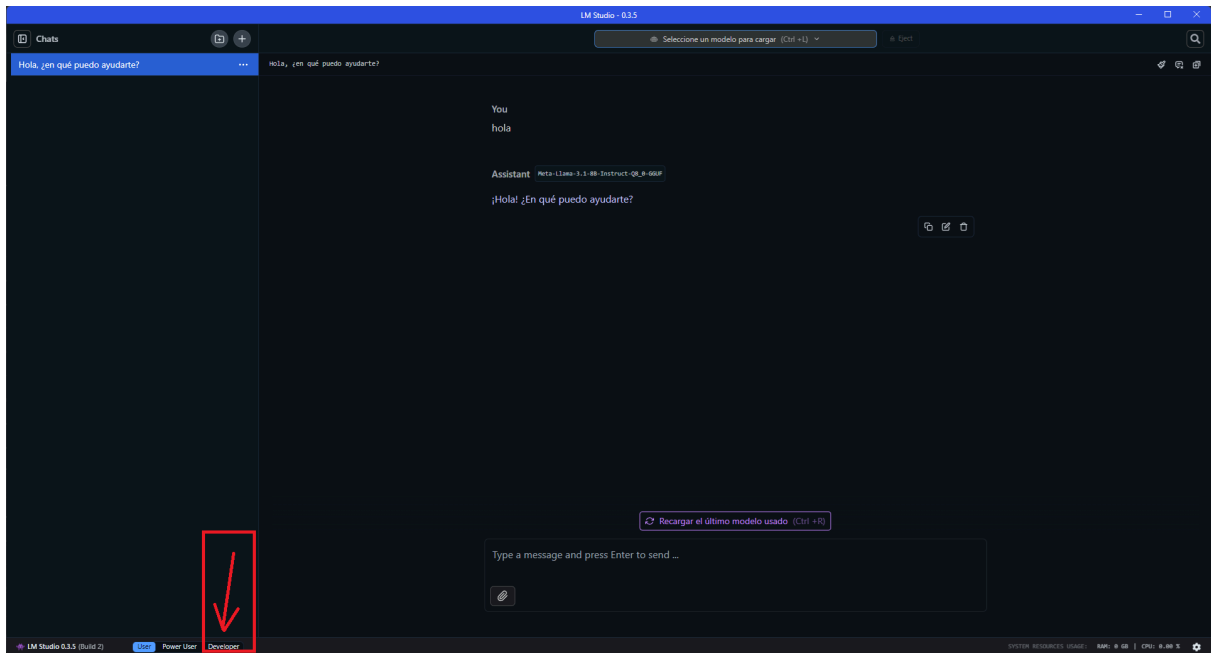
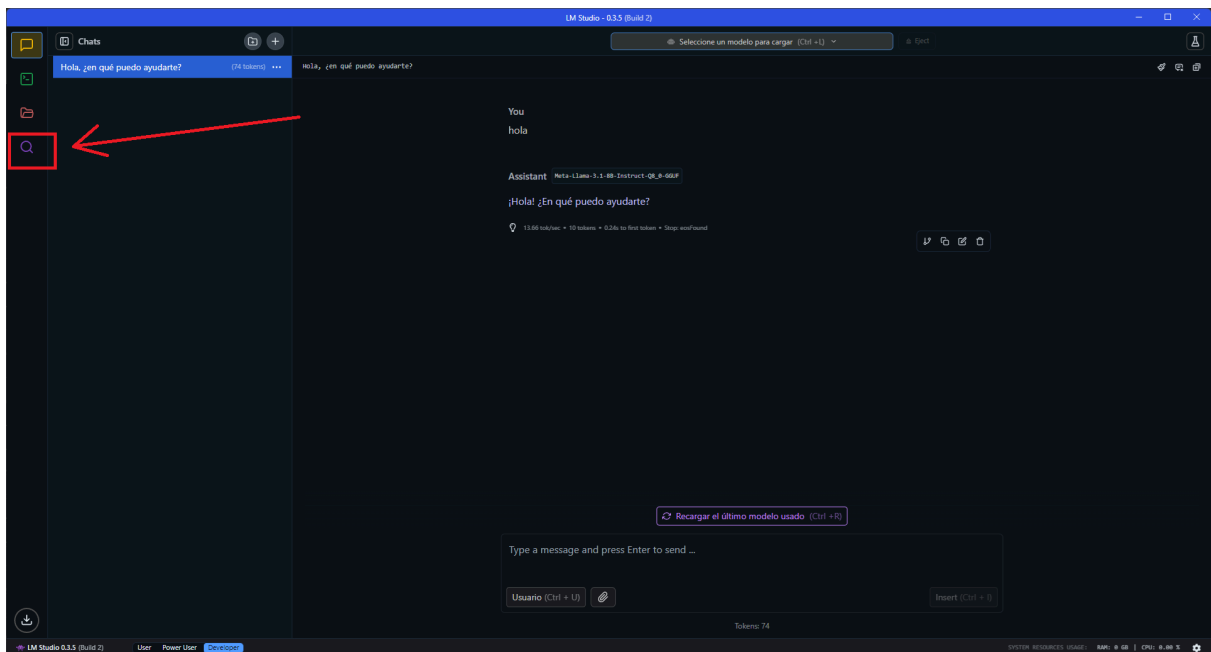


Paso 1: modo “developer”:



Paso 2: Buscamos un modelo para descargar.



Paso 3: descargar un modelo

The screenshot shows the LM Studio Mission Control interface. The search bar at the top contains 'llama 3.1'. The search results list several models, with 'Meta Llama 3.1 8B' highlighted. The details panel on the right shows the model's architecture (Llama), parameters (8B), and status (222 likes, 41511 downloads). It also mentions that it requires LM Studio version 0.2.29. The download button is labeled 'Download 8.54 GB'.

Model Search: Ctrl + M
Runtimes: Ctrl + R

Showing 2817 models

Mejor Coincidencia

Meta Llama 3.1 8B

LM Studio Staff Pick

The latest in Meta's long-running Llama series, Llama 3.1 is another jack of all trades and master of some, now in 8 languages and up to 128k tokens.

Architecture: Llama
Params: 8B
Stats: 222 likes, 41511 downloads
Last updated: 115 days ago

6 download options available

Q8_0 Meta Llama 3.1 8B Instruct 8B 8.54 GB

Model Readme

Pulled from the model's repository

Community Model> Llama 3.1 8B Instruct by Meta

LM Studio Community models highlights program. Highlighting new & noteworthy models by the community. Join the conversation on Discord.

Model creator: meta-llama
Original model: Meta-Llama-3.1-8B-Instruct
GGUF quantization: provided by bartowski based on llama.cpp release b3472

Important: Requires LM Studio version 0.2.29, available now [here!](#)

Model Summary:

Llama 3.1 is an update to the previously released family of Llama 3 models. It is

Cancel Download 8.54 GB

En nuestro caso tenemos nuestras propia versiones:

The screenshot shows the LM Studio Mission Control interface. The search bar at the top contains 'Anrom97'. The search results list two models, with 'LLama-3.2-WIS-v1.0-GGUF' highlighted. The details panel on the right shows the model's repository (Anrom97/LLama-3.2-WIS-v1.0-GGUF), stats (0 likes, 27 downloads), and that it was downloaded. It also mentions that it was trained 2x faster with Unsloth and Huggingface's TRL library. The download button is labeled 'Use in New Chat'.

Model Search: Ctrl + M
Runtimes: Ctrl + R

Showing 2 models

Mejor Coincidencia

LLama-3.2-WIS-v1.0-GGUF

Repository: Anrom97/LLama-3.2-WIS-v1.0-GGUF

Stats: 0 likes, 27 downloads
Last updated: 2 days ago

One download option available

Q8_0 Unsloth 3.42 GB

Model Readme config.json

Pulled from the model's repository

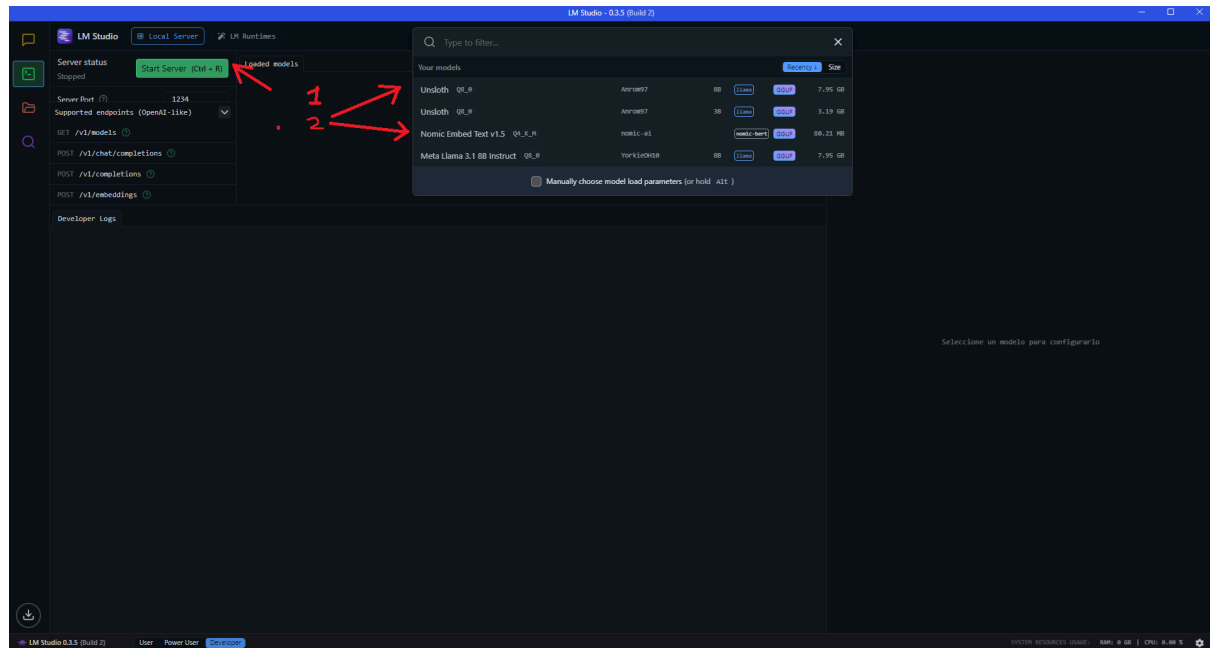
Uploaded model

- Developed by: Anrom97
- License: apache-2.0
- Finetuned from model : unsloth/llama-3.2-3b-instruct-bnb-4bit

This llama model was trained 2x faster with Unsloth and Huggingface's TRL library.

Cancel Use in New Chat

Paso 4: Levantar el servidor.



Primero: Cargamos el modelo (llama 3.1 WIS).

Segundo: Cargamos el modelo de embedding (Nomic Embed Text 1.5)

Tercero: Oprimimos en "Start Server".

Correr el chatbot:

Librerías necesarias (instalar antes de correr los scripts):

```
pip install chromadb==0.5.20
```

```
pip install streamlit==1.38.0
```

```
pip install python-docx==1.1.2
```

```
pip install pypdf==5.1.0
```

```
pip install PyPDF2==3.0.1
```

```
pip install langchain==0.3.12
```

```
pip install langchain-community==0.3.12
```

```
pip install langchain-huggingface==0.1.2
```

```
pip install langchain-chroma==0.1.4
```

```
pip install langchain-core==0.3.25
```

```
pip install langchain-openai==0.1.14
```

```
pip install langchain-text-splitters==0.3.3
```

```
pip install langchainhub==0.1.21
```

```
pip install sentence-transformers==2.7.0
```

```
pip install torch==2.4.1
```

```
pip install tqdm==4.66.5
```

```
pip install requests==2.32.3
```

```
pip install numpy==1.26.4
```

```
pip install scikit-learn==1.5.2
```

También se puede instalar todas las dependencias de una sola vez utilizando el archivo requirements.txt: **pip install -r requirements.txt**

Crear la base de datos de vectores de RAG:

```
python create_database.py
```

Correr el chatbot (Es necesario tener el servidor de LLMStudio levantado):

```
streamlit run app.py
```