**輕鬆本地佈署google-gemma 與 ollama安裝**

<https://medium.com/ricos-note/輕鬆本地佈署google-gemma-577d002c08c0>

安裝教學影片: https://youtu.be/1034tFMRjvM

**[install ollama]:**

**至**[**https://ollama.com/download**](https://ollama.com/download)**下載windows版. 依指示安裝.**

**一張含有 文字, 螢幕擷取畫面, 字型, 標誌 的圖片

AI 產生的內容可能不正確。**

**裝好後看到ollama在常駐在windows右下角工具列中..**

**一張含有 螢幕擷取畫面, 小工具, 多媒體, 電子裝置 的圖片

AI 產生的內容可能不正確。**

**[serve]**

**運行 Ollama 時會佔用 Port 11434 ，目的是為了後續可以執行 API Service 作預備；如果想要更改 port 號，以 macOS 為例子要使用 launchctl setenv 來設定 ollama 的環境變數並重啟（詳情可以參考 ollama GitHub**[**FQA**](https://github.com/ollama/ollama/blob/main/docs/faq.md)**）。**

**ollama serve**

**或**

**下載gemma:2b模型. 順便啓動.**

**ollama run gemma:2b**

**[建立專門env來跑**opneWebUI**]**

Using anaconda to create a vm for open-webui: python 3.11 version

Ex: create vm: opneWebUI

Using anaconda to install CMD environment, Under cmd:

**python.exe -m pip install --upgrade pip**

**Installation via Python pip 🐍**

Open WebUI can be installed using pip, the Python package installer. Before proceeding, ensure you're using **Python 3.11** to avoid compatibility issues.

1. **Install Open WebUI**: Open your terminal and run the following command to install Open WebUI:

**pip install open-webui**

1. **Running Open WebUI**: After installation, you can start Open WebUI by executing:

**open-webui serve**

第一次啓動. 下載主要模型.

一張含有 文字, 螢幕擷取畫面, 字型, 設計 的圖片

AI 產生的內容可能不正確。

This will start the Open WebUI server, which you can access at [http://localhost:8080](http://localhost:8080/)

一張含有 文字, 螢幕擷取畫面, 軟體, 多媒體軟體 的圖片

AI 產生的內容可能不正確。

第一次要先設定一個管理員帳號, 才能進入

使用教學:

https://www.youtube.com/watch?v=p3uwtvy8uVI

一張含有 文字, 螢幕擷取畫面, 軟體, 電腦圖示 的圖片

AI 產生的內容可能不正確。

按管理ollama api連線右邊的下載按鈕.

一張含有 文字, 螢幕擷取畫面, 軟體, 電腦圖示 的圖片

AI 產生的內容可能不正確。

再按點選這裡

依gpu ram大小選擇 一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

AI 產生的內容可能不正確。, 若模型超過GPU RAM大小 或沒有GPU.則會以CPU來執行.一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

AI 產生的內容可能不正確。

把名字 gemma3:12b 拷貝下來. 貼回原畫面.按右邊下載.

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

AI 產生的內容可能不正確。

下載中.完成即可交談使用.

一張含有 文字, 螢幕擷取畫面, 軟體, 電腦圖示 的圖片

AI 產生的內容可能不正確。

多模態交談: 可語音輸入或上傳圖檔

一張含有 文字, 輪, 輪胎, 車輛 的圖片

AI 產生的內容可能不正確。

**Ps. 也可自行下載**[**gemma3**](https://ollama.com/library/gemma3) **再匯入**

https://huggingface.co/google/gemma-3-4b-it/tree/main

或透過 ollama 下載

ollama pull gemma 3:12b

**[Ollama 模型管理指令]**

**list**

下載完成後我們可以用 list 來查看目前本機上有多少個 LLM，下載完的模型都會在 .ollama /models/manifests/registry.ollama.ai/library/ 中。

ollama list  
  
// model storage path   
cd .ollama /models/manifests/registry.ollama.ai/library/

**show**

下載完成後我們可以用 show 搭配不同參數來查看模型的資訊，包括

1. license：顯示模型所使用的 license
2. parameters：模型可以設定的參數
3. template：模型的 System Prompt 和 input pattern
4. modelfile：顯示除了 license 的所有模型訊息

// '--license', '--modelfile', '--parameters' or '--template'  
  
ollama show --license gemma:2b  
ollama show --modelfile gemma:2b  
ollama show --parameters gemma:2b  
ollama show --template gemma:2b

**run**

接下來我們用 run 指命來執行剛下載好的模型：

ollama run gemma:2b

**[其它參十考]**

**Install and Run Google's Gemma 3: Strongest AI Model for a Single GPU - Installation Instructions**

https://youtu.be/jSXTE2ApI0s

**Google Gemma 3 Vision AI Model on Linux Ubuntu - Install and Run Locally**

<https://youtu.be/NuTAwdY2i1o>