Infoware Edu Prototype (Demo Document)

Project: Slides & Short Video Generator
Intern: Mauli Patel
Organization: Infoware Edu Prototype

Objective

Automatically convert any input **PDF (chapter, article, or resume)** into:

- A short, readable **slide deck (slides.pptx)**, and
- A **30-90s explainer video (video.mp4)** with transitions and Canva-like visuals.

How It Works (The Pipeline)

1. **Extract & OCR**
   → Reads the PDF using pdfplumber.
   → If it's image-based, runs OCR using pytesseract to extract text.
2. **Summarize & Segment**
   → Detects key headings and sections (like Education, Skills, Projects).
   → Summarizes each part into short, readable text chunks.
3. **Slide Builder**
   → Uses python-pptx to automatically make 6–12 clean slides with titles and content.
4. **Video Maker**
   → Uses moviepy and Pillow to turn slides into a short animated video with transitions.

Final Output:

- outputs/slides.pptx → Slide deck
- outputs/video.mp4 → Short explainer video

Core Modules

Module	Purpose
summarizer.py	Extracts text from PDFs, handles OCR for image-based files, and groups content into sections
run_pipeline.py	Main orchestrator, runs all steps and saves outputs
video_maker.py	Generates Canva-style visuals and transitions for video.mp4

How to Run (Local Setup)

1. **Install requirements**

pip install -r requirements.txt

2. **Place your PDF in the project folder**

(e.g., resume.pdf or chapter1.pdf)
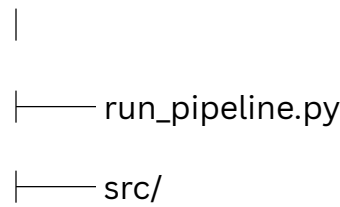
3. **Run the pipeline**

python run_pipeline.py --input resume.pdf --outdir outputs

4. **Output files**

File	Description
outputs/slides.pptx	Auto-generated slide deck
outputs/video.mp4	Animated short video version

Folder Snapshot

infoware-edu-prototype/

|

├────── run_pipeline.py

├────── src/

```
│       ├──── summarizer.py
│       ├──── video_maker.py
│
├──── sample.pdf
├──── outputs/
│       ├──── slides.pptx
│       ├──── video.mp4
│
└──── README.md
```

Example Run

Input: Mauli_Patel_Resume.pdf
Output:

- slides.pptx → 10 concise slides (Education, Skills, Projects, etc.)
- video.mp4 → Animated 40s visual explainer

Summary

The pipeline automates the creation of educational explainers from any document.
It integrates **PDF parsing, NLP summarization, PowerPoint generation, and video assembly,** fully offline and modular.

Note

In this demo, I've used a **resume PDF** as an example input, but you can use **any PDF (chapter, article, or notes)** to test the pipeline. The script will automatically extract text, generate slides, and make a short explainer video from it.