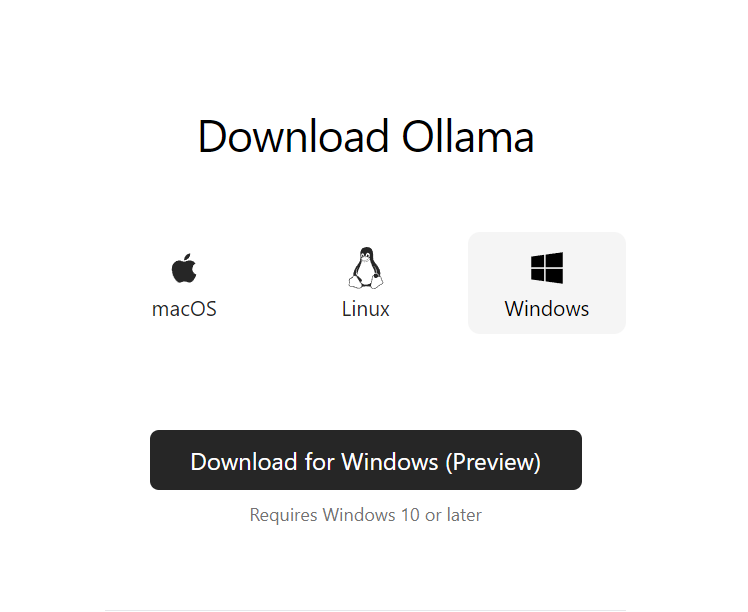
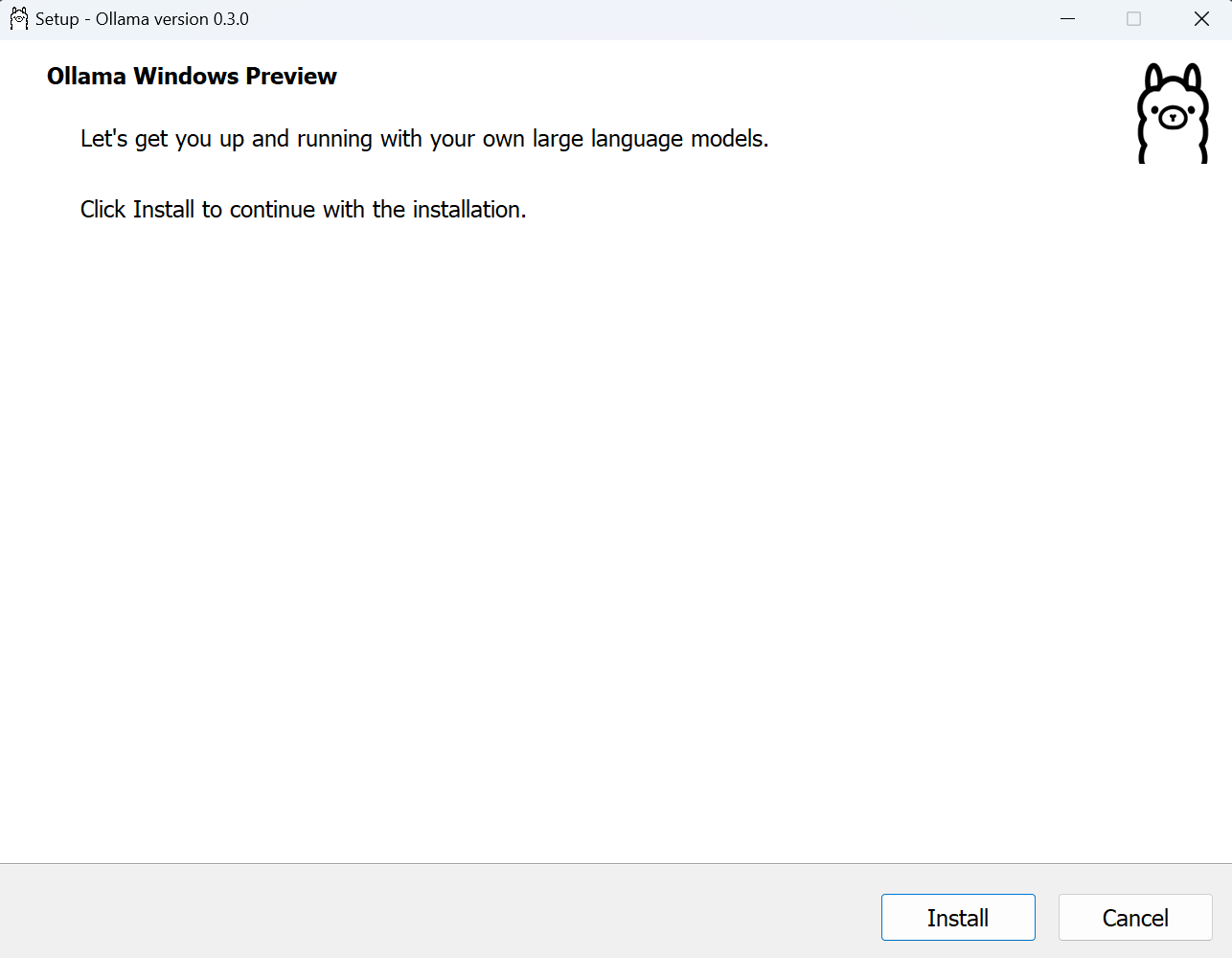
LLM installation and running for CDAC

Step 1: Download Ollama from (<https://ollama.com/download>)

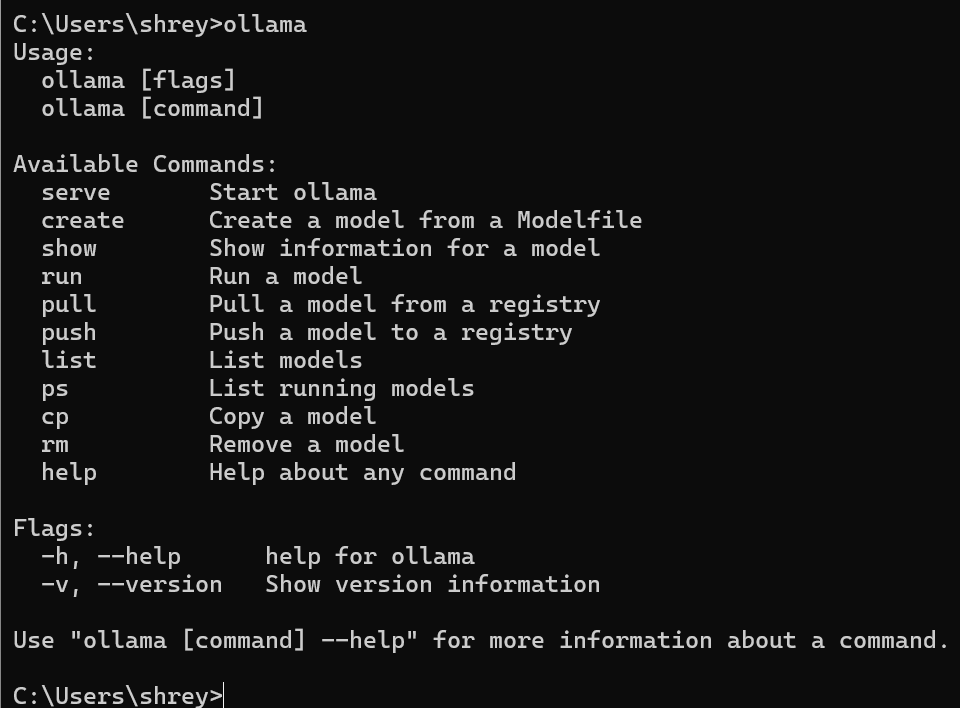


Step 2: Install Ollama



Step 3: confirm the installation

* Open CMD and type ollama and if it shows usage and Available Commands, it is installed correctly.

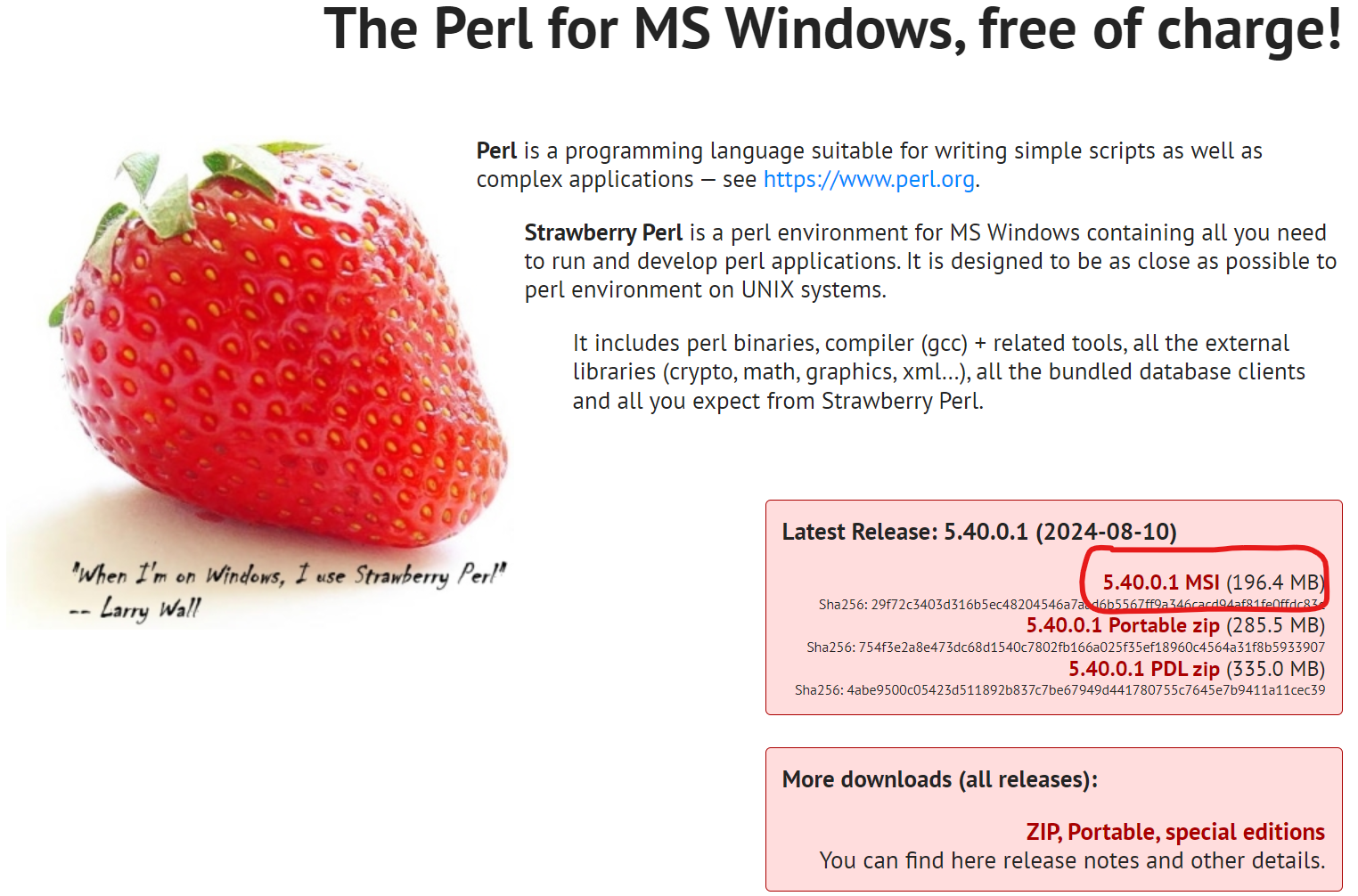


Step 4: Start Ollama using command (ollama serve)

If ollama is listening its already started

Step5 : download llama3.1 using command:(ollama pull llama3.1)

Step 6:install strawberry perl(<https://strawberryperl.com/>)



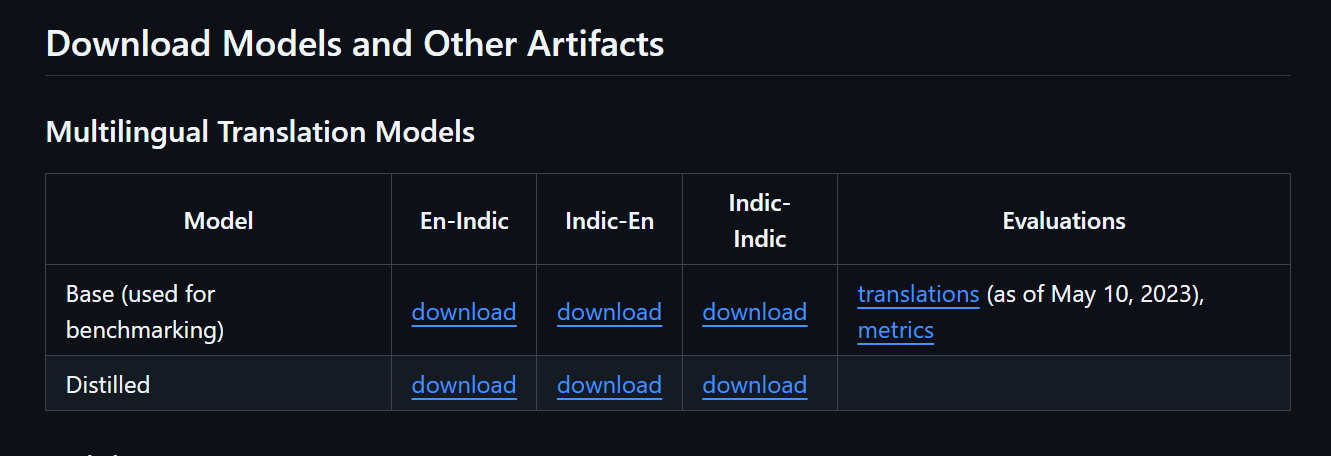
Step 7: install CUDA toolkit(<https://developer.nvidia.com/cuda-downloads>)

Install CUDA as per system specification.

Step 8: From AI4Bharat/IndicTrans2 github repository (https://github.com/AI4Bharat/IndicTrans2)

download all 3 Base models for translation and save in “models” folder

* Indic-en
* En-Indic
* Indic-Indic



Step 8 : On vscode install all Libraries using requirements.txt :

cffi==1.17.1

charset-normalizer==3.3.2

cryptography==43.0.1

fpdf==1.7.2

pdfminer.six==20231228

pdfplumber==0.11.4

pillow==10.4.0

pycparser==2.22

PyPDF2==3.0.1

pypdfium2==4.30.0

typing\_extensions==4.12.2

fastapi==0.95.1

uvicorn==0.22.0

langchain==0.0.288

langchain-community==0.0.1

torch

nltk

sentence-transformers

faiss-cpu

fastapi[all]

transformers==4.30.2

pydantic==1.10.2

jinja2==3.1.2

sentencepiece

indicnlp

sacremoses

Step 9 :Run main.py as service

uvicorn main:app --reload