AI-Powered CV Summary

Overview

The AI-powered CV Summary feature parses a PDF CV, identifies major sections (e.g., Work Experience, Education, Skills), and leverages **Gemini** (an LLM-based summarization service) to generate a concise, human-readable summary of the applicant's background.

1. Functional Requirements

ID	Requirement
FR 1.1	The system shall accept a PDF file as input via the command-line interface.
FR 1.2	The system shall extract text from the PDF while preserving basic structure (e.g., headings, bulleted lists) as much as possible.
FR 2.1	The system shall analyze the extracted text to identify major CV sections (e.g., Education, Work Experience, Skills, Contact Info).
FR 2.2	The system may rely on keyword- or heuristic-based detection (e.g., matching "Education," "Experience," etc.) to locate key sections.
FR 3.1	The system shall use the **Gemini LLM** to generate a concise summary of the candidate's background.
FR 3.2	The summary shall emphasize: top skills/areas of expertise, years of experience, highest level of education, and notable achievements.
FR 4.1	The system shall print the final summary to the command line.
FR 4.2	The system may provide an option to export the generated summary to a text file (e.g., summary.txt).
FR 5.1	The user should be able to specify a maximum word/character count for the summary (e.g., --max-words 100).
FR 5.2	The user may provide a specific focus area (e.g., "leadership," "technical skills") to guide the Gemini-based summarization if needed.

2. Non-Functional Requirements

ID	Requirement
NFR 1.1	**Performance**: The system should handle a typical 2–4 page CV and produce a summary in under 5 seconds on average hardware.
NFR 1.2	**Performance**: Memory usage should remain within reasonable bounds (no excessive in-memory expansions).
NFR 2.1	**Usability**: The command-line interface must be simple to invoke (e.g., python main.py --pdf candidate_cv.pdf --max-words 100).
NFR 2.2	**Usability**: Error messages must be clear and informative (e.g., “File not found,” “Unsupported PDF format”).
NFR 3.1	**Reliability**: The PDF parsing module should handle minor formatting issues or unusual layouts without crashing.
NFR 3.2	**Reliability**: If the system fails to detect key sections, it should still generate a fallback summary (though it may omit missing sections).
NFR 4.1	**Maintainability**: The code shall adhere to PEP 8 (Python style guide).
NFR 4.2	**Maintainability**: The code should be modular and well-documented, with clear separation of concerns (e.g., PDF parsing, Gemini integration, CLI parsing).
NFR 5.1	**Security**: No personal user data (beyond the CV itself) should be stored or transmitted without explicit user consent.
NFR 5.2	**Security**: If using Gemini’s API, credentials or tokens must be handled securely (e.g., environment variables), not hard-coded in the source.

3. Acceptance Criteria

ID	Criterion
AC 1.1	**Accuracy & Completeness**: The summary must mention the candidate’s primary field/role, highest educational attainment (if available), and notable skills. It must avoid unrelated or incorrect details.
AC 1.2	**Accuracy & Completeness**: If the candidate has achievements or certifications listed, these should be reflected or referenced in the summary (if space allows).
AC 2.1	**Readability & Style**: The summary must be written in clear, grammatical English.
AC 2.2	**Readability & Style**: The summary must be limited to the maximum length specified (defaulting to 100 words if no limit is provided).
AC 3.1	**Section Coverage**: If sections like Education, Experience, and Skills are detected, they must be referenced in the generated summary. Missing sections should be skipped gracefully.
AC 4.1	**CLI Functionality**: Running the CLI with correct flags must produce an output summary within 5 seconds for a 2–4 page CV on typical hardware.
AC 4.2	**CLI Functionality**: Invalid inputs (e.g., missing PDF, unreadable file) should produce an informative error message rather than cause a crash.
AC 5.1	**Test Cases**: At least one automated **unit test** must confirm that the summarizer extracts known sections from a sample CV.
AC 5.2	**Test Cases**: At least one **system-level test** must confirm end-to-end functionality (PDF → Gemini Summarization → CLI output).