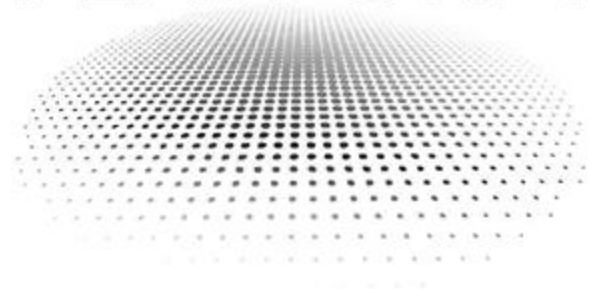


# NATIONAL DATA PLATFORM



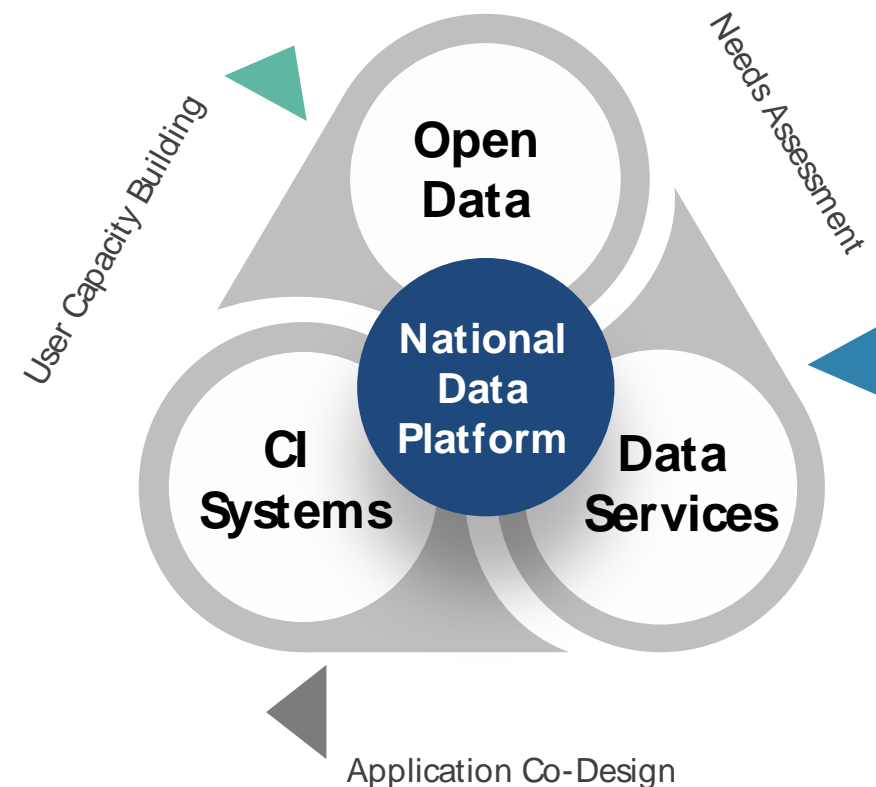
## National Data Platform Pilot: Services for Equitable Open Access to Data

[nationaldatapatform.org](https://nationaldatapatform.org)

National Data Platform is a federated and extensible data and service ecosystem to promote collaboration, innovation and equitable use of data on top of existing cyberinfrastructure capabilities.

NDP enables AI-integrated science workflows that foster discovery, decision-making, policy formation and societal impact related to wildfire, climate, earthquake and food security among others.

Link to the award abstract: [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=2333609](https://www.nsf.gov/awardsearch/showAward?AWD_ID=2333609)



UC San Diego



University of Colorado  
Boulder

**SDSC**  
SAN DIEGO SUPERCOMPUTER CENTER

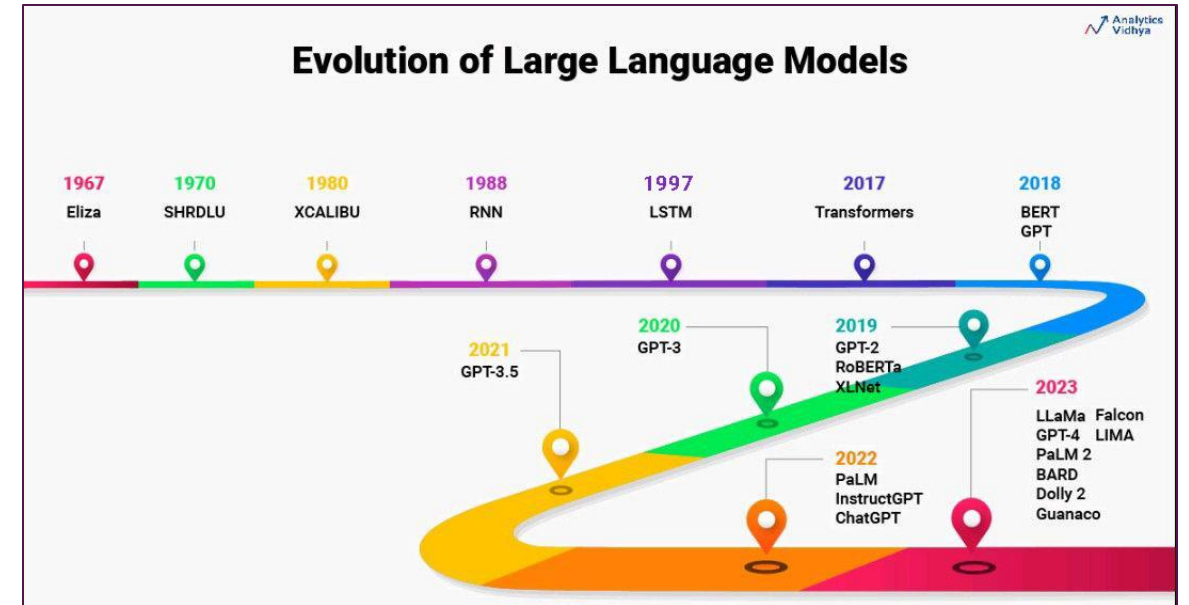
**SCI**  
www.sci.utah.edu

 **EarthScope**  
Consortium

# An NDP Service Example:

## Generative AI and Large Language Models (LLM)

- Huge generative potential
- Ability to create human-like outputs
- Integration with complex models
- Libraries advanced technologies
  - e.g., GPT, Prompt Engineering, and vector storage
- Shortcomings on domain expertise
- Need domain-specific LLMs
  - with human-curated data and controlled knowledge



Source: <https://www.analyticsvidhya.com/blog/2023/07/beginners-guide-to-build-large-language-models-from-scratch>

“Generative AI helped workers avoid awful ideas, but it also led to more average ideas”

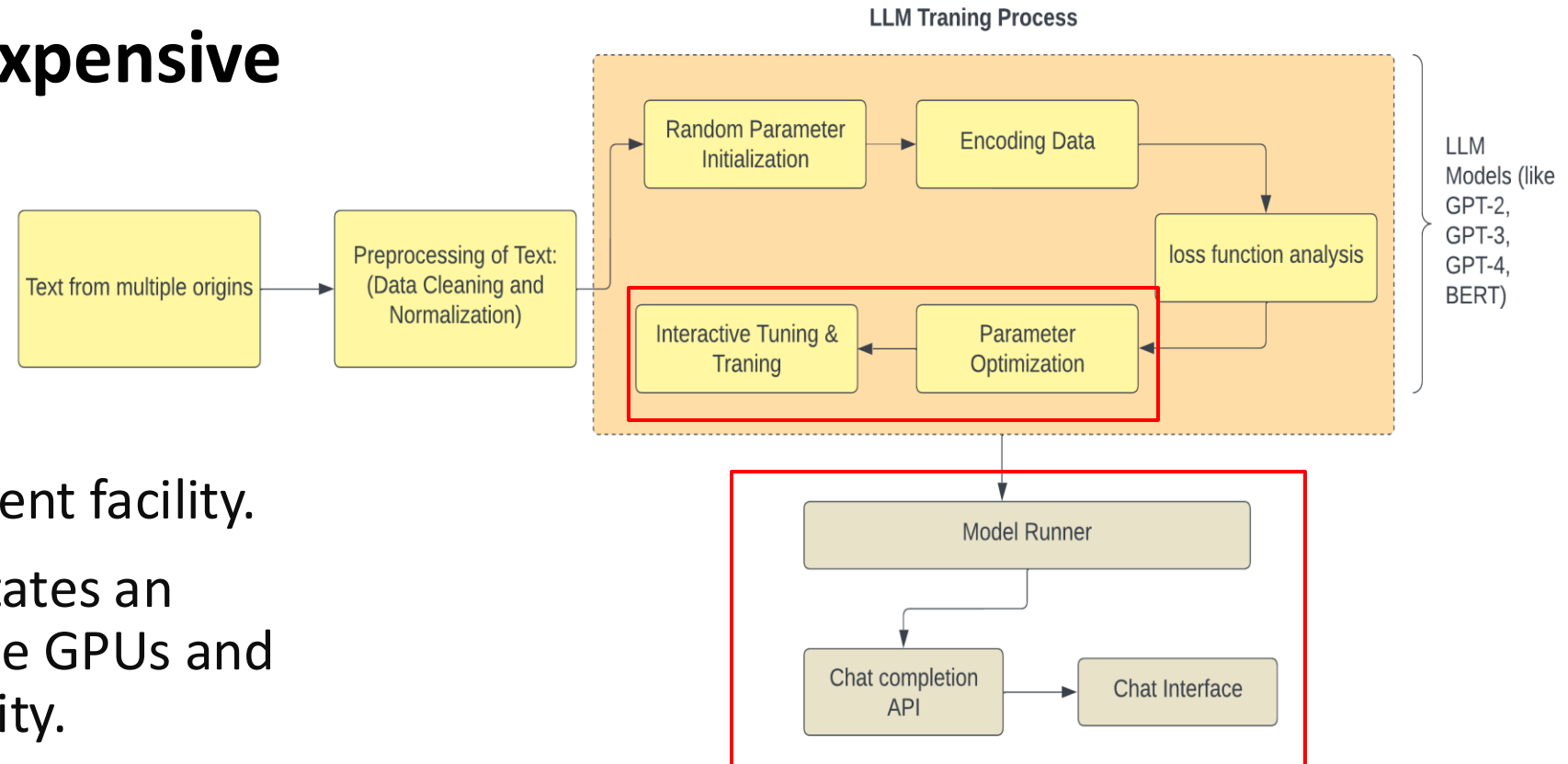
- Harvard Business Report (March – April

2024)

# Accessing and Using LLMs is an Equity Issue

## LLM Deployment is Expensive

- Even tuning an LLM can incur substantial costs, necessitating 4-5 AT100 GPUs, expansive nodes, and an equipped deployment facility.
- Operating an LLM necessitates an infrastructure with multiple GPUs and substantial memory capacity.



# LLM Deployments

## Commercial LLM

### Pros:

- Ready to use
- Large knowledge base
- Low latency and distributed deployment
- Robust API with security

### Cons:

- Lack of domain-specific knowledge
- High costs
- Information security and privacy concerns

## Community LLM

### Pros:

- Domain-specific knowledge
- Average latency
- Low costs
- Community-owned

### Cons:

- Need infrastructure
- Dedicated tech and knowledge team
- Community volunteers to manage everything
- Privacy concerns

## Private LLM

### Pros:

- Usage mode is private access
- No knowledge control and security issues

### Cons:

- Required hardware and software run models

# NDP

## LLM as a Service



- Tailored Model Selection
- Enhanced Data Control
- Privacy and Security
- Cost Efficiency
- OpenAI API and LangChain Support

## LLM Client Service

- Use an existing model
- Add context with domain-specific documents



## LLM Training Service

- Fine-tune an existing model to create a new model
- Use a larger corpus for training
- Deploy as a service

# NDP LLM as a Service

## Alignment with NAIRR Objectives

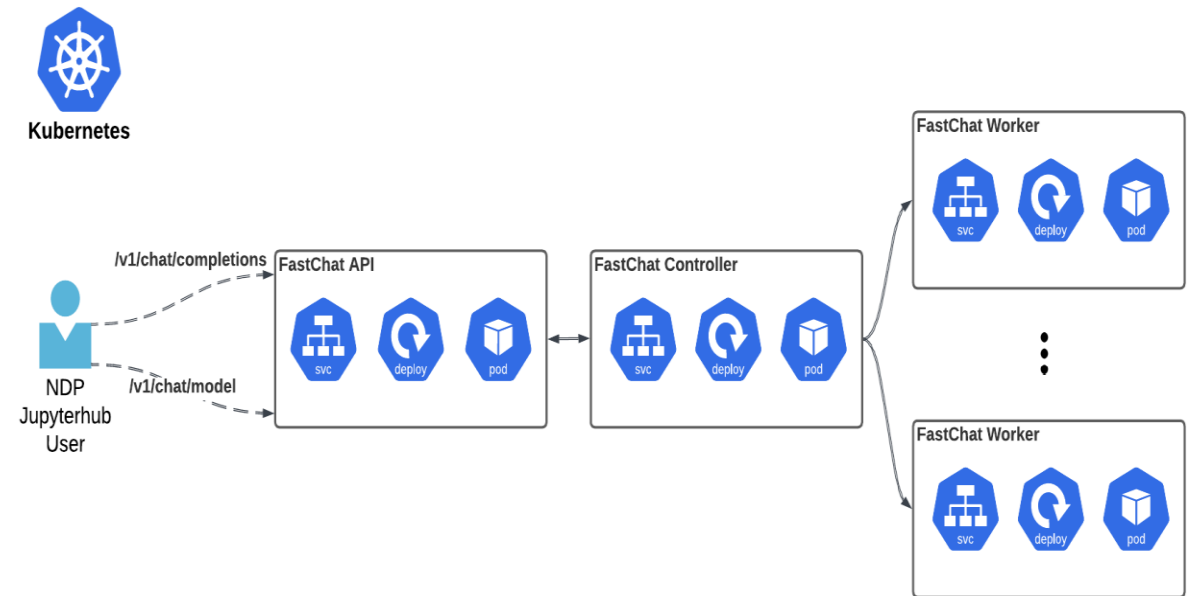
- **Capacity** to support many users with a spectrum of backgrounds
- **Capabilities**
  - Ability to train (*and use*) resource-intensive AI models on CI resources
  - Ability to make use of a mix of computational resources
  - Option to select which resources to use through a range of mechanisms, including ... optionally interactive “notebook”-like environments
  - A NAIRR system should include at least one large-scale machine-learning supercomputer capable of training 1 trillion-parameter models



In today's tutorial we are using a model  
with **7B** parameters running on NRP

# NDP LLM Deployment Architecture

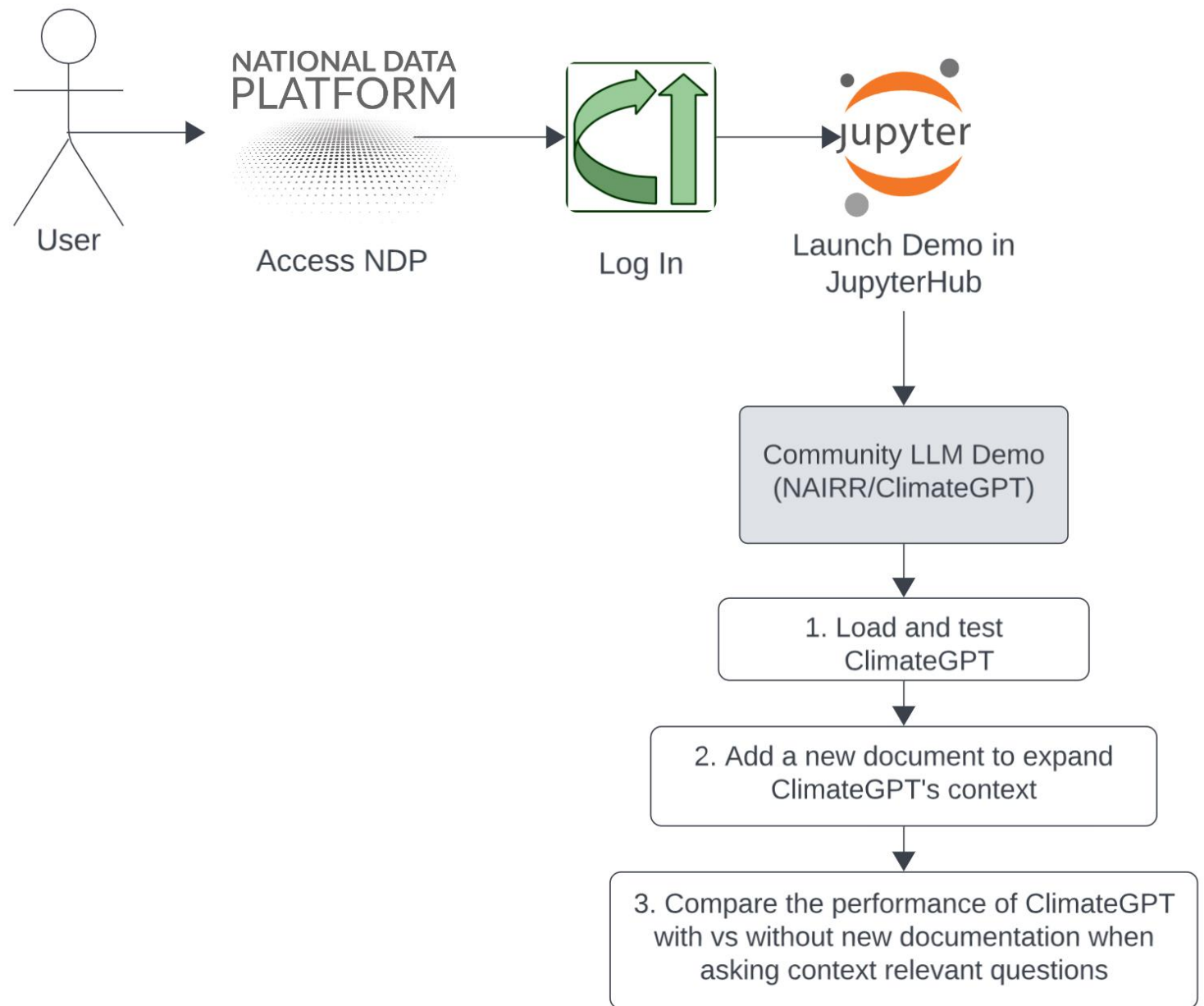
- FastChat
  - Open source LLM execution library
  - Deployed on Nautilus
    - API Server
    - Controller
    - Worker (serves different or the same LLMs)
- Currently all workers are serving the following LLMs
  - eci-io/climategpt-7b,
  - ECarbenia/grimoiresigils
  - text-embedding-ada-002



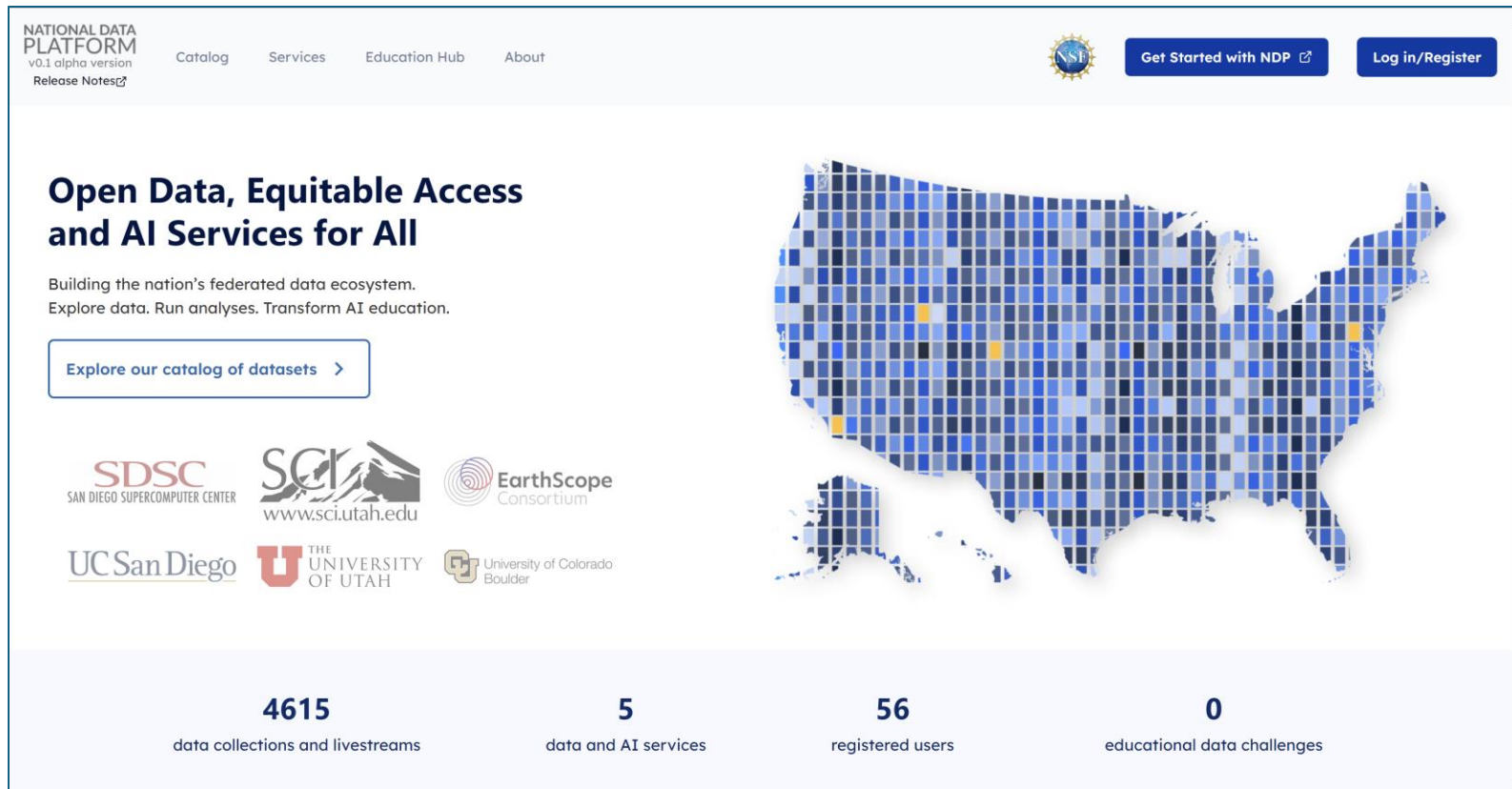
# NDP LLM-as-a-Service on NRP Tutorial



# Tutorial Steps




# Step 1: Go to nationaldataplatform.org



The screenshot shows the homepage of the National Data Platform (NDP). The header includes the site name 'NATIONAL DATA PLATFORM v0.1 alpha version' with a 'Release Notes' link, navigation links for 'Catalog', 'Services', 'Education Hub', and 'About', a 'Get Started with NDP' button, and a 'Log in/Register' button. The main content area features the headline 'Open Data, Equitable Access and AI Services for All' with a subtext 'Building the nation's federated data ecosystem. Explore data. Run analyses. Transform AI education.' and a button 'Explore our catalog of datasets'. Below this are logos for partner organizations: SDSC (San Diego Supercomputer Center), SCRI (www.sci.utah.edu), EarthScope Consortium, UC San Diego, The University of Utah, and the University of Colorado Boulder. A large map of the United States is composed of a grid of blue and white squares, with a few yellow squares. At the bottom, a statistics bar displays: 4615 data collections and livestreams, 5 data and AI services, 56 registered users, and 0 educational data challenges.

NATIONAL DATA PLATFORM  
v0.1 alpha version  
[Release Notes](#)




[Catalog](#) [Services](#) [Education Hub](#) [About](#)




 [Get Started with NDP](#) [Log in/Register](#)

## Open Data, Equitable Access and AI Services for All

Building the nation's federated data ecosystem.  
Explore data. Run analyses. Transform AI education.

[Explore our catalog of datasets](#)

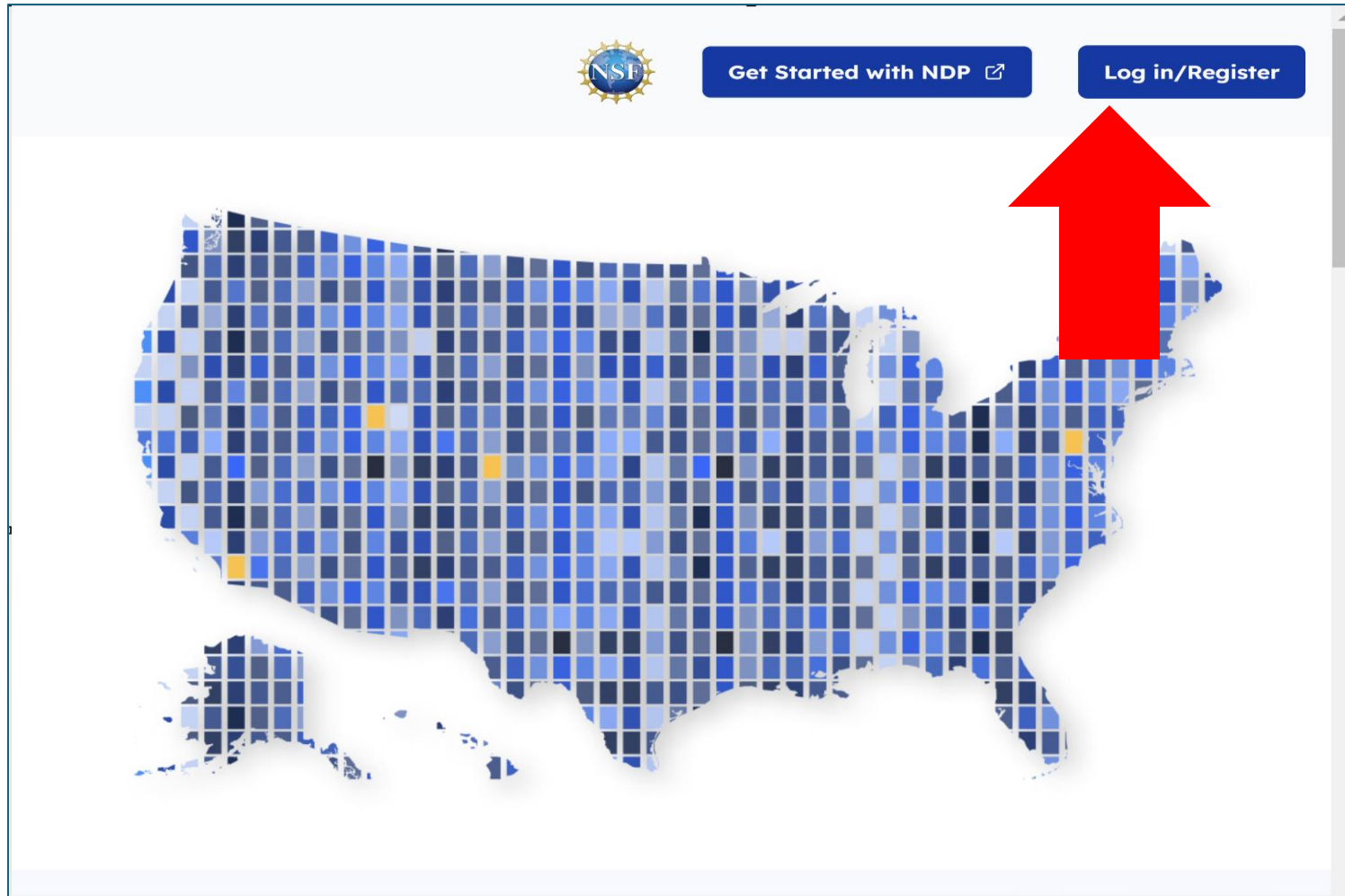
**4615**  
data collections and livestreams

**5**  
data and AI services

**56**  
registered users

**0**  
educational data challenges

## Step 2: Click on Log in/Register



# Step 3: Select CI Logon

You are using v0.1 alpha version of the NDP platform. Please report any issues [here](#).

### LOG IN


USERNAME or EMAIL

PASSWORD

☐ Remember me [Forgot Password?](#)

SIGN IN

Or sign in with

 CI Logon



**Step 4: Click on the Select an identity Provider dropdown and search your institution. Click on Log On.**

Select an Identity Provider

ORCID ▼ ?

☐ Remember this selection ?

Log On

By selecting "Log On", you agree to the [privacy\\_policy](#).

# Step 5: Log in with your institutional credentials

SINGLE SIGN-ON

UC San Diego

Passwords and Access   Enroll in Two-Step Login   Get Help

**Signing on using:** Active Directory

User name (or email address)

Or sign on with:

Active Directory

Password:

[Reset password](#)

LOGIN

**i** Sign out and close your browser when you're finished.

# Step 6: In your dashboard, click on JupyterHub

The screenshot displays the National Data Platform (NDP) dashboard. On the left is a sidebar with navigation links: My Dashboard, My Uploads, Analysis Hub, Catalogs, and Education Hub. The main content area is titled 'My Dashboard' and includes a welcome message for Pedro Antonio Ramonetti Vega. Below this is a 'Quick Explore' section with four buttons: 'Explore Data Catalog', 'Upload Data to Catalog', 'JupyterHub', and 'Education Gateway'. A large red arrow points to the 'JupyterHub' button. To the right of the 'Quick Explore' section is a 'Your Profile' box showing the user's name, email, and last login time. A small 'PV' dropdown menu is visible in the top right corner of the dashboard.

**NATIONAL DATA PLATFORM**  
v0.1 alpha version  
[Release Notes](#)

My Dashboard

Welcome back, Pedro Antonio Ramonetti Vega!

January 19, 2024 More details ironed out. Alpha version of NDP is live!

January 12, 2024 User dashboard drafted. JupyterHub integrated.

January 5, 2024 String search integrated.

December 22, 2023 CKAN data catalog is integrated into NDP. Large data and streaming uses cases are underway. Education gateway flow is being laid out.

December 8, 2023 Landing page complete with Keycloak login. The foundation of the catalog of datasets, search, and education gateway is set.

November 2023 Outlined user personas, assessed the needs of our users, and iron out the details of the initial features of NDP.

**Quick Explore**

Explore Data Catalog

Upload Data to Catalog

JupyterHub

Education Gateway

**Your Profile**

**Pedro Antonio Ramonetti Vega**  
pramonettivega@ucsd.edu  
Last Logged In: January 18, 2024 8:45 PM

# Step 7: Go to Image and select LLM Service Client. Click on Start.

Region  
Any

GPUs  
0

Cores  
1

RAM, GB  
16

GPU type  
NVIDIA GeForce GTX 1080 Ti

☐ /dev/shm for pytorch

Image  
Minimal NDP Starter Jupyter Lab

- Minimal NDP Starter Jupyter Lab
- Physics Guided Machine Learning Starter Code
- SAGE Pilot Streaming Data Starter Code
- EarthScope Consortium Streaming Data Starter Code
- NAIRR Pilot - NASA Harmonized Landsat Sentinel-2 (HLS) Starter Code
- LLM Training (CUDA 12.3, tested with 1 GPU, 12 cores, 64GB RAM, NVIDIA A100-80GB)
- LLM Service Client (Minimal, No CUDA)**

will be deleted

Start

Today's Demo Tutorial running on NRP/Nautilus

- **LLM Service Client:** Built for question and answer using pre-trained LLM model for ClimateGPT
  - Load any model and add your documents
- **LLM Training:** Built for customizing models for your domain
  - Update and fine-tune an existing model permanently for your domain using a large corpus of documents
  - Contact [ndp@sdsc.edu](mailto:ndp@sdsc.edu) for more information on training services



# Step 8: Wait for the server to launch

NATIONAL DATA  
PLATFORM

HomeToken

pramonettivega@ucsd.eduLogout

Your server is starting up.

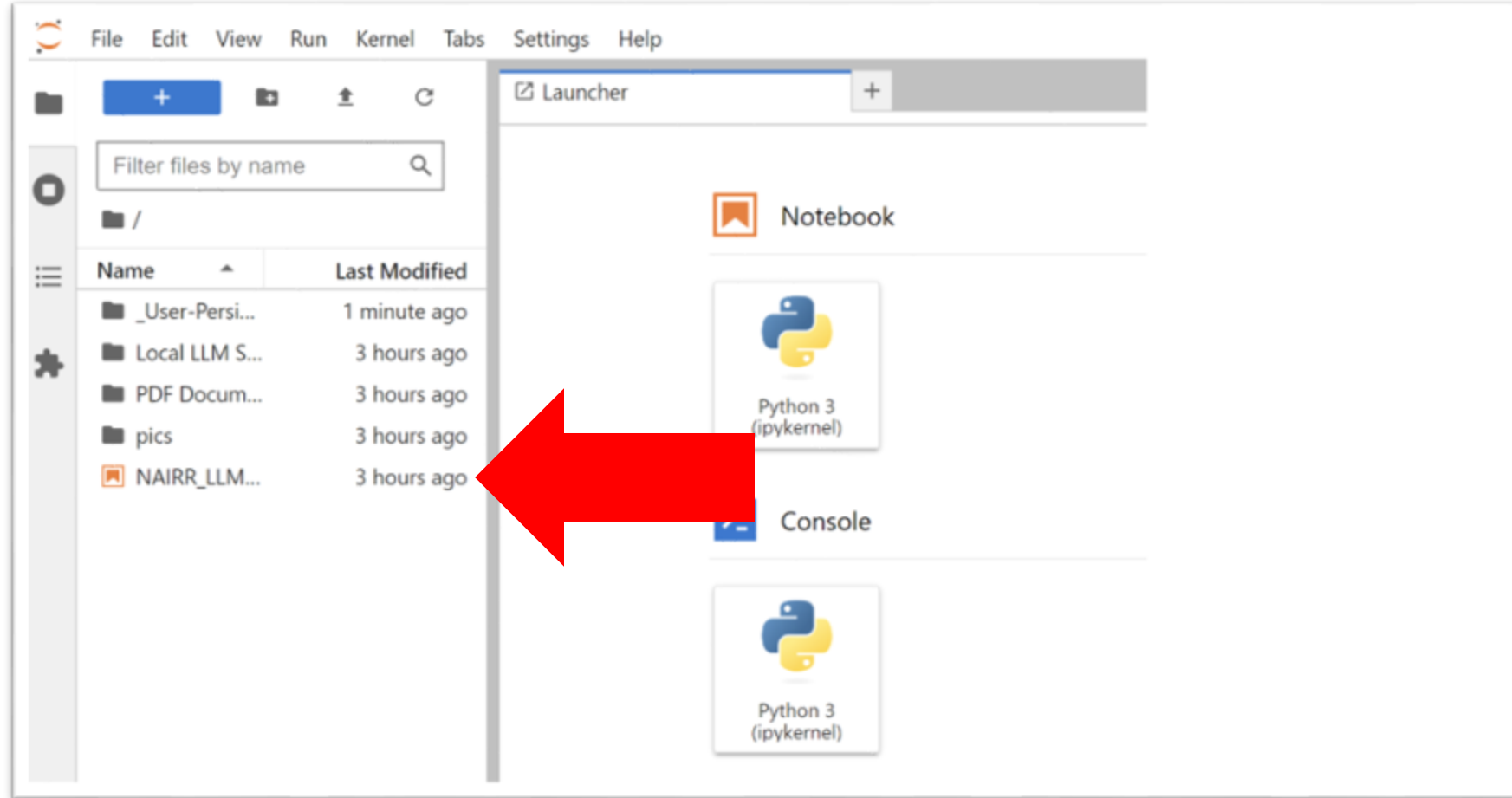
You will be redirected automatically when it's ready for you.

2024-03-18T19:18:20Z [Normal] AttachVolume.Attach succeeded for volume "pvc-dabaeab0-9c6c-4e2e-8f0c-d8dc3397bfcf"

Event log

**It will take a couple of minutes.**

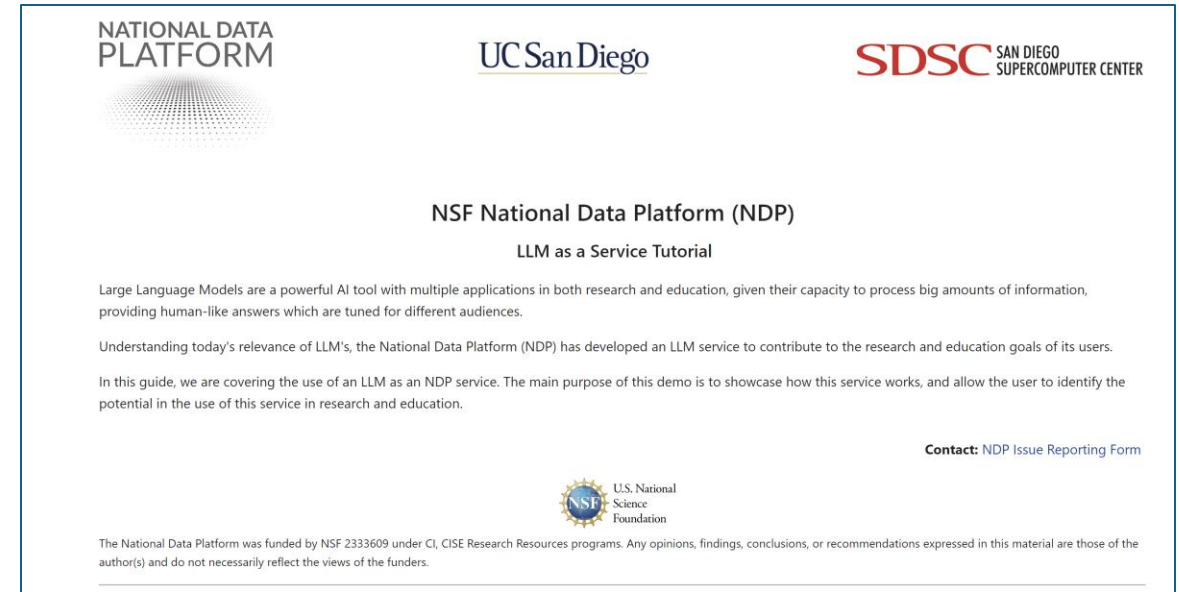
# Step 9: Select `NAIRR_LLM_chat.ipynb`



# Step 10: Hands On

## Part 1: Explore Q&A using ClimateGPT

- Run cells under "A. Set Up"
  - Run using Shift+Enter on the cell
  - Functions built
    - make\_query, run\_conversation and save\_conversation\_to\_file
- Run cells under "B. Running the conversation"
  - Use example questions or ask your own
  - **Type ENTER after each question. DO NOT TYPE SHIFT+ENTER**
  - Answer might take some time



## Part 2: Add New Document and Compare Q&A Results

- Run cells under "C. Adding new context to ClimateGPT"
  - Loading the document can take up to a couple of minutes
  - Run queries and compare results

3. Go to the next cell *POST*, and repeat the process. Now the model will provide an answer relying on the document.

## PRE-Context

```
[6]: # PRE - In this cell, we will ask the question without giving any context to ClimateGPT
conversation()
```

Question (q=quit, s=save previous answer): What are NAIRR's goals with respect to human capital?

Answer: The North American Interfraternity Conference's (NAIRR) goals with respect to human capital are focused on promoting the growth and development of fraternity and sorority communities. Specifically, NAIRR aims to:

1. Enhance leadership development: NAIRR provides educational programs, resources, and guidance to help fraternity and sorority members develop leadership skills and become effective leaders in their communities.
2. Foster brotherhood and sisterhood: NAIRR supports the development of meaningful relationships among fraternity and sorority members through various programs and initiatives, aiming to cultivate a strong sense of brotherhood and sisterhood within these communities.
3. Promote diversity, equity, and inclusion: NAIRR recognizes the importance of creating a welcoming and inclusive environment for all members of fraternity and sorority communities. They work to support diversity, equity, and inclusion efforts in these communities.
4. Improve risk management: NAIRR provides resources and guidance to help fraternity and sorority communities manage risks effectively, ensuring a safe and positive experience for members.
5. Support the overall success of fraternity and sorority communities: NAIRR aims to help fraternity and sorority communities thrive by providing resources, advocacy, and support to member organizations.

By focusing on these goals, NAIRR works to create well-rounded, engaged, and successful fraternity and sorority communities that benefit members and contribute positively to higher education and society as a whole.

Question (q=quit, s=save previous answer): q

Looking at the responses, we can confirm the model is given proper answers which are constructed taking the added document as context.

## POST-Context

```
[*]: # POST - We add True to indicate the model to make use of the new document
conversation(True)
```

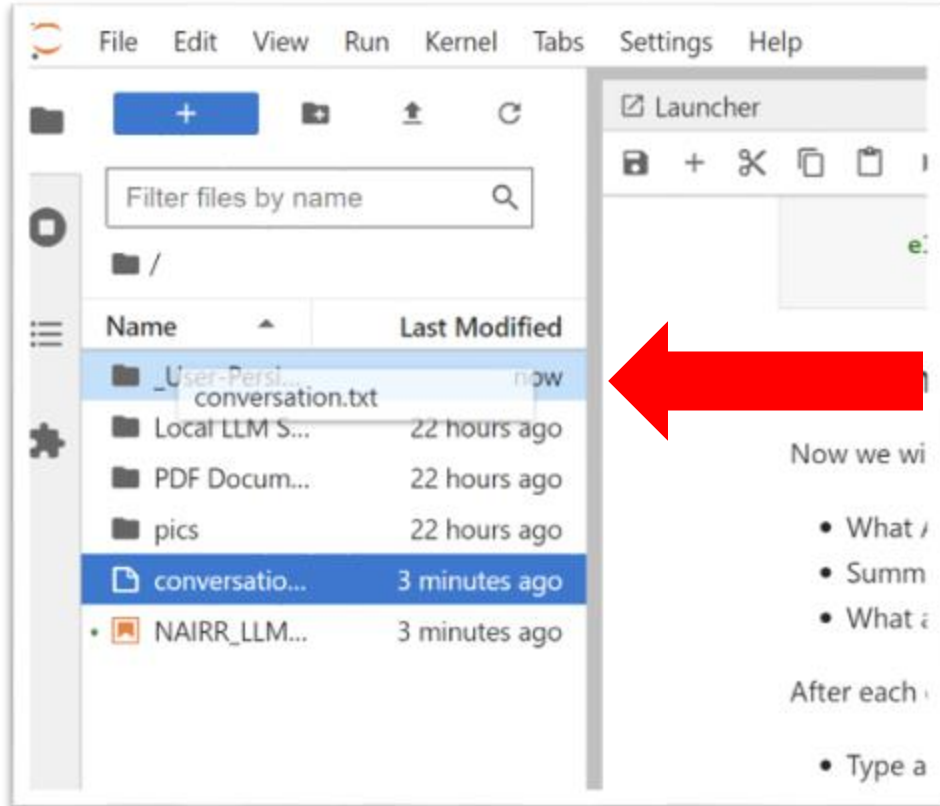
Question (q=quit, s=save previous answer): What are NAIRR's goals with respect to human capital?

Answer: The NAIRR aims to provide a comprehensive set of AI tools and resources for federal government use, and it recognizes the importance of human capital in the successful implementation and use of these tools. The Operating Entity is responsible for ensuring that the NAIRR includes appropriate training and support materials for users, and for sharing responsibility with resource providers for providing this training and support. The NAIRR also aims to incorporate lessons learned from its implementation and use to improve its governance, management, and operations over time.

Question (q=quit, s=save previous answer):

# Screenshot of a pre-context and post-context response

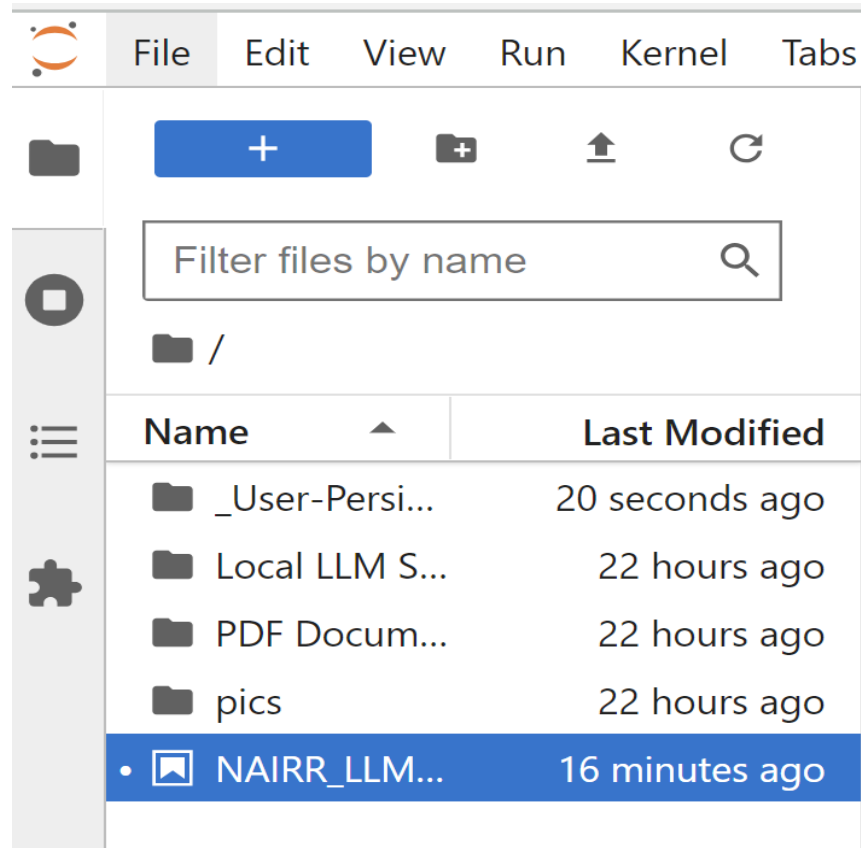
# Step 11: Save your notebook and outputs



To save your notebook and or output, you can:

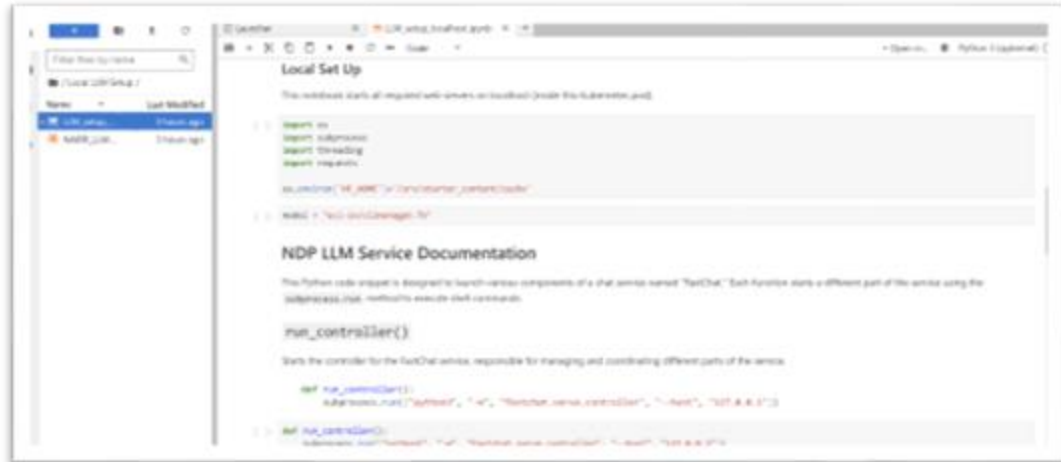
- Select the files and drag them into \_User-Persistent-Storage\_
- Copy-Paste the files into \_User-Persistent-Storage\_
- Download them locally

# Local LLM: We provide the set-up code to host your own LLM within your server



To make this service work, it is necessary to reserve a server with at least one GPU instance.

# Other NDP LLM Notebooks



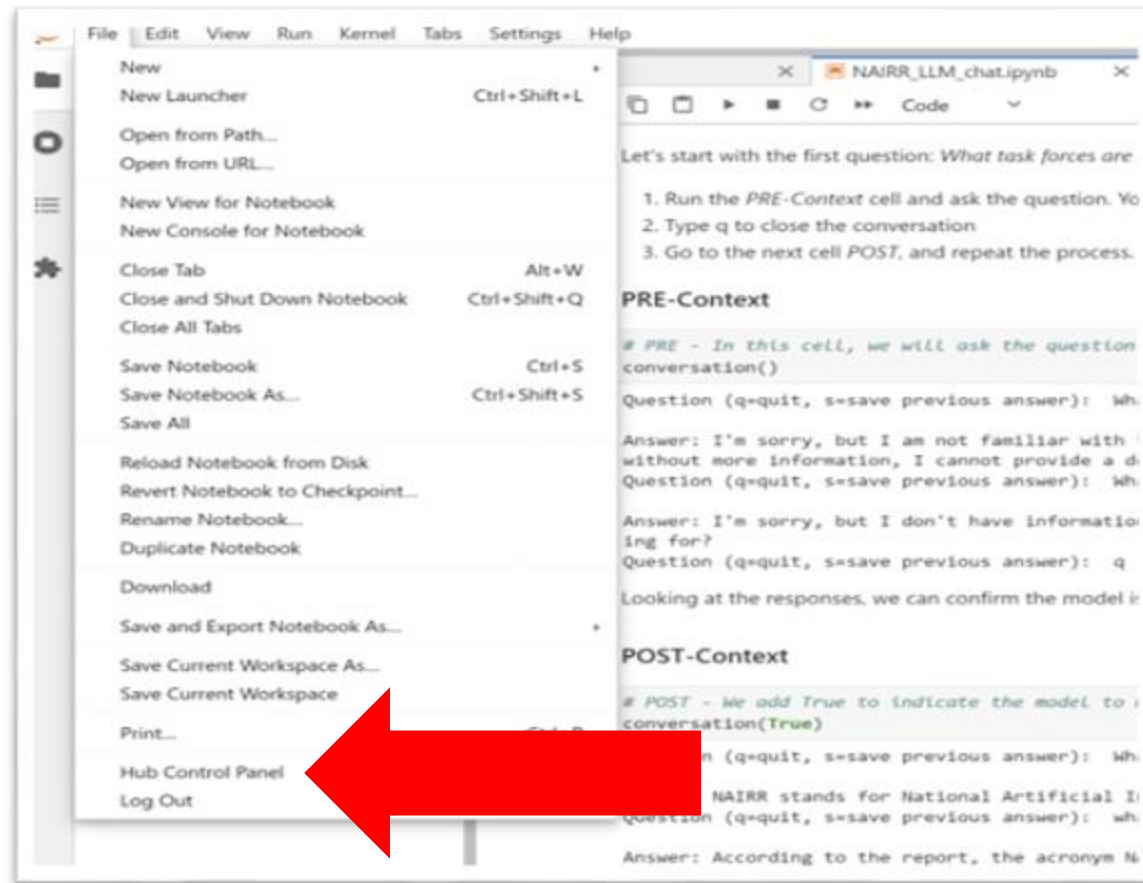
**LLM\_setup\_localhost.ipynb:** This notebook allows users to start their own host API server, and to load their model from HuggingFace into their server.



**NAIRR\_LLM\_chat\_localhost.ipynb:** The code explored today, which connects to the localhost server instead of Community's LLM server. Users can connect and start interacting with their model.

Contact [ndp@sdsc.edu](mailto:ndp@sdsc.edu) for support

# Step 12: Click on *File* and select *Hub Control Panel*





# Step 13. Stop your Server

Stop My Server

My Server

## Named Servers

In addition to your default server, you may have additional server(s) with names. This allows you to have more than one server running at the same time.

Server name	URL	Last activity	Actions
<input type="text" value="Name your server"/>			<a href="#">Add New Server</a>