✅ HepaBot LLM - Setup Guide (for Windows)

# 1. Install Python

* Download Python 3.9+ from: https://www.python.org/downloads/
* During installation, ensure you check 'Add Python to PATH'

# 2. Install System Dependencies

* ✅ Visual C++ Build Tools:
* - Download: https://visualstudio.microsoft.com/visual-cpp-build-tools/
* - Select 'Desktop development with C++' during installation.
* ✅ FFmpeg:
* Option 1 - Manual:
* - Download: https://www.gyan.dev/ffmpeg/builds/
* - Extract ffmpeg-release-essentials.zip to C:\ffmpeg
* - Add C:\ffmpeg\bin to your system PATH
* Option 2 - Chocolatey:
* - Run: choco install ffmpeg

# 3. Clone Project & Set Up Environment

* git clone https://github.com/IqraMuzaffar/HepaBot-LLM
* cd HepaBot-LLM
* python -m venv venv
* venv\Scripts\activate

# 4. Install Python Dependencies

* pip install --upgrade pip
* pip install -r requirements.txt
* If missing, use this list:
* - streamlit
* - langchain
* - langchain-ollama
* - pymupdf
* - faster-whisper
* - tqdm
* - pandas
* - reportlab
* - chromadb
* - transformers
* - torch
* - git+https://github.com/m-bain/whisperx.git
* Alternatively: pip install git+https://github.com/m-bain/whisperx.git

# 5. Set Up Ollama + LLM

* Download Ollama: https://ollama.com/download
* Run: ollama pull llama3
* If needed: ollama create llama3.2 -f <path-to-Modelfile>

# 6. Place Fine-Tuned BioClinicalBERT Model

* Place model in: models/finetuned\_clinicalbert/
* Should include:
* - config.json
* - pytorch\_model.bin or model.safetensors
* - tokenizer\_config.json
* - vocab.txt
* - special\_tokens\_map.json

# 7. Launch the App

* streamlit run app.py
* Visit: http://localhost:8501