

# Indian Legal Advisor (RAG) - Local Setup & PDF Upload Guide

This PDF explains how to upload and use PDF documents in the Local RAG (Retrieval-Augmented Generation) project.

## 1) Requirements

Make sure you have these packages installed:

- google-genai
- gradio
- faiss-cpu (or system faiss on Arch Linux)
- pypdf
- python-docx
- python-dotenv
- numpy

Install using:

```
pip install -r requirements.txt
```

## 2) Gradio File Upload Supports PDF

Your app.py should include:

```
files = gr.File(file_types=[".pdf", ".docx", ".txt"],  
file_count="multiple")
```

## 3) PDF Text Extraction

Your src/extractor.py should include:

```
from pypdf import PdfReader  
  
and:  
  
reader = PdfReader(file_path)  
  
for page in reader.pages:  
  
text += page.extract_text() or ""
```

## 4) Common Issue: Scanned PDFs

If a PDF is scanned (image-based), pypdf cannot extract text and returns empty output.

Solution: use OCR (optional) with Tesseract.

## 5) Recommended Safety Check

In app.py, after extracting text, skip empty text:

```
if not text.strip():  
    continue
```

## 6) Run Locally

Commands:

```
python -m venv .venv  
source .venv/bin/activate  
pip install -r requirements.txt  
python app.py
```

## 7) Testing

Upload a PDF, click Process Documents, then ask:

- What is Section 10 of the Indian Contract Act?
- What is cheating under IPC Section 420?

## Disclaimer

This project provides general legal information and is not legal advice.