

■ POV: AI-Powered PDF Summarizer

■ Problem

- Professionals, researchers, and students deal with long documents (100+ pages).
- Extracting key insights is time-consuming.
- Existing tools often fail with large PDFs or produce generic summaries.

■ Opportunity

- Build a system that can ingest multiple PDFs (or one large one).
- Automatically extract important sentences, keywords, and insights.
- Condense hundreds of pages into 2–10 concise pages.
- Output in a structured PDF summary that can be shared or stored.

■ Target Users

- Students & Researchers → Summarize textbooks, academic papers.
- Business Analysts → Summarize reports, contracts.
- Lawyers & Legal Teams → Condense case files and legal documents.
- General Users → Summarize ebooks, articles, or notes.

■■ Solution Approach

1. Upload → User uploads 1 or multiple PDFs.
2. Extract → Extract raw text using pdfplumber.
3. Chunk & Process → Split into smaller sections for processing.
4. Summarize → Use AI models (Hugging Face BART / Gemini API).
5. Output → Create a new PDF with shortened key insights.

■ Key Features

- Handles large documents (100+ pages).
- Works with single or multiple PDFs.
- Outputs short, structured summaries (2–10 pages).
- AI-powered → ensures context-aware, human-like summaries.
- Export as shareable PDF.

■ Benefits

- Saves time: hours of reading → minutes of scanning.
- Improves decision-making: highlights only important parts.
- Flexible: usable across education, business, law, research.
- Scalable: can be extended into a web or mobile app.

■ Tech Stack

- Backend → Python, pdfplumber, Hugging Face Transformers.
- AI Models → BART / T5 / Longformer / Gemini API.
- Frontend (optional) → Flask / Streamlit for file upload & preview.
- Output → ReportLab (PDF generator).