# AskMyPDF – Chat with any PDF Instantly

AskMyPDF is an intelligent chatbot that allows you to upload any PDF document—like resumes, research papers, books, or notes—and ask natural language questions about its content. It uses LangChain, Hugging Face LLMs, and Astra DB Vector Search to accurately extract, understand, and respond to queries using the actual document content.  
“No more scrolling through hundreds of pages—just ask your PDF.”

## 🚀 Features

* ✅ Upload any PDF (resume, research paper, notes, textbook)
* ✅ Automatically extract and process text
* ✅ Smart chunking and embedding using MiniLM
* ✅ Store and retrieve document context using Astra DB
* ✅ Dynamic question answering using Zephyr-7B LLM
* ✅ Supports all types of documents with new vector tables per file
* ✅ Simple command-line interface (Colab/terminal ready)

## 🧠 How It Works

PDF Upload → Text Extraction → Chunking → Embedding → Store in Astra DB  
 ↓ ↓  
 Ask a Question → Search for relevant chunks  
 ↓ ↓  
 Prompt LLM with context → Generate Answer ✔️

## 🛠️ Tech Stack

PDF Extraction: pdfplumber

Chunk Splitting: LangChain CharacterTextSplitter

Embedding Model: sentence-transformers/all-MiniLM-L6-v2

Vector DB: Astra DB (via CassIO + LangChain)

Language Model (LLM): HuggingFaceH4/zephyr-7b-beta

Framework: LangChain

Platform: Google Colab / Python

## 📦 Installation

Install required Python packages:

pip install cassio langchain langchain-community langchain-huggingface datasets sentence-transformers pdfplumber

## 🔐 Required API Keys

Create `.env` or manually set these variables:

- ASTRA\_DB\_APPLICATION\_TOKEN  
- ASTRA\_DB\_ID  
- HUGGINGFACEHUB\_API\_TOKEN

## 📂 Usage

1. Upload a PDF  
2. Run the chatbot using the Colab notebook or script  
3. Ask questions and receive intelligent answers based on the PDF content.

## 📌 Project Status

Fully functional and ready for demo use. Supports all PDF types and interactive question answering.

## 🧑‍💻 Author

Created by Bhagwati Jadhav