LICENSE **MIT** **in LINKEDIN**
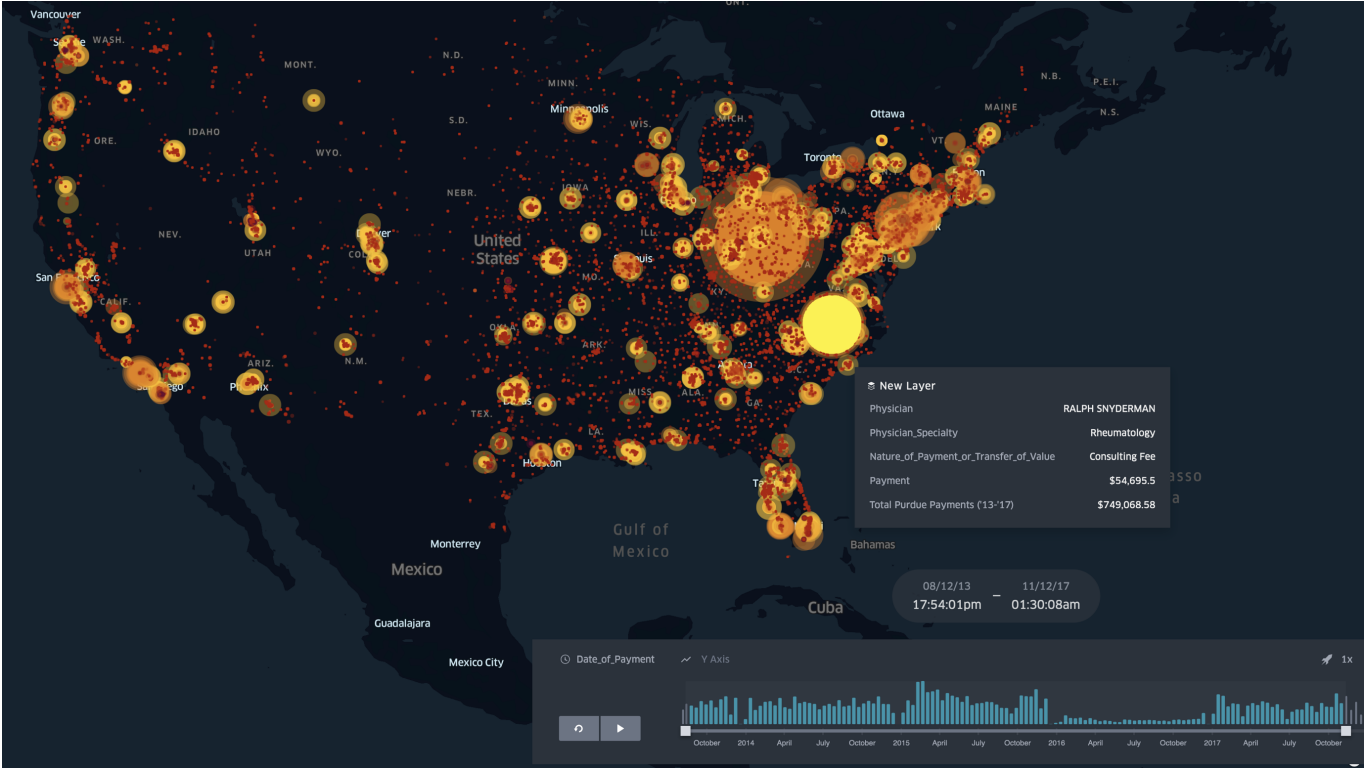
▶ Table of Contents

A Local LLM Research Assistant

Imagine being able to rescue your research assistant from manually classifying 24 thousand entries by hand. Better yet, imagine being able to conduct research that you're passionate about even when you have limited time and no research assistants at all!

Ollama on your laptop is a great option if you have a limited budget, concerns about putting your data in the cloud, or just want to tinker with large language models (LLMs) without any platform lock-in. Mistral open-sourced their excellent 7 billion parameter model under an Apache license. The pair make a powerful combination.

This repo processes author affiliations and returns the nation & city, states

Data: PubMed metadata on JAMA articles published between 2013 and 2023 downloaded as a .nbib citation manager file Tools: Python, Ollama, LangChain, Pandas, and Mistral 7b Instruct

Results: 24,184 affiliations were processed. Manual evaluation of a random sample of 2,000 affiliations showed 13 errors (0.0.65%), though the results also included 14 examples of successful processing that depended on context that did not appear in the affiliation (e.g. "senior policy service professor at george washington university school of nursing's center for health policy and media." →"Washington, dc").

(back to top)

Built With

- Ollama

- [Mistral.AI](#)
- [LangChain](#)
- [HuggingFace](#)

<div align="right">

([back to top](#))

</div>

Getting Started

To get started:

- download [Ollama](#), open it, and leave it running(Windows users may only have a preview version depending on development timeline).
- Set up a venv with the libraries from the requirements.txt file.
- If you haven't already pulled the Mistral-Instruct model, uncomment the line "!ollama pull mistral:instruct" in th LLM_geography_extractor.py file
- Run LLM_geography_extractor.py

Prerequisites

This is an example of how to list things you need to use the software and how to install them.

- Python 3.9+
- LangChain
- HuggingFace
- Ollama

Usage

This is a quick demo of using Mistral 7b to extract geographic information from medical journal authors listed intstitutional affiliations. For example: Input: "department of medicine, university of pennsylvania." Response: "philadelphia, pa"

The dataset comes from a .nbib citation manager file drawn from a PubMed search for all articles published in JAMA between 2013 and 2023.

Contributing

Contributions are what make the open source community such an amazing place to learn, inspire, and create. Any contributions you make are **greatly appreciated**.

If you have a suggestion that would make this better, please fork the repo and create a pull request. You can also simply open an issue with the tag "enhancement". Don't forget to give the project a star! Thanks again!

1. Fork the Project
2. Create your Feature Branch (`git checkout -b feature/AmazingFeature`)
3. Commit your Changes (`git commit -m 'Add some AmazingFeature'`)
4. Push to the Branch (`git push origin feature/AmazingFeature`)
5. Open a Pull Request

<div align="right">

([back to top](#))

</div>

License

Distributed under the MIT License. See `LICENSE.txt` for more information.

([back to top](#))

Contact

Your Name - alexrich@duck.com

Project Link: https://github.com/your_username/repo_name

([back to top](#))

Acknowledgments

Greate resources and people whose work make this all possible:

- [Ollama](#)
- [Mistral.AI](#)
- [LangChain](#)
- [HuggingFace](#)

([back to top](#))