1. Install Dependencies

You'll need:

```
pip install langchain langchain-community langchain-ollama beautifulsoup4 requests chromadb unstructured
```

```
In [8]: %pip install langchain langchain-community langchain-ollama beautifulsoup4 requests chromadb unstructured
```

```
Requirement already satisfied: langchain in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packag
es (0.3.21)
Requirement already satisfied: langchain-community in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/s
ite-packages (0.3.20)
Requirement already satisfied: langchain-ollama in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (0.3.0)
Requirement already satisfied: beautifulsoup4 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-p
ackages (4.13.3)
Requirement already satisfied: requests in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-package
s (2.32.3)
Requirement already satisfied: chromadb in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-package
s (0.6.3)
Requirement already satisfied: unstructured in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (0.17.2)
Requirement already satisfied: langchain-core<1.0.0,>=0.3.45 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/py
thon3.12/site-packages (from langchain) (0.3.49)
Requirement already satisfied: langchain-text-splitters<1.0.0,>=0.3.7 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/ve
nv/lib/python3.12/site-packages (from langchain) (0.3.7)
Requirement already satisfied: langsmith<0.4,>=0.1.17 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from langchain) (0.3.15)
Requirement already satisfied: pydantic<3.0.0,>=2.7.4 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from langchain) (2.10.6)
Requirement already satisfied: SQLAlchemy<3,>=1.4 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from langchain) (2.0.39)
Requirement already satisfied: PyYAML>=5.3 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pack
ages (from langchain) (6.0.2)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from langchain-community) (3.11.13)
Requirement already satisfied: tenacity!=8.4.0,<10,>=8.1.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/pyth
on3.12/site-packages (from langchain-community) (9.0.0)
Requirement already satisfied: dataclasses-json<0.7,>=0.5.7 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/pyt
hon3.12/site-packages (from langchain-community) (0.6.7)
Requirement already satisfied: pydantic-settings<3.0.0,>=2.4.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/
python3.12/site-packages (from langchain-community) (2.8.1)
Requirement already satisfied: httpx-sse<1.0.0,>=0.4.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.
12/site-packages (from langchain-community) (0.4.0)
Requirement already satisfied: numpy<3,>=1.26.2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (from langchain-community) (2.2.4)
Requirement already satisfied: ollama<1,>=0.4.4 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (from langchain-ollama) (0.4.7)
Requirement already satisfied: soupsieve>1.2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from beautifulsoup4) (2.6)
Requirement already satisfied: typing-extensions>=4.0.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from beautifulsoup4) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from requests) (3.4.1)
```

```
Requirement already satisfied: idna<4,>=2.5 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from requests) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from requests) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from requests) (2025.1.31)
Requirement already satisfied: build>=1.0.3 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from chromadb) (1.2.2.post1)
Requirement already satisfied: chroma-hnswlib==0.7.6 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from chromadb) (0.7.6)
Requirement already satisfied: fastapi>=0.95.2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-
packages (from chromadb) (0.115.11)
Requirement already satisfied: uvicorn[standard]>=0.18.3 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from chromadb) (0.34.0)
Requirement already satisfied: posthog>=2.4.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-p
ackages (from chromadb) (3.23.0)
Requirement already satisfied: onnxruntime>=1.14.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/s
ite-packages (from chromadb) (1.21.0)
Requirement already satisfied: opentelemetry-api>=1.2.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from chromadb) (1.31.1)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-grpc>=1.2.0 in /Users/nychanthrith/Documents/Chantharith/ME-Cha
tbot/venv/lib/python3.12/site-packages (from chromadb) (1.31.1)
Requirement already satisfied: opentelemetry-instrumentation-fastapi>=0.41b0 in /Users/nychanthrith/Documents/Chantharith/ME-Cha
tbot/venv/lib/python3.12/site-packages (from chromadb) (0.52b1)
Requirement already satisfied: opentelemetry-sdk>=1.2.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from chromadb) (1.31.1)
Requirement already satisfied: tokenizers>=0.13.2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from chromadb) (0.21.0)
Requirement already satisfied: pypika>=0.48.9 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-p
ackages (from chromadb) (0.48.9)
Requirement already satisfied: tqdm>=4.65.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from chromadb) (4.67.1)
Requirement already satisfied: overrides>=7.3.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (from chromadb) (7.7.0)
Requirement already satisfied: importlib-resources in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/s
ite-packages (from chromadb) (6.5.2)
Requirement already satisfied: grpcio>=1.58.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-p
ackages (from chromadb) (1.71.0)
Requirement already satisfied: bcrypt>=4.0.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from chromadb) (4.3.0)
Requirement already satisfied: typer>=0.9.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from chromadb) (0.15.2)
Requirement already satisfied: kubernetes>=28.1.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from chromadb) (32.0.1)
Requirement already satisfied: mmh3>=4.0.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pack
ages (from chromadb) (5.1.0)
```

```
Requirement already satisfied: orjson>=3.9.12 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-p
ackages (from chromadb) (3.10.15)
Requirement already satisfied: httpx>=0.27.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from chromadb) (0.28.1)
Requirement already satisfied: rich>=10.11.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from chromadb) (13.9.4)
Requirement already satisfied: chardet in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages
(from unstructured) (5.2.0)
Requirement already satisfied: filetype in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-package
s (from unstructured) (1.2.0)
Requirement already satisfied: python-magic in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from unstructured) (0.4.27)
Requirement already satisfied: lxml in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (f
rom unstructured) (5.3.1)
Requirement already satisfied: nltk in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (f
rom unstructured) (3.9.1)
Requirement already satisfied: emoji in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages
(from unstructured) (2.14.1)
Requirement already satisfied: python-iso639 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from unstructured) (2025.2.18)
Requirement already satisfied: langdetect in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packa
ges (from unstructured) (1.0.9)
Requirement already satisfied: rapidfuzz in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packag
es (from unstructured) (3.12.2)
Requirement already satisfied: backoff in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages
(from unstructured) (2.2.1)
Requirement already satisfied: unstructured-client in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/s
ite-packages (from unstructured) (0.31.4)
Requirement already satisfied: wrapt in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages
(from unstructured) (1.17.2)
Requirement already satisfied: psutil in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages
(from unstructured) (7.0.0)
Requirement already satisfied: python-oxmsg in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from unstructured) (0.0.2)
Requirement already satisfied: html5lib in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-package
s (from unstructured) (1.1)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.
12/site-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (2.5.0)
Requirement already satisfied: aiosignal>=1.1.2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (25.1.0)
Requirement already satisfied: frozenlist>=1.1.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/sit
e-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/s
ite-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (6.1.0)
```

```
Requirement already satisfied: propcache>=0.2.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (0.3.0)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/sit
e-packages (from aiohttp<4.0.0,>=3.8.3->langchain-community) (1.18.3)
Requirement already satisfied: packaging>=19.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-
packages (from build>=1.0.3->chromadb) (24.2)
Requirement already satisfied: pyproject_hooks in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-
packages (from build>=1.0.3->chromadb) (1.2.0)
Requirement already satisfied: marshmallow<4.0.0,>=3.18.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/pytho
n3.12/site-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community) (3.26.1)
Requirement already satisfied: typing-inspect<1,>=0.4.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from dataclasses-json<0.7,>=0.5.7->langchain-community) (0.9.0)
Requirement already satisfied: starlette<0.47.0,>=0.40.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from fastapi>=0.95.2->chromadb) (0.46.0)
Requirement already satisfied: anyio in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages
(from httpx>=0.27.0->chromadb) (4.8.0)
Requirement already satisfied: httpcore==1.* in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from httpx>=0.27.0->chromadb) (1.0.7)
Requirement already satisfied: h11<0.15,>=0.13 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-
packages (from httpcore==1.*->httpx>=0.27.0->chromadb) (0.14.0)
Requirement already satisfied: six>=1.9.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packa
ges (from kubernetes>=28.1.0->chromadb) (1.17.0)
Requirement already satisfied: python-dateutil>=2.5.3 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from kubernetes>=28.1.0->chromadb) (2.9.0.post0)
Requirement already satisfied: google-auth>=1.0.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from kubernetes>=28.1.0->chromadb) (2.38.0)
Requirement already satisfied: websocket-client!=0.40.0,!=0.41.*,!=0.42.*,>=0.32.0 in /Users/nychanthrith/Documents/Chantharith/
ME-Chatbot/venv/lib/python3.12/site-packages (from kubernetes>=28.1.0->chromadb) (1.8.0)
Requirement already satisfied: requests-oauthlib in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/sit
e-packages (from kubernetes>=28.1.0->chromadb) (2.0.0)
Requirement already satisfied: oauthlib>=3.2.2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-
packages (from kubernetes>=28.1.0->chromadb) (3.2.2)
Requirement already satisfied: durationpy>=0.7 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-
packages (from kubernetes>=28.1.0->chromadb) (0.9)
Requirement already satisfied: jsonpatch<2.0,>=1.33 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/
site-packages (from langchain-core<1.0.0,>=0.3.45->langchain) (1.33)
Requirement already satisfied: requests-toolbelt<2.0.0,>=1.0.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/
python3.12/site-packages (from langsmith<0.4,>=0.1.17->langchain) (1.0.0)
Requirement already satisfied: zstandard<0.24.0,>=0.23.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python
3.12/site-packages (from langsmith<0.4,>=0.1.17->langchain) (0.23.0)
Requirement already satisfied: coloredlogs in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pack
ages (from onnxruntime>=1.14.1->chromadb) (15.0.1)
Requirement already satisfied: flatbuffers in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pack
ages (from onnxruntime>=1.14.1->chromadb) (25.2.10)
Requirement already satisfied: protobuf in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-package
s (from onnxruntime>=1.14.1->chromadb) (5.29.4)
```

```
Requirement already satisfied: sympy in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages
(from onnxruntime>=1.14.1->chromadb) (1.13.3)
Requirement already satisfied: deprecated>=1.2.6 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/sit
e-packages (from opentelemetry-api>=1.2.0->chromadb) (1.2.18)
Requirement already satisfied: importlib-metadata<8.7.0,>=6.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/p
ython3.12/site-packages (from opentelemetry-api>=1.2.0->chromadb) (8.6.1)
Requirement already satisfied: googleapis-common-protos~=1.52 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/p
ython3.12/site-packages (from opentelemetry-exporter-otlp-proto-grpc>=1.2.0->chromadb) (1.69.2)
Requirement already satisfied: opentelemetry-exporter-otlp-proto-common==1.31.1 in /Users/nychanthrith/Documents/Chantharith/ME-
Chatbot/venv/lib/python3.12/site-packages (from opentelemetry-exporter-otlp-proto-grpc>=1.2.0->chromadb) (1.31.1)
Requirement already satisfied: opentelemetry-proto==1.31.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/pyth
on3.12/site-packages (from opentelemetry-exporter-otlp-proto-grpc>=1.2.0->chromadb) (1.31.1)
Requirement already satisfied: opentelemetry-instrumentation-asgi==0.52b1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbo
t/venv/lib/python3.12/site-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.52b1)
Requirement already satisfied: opentelemetry-instrumentation==0.52b1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/ven
v/lib/python3.12/site-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.52b1)
Requirement already satisfied: opentelemetry-semantic-conventions==0.52b1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbo
t/venv/lib/python3.12/site-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.52b1)
Requirement already satisfied: opentelemetry-util-http==0.52b1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/
python3.12/site-packages (from opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (0.52b1)
Requirement already satisfied: asgiref~=3.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from opentelemetry-instrumentation-asgi==0.52b1->opentelemetry-instrumentation-fastapi>=0.41b0->chromadb) (3.8.1)
Requirement already satisfied: monotonic>=1.5 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-p
ackages (from posthog>=2.4.0->chromadb) (1.6)
Requirement already satisfied: distro>=1.5.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pa
ckages (from posthog>=2.4.0->chromadb) (1.9.0)
Requirement already satisfied: annotated-types>=0.6.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from pydantic<3.0.0,>=2.7.4->langchain) (0.7.0)
Requirement already satisfied: pydantic-core==2.27.2 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from pydantic<3.0.0,>=2.7.4->langchain) (2.27.2)
Requirement already satisfied: python-dotenv>=0.21.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from pydantic-settings<3.0.0,>=2.4.0->langchain-community) (1.0.1)
Requirement already satisfied: markdown-it-py>=2.2.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from rich>=10.11.0->chromadb) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.
12/site-packages (from rich>=10.11.0->chromadb) (2.19.1)
Requirement already satisfied: greenlet!=0.4.17 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (from SQLAlchemy<3,>=1.4->langchain) (3.1.1)
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/pyt
hon3.12/site-packages (from tokenizers>=0.13.2->chromadb) (0.29.2)
Requirement already satisfied: click>=8.0.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-pac
kages (from typer>=0.9.0->chromadb) (8.1.8)
Requirement already satisfied: shellingham>=1.3.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from typer>=0.9.0->chromadb) (1.5.4)
Requirement already satisfied: httptools>=0.6.3 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site
-packages (from uvicorn[standard]>=0.18.3->chromadb) (0.6.4)
```

Requirement already satisfied: uvloop!=0.15.0,!=0.15.1,>=0.14.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from uvicorn[standard]>=0.18.3->chromadb) (0.21.0)
Requirement already satisfied: watchfiles>=0.13 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from uvicorn[standard]>=0.18.3->chromadb) (1.0.4)
Requirement already satisfied: websockets>=10.4 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from uvicorn[standard]>=0.18.3->chromadb) (15.0.1)
Requirement already satisfied: webencodings in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from html5lib->unstructured) (0.5.1)
Requirement already satisfied: joblib in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from nltk->unstructured) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from nltk->unstructured) (2024.11.6)
Requirement already satisfied: olefile in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from python-oxmsg->unstructured) (0.47)
Requirement already satisfied: aiofiles>=24.1.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from unstructured-client->unstructured) (24.1.0)
Requirement already satisfied: cryptography>=3.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from unstructured-client->unstructured) (44.0.2)
Requirement already satisfied: eval-type-backport>=0.2.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from unstructured-client->unstructured) (0.2.2)
Requirement already satisfied: nest-asyncio>=1.6.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from unstructured-client->unstructured) (1.6.0)
Requirement already satisfied: pypdf>=4.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from unstructured-client->unstructured) (5.4.0)
Requirement already satisfied: typing-inspection>=0.4.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from unstructured-client->unstructured) (0.4.0)
Requirement already satisfied: cffi>=1.12 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from cryptography>=3.1->unstructured-client->unstructured) (1.17.1)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from google-auth>=1.0.1->kubernetes>=28.1.0->chromadb) (5.5.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from google-auth>=1.0.1->kubernetes>=28.1.0->chromadb) (0.4.1)
Requirement already satisfied: rsa<5,>=3.1.4 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from google-auth>=1.0.1->kubernetes>=28.1.0->chromadb) (4.9)
Requirement already satisfied: filelock in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers>=0.13.2->chromadb) (3.17.0)
Requirement already satisfied: fsspec>=2023.5.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from huggingface-hub<1.0,>=0.16.4->tokenizers>=0.13.2->chromadb) (2024.12.0)
Requirement already satisfied: zipp>=3.20 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from importlib-metadata<8.7.0,>=6.0->opentelemetry-api>=1.2.0->chromadb) (3.21.0)
Requirement already satisfied: jsonpointer>=1.9 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from jsonpatch<2.0,>=1.33->langchain-core<1.0.0,>=0.3.45->langchain) (3.0.0)
Requirement already satisfied: mdurl~=0.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->chromadb) (0.1.2)
Requirement already satisfied: sniffio>=1.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packages (from anyio->httpx>=0.27.0->chromadb) (1.3.1)

```
Requirement already satisfied: mypy-extensions>=0.3.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.1
2/site-packages (from typing-inspect<1,>=0.4.0->dataclasses-json<0.7,>=0.5.7->langchain-community) (1.0.0)
Requirement already satisfied: humanfriendly>=9.1 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from coloredlogs->onnxruntime>=1.14.1->chromadb) (10.0)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/si
te-packages (from sympy->onnxruntime>=1.14.1->chromadb) (1.3.0)
Requirement already satisfied: pycparser in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/site-packag
es (from cffi>=1.12->cryptography>=3.1->unstructured-client->unstructured) (2.22)
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in /Users/nychanthrith/Documents/Chantharith/ME-Chatbot/venv/lib/python3.12/
site-packages (from pyasn1-modules>=0.2.1->google-auth>=1.0.1->kubernetes>=28.1.0->chromadb) (0.6.1)

[notice] A new release of pip is available: 23.2.1 -> 25.0.1
[notice] To update, run: pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

2. Load Data (PDFs & Websites)

We need to extract text from both **PDF Documents** and **webpage**

2.1 Extract Text from PDFs

Modify your existing LangChain PDF loader to handle multiple PDFs:

In [13]:
```python
from langchain_community.document_loaders import UnstructuredPDFLoader

def load_pdfs(pdf_paths):
    all_data = []
    for path in pdf_paths:
        loader = UnstructuredPDFLoader(path)
        all_data.extend(loader.load())
    return all_data

pdf_file = ["/Users/nychanthrith/Documents/Chantharith/ME-Chatbot/ai-model/data/raw_pdfs/Institute of Technology of Cambodia.pd
pdf_data = load_pdfs(pdf_file)
print(f"Load {len(pdf_data)} documents from PDFs")
```

```
Load 1 documents from PDFs
```

2.2 Scrape Websites

For retrieving data from ITC, RUPP, CADT, and UHS websites, use `BeautifulSoup`:

In [14]:
```python
import requests
from bs4 import BeautifulSoup
```

```python
def scrape_website(url):
    response = requests.get(url)
    if response.status_code == 200:
        soup = BeautifulSoup(response.text, "html.parser")
        return soup.get_text()
    return ""

urls = [
    "https://itc.edu.kh/",
    "https://www.rupp.edu.kh/",
    "https://cadt.edu.kh/",
    "https://uhs.edu.kh/"
]

web_data = [scrape_website(url) for url in urls]
print(f"Scraped {len(web_data)} websites")
```

```
Scraped 4 websites
```

3. Process and Store Data

3.1 Split Text into Chunks

Since PDFs and websites have long text, we split them for better retrieval.

In [17]:
```python
print(f"Type of pdf_data: {type(pdf_data)}")  # Should be a list of Document objects
print(f"Type of web_data: {type(web_data)}")  # Should be a list of Document objects
```

```
Type of pdf_data: <class 'list'>
Type of web_data: <class 'list'>
```

In [26]:
```python
from langchain_text_splitters import RecursiveCharacterTextSplitter
from langchain_core.documents import Document

def split_text(data):
    # Ensure input is a list of LangChain Documents
    if isinstance(data, str):
        data = [Document(page_content=data)]
    elif isinstance(data, list) and all(isinstance(item, str) for item in data):
        data = [Document(page_content=item) for item in data]  # Corrected 'd' to 'item'

    text_splitter = RecursiveCharacterTextSplitter(chunk_size=2000, chunk_overlap=300)
    return text_splitter.split_documents(data)
```

```python
# ✅ Convert PDF data to LangChain Documents if it's raw text
if isinstance(pdf_data, str):
    pdf_data = [Document(page_content=pdf_data)]
elif isinstance(pdf_data, list) and all(isinstance(item, str) for item in pdf_data):  # Corrected 'd' to 'item'
    pdf_data = [Document(page_content=item) for item in pdf_data]

# ✅ Convert Web data to LangChain Documents if needed
if isinstance(web_data, str):
    web_data = [Document(page_content=web_data)]
elif isinstance(web_data, list) and all(isinstance(item, str) for item in web_data):  # Corrected 'd' to 'item'
    web_data = [Document(page_content=item) for item in web_data]

# Now process
pdf_chunks = split_text(pdf_data)
web_chunks = split_text(web_data)

# Combine all chunks
all_chunks = pdf_chunks + web_chunks
print(f"Total chunks created: {len(all_chunks)}")
```

```
Total chunks created: 20
```

3.2 Store Data in a Vector Database

We store the chunks using **ChromaDB**:

In [27]:
```python
from langchain_community.vectorstores import Chroma
from langchain_ollama import OllamaEmbeddings

vector_db = Chroma.from_documents(
    documents=all_chunks,
    embedding=OllamaEmbeddings(model="nomic-embed-text"),
    collection_name="exam-chatbot"
)

print("Vector database created successfully")
```

```
Vector database created successfully
```

4. Build the Chatbot

4.1 Setup LLM & Retrieval

In [28]:
```python
from langchain_ollama.chat_models import ChatOllama
from langchain.retrievers.multi_query import MultiQueryRetriever
```

```python
from langchain.prompts import ChatPromptTemplate, PromptTemplate

llm = ChatOllama(model="llama3.2")

retriever = MultiQueryRetriever.from_llm(vector_db.as_retriever(), llm)
```

4.2 Create RAG Chain

In [29]:
```python
from langchain_core.runnables import RunnablePassthrough
from langchain_core.output_parsers import StrOutputParser

template = """Answer based ONLY on the following context:
{context}
Question: {question}
"""

prompt = ChatPromptTemplate.from_template(template)

chain = (
    {"context": retriever, "question": RunnablePassthrough()}
    | prompt
    | llm
    | StrOutputParser()
)
```

4.3 Create Chat Function

In [32]:
```python
from IPython.display import display, Markdown

def chat_with_bot(question):
    english_instruction = "Please respond in English: "
    response = chain.invoke(english_instruction + question)
    display(Markdown(response))

# Test Example
chat_with_bot("What are the admission requirements for ITC")
```

The Institute of Technology of Cambodia (ITC) offers a variety of training programs, but it appears that they also have an academic component with admission requirements. Based on the available information, here are some general admission requirements for ITC:

1. **Foundation Year**: For students who want to pursue a degree at ITC, they typically need to complete the Foundation Year program. The requirements for this program include:
   - Age: 17-25 years old
   - Academic qualifications: High school diploma or equivalent (e.g., A-levels, IB)
   - English proficiency: TOEFL or IELTS with a minimum score of 500 or 5.0, respectively
2. **Undergraduate Programs**: For students who want to pursue an undergraduate degree at ITC, they typically need to meet the following requirements:
   - Age: 17-25 years old (for domestic students) or 18-30 years old (for international students)
   - Academic qualifications: High school diploma or equivalent (e.g., A-levels, IB), plus specific prerequisite courses for each program
   - English proficiency: TOEFL or IELTS with a minimum score of 500 or 5.0, respectively
   - Other requirements may include:
     - SAT or ACT scores for international students
     - Letters of recommendation from teachers or mentors
     - Personal statement or essay
3. **Postgraduate Programs**: For students who want to pursue a postgraduate degree at ITC, they typically need to meet the following requirements:
   - Age: 25-40 years old (for domestic students) or 26-35 years old (for international students)
   - Academic qualifications: Bachelor's degree from an accredited institution
   - English proficiency: TOEFL or IELTS with a minimum score of 500 or 5.0, respectively
   - Other requirements may include:
     - Master's degree transcripts and certificates
     - Letters of recommendation from teachers or mentors
     - Personal statement or essay

Please note that these are general requirements and may vary depending on the specific program and admission cycle. I recommend checking the official website of ITC or contacting their admissions office directly for more information and up-to-date requirements.

Additionally, the Institute of Technology of Cambodia also offers various scholarships and financial aid options to support students with limited financial resources.