1. **Title:** AI-Powered Skill Gap Analyzer

---

2. Project Statement:

In the rapidly evolving job market, identifying mismatches between candidate resumes and job descriptions is crucial for both job seekers and recruiters. This application uses Natural Language Processing (NLP) to extract skills from resumes and job descriptions, compares them, and identifies gaps. The system provides intelligent recommendations for upskilling by highlighting missing or underrepresented skills, improving job readiness.

---

3. Outcomes:

The system will extract relevant skills from resumes and job descriptions using Named Entity Recognition (NER) models.
It will compute semantic similarity scores between extracted skills using Sentence-BERT.
The application will identify skills that are either missing or weakly aligned with job requirements.
It will provide personalized recommendations for upskilling based on identified skill gaps.
A role-based dashboard will be implemented to support both job seekers and recruiters.
The system will include functionality to upload and parse resumes and job descriptions in various file formats.
Users will be able to export skill gap analysis reports in PDF or CSV formats.

---

4. Modules to be Implemented:

1. Data Ingestion and Parsing

- The system will allow users to upload resumes and job descriptions in PDF, DOCX, or TXT formats.
- Uploaded documents will be parsed, and the plain text content will be extracted, normalized, and cleaned to remove noise.

2. Skill Extraction using NLP

- Natural Language Processing techniques, using spaCy and custom NER models, will be used to identify and extract skill entities.
- Preprocessing steps will be applied to isolate relevant keywords, technical competencies, and soft skills.

3. Skill Gap Analysis and Similarity Matching

- Extracted skills will be encoded using Sentence-BERT embeddings from Hugging Face. Cosine similarity will be used to compare resume and job description skills, and the system will identify mismatches.
- Skill gaps will be ranked based on importance and categorized to aid decision-making.

4. Visualization and Dashboard

- A Streamlit-based dashboard will be developed to enable uploading and comparing documents.
- The dashboard will visually represent the gap analysis and display skill match percentages. Users will be able to export the analysis as reports in PDF or CSV format.

---

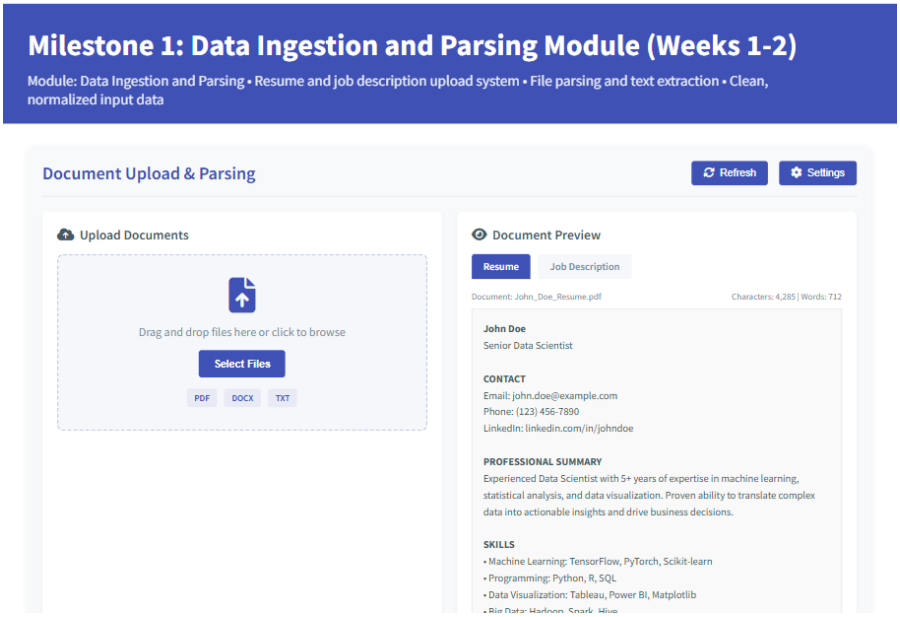
5. Week-wise Module Implementation and High-Level Requirements with Output Screenshots:

Milestone 1: Weeks 1–2

Module: Data Ingestion and Parsing
The resume and job description upload system will be implemented, supporting various document types.
File parsing and text extraction functionality will be completed to ensure clean, normalized input data.

Output Screenshots:

- Upload interface
- Parsed document preview

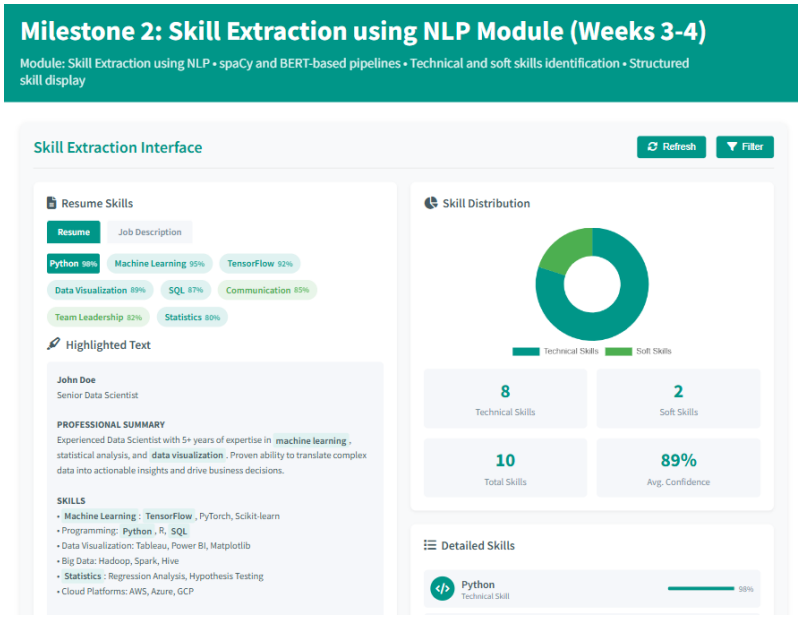


Milestone 2: Weeks 3–4

Module: Skill Extraction using NLP
The spaCy-based and BERT-based pipelines for skill extraction will be developed and integrated. Both technical and soft skills will be identified and displayed in structured form.

Output Screenshots:

- Extracted skill sets from resume and job description
- Skill tag visualization



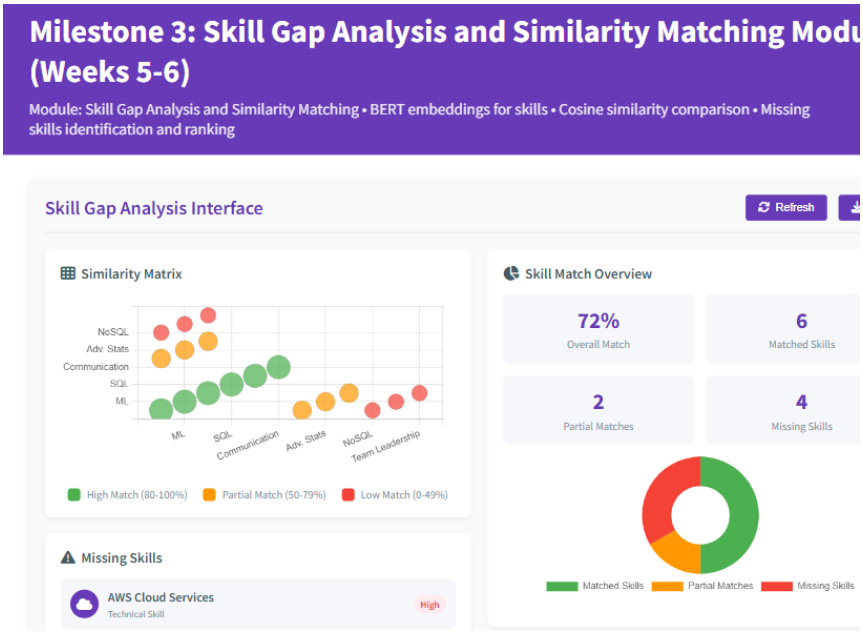
Milestone 3: Weeks 5–6

Module: Skill Gap Analysis and Similarity Matching
The system will generate BERT embeddings for skills and compute similarity scores between them. A list of missing or partially matched skills will be generated and displayed.

Output Screenshots:

- Skill gap report showing matched and unmatched skills

- Similarity matrix visualized as a heatmap



Milestone 4: Weeks 7–8

Module: Dashboard and Report Export
A Streamlit interface will be finalized for end-to-end comparison and result presentation. The system will provide interactive graphs, scores, and downloadable reports.

Output Screenshots:

- Final dashboard user interface
- PDF and CSV report download options

---

6. Evaluation Criteria:

Milestone 1 Evaluation (Week 2):

The file upload functionality and text extraction pipeline must be fully operational.
Resumes and job descriptions must be converted into structured and clean text for processing.

Milestone 2 Evaluation (Week 4):

The system must accurately extract skills using trained NLP models.
NER and keyword extraction must work consistently for various document types.

Milestone 3 Evaluation (Week 6):

Skill gap analysis using BERT-based similarity scoring must identify accurate and contextually relevant gaps.
Missing skills must be clearly highlighted and classified.

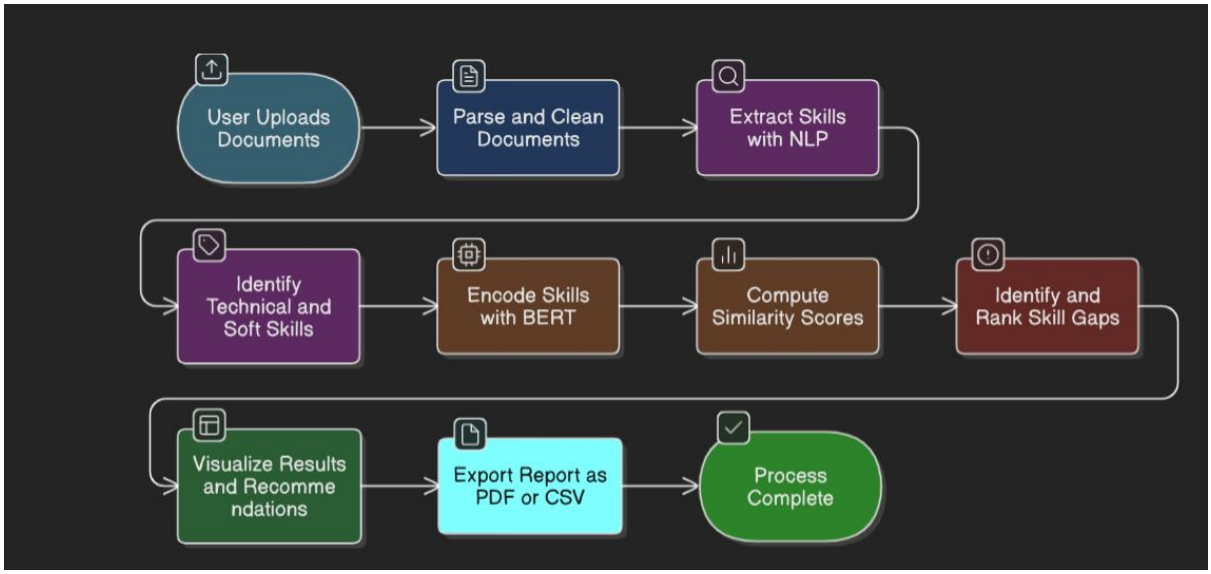
Milestone 4 Evaluation (Week 8):

The dashboard must be interactive, visually clear, and allow for skill gap interpretation.
Report generation and export features must function as intended.

---

7. Workflow Diagram:

(To be inserted – representing the process flow: Resume and JD Upload → NLP Preprocessing → Skill Extraction → Skill Matching → Visualization → Report Generation)



---

8. Architecture Diagram:

(Input Layer → NLP Pipeline → Embedding Layer → Matching Logic → Streamlit Dashboard → Export Service)