## **Project: PDF Summarization Tool**

## **🧠 Overview**

A web-based tool built with **Streamlit** that allows users to **upload PDF files** and get a **text summary** using Hugging Face transformer models like *facebook/bart-large-cnn*.

## **🚀 Features**

* Upload and extract text from any PDF.
* Chunk large texts for transformer summarization.
* Summarize using pretrained models from Hugging Face (transformers).
* Display the summarized text in the browser.
* Handles large PDFs with multi-chunk processing.
* Works offline after the model is downloaded.

## **🗂️ Project Structure**

pdf-summarizer-app/

├── app.py

├── requirements.txt

## 

## 

## **🧰 Requirements**

* Python 3.8+
* pip
* Optional: virtual environment (venv)

## **▶️ Running the App**

streamlit run app.py

Then open<http://localhost:8501> in your browser.

## 

## **🧾 requirements.txt**

streamlit

transformers

torch

PyMuPDF

## **🧪 Usage Instructions**

1. Launch the app.
2. Upload a .pdf file using the file uploader.
3. The app extracts the text from all pages.
4. It breaks the text into chunks (to handle model token limits).
5. Each chunk is summarized using Hugging Face transformers.
6. The full summary is displayed on the page.

## **💡 Notes**

* Default model: *facebook/bart-large-cnn*
* You can switch to a faster model like **sshleifer/distilbart-cnn-12-6** for performance
* Streamlit caches the summarizer and model after the first run
* May be slow on large PDFs if using CPU