# FOUNDATION MODELS AND AGENTS FOR PHYSICS

Nesar Ramachandra
nramachandra@anl.gov
CPS Division, Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

Argonne NATIONAL LABORATORY

# INTRODUCTION

Argonne NATIONAL LABORATORY

- What are Foundation Models (FMs)? What makes them particularly interesting or powerful?
- What does it take to train a Large Language Model (LLM)? How can we align the LLMs to our science domains?
- **How much productivity can we harness out of LLMs? Can we go beyond text-type data? Can we use a system of AI-agents/experts in solving physics problems?**
- All the applications here are outside of nuclear physics, however they are adaptable.

*Disclaimer: Some of the opinions are my own.

# HOW ARE FOUNDATION MODELS DIFFERENT?

Argonne NATIONAL LABORATORY

**Traditional AI:**

- *Parameters:* O(1e2-1e8)
- *Data:* Specialized, annotated
- *Flexibility:* Meant for specific tasks (classification, regression, generative, un-masking, …)
- *Training Approach:* Supervised, unsupervised, or transfer learning.
- *Applications:* Suited for narrowly defined problems with clear objectives.

**Foundation models:**

- *Parameters:* O(1e8 to 1e11)
- *Data:* Extensive, datasets from various domains
- *Flexibility:* Highly versatile, can be fine-tuned different tasks.
- *Training Approach:* Generic pre-training followed by alignments
- *Applications:* Suitable for multi-faceted problems requiring nuanced understanding

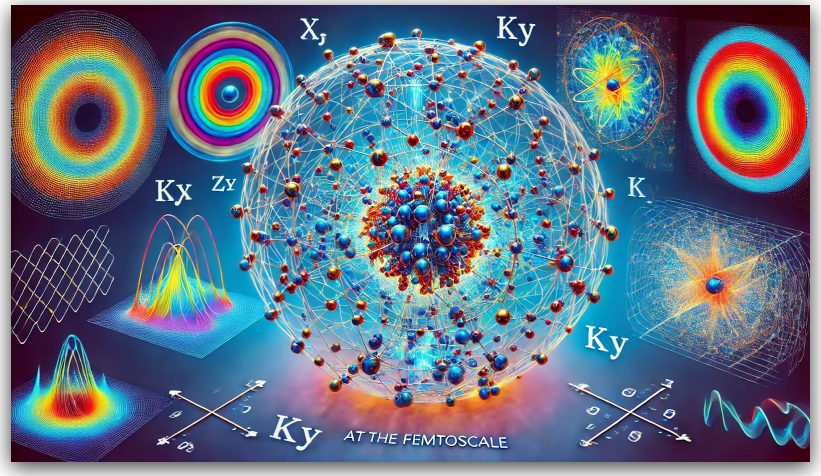


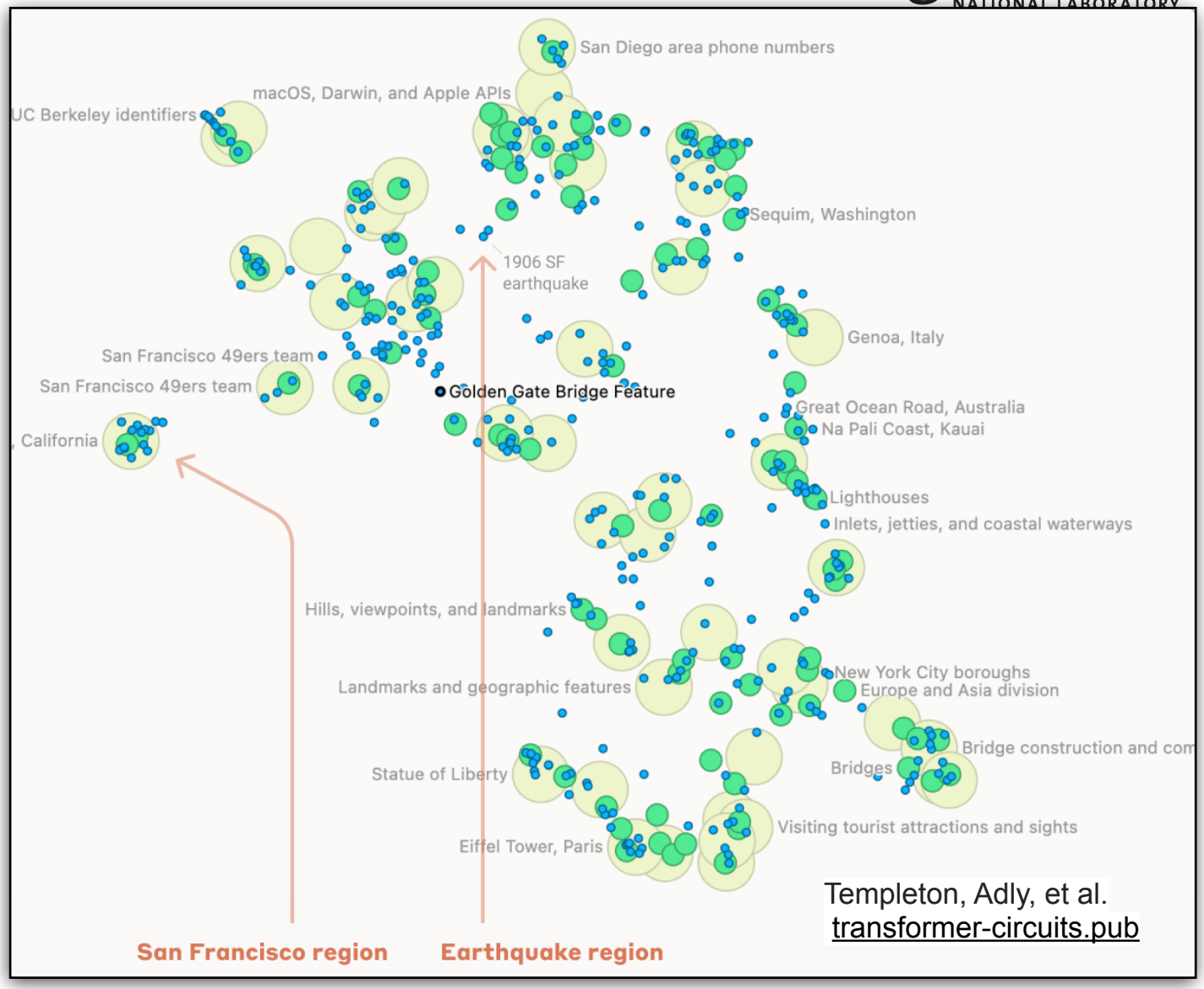

Image credits: Emil Constantinescu & ChatGPT

Gemini

code llama

# DO FOUNDATION MODELS GENERALIZE?

- Majority of the foundation models are language based.
- Emergent behaviors seen with larger models.
- Questions about whether this is the path towards AGI, whether there is a reasoning/thinking involved at inference.
- Regardless, FMs are technological paradigm shifts, and a highly useful knowledge base.




Templeton, Adly, et al.
transformer-circuits.pub

**Latent space of LLMs showing emergent behavior**

# STAGES IN TRAINING AN LLM

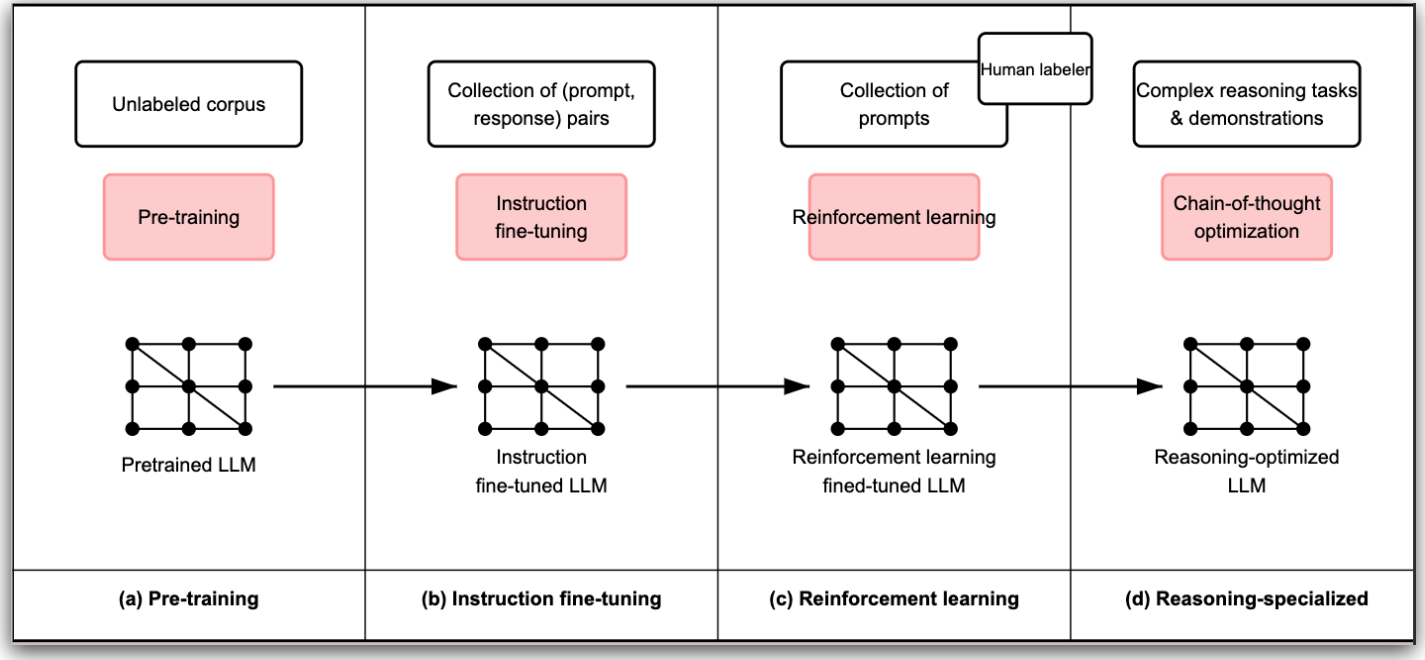
- **Data preparation**
  - Both labeled and unlabeled datasets.
- **Pre-training**
  - Computationally expensive (10-100 TB data, 10+ million GPU hours, $100 million)
- **Tuning**
  - Smaller task, includes fine-tuning or instruction tuning — often with supervised datasets.
- **Reinforcement learning**
  - Safety guard-rails
  - Human-in-the-loop
- **Inference improvements**
  - Can include reinforcement learning, Chain-of-thought optimization, reasoning tokens and feedback preference optimizations.
  - Heavily utilized in “Reasoning” class models (Open AI o1, o3, …).
- **Benchmarking and validation**
- **Deployment and API**

Sequence of tokens: $U = \{u_1, u_2, u_3, \ldots, u_N\}$

Minimize: $L(U) = -\sum_{i=1}^{N} log(p(u_i \mid u_{i-k}, \ldots, u_{i-1}, \Theta))$

(Conditional probability of i-th token given k preceding tokens)

Transformer-based architecture and attention are crucial for scaling training to billions of tokens

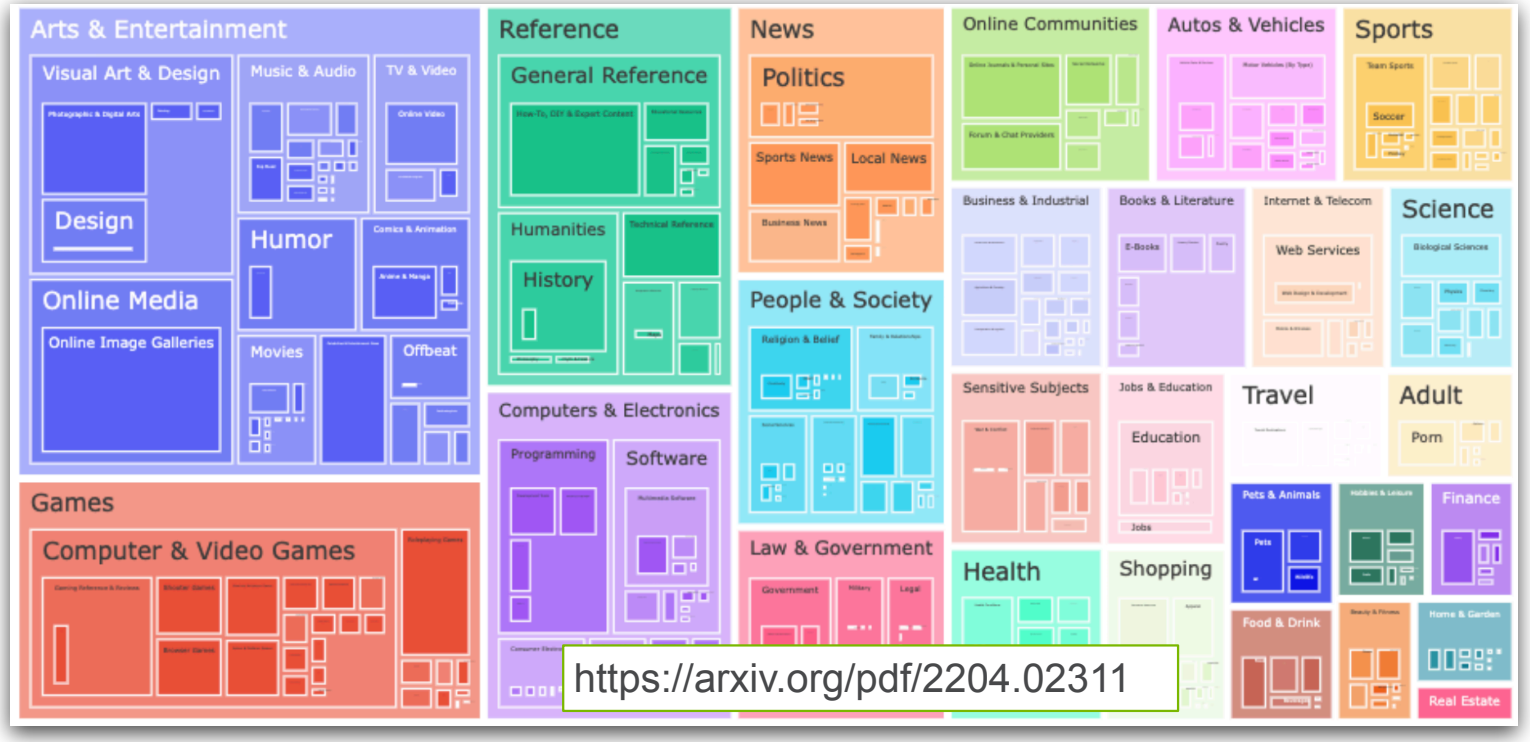



**LLM construction stages: pre-training to reasoning abilities**


U.S. DEPARTMENT OF ENERGY | Office of Science

# ALIGNING MODELS TO SCIENCE

- **Fine-tuning**
  - Smaller, specific scientific datasets (e.g: AstroLLaMa)
- **Full model building**
  - Expensive pre-training from scratch (e.g: AuroraGPT at Argonne)
    - Scientific datasets
    - Scientifically rigorous benchmarks
- **Continual pre-training**
  - Middle ground.
  - Less annotated but more domain-specific data




**Data-corpus division in LLMs**

## Fine-tuning model performance

**Completed by LLaMA-2**

The Magellanic Stream (MS) - an enormous ribbon of gas spanning 140◦ of the southern sky trailing the Magellanic Clouds - has been exquisitely mapped in the five decades since its discovery. However, despite concerted efforts, no stellar counterpart to the MS has been conclusively identified. This stellar stream would reveal the distance and 6D kinematics of the MS, constraining its formation and the past orbital history of the Clouds. We find a stellar stream in the MS, with a 6D velocity of 100 km s−1, and a distance of 100 kpc. The stream is 100 pc wide and 1000 pc long, and is consistent with the MS. The stream is 1000 times more massive than the Magellanic Clouds, and is likely to be a tidal stream from the Large Magellanic Cloud.

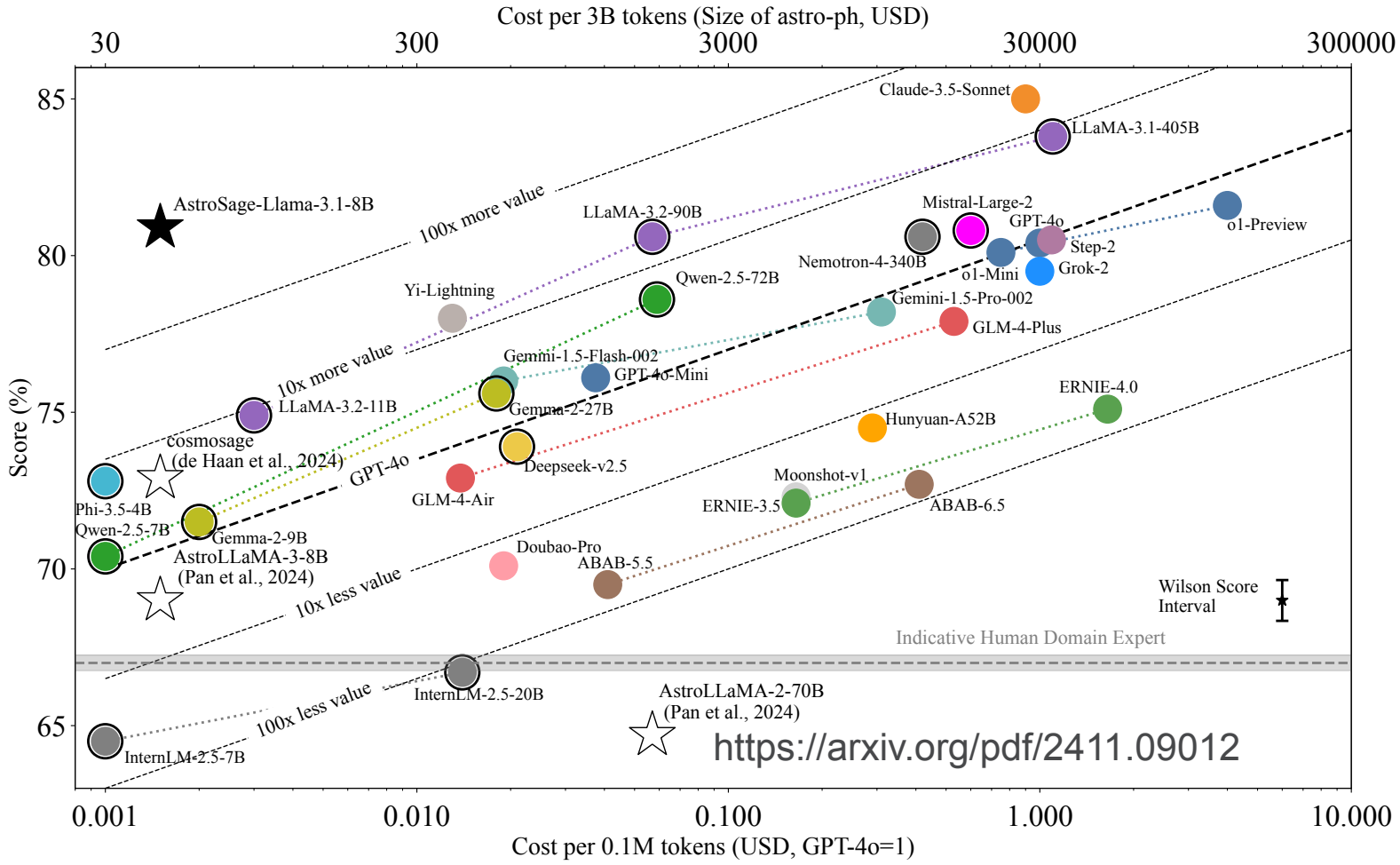
**Completed by AstroLLaMA**

The Magellanic Stream (MS) - an enormous ribbon of gas spanning 140◦ of the southern sky trailing the Magellanic Clouds - has been exquisitely mapped in the five decades since its discovery. However, despite concerted efforts, no stellar counterpart to the MS has been conclusively identified. This stellar stream would reveal the distance and 6D kinematics of the MS, constraining its formation and the past orbital history of the Clouds. We 1) present a new method to select MS candidate stars from wide field optical imaging, 2) apply this method to the MS, 3) compare the results to existing catalogs, and 4) identify the most promising MS candidate stars. We apply our method to 2547 stars in the Gaia EDR3 catalog that have Gaia proper motions consistent with the MS and use these as a pilot sample to optimize our method. We then apply our method to the entire Gaia EDR3 catalog to identify 46,000 stars that are candidates for the MS, and we use the Gaia-ESO Survey to identify 13,000 of these stars that are likely to be MS candidates. We identify 266 stars with (G-K) colours and [Fe/H] that are consistent with previous estimates of the MS. These stars have proper motions and radial velocities that are consistent with the MS, and we place them on a 3D map of the MS. We find a previously unknown stellar stream, which we call the Southeastern Stream (SES), in the direction of the Clouds and MS. The SES has an M31 distance of 32.2+/-0.5 kpc and a 6D velocity of 136+/-10 km/s. The SES is about 20 times brighter in stars than the MS and about 40 times brighter in stars than the SMC. The SES is also 2.7 times more metal-rich than the MS, which suggests that it is a separate population.

https://arxiv.org/pdf/2309.06126

# LLM-FOR ASTROPHYSICS: ASTROSAGE

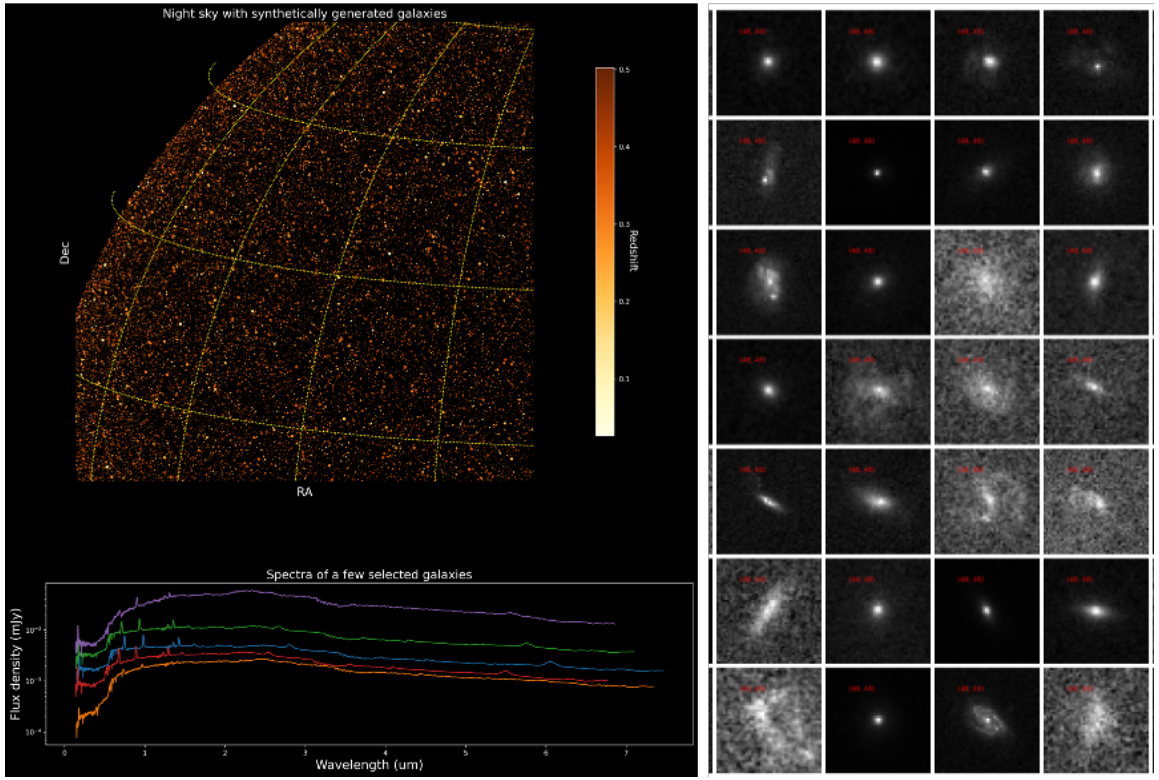
- **Continual pre-training + Fine-tuning**
  - Texts from astrophysics papers as training tokens. (20 GB data)
  - 2nd step of fine-tuning the models to answer astrophysics questions (10 GB data)
  - Built on open-source Llama-8B.
  - For domain-specific tasks, as powerful as GPT-4o, with fraction of the training cost (30k GPU hours).




**Model cost vs performance: AstroSage vs rest.**

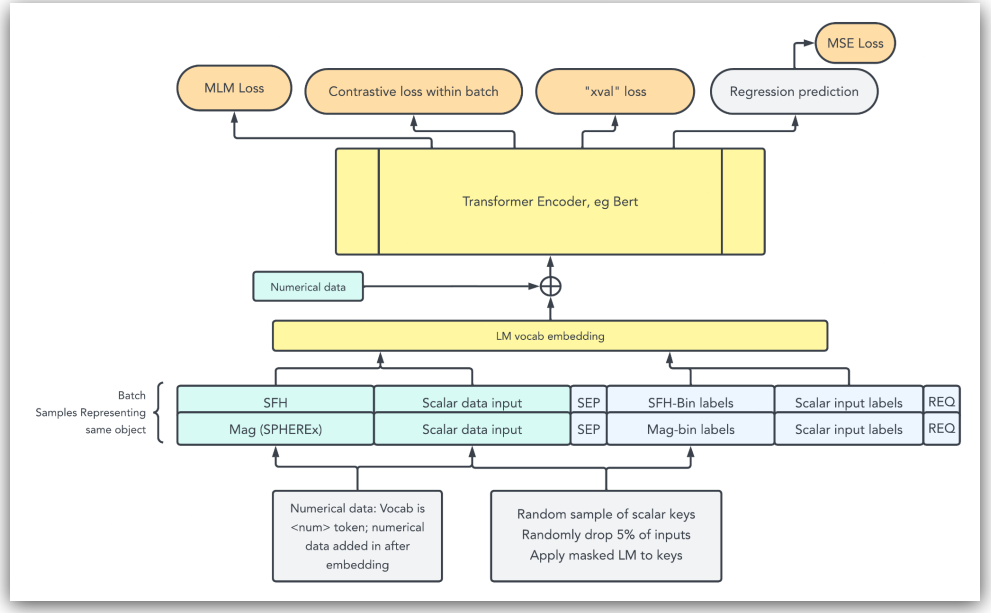
# BEYOND-TEXT: FULL MODEL BUILDING

- Foundation models for non-text scientific data formats:
  - For science datasets: Star-foundation models, AstroCLIP, OmniLearn
- Numbers be processed as numbers and not as tokenized representations: xVal
- Multi-modality in object expressions, not just data-types.
  - Distinct from industry trend: videos, subtitles, plots and captions, songs and lyrics
  - Nuclear physics: particle trajectories (graphs), events (scalars), catalog values, metadata, instrument settings.
- Variety of transformers
  - Vision Transformers, graph transformers
- Latent space modeling using Contrastive learning

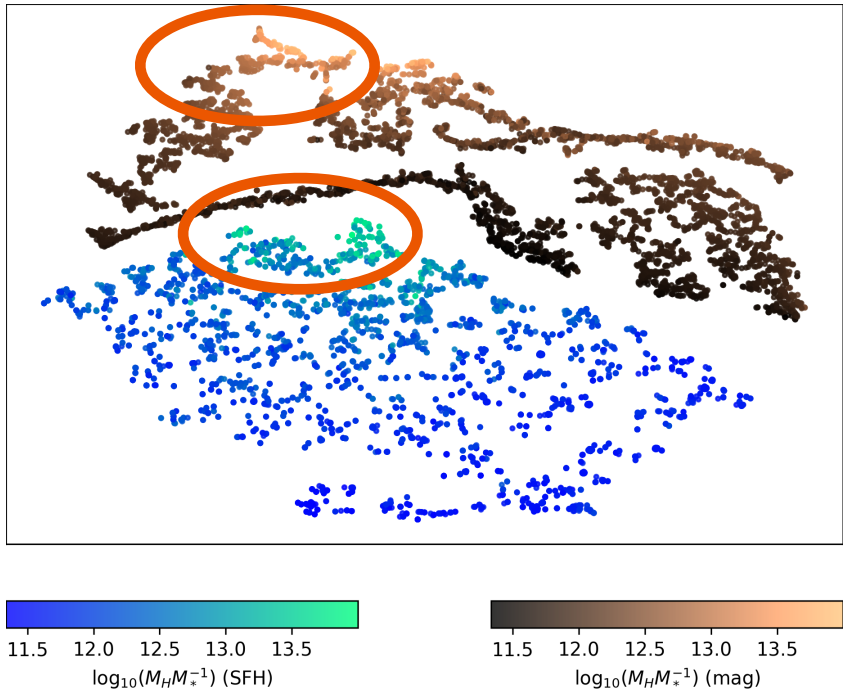



**Multiple modalities in astrophysics data. Galaxy distribution in the sky (left), randomly chosen spectra (bottom left), images (right), catalog values (not shown)**

# DOMAIN FOUNDATION MODELS

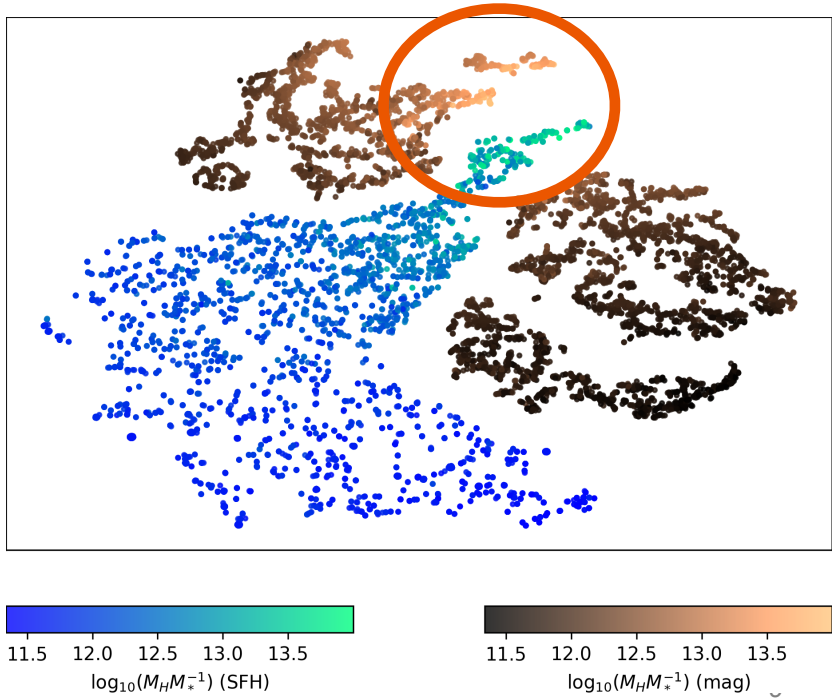



- Training different (encoder-decoder) transformers with different data modalities. Numerical tokenization used.
- Contrastive loss to connect simulation modes with observed/experimental modes
  - Changes embedding space



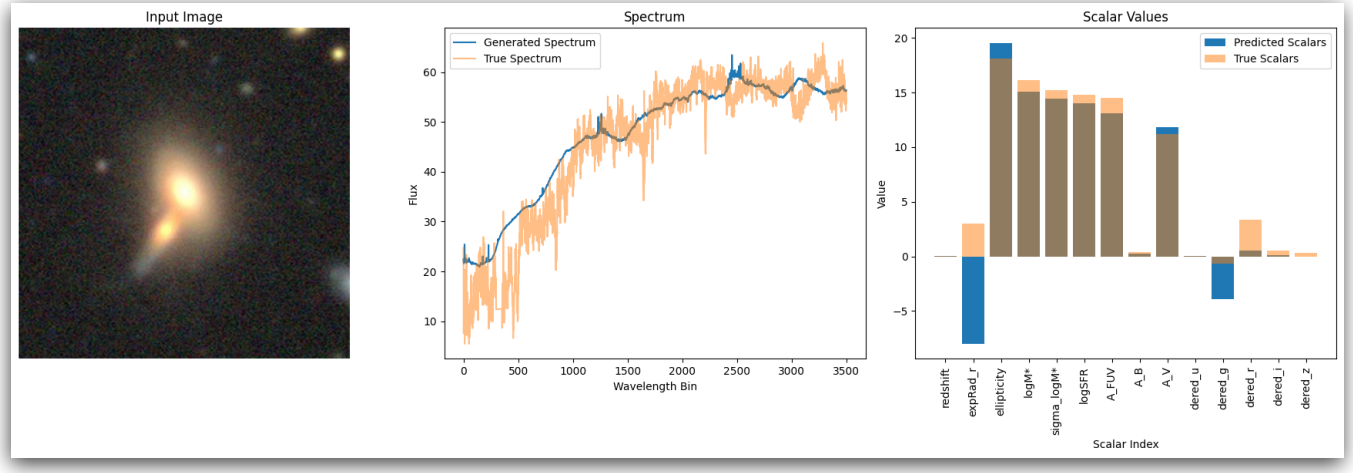

DOI: 10.22541/essoar.172675995.55091022/v1

**Latent space of Foundation models showing emergent behavior: Different modalities for the same object are correlated**

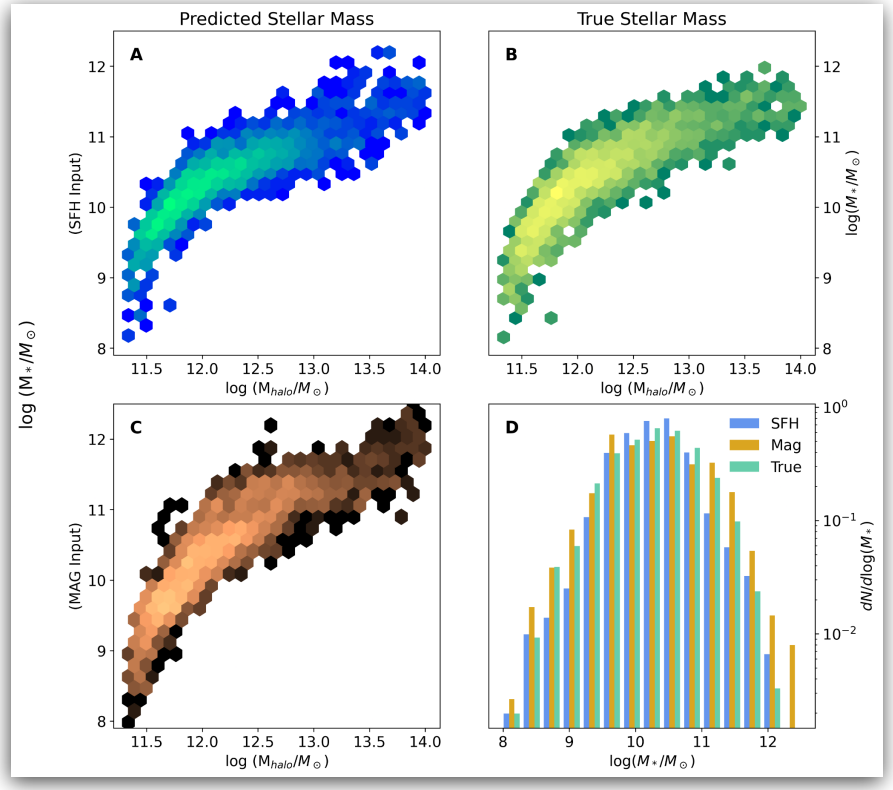


# DOMAIN FOUNDATION MODELS

- Training is general purpose, deployment is task-specific.
- Flexibility in deployment: queries dictate latent space access.
- Compatibility wrto datasets in multiple domains
- DFMs can be joined with existing LLMs for contexts along with knowledge base access




**Flexible input/output combinations: Single input predicting multiple outputs.**



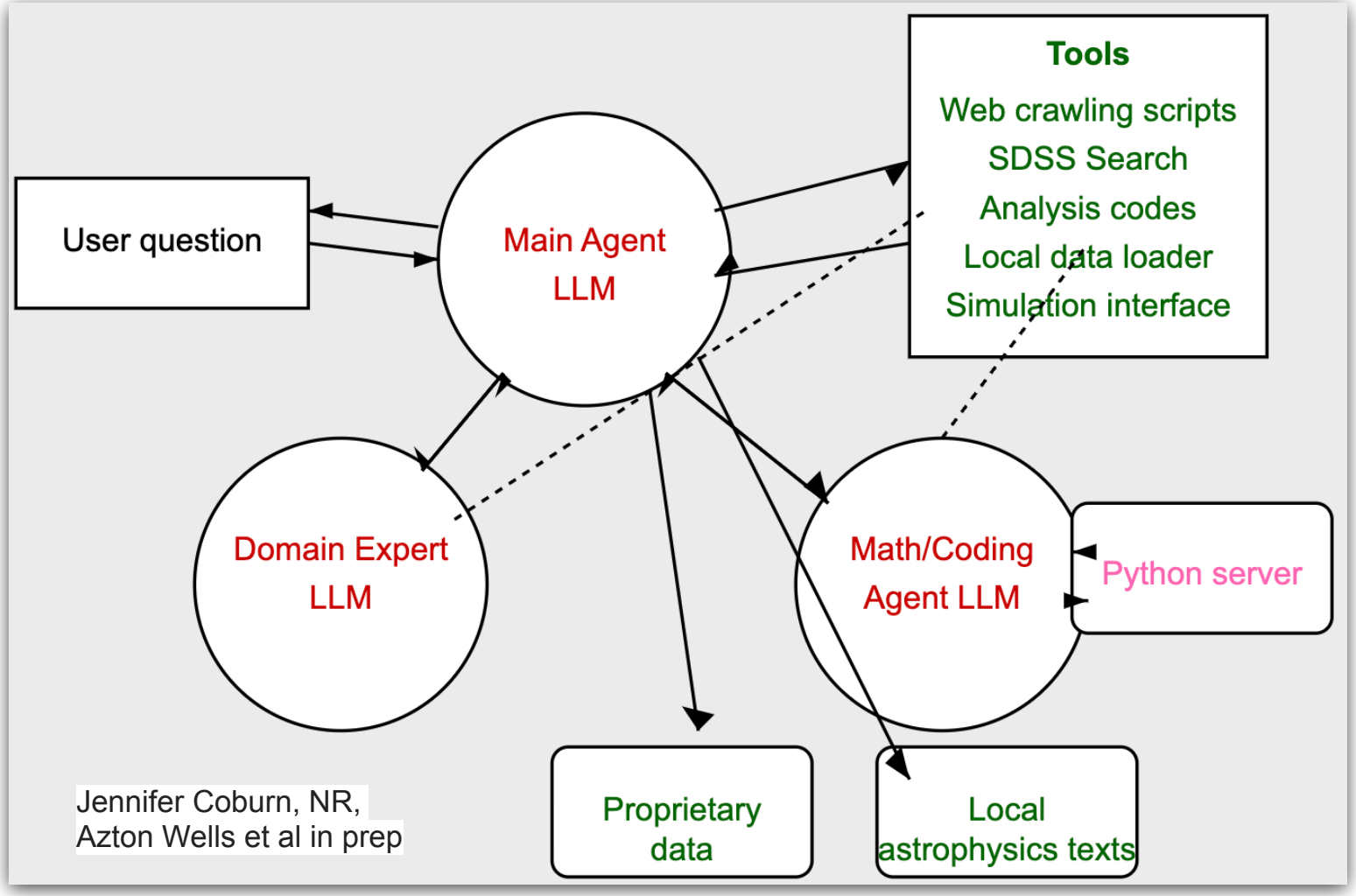

**Flexible input/output combinations: Multiple inputs used to predict the same output**

DOI: 10.22541/essoar.172675995.55091022/v1

# MULTI-AGENT FRAMEWORKS

**Ingredients: tools, servers, and databases**

- LLMs or Reasoning models with tool access (web search, data base reading)
- Multiple expert interactions (Roles: Manager, domain-expert, coding expert, critic)
- Local datasets (texts and numerical data) access
  - Recent datasets, codes, publications
  - Proprietary data
- Local simulation interface access.
- Python server to run codes, feed results back to LLMs




**Agentic framework of Dr. MACS (Multi-Agent Collaborative System)**

# DR MACS IN ACTION: SCIENCE QUESTION


New run

What is the average brightness of galaxies with stellar mass between 10^10 and 10^12 solar masses and within the redshift range 0.1<z<0.2?

HfApiModel - Qwen/Qwen2.5-Coder-32B-Instruct

User inputs a question

# DR MACS IN ACTION: MAIN AGENT

Step 1

*Output message of the LLM:*

```
Thought: To solve this task, I need to retrieve data for galaxies within the specified stellar mass and redshift range from
an astronomical catalog. Then, I will calculate the average brightness of these galaxies. The CodingAgent can help with data
retrieval from astronomical catalogs, so I will use it for this task.
Code:
```py
CodingAgent('query string: Retrieve data for galaxies with stellar mass between 10^10 and 10^12 solar masses and within the
redshift range 0. 1<z<0. 2 from the SDSS galaxy catalog')
```<end_code>
```

Main Agent plans the tasks

New run

```
You're a helpful agent named 'CodingAgent'.
You have been submitted this task by your manager.
---
Task:
query string: Retrieve data for galaxies with stellar mass between 10^10 and 10^12 solar masses and within the redshift
range 0. 1<z<0. 2 from the SDSS galaxy catalog
---
You're helping your manager solve a wider task: so do not just provide a one-line answer, instead give as much
information as possible to give them a clear understanding of the answer.

Your final_answer WILL HAVE to contain these parts:
### 1. Task outcome (short version):
### 2. Task outcome (extremely detailed version):
### 3. Additional context (if relevant):

Put all these in your final_answer tool, everything that you do not pass as an argument to final_answer will be lost.
And even if your task resolution is not successful, please return as much context as possible, so that your manager can
act upon this feedback.
```

HfApiModel - Qwen/Qwen2.5-Coder-32B-Instruct

Main Agent delegates the tasks, provides instructions

# DR MACS IN ACTION: SECONDARY AGENT

Argonne NATIONAL LABORATORY

*Output message of the LLM:*

````
Thought: To solve the task, I will use the `search_sdss_galaxies` tool to retrieve galaxies with the specified stellar mass
and redshift constraints. I will start by calling the tool with the appropriate filters, such as `max_brightness` and
`redshift`.

Code:
```py
galaxies = search_sdss_galaxies(min_redshift=0.1, max_redshift=0.2, num_entries=100)
print(galaxies)
```<end_code>

Observation:
````

**Agent wants to execute the following code:**

```python
galaxies = search_sdss_galaxies(min_redshift=0.1, max_redshift=0.2, num_entries=100)
print(galaxies)
```

Press Enter to execute, or type 'no' to abort:

Coding Agent goes over the tools provided, decides on the inputs

Coding Agent asks the user permission to execute the code

# DR MACS IN ACTION: DATA DOWNLOAD

Argonne NATIONAL LABORATORY

**Execution logs:**

| | objID | ra | dec | redshift | expRad_r | ... | run2d | ellipticity | distance | M_r | stellar_mass |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1237648720141746438 | 145.186672 | -0.893394 | 0.142611 | 2.528442 | ... | 26 | 0.533659 | 610.768460 | -21.283233 | 1.214124e+11 |
| 1 | 1237648720141746438 | 145.186672 | -0.893394 | 0.142567 | 2.528442 | ... | 26 | 0.533659 | 610.579162 | -21.282560 | 1.213372e+11 |
| 2 | 1237648720144040025 | 150.470354 | -0.878049 | 0.136434 | 2.447338 | ... | 26 | 0.317326 | 584.310919 | -23.071740 | 6.304723e+11 |
| 3 | 1237648720144040175 | 150.383283 | -0.939088 | 0.135132 | 2.838660 | ... | 26 | 0.372125 | 578.736492 | -21.052064 | 9.812863e+10 |
| 4 | 1237648720144040025 | 150.470354 | -0.878049 | 0.136351 | 2.447338 | ... | 26 | 0.317326 | 583.959305 | -23.070433 | 6.297137e+11 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95 | 1237648721227022436 | 171.492665 | -0.094285 | 0.102538 | 2.096871 | ... | 26 | 0.223799 | 439.144130 | -21.718615 | 1.813077e+11 |
| 96 | 1237648720147448142 | 158.213610 | -0.911631 | 0.115113 | 3.056103 | ... | 26 | 0.377400 | 492.999275 | -21.141671 | 1.065709e+11 |
| 97 | 1237648720679927941 | 148.133192 | -0.524590 | 0.145860 | 2.376235 | ... | 26 | 0.160101 | 624.681399 | -22.819223 | 4.996422e+11 |
| 98 | 1237648720156557509 | 179.071941 | -0.975031 | 0.109721 | 2.891736 | ... | 26 | 0.747597 | 469.906262 | -20.620606 | 6.594965e+10 |
| 99 | 1237648720156557509 | 179.071941 | -0.975031 | 0.109727 | 2.891736 | ... | 26 | 0.747597 | 469.934528 | -20.620737 | 6.595758e+10 |

[100 rows x 22 columns]

First set of data downloaded from online database

# DR MACS IN ACTION: IDENTIFYING ISSUES

Argonne NATIONAL LABORATORY

```
        AND s.z > 0.1 AND s.z < 0.2 AND p.dered_i < 18 AND p.expRad_r > 2.0 AND p.ra IS NOT NULL AND p.dec IS NOT NULL AND s
.z IS NOT NULL AND p.expRad_r IS NOT NULL AND p.dered_i IS NOT NULL
Code execution failed at line 'print(filtered_galaxies[['id', 'ra', 'dec', 'redshift', 'stellar_mass']])' due to:
KeyError:"['id'] not in index"
[Step 1: Duration 23.71 seconds| Input tokens: 6,566 | Output tokens: 323]
                                    Step 3
Output message of the LLM:
Thought: Given the error indicating that the expected columns are not in the output, it appears the `search_sdss_galaxies`
tool might not return all the columns we assumed. We need to refine how we retrieve and filter the data. I will first
retrieve the data without any specific columns and then filter it based on redshift and stellar mass. If the stellar mass
data is not available directly, I'll need to filter only on redshift and provide a detailed report on the available data and
whether we can get the required information.

Let's fetch the data again without specifying the columns and analyze it.
Code:
```py
galaxies = search_sdss_galaxies(min_redshift=0.1, max_redshift=0.2, num_entries=100)
print(galaxies)
```<end_code>

Observation:

**Agent wants to execute the following code:**

```python
galaxies = search_sdss_galaxies(min_redshift=0.1, max_redshift=0.2, num_entries=100)
print(galaxies)
```

Press Enter to execute, or type 'no' to abort:
```

Plain tool calling not providing desired output, Coding agent figuring out a workaround

# DR MACS IN ACTION: DEBUG AND RE-RUN

Argonne NATIONAL LABORATORY

✅ Proceeding with execution.

**Executing parsed code:**

```
# Filter galaxies based on redshift range 0.1 < z < 0.2
filtered_galaxies = galaxies[(galaxies['redshift'] > 0.1) & (galaxies['redshift'] < 0.2)]

# Check if the stellar_mass column is available and contains valid values
if 'stellar_mass' in filtered_galaxies.columns and not filtered_galaxies['stellar_mass'].isnull().all():
    # Filter galaxies based on stellar mass range 10^10 to 10^12 solar masses
    final_filtered_galaxies = filtered_galaxies[(filtered_galaxies['stellar_mass'] >= 1e10) &
(filtered_galaxies['stellar_mass'] <= 1e12)]
else:
    final_filtered_galaxies = filtered_galaxies
    print("Stellar mass data is not available or contains no valid values.")

# Print the final filtered galaxies
print(final_filtered_galaxies[['ra', 'dec', 'redshift', 'stellar_mass']])
```

**Execution logs:**

| | ra | dec | redshift | stellar_mass |
|---|---|---|---|---|
| 0 | 165.686010 | 0.683687 | 0.123009 | 8.674490e+10 |
| 1 | 165.734059 | 0.661327 | 0.107443 | 1.966863e+11 |
| 2 | 145.186672 | -0.893394 | 0.142611 | 1.214124e+11 |
| 3 | 145.186672 | -0.893394 | 0.142567 | 1.213372e+11 |
| 4 | 153.831485 | -1.009414 | 0.164296 | 1.928532e+11 |
| .. | ... | ... | ... | ... |
| 95 | 174.744577 | -0.848909 | 0.130069 | 1.550161e+11 |
| 96 | 150.381782 | -0.174238 | 0.105979 | 1.667750e+11 |
| 97 | 163.384830 | -0.960156 | 0.126585 | 1.575075e+11 |
| 98 | 148.033231 | -0.562277 | 0.134783 | 9.866342e+10 |
| 99 | 150.381782 | -0.174238 | 0.105973 | 1.667552e+11 |

[100 rows x 4 columns]

Out: None

[Step 3: Duration 28.53 seconds| Input tokens: 16,947 | Output tokens: 827]

Step 5

Code corrected, run, new data downloaded

# DR MACS IN ACTION: FINAL STEPS

```
### 3. Additional context (if relevant):
The Petrosian absolute magnitude is a measure of the intrinsic brightness of a galaxy, corrected for interstellar absorption.
It is commonly used in galaxy astronomy as it provides a consistent way to compare the brightness of galaxies over different
distances and environments.
[Step 1: Duration 22.81 seconds| Input tokens: 6,508 | Output tokens: 527]
Out: Here is the final answer from your managed agent 'codi[illegible]':
### 1. Task outcome (short version):
Retrieved Petrosian absolute magnitude for 100 galaxies from the SDSS galaxy catalog.
### 2. Task outcome (extremely detailed version):
Retrieved Petrosian absolute magnitude data for 100 galaxies. The data includes the following columns:
['objID', 'ra', 'dec', 'redshift', 'expRad_r', 'expAB_r', 'expPhi_r', 'dered_u', 'dered_g', 'dered_r', 'dered_i', 'dered_z',
'mjd', 'plate', 'fiberID', 'specObjID', 'instrument', 'run2d', 'ellipticity', 'distance', 'M_r', 'stellar_mass']
The Petrosian absolute magnitude values for the first 10 galaxies are:
0   -20.584366
1   -20.583812
2   -21.303724
3   -21.304331
4   -21.282560
5   -21.283233
6   -21.887198
7   -21.888559
8   -21.643672
9   -21.642978
Name: M_r, dtype: float64
The minimum Petrosian absolute magnitude is: -23.071740008457887
The maximum Petrosian absolute magnitude is: -16.843699423068873
The mean Petrosian absolute magnitude is: -21.1615320371003
The standard deviation of Petrosian absolute magnitude is: 1.0570285824448769
```

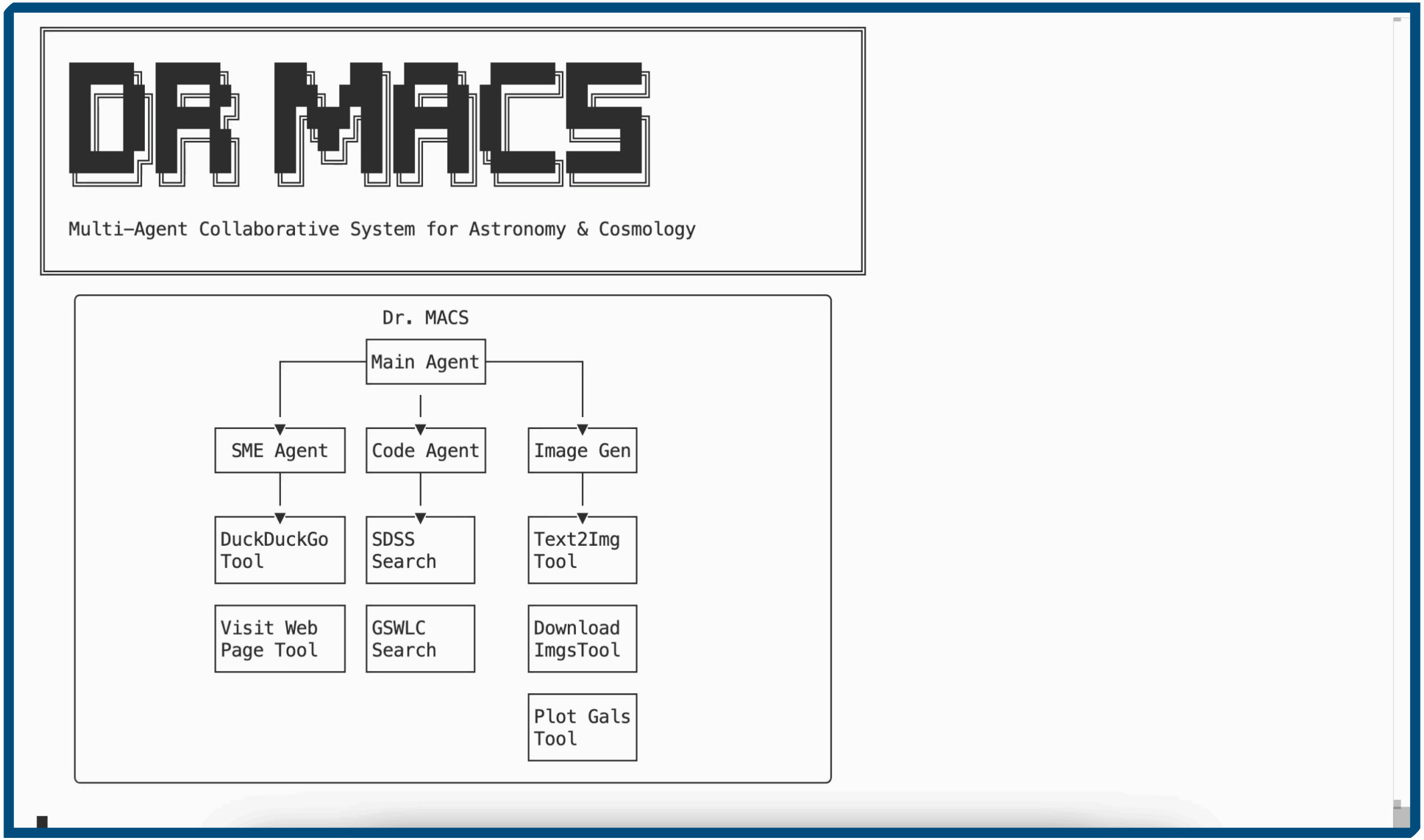
Agent deciding on extracting the data, provides explanations

Agent selecting the relevant data, performing data analysis

```
Press Enter to execute, or type 'no' to abort:
✅ Proceeding with execution.
 — Executing parsed code: ——————————————————
  final_answer(average_brightness)
 ——————————————————————————————————————————
Out - Final answer: -21.1615320371003
```

Final Answer from Main Agent in <1.5 minutes, with full logs of the reasoning steps, codes, issues and notes

# DR MACS IN ACTION: FULL RUN



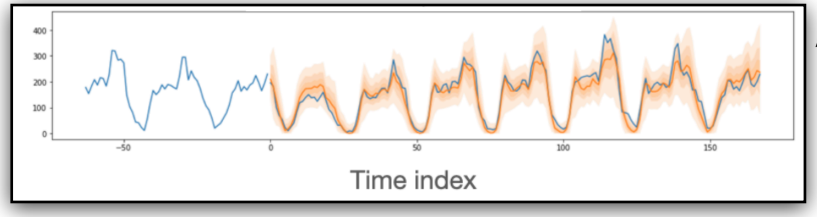
# CONCLUSIONS AND FUTURE OUTLOOK

- The foundation models have finally facilitated a truly large and deep representations of some of the largest datasets.
- The FMs are mostly trained for language tasks. Encoding non-text information (either directly or indirectly) is the next important step for scientific impact.
- Agentic frameworks allow for automated explorations of data based on hypothesis and planning, code deployment and feedback loops — all geared towards solving a single problem.
- Exciting times ahead, keeping up with the progress is key.

Questions?

# BACKUP: BUILDING BLOCKS OF FOUNDATION MODELS

- Transformer blocks are the fundamental blocks, self-attention is essential functionality:
  - From "**Attention is all you need:** https://arxiv.org/abs/1706.03762"
- Successor to Recurrent Neural Networks (RNNs), designed to predict next occurrence in a time series.
- Self attention is key to process large amount of data at once.
- Associated 'technical' details: Tokens, Token embeddings, Position encoding, Vocabulary, Context, Context window lengths, Multi-head attention.
- The reasoning-class models (OpenAI O-models and others) also use reinforcement learning in strategizing steps using reward mechanisms.




Sequence of tokens: $U = \{u_1, u_2, u_3, \ldots, u_N\}$

Minimize: $L(U) = \sum_{i=1}^{N} log(p(u_i \mid u_{i-k}, \ldots, u_{i-1}, \Theta))$

(Conditional probability of i-th token given k preceding tokens)

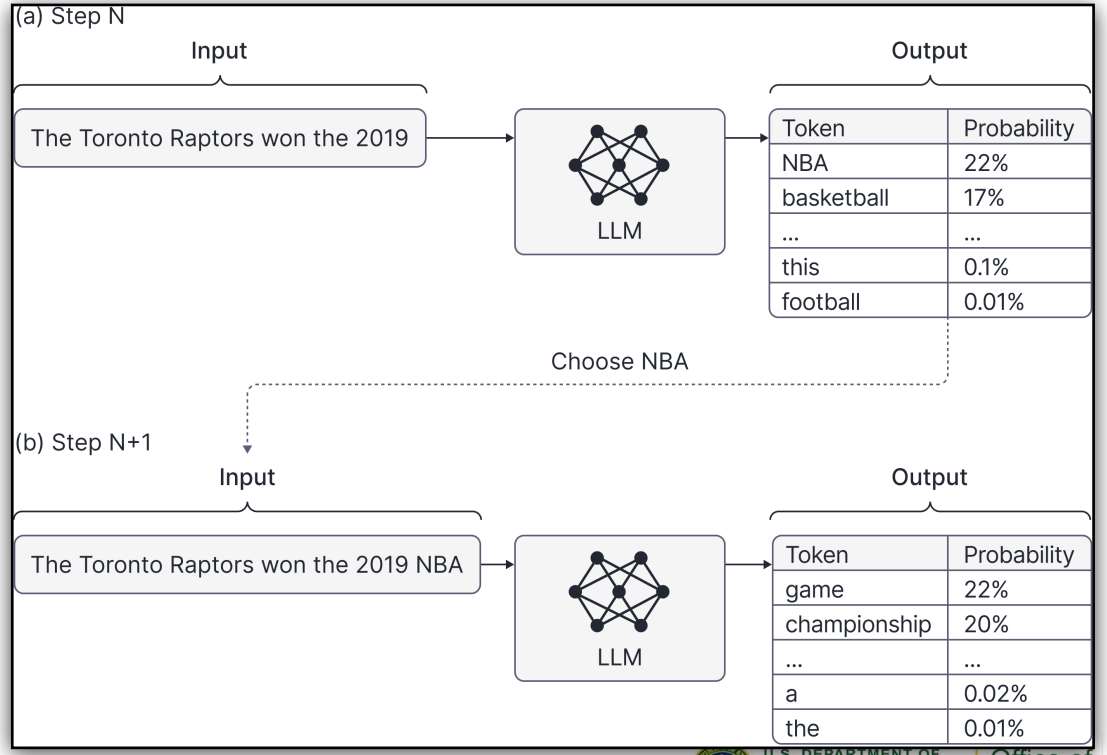



Image source:borealisai

# BACKUP: PHYSICS BENCHMARKS FOR LANGUAGE MODELS

- Multiple choice questions and Free-form questions can be used to evaluate and compare models
- Compared to biology/chemistry, available LLM benchmarks for physics/astronomy is inadequate
  - Both in terms of quantity and quality

| |
|---|
| Where do most short-period comets come from and how do we know? |
| Say the pupil of your eye has a diameter of 5 mm and you have a… |
| Why isn't there a planet where the asteroid belt is located? |
| Why is Mars red? |

mmlu-astronomy

| | |
|---|---|
| Electromagnetic radiation provides a means to probe aspects of the physical universe. Which of the following statements regarding radiation spectra is NOT correct? | [ "Lines in the infrared, visible, and ultraviolet regions of the spectrum reveal primarily the nuclear structure of the sample.", "The wavelengths identified in an absorption spectrum of an element are among those in its emission spectrum.", "Absorption spectra can be used to determine which elements are present in distant stars.", "Spectral analysis can be used to identify the composition of galactic dust." ] |
| One end of a horizontal, massless spring is attached to a wall. A mass of 0.30 kg is attached to the other end of the spring and rests on a table. The mass is displaced… | [ "0.24 mJ", "0.38 mJ", "0.48 mJ", "0.75 mJ" ] |

mmlu-college_physics-original-neg

| | |
|---|---|
| Things that are equivalent according to the equivalence principle are | [ "space and time.", "a traveling twin and a stay-at-home twin.", "gravity and acceleration.", "mass and energy." ] |
| Colors in a soap bubble result from light | [ "converted to a different frequency", "deflection", "interference", "polarization" ] |

mmlu-conceptual_physics-dev

# BACKUP: BENCHMARKING FOR AURORA-GPT


- Benchmark development may be crucial for a Science-focussed GPT comparison with existing LLMs.
- Benchmarking team at the Aurora-GPT collaboration has released a web-form to collect science questions of interest — with real-time evaluation from multiple LLMs.
- Goal is to collect O(1000) questions across scientific fields — **HELP needed! (and potential collaboration opportunities)**
  - High-quality.
  - Should represent what the science community wants out of an LLM.
  - Should not be exposed to current LLMs.

***https://web.cels.anl.gov/projects/auroragptquestions/ui/index.html***

**Trial Q:For the force between quarks, which of the following statements is true?**

**a)** The force follows an inverse square law
**b)** Approximately an inverse square law but asymptomatically weaker at short distances
**c)** Approximately an inverse square law but asymptomatically stronger at short distances
**d)** Approximately an Inverse linear relationship, but goes asymptomatically free at large distances.
**e)** Asymptomatically free at large distances and small distances, inverse relationship in between.

# BACKUP: SEMI-AUTOMATED BENCHMARKS

Argonne NATIONAL LABORATORY

- Existing LLMs can be used to construct questions from a small number of papers (review papers, white papers).
  - Deeper than current MCQs, yet not hyper-specific to papers.
  - Scaling to O(10000) questions is relatively more straightforward.

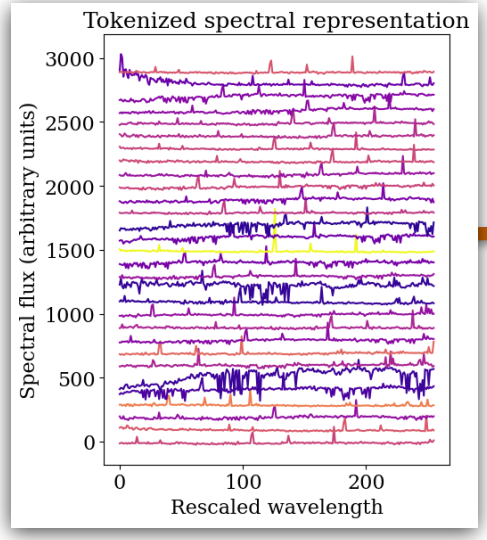
```
"question": "What is the predominant model
    for cosmic ray (CR) transport in the
    Galaxy at energies below about 10^17 eV?",
"distractors": [
    "Convective wind model with uniform
        source distribution",
    "Static halo model with isotropic source
        emission",
    "Leaky-box model with constant escape
        time"
],
"correct_answer": "Diffusion model with
    possible inclusion of convection",
```

```
"question": "If a new particle was discovered
    that interacts with the weak force but
    not the electromagnetic or strong force,
    what could be inferred about its
    properties?",
"distractors": ["It must have a fractional
    electric charge", "It must be massless",
    "It could be a candidate for dark
    matter", "It must be a type of quark"],
"correct_answer": "It could be a candidate
    for dark matter",
```

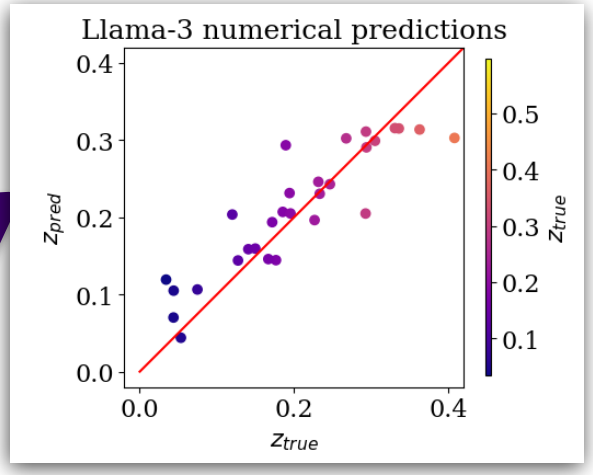
```
"question": "If gauge bosons were discovered
    to have non-zero rest mass without the
    need for a Higgs mechanism, what would be
    the most likely implication?",
"distractors": ["The existence of a fifth
    fundamental force", "A revision of the
    Standard Model to include mass terms for
    gauge bosons", "The invalidation of gauge
    symmetry", "The discovery of new gauge
    bosons beyond the Standard Model"],
"correct_answer": "A revision of the Standard
    Model to include mass terms for gauge
    bosons",
```

# BACKUP: FOUNDATION MODELS BEYOND TEXT: FINE-TUNING

- Missing features in current LLMs:
  - Support for non-text data like scalar numbers, vectors, images, trees, unstructured datasets.
- Can pre-trained LLMs be re-purposed for numerical tasks?
  - It depends!
  - Degradation/forgetting is expected




Fine-tuned LLaMa




```
Q: What type of galaxy is expected at redshift of z=0.211? List the telescope
    surveys where such a galaxy can be observed.

A galaxy at redshift z=0.211 is expected to be a luminous infrared galaxy
    (LIRG) or an ultraluminous infrared galaxy (ULIRG) due to the fact that
    the universe was in its peak star formation period during that epoch. The
    galaxy is likely to be a merger or a strongly interacting system, and it
    could be a dusty galaxy with a high rate of star formation.

The galaxy can be observed in the following telescope surveys:

1. Sloan Digital Sky Survey (SDSS): The galaxy can be observed in the SDSS-IV
    survey, which covers the sky in five bands (u, g, r, i, and z) and has a
    limiting magnitude of 22.5 in the r-band.
2. Dark Energy Survey (DES): The galaxy can be observed in the DES survey,
    which covers the sky in five bands (g, r, i, z, and Y) and has a limiting
    magnitude of 24.5 in the i-band.
3. Hyper Suprime-Cam (HSC) Survey: The galaxy can be observed in the HSC
    survey, which covers the sky in five bands (g, r, i, z, and Y) and has a
    limiting magnitude of 26.5 in the i-band.
4. Subaru Hyper Suprime-Cam (HSC) Survey: The galaxy can be observed in the
    Subaru HSC survey, which covers the sky in five bands (g, r, i, z, and Y)
    and has a limiting magnitude of 26.5 in the i-band.
```

NR, Yuan-Sen Ting (in prep.)

# BACKUP: AURORA-GPT — OPEN SCIENCE FOUNDATION MODELS

- Initially with textual interface that would respond to scientific prompts
- Single model or Mixture/Combination of Experts (scale: ~1 T parameter)

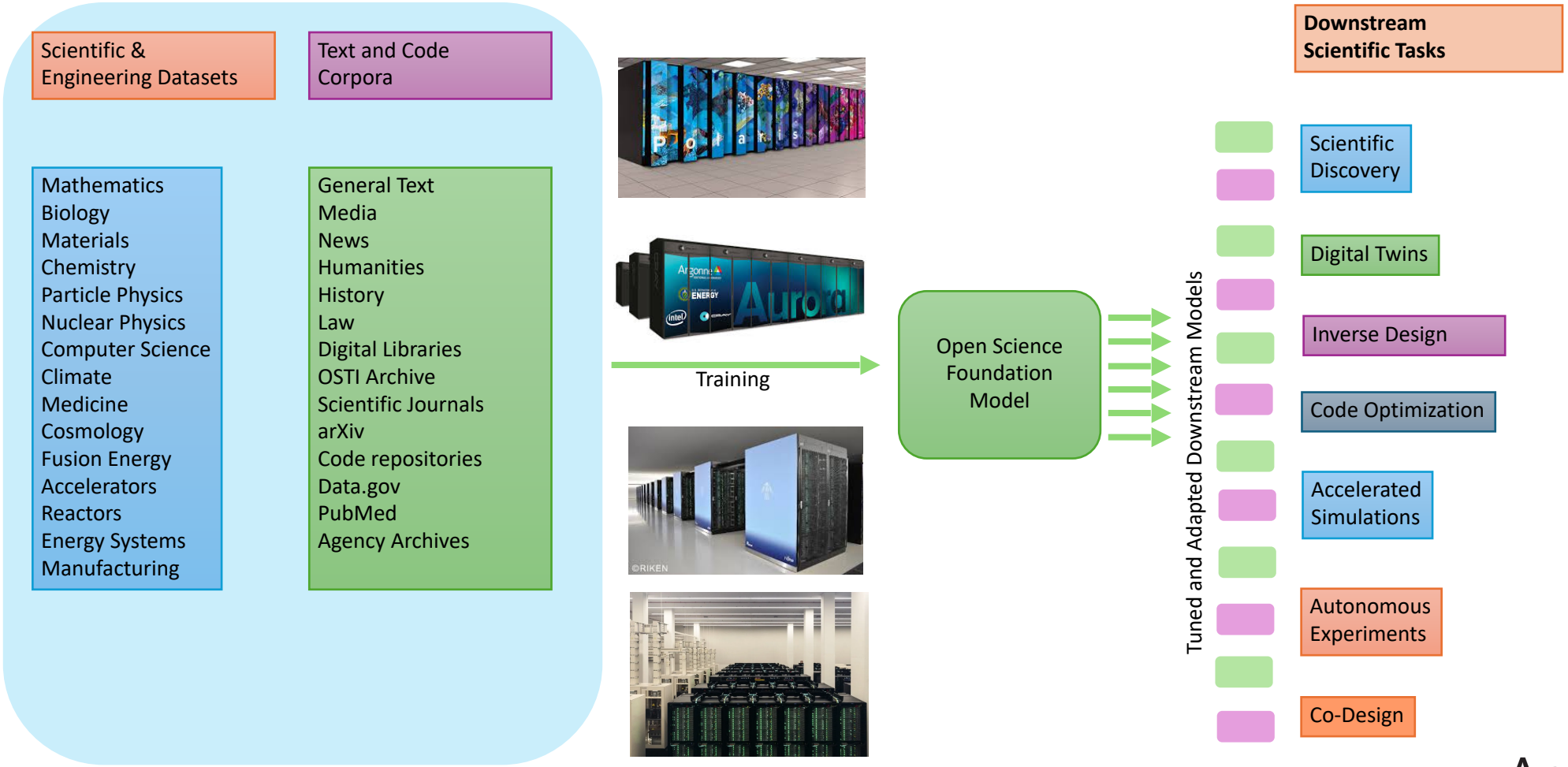



Figure inspired by Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258* (2021).

Argonne NATIONAL LABORATORY