

INSTRUCTIONS FOR VERSION 1.0

ChromaDB Plugin for LM Studio

The ChromaDB Plugin for [LM Studio](#) adds a vector database to LM Studio utilizing ChromaDB! Tested on a 1000 page legal treatise

COMPATIBLE with [Python 3.10](#).

Link to [Medium article](#)

Table of Contents

1. [Installation Instructions](#)
2. [Usage Guide](#)
3. [Important Notes](#)
4. [Feedback](#)
5. [Final Notes](#)
6. [Clearing Database](#)

Installation Instructions

- **Step 1:** Download all the files in this repository and put them into a directory.
- **Step 2:** Install [CUDA 11.8](#) if it's not already installed.
- **Step 3:** Go to the folder where my repository is located, open a command prompt and run to create a [virtual environment](#):

```
python -m venv .
```

- **Step 4:** Then run to activate the virtual environment:

```
.\Scripts\activate
```

- **Step 5:** Then run to make sure "[pip](#)" is updated. If you don't, it might not install the subsequent libraries:

```
python -m pip install --upgrade pip
```

- **Step 6:** Then run to install PyTorch with CUDA support:

```
pip install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu118
```

- **Step 7:** Then run to install the rest of the Python libraries via the [requirements.txt](#):

```
pip install -r requirements.txt
```

[Back to top](#)

Usage Guide

- **Step 1:** In the same command prompt run:

```
python gui.py
```

- **Step 2:** Click the "Choose Documents" button and choose one or more documents to include in the vector database.
 - **Note:** Only PDFs with OCR done on them will work as of Version 1. A folder named "Docs_to_DB" will be created and populated.
- **Step 3:** Click the "Create ChromaDB" button. A folder named "Vector_DB" will be created if it doesn't already exist and the DB will be created in there (see notes below).
 - **Note:** A message will appear with instructions on how to monitor CUDA usage. Please follow them.
 - **Note:** The embedding model will be downloaded to your cache folder if it's not downloaded already. Once downloaded, the vector database will automatically be created. Watch your CUDA usage to verify that it's working - pretty awesome! The database is fully-created when CUDA usage drops to zero.
- **Step 4:** Open LM Studio, select a model, and click "Start Server."
 - **Note:** The formatting of the prompt in my scripts is specifically geared to work with any Llama2 "chat" models. Any others might not work if they provide an intelligible response at all. This can be addressed in future versions.
- **Step 5:** Enter your query and click the "Submit Query" button and be amazed at the response you get.
 - **Note:** If will give an error if you don't start the server before clicking "Submit Query."
 - **Note:** For [extra entertainment](#), watch LM Studio server's log window to watch it interact with the vector database!

[Back to top](#)

Important Notes

- **Compatibility:** This is a personal project and was specifically tested using CUDA 11.8 and the related PyTorch installation.
- **Embedding Model:** This plugin uses "hkunlp/instructor-large" as the embedding model. Look [here](#) for more details. If people express an interest, I'll likely include other embedding models in future versions!

[Back to top](#)

Feedback

My motivation to improve this beyond what I personally use it for is directly related to people's interest and suggestions. All feedback, positive and negative, is welcome! I can be reached at the LM Studio discord server or "bbc@chintellalaw.com".

[Back to top](#)

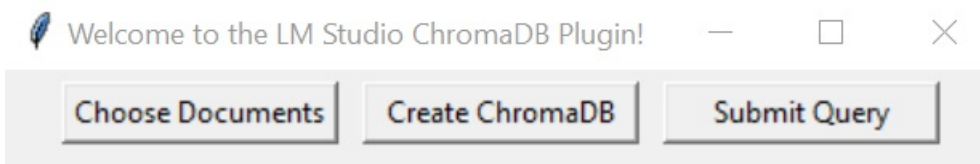
Final Notes

- **Note:** I only tested this on Windows 10 but can possibly expand on this in later versions.
- **Note:** Please be aware that when you click "Create Database" as well as "Submit Query" the GUI will hang. Just wait...it'll resume. This is a minor characteristic of the scripts that can easily be fixed in future versions.
- **Note:** Everytime you want to use the program again, enter the folder, activate the virtual enviroment using `.\Scripts\activate` and run `python gui.py`.

Clearing Database

- Simply delete all the files in the "Docs_to_DB" and "Vector_DB" folders.

[Back to top](#)



What's the deadline to hold a preliminary protective hearing in a dependency case?

According to Ga. Juv. Prac. & Proc. § 6:29, the preliminary protective hearing must be held promptly after a child is removed from their home and no later than 72 hours after the child is placed in foster care. If the 72-hour time frame expires on a weekend or legal holiday, the hearing shall be held on the next day which is not a weekend or legal holiday. Failure to conduct the hearing within the lawful time frames mandates dismissal of the petition without prejudice.