

Using Local LLMs and RAG for Content Promotion

By: Sean Twiehaus

Disclaimer

I am not an expert



SLUUG on Social Platforms

YouTube



Saint Louis Unix Users Group and Linux

@sluugOrg · 72 subscribers · 72 videos

The Saint Louis Unix Users Group (SLUUG) is an IRS 501(c)(6) designated not-for-profit profes... >

sluug.org and 6 more links

Subscribed

Home Videos Live Community

Latest Popular Oldest



2021-12-16 STLLINUX ~ St. Louis Linux Users Group

19 views · 2 weeks ago



2021-12-08 SLUUG ~ St. Louis Unix Users Group

11 views · 2 weeks ago



2021-11-18 STLLINUX ~ St. Louis Linux Users Group

19 views · 3 weeks ago



2021-11-10 SLUUG ~ St. Louis Unix Users Group

27 views · 3 weeks ago



2021-10-21 STLLINUX ~ St. Louis Linux Users Group

22 views · 3 weeks ago



2021-10-13 SLUUG ~ St. Louis Unix Users Group

19 views · 3 weeks ago



2021-10-13 SLUUG ~ St. Louis Unix Users Group - gits0 Demo

19 views · 3 weeks ago



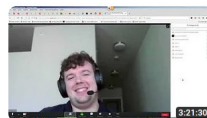
2021-09-16 STLLINUX ~ St. Louis Linux Users Group

6 views · 3 weeks ago



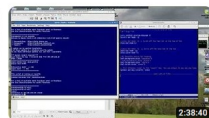
2021-09-08 SLUUG ~ St. Louis Unix Users Group

2 views · 3 weeks ago



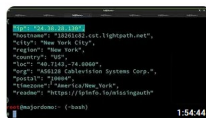
2021-08-19 STLLINUX ~ St. Louis Linux Users Group

17 views · 3 weeks ago



2021-08-11 SLUUG ~ St. Louis Unix Users Group

105 views · 3 weeks ago



2021-07-22 STLLINUX ~ St. Louis Linux Users Group

27 views · 3 weeks ago

Twitter

← St. Louis Linux/Unix Users Groups

21 posts



Edit profile

St. Louis Linux/Unix Users Groups

@SLUUG_Org

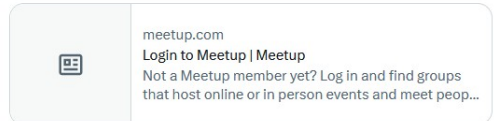
The St. louis Linux Unix Users Group (SLUUG) is a non-profit, volunteer group that provides education, information and/or support for Linux users.

📍 St. Louis, Missouri 🌐 sluug.org 📅 Joined June 2022

35 Following 8 Followers

Posts Replies Highlights Articles Media Likes

 **St. Louis Linux/Unix Users Groups** @SLUUG_Org · Feb 8, 2023 ...
Join us (Virtually) for the St. Louis Unix Users Group. @LVLammert will be presenting "btrFS Hands On" and @seantwie03 will be demoing YADM (yadm.io)



🗨️ 1 🍷 16 📌 🔄

Run Large Language Models Locally



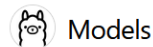
Get up and running with large language models, locally.

Run [Llama 2](#), [Code Llama](#), and other models.
Customize and create your own.

Download ↓

Available for macOS, Linux, and
Windows (preview)

[Ollama](#)



Filter by name...

Featured

gemma

Gemma is a family of lightweight, state-of-the-art open models built by Google DeepMind.

↓ 209.1K Pulls ↻ 69 Tags ⌚ Updated 13 days ago

llama2

Llama 2 is a collection of foundation language models ranging from 7B to 70B parameters.

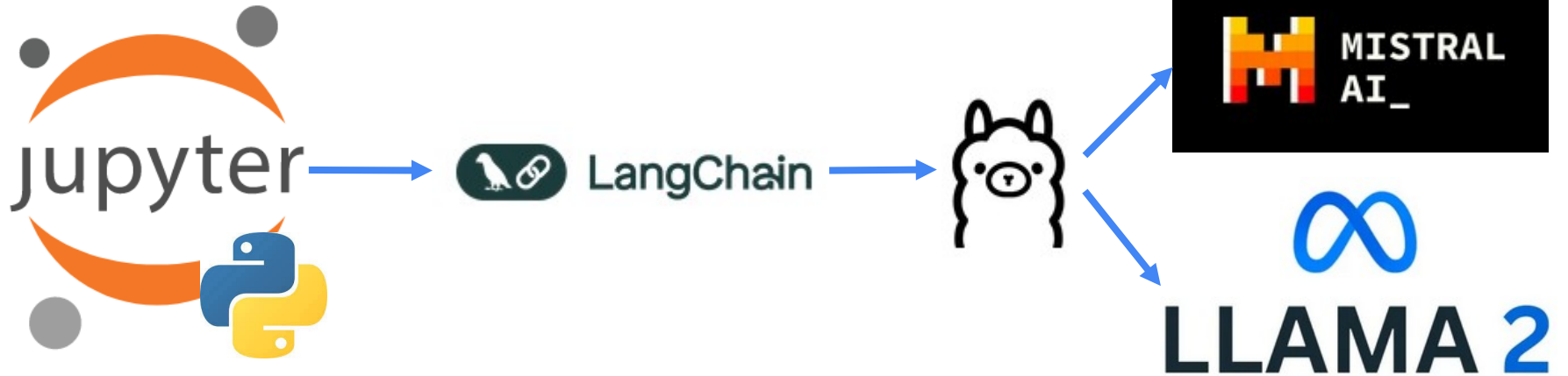
↓ 781.2K Pulls ↻ 102 Tags ⌚ Updated 5 weeks ago

mistral

The 7B model released by Mistral AI, updated to version 0.2.

↓ 349.9K Pulls ↻ 53 Tags ⌚ Updated 2 months ago

Call Local LLM From Jupyter Notebook Using LangChain



Demo - The Problem with LLMs - 1

```
lang-chain-demo\sluug_2021 on master [?]
```

```
> ollama run gemma
```

```
>>> What was the presentation to the St. Louis UNIX Users  
Group on February 10th, 2021 about?
```

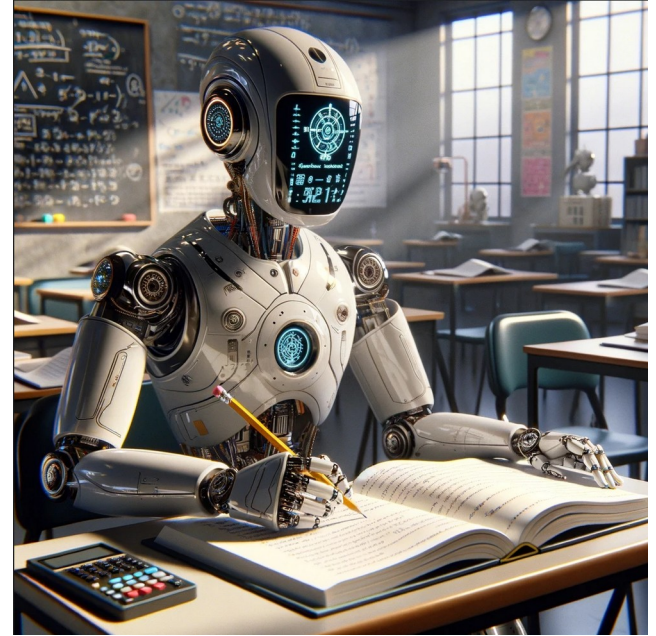
```
Answer: I do not have access to real-time information,  
therefore I cannot answer the question regarding the  
presentation to the St. Louis UNIX Users Group on  
February 10th, 2021.
```

The Problem with Large Language Models (LLMs) - 1

- Trained on a broad spectrum of data
- Does not include **newer**, proprietary, or domain-specific information.

Prompt Engineering / Augmented Generation

- Augmented Generation is like an open book test for the LLM.
- Is equivalent to copy/pasting text into the Chat GPT Input box before typing your question.



Call to Action - Volunteers Wanted

- LLMs are useful
- LLMs can benefit the LUG
- Volunteers Wanted
 - Turn this Proof of Concept into a Realized Workflow
 - MVP - Create a Script to generate and email
 - Open Source!
 - Continue to Enhance
 - Stable Diffusion → YT Thumbnails
 - ...
- Mine DISCUSS for future presentation ideas

Retrieval-Augmented Generation (RAG)



Demo - The Problem with LLMs - 2

```
llm.invoke(system_prompt + doc_str + "What was the BASE presentation to the St. Louis UNIX Users Group on February 10th, 2021?")
```

✓ 8.4s

Python

' A Gentle Introduction To git - This presentation covered the basics of using Git as a version control system, including the working folder, the staging area, and the repository. It also provided an overview of common Git commands for initializing a repository, making and reverting commits, and branching and merging. The goal was to give attendees a solid foundation in using Git for version control.'

The Problem with Large Language Models (LLMs) - 2

- Limited Context Window
 - Hard Limitation

Retrieval-Augmented Generation (RAG)

- Break text into chunks
- Retrieve chunks relevant to the Question
- Send relevant chunks along with Prompt
- Relevant chunks “should” fit in Context



Additional Resources

- Enterprise Grade RAG - [azure-search-openai-demo](#)
 - Hybrid RAG
 - Semantic
 - Keyword Search
 - Half a dozen Azure cloud services

Call to Action - Volunteers Wanted

- LLMs are useful
- LLMs can benefit the LUG
- Volunteers Wanted
 - Turn this Proof of Concept into a Realized Workflow
 - MVP - Create a Script to generate and email
 - Open Source!
 - Continue to Enhance
 - Stable Diffusion → YT Thumbnails
 - ...
- Mine DISCUSS for future presentation ideas