**Model:**

To study the file and answer question you need **retrieval-augmented generation (RAG) model.**

we use ollama

* Ollama runs **completely offline**
* No API calls or cloud processing

**1.Install Ollama**

* Download from <https://ollama.com>
* Or type following cmd   
  for windows->winget install Ollama.Ollama  
  for mac->brew install ollama  
  for linux->curl -fsSL https://ollama.com/install.sh | sh
* Install and open Terminal or Command Prompt

2️.**Download a Model Locally**  
Run the following command:

> ollama pull mistral

Or

>ollama pull gemma:2b

*(You can replace mistral with llama3, phi, etc.)*

**3.run the python file which is in the same folder as this file by using the command**

>python localRAG.py

**4.enter the text file location.**

**5.ask questions or type “exit” for exit.**