



July 28, 2026

Using LLM Chat to Simplify User Interactions with GDEX data

Anna Scribner

SIParCS Intern

Nathan Hook, Harsha Hampapura, Jason Cunning, Eric Nienhouse, Chia-Wei Hsu



Table of Contents



- 1. Background**
- 2. The Problem**
- 3. My Project**
- 4. Demonstration**
- 5. Challenges**
- 6. Future Work**



Have you ever searched for something and felt overwhelmed by all the information?



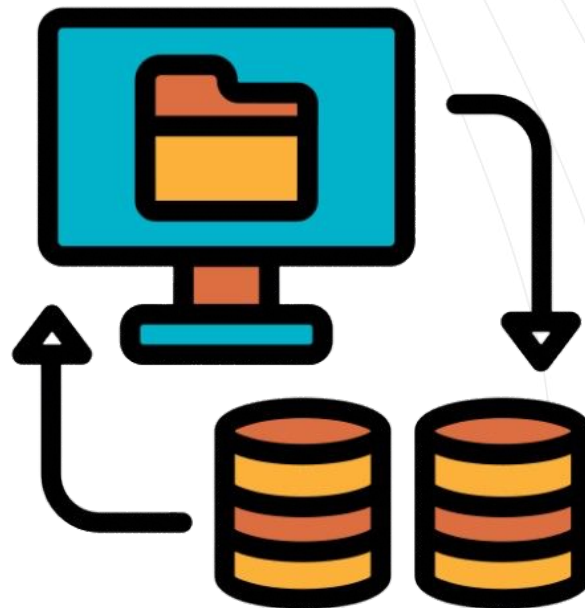
The Problem



1.53 TB Dataset,
Global Hourly ~ 0.5 -degree Land
Surface Air Temperature
Dataset

306,000 Photos

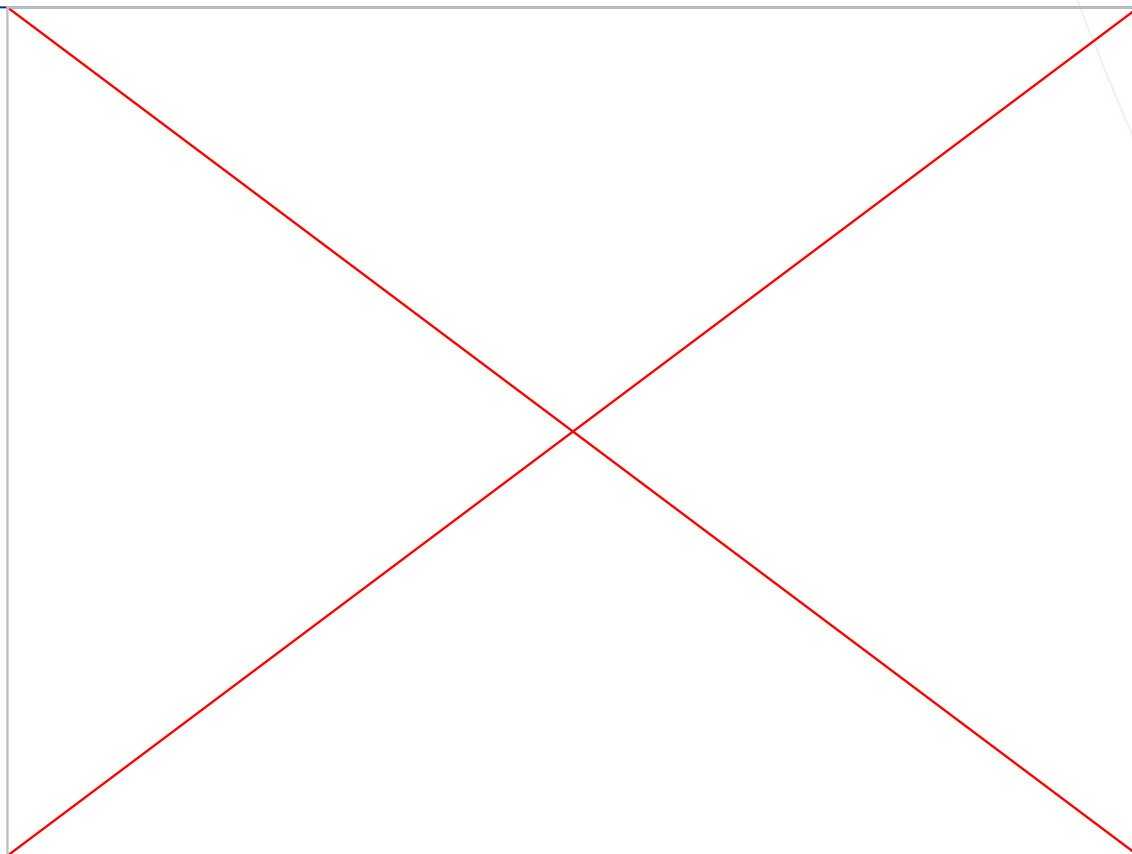
Assuming 5MB per photo



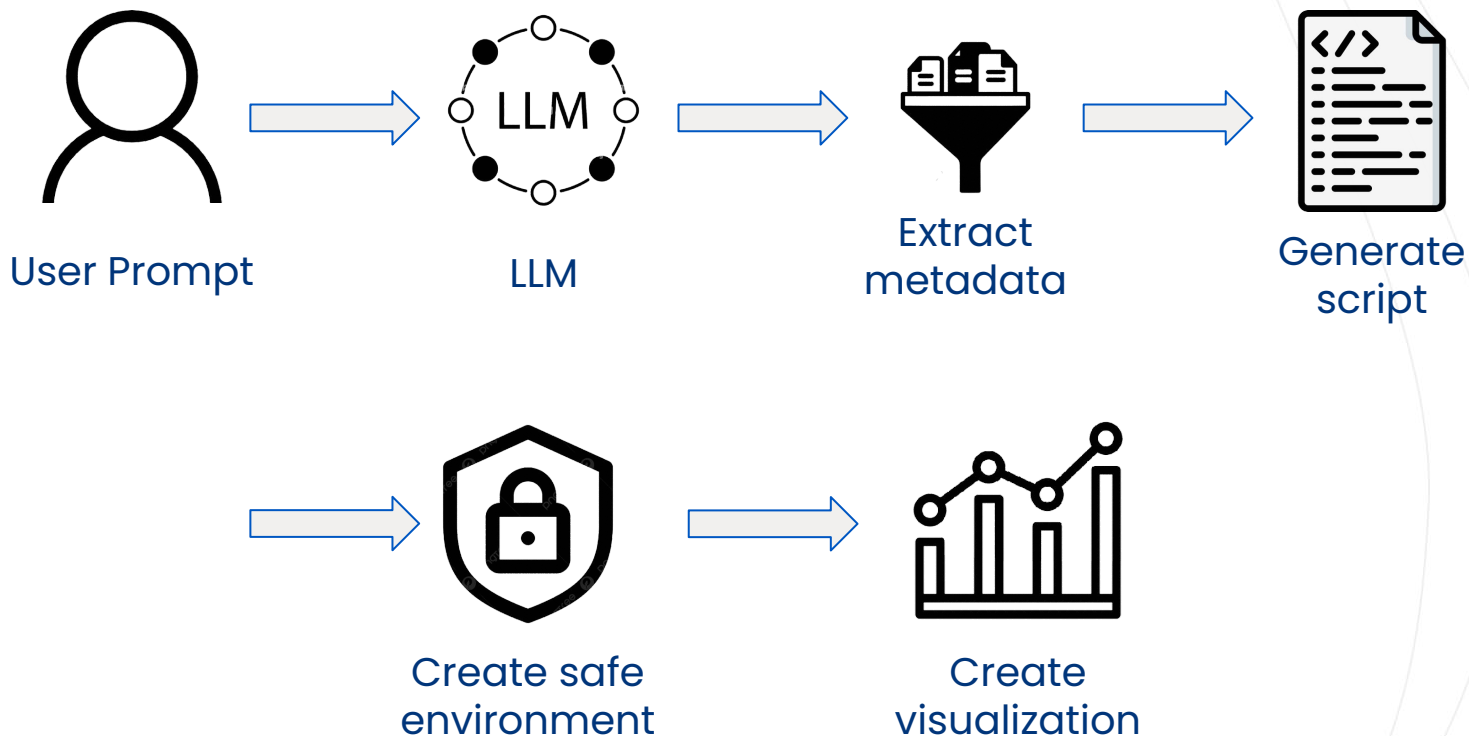
Getting useful insights from the data is hard.

- Explore easier access to scientific datasets
- Experiment with large language models (LLMs)
- Build a natural language search prototype

Demonstration



How it works



Why does this matter?



- Makes scientific data more accessible
- Provides a judgment-free environment
- Reduces the amount of coding

- LLM choosing incorrect dates
- Finding the correct prompts
- Template vs Generated code
- extracting variables
- handling failures
- Technical difficulties (e.g docker inside of docker)

- Continue improving the reliability of LLM generated code and metadata extraction.
- Making the request conversational.
- Include statistics analysis.
- Real code generation
- Data validation

Acknowledgements



Administration Team:

- Gabrielle Forbes, Virginia Do, Jerry Cycone, Jessica Amaya

Mentors

- Nathan Hook, Harsha Hampapura, Jason Cuning, Eric Nienhouse, Chia-Wei Hsu.

North Seattle College

CISL Team

Sage Team

SIParCS Interns

NSF NCAR



Questions?

Thank you very much for your time!

Anna Scribner

