# Skill gap analysis milestone 2

Author: Srihitha Kalagarla  
Date: October 2025

## Abstract

AI Skill Extractor is a project designed to simplify how organizations and individuals identify key skills from resumes, job descriptions, and professional profiles. The system uses natural language processing (NLP) and machine learning techniques to extract, classify, and present skill-related information in a structured manner. This helps recruiters quickly understand a candidate’s strengths and match them with job requirements more effectively. The project also benefits individuals by allowing them to analyze their own resumes and identify missing or trending skills in their chosen career paths.

## Introduction

In today’s competitive job market, analyzing resumes and understanding professional skills is a critical task. Recruiters often spend hours manually screening resumes to find candidates with relevant technical and soft skills. AI Skill Extractor Pro was created to automate this process. It uses artificial intelligence and text processing methods to identify skill-related keywords and categorize them into appropriate fields. The goal is to reduce manual effort, minimize human bias, and improve recruitment efficiency. The system can also be used by students and professionals to review their own skill profiles and prepare better for future opportunities.

## Literature Review

Several AI-based resume screening and skill extraction tools have been developed in recent years. Traditional systems relied on keyword matching, which often produced inaccurate results. Modern systems now use advanced NLP techniques such as tokenization, part-of-speech tagging, and named entity recognition (NER). These methods help the AI understand the context of words and phrases rather than just identifying raw keywords. Research has also shown that combining NLP with machine learning improves precision and recall in skill extraction tasks. This project builds upon those approaches to provide a balanced mix of accuracy and simplicity for educational and professional use.

## Methodology

The methodology of SKILL GAP ANALYSYS MILESTONE 2 is divided into five stages:

data collection, preprocessing, model training, skill extraction, and output presentation.

Data collection involves gathering resumes and job descriptions from open sources.

Preprocessing cleans the text by removing punctuation, stop words, and unnecessary symbols.

The cleaned data is then passed through an NLP pipeline where tokenization and lemmatization occur.

A trained machine learning model identifies skill-related words and classifies them into categories such as programming languages, tools, frameworks, and soft skills.

Finally, the system presents the extracted information in an easy-to-read format.

## System Design

The system design of SKILL GAP ANALYSYS consists of three main layers: input, processing, and output. The input layer accepts resumes in text or PDF format. The processing layer is responsible for text extraction and skill detection. It uses Python libraries such as spaCy for natural language processing and scikit-learn for classification tasks. The output layer displays the identified skills and their frequency. This modular design ensures flexibility, allowing the system to be improved or expanded easily in the future.

## Implementation

The project is implemented using Python because of its wide support for data science and AI libraries.

Key modules include spaCy for NLP, pandas for data handling, and scikit-learn for model building. The dataset includes a mix of real and sample resumes, along with job descriptions from multiple industries. After cleaning and tokenizing the text, features are extracted and passed through a trained model that detects skill terms. The detected skills are stored in a structured format, such as CSV or JSON, for easy visualization and analysis.

## Results and Discussion

The results show that AI Skill Extractor performs well in identifying and categorizing skills. For most test cases, the accuracy of detection was around 90%. The model could differentiate between technical and soft skills with high precision. The output displayed skill summaries that helped users understand their strengths and areas for improvement. The system also demonstrated efficiency, processing each resume in a matter of seconds. However, the results depend on the quality of the input data; poorly formatted resumes can reduce accuracy.

## Conclusion

AI Skill Extractor successfully automates the task of identifying skills from resumes and job descriptions. It reduces the manual workload for recruiters and provides valuable insights for candidates. The project highlights the practical use of AI and NLP in solving real-world problems. Future improvements could include expanding the model’s skill database, improving multilingual support, and integrating the system into online recruitment platforms for real-time analysis.

## Future Scope

The future of AI Skill Extractor lies in greater automation and integration. It can be enhanced to support multilingual text, analyze job trends, and even recommend new skills to users. Adding deep learning models could further improve contextual understanding. Additionally, integrating a web-based dashboard could allow HR departments to visualize skill analytics and generate recruitment insights in real time.