technical, soft, and tool-based skills using NLP.  
  
• Compute similarity using semantic embeddings.  
  
• Visualize skill matches, gaps, and recommendations.  
  
• Generate exportable reports (PDF and CSV) for analysis.

# 4. System Architecture

The system consists of four main modules:  
  
1. Data Ingestion & Parsing – Upload and clean text from resumes and JDs.  
  
2. Skill Extraction – Identify and categorize skills using NLP and pattern matching.  
  
3. Skill Gap Analysis – Compare resume and JD skills using Sentence-BERT + cosine similarity.  
  
4. Visualization & Dashboard – Display results with radar charts, heatmaps, and export options.

# 5. Module Descriptions

## 5.1 Data Ingestion & Parsing

This module handles uploading and parsing resumes and job descriptions in multiple formats   
(PDF, DOCX, TXT). It uses PyPDF2, docx, and OCR for extraction, then normalizes text by   
removing personal info, headers, and noise using remove\_personal() and normalize\_text().

## 5.2 Skill Extraction

This module uses spaCy, regex, and custom NER to extract skills from cleaned text. It identifies   
technical, soft, and tool-based skills with confidence scores and displays them as tag clouds   
and categorized lists.

## 5.3 Skill Gap Analysis

This module encodes extracted skills using Sentence-BERT embeddings and computes cosine similarity   
scores to measure how closely resume skills match job description skills. It classifies matches as   
Strong, Partial, or Missing and ranks them by importance.

## 5.4 Visualization & Dashboard

This module provides an interactive Streamlit interface with multiple tabs for overview, skills,   
gap analysis, and report export. It visualizes results with radar charts, heatmaps, and bar graphs.   
Users can download analysis as PDF or CSV reports.

# 6. Technologies Used

Frontend/UI: Streamlit, HTML, CSS  
  
Text Parsing: PyPDF2, docx, pytesseract (OCR)  
  
NLP & ML: spaCy, Sentence-BERT, scikit-learn  
  
Data Handling: Pandas, NumPy  
  
Visualization: Plotly, Seaborn, Matplotlib

# 7. Results

The system successfully identifies strong, partial, and missing skills between resume and job description.  
Example results:  
  
• Overall Skill Match: 78%  
  
• Strong Matches: Python, SQL, Machine Learning  
  
• Partial Matches: AWS, Tableau  
  
• Missing Skills: TensorFlow, CI/CD, Jenkins

# 8. Conclusion & Future Work

The AI Skill Gap Analyzer Pro provides an accurate and automated method for understanding skill differences   
between resumes and job descriptions. It benefits both candidates and recruiters by saving time and improving   
the quality of hiring decisions.  
  
Future work includes integrating LinkedIn APIs for live skill suggestions, adding multilingual support,   
and recommending online courses for missing skills.