

NSF-Simons AI Institute for Cosmic Origins: Onwards into Year 2

Stella Offner, Oct 7

UT Austin, Director of CosmicAI

National Academy Of Sciences Committee on A&A Fall Meeting



CosmicAI Mission

CosmicAI aims to develop transformative AI methods to meet pressing Astronomical challenges and tackle outstanding questions about our cosmic origins through research in four fundamental AI themes: trustworthiness, efficiency, interpretability, and robustness.

CosmicAI will serve as a nexus of collaboration to increase the accessibility of astronomy and AI data and methods through open-source AI-powered tools, data sharing, and AI educational initiatives.



Who we are: Nexus of Collaboration

★ Lead Institutions

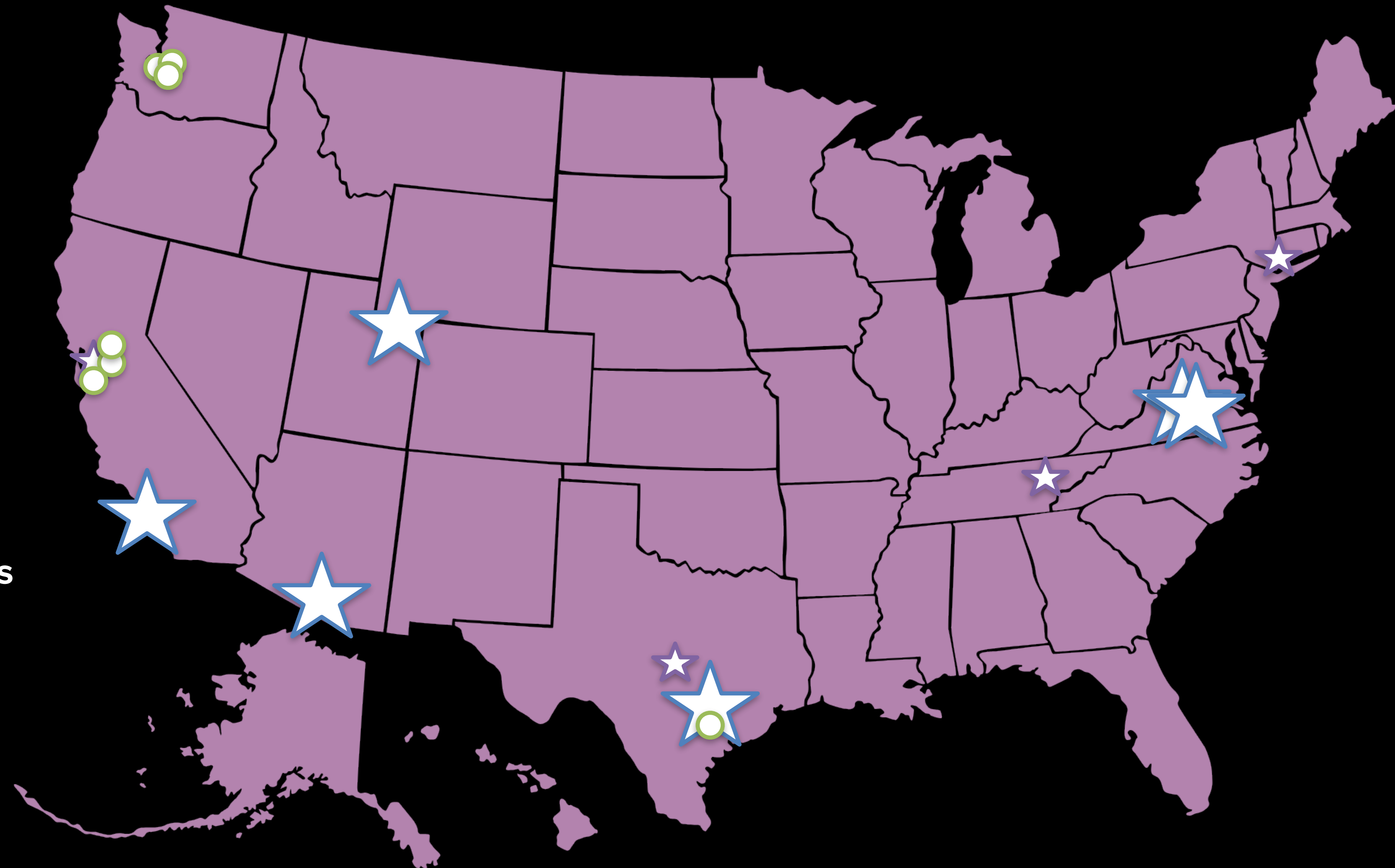
- University of Texas at Austin
- University of Virginia
- University of Utah
- University of California Los Angeles
- National Radio Astronomy Observatory
- National Optical-Infrared Astronomy Research Laboratory

★ Academic Partners & Labs

- SLAC National Accelerator Laboratory
- University of Texas Arlington
- University of Tennessee
- New York University

● Industry/Non-Profit Partners

- Ai2
- Microsoft
- Intel
- NVIDIA
- Sony AI
- Amazon
- SparkCognition



Executive Leadership



Stella Offner
UT Austin
PI and Director
Accelerated Universe,
Astro Lead



Matt Lease
UT Austin
Co-Director
Explorable Universe,
AI Lead



Jessy Li
UT Austin
Explorable Universe,
AI Lead



Stéphanie Juneau
NOIRLab
Explorable Universe,
Astro Lead



Adam Bolton
SLAC Stanford
Co-PI
Explorable Universe
Astro Lead



Arya Farahi
UT-Austin
Co-PI
Explainable Universe,
AI Lead



Paul Torrey
University of Virginia
Explainable Universe,
Astro Lead



Jeff Phillips
University of Utah
Observable Universe,
AI Lead



Eric Murphy
NRAO
Co-PI
Observable Universe,
Astro Lead



George Biros
UT Austin
Accelerated Universe,
AI Lead

Research Themes

Artificial Intelligence | Astronomy

Explorable Universe

Advance generative AI **trustworthiness** by developing open AI foundation models for astronomy

Observable Universe

Develop **efficient** AI techniques to expedite observing pipelines and discovery in big datasets

Accelerated Universe

Develop fast **robust** AI surrogates to model multi-physics, multi-scale astrophysical systems

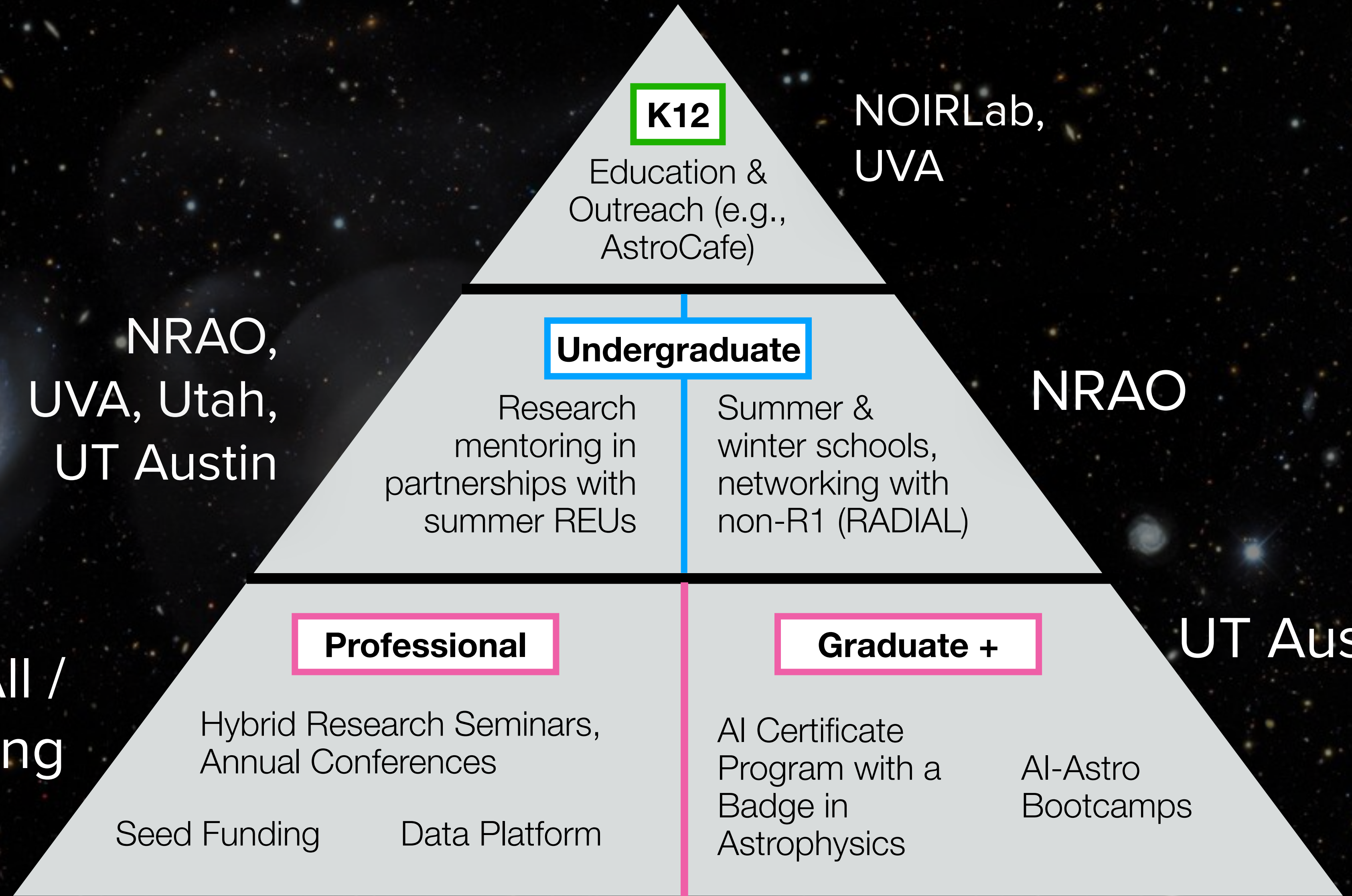
Explainable Universe

Extend **interpretable** AI methods and apply causal reasoning to analyze large-scale simulations

Broader Impacts

Education +
Broadening
Participation

All /
rotating



NOIRLab,
UVA

K12
Education &
Outreach (e.g.,
AstroCafe)

NRAO,
UVA, Utah,
UT Austin

Undergraduate

Research
mentoring in
partnerships with
summer REUs

Summer &
winter schools,
networking with
non-R1 (RADIAL)

NRAO

Professional

Hybrid Research Seminars,
Annual Conferences

Seed Funding

Data Platform

Graduate +

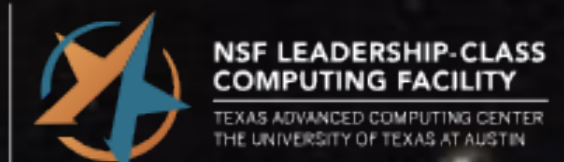
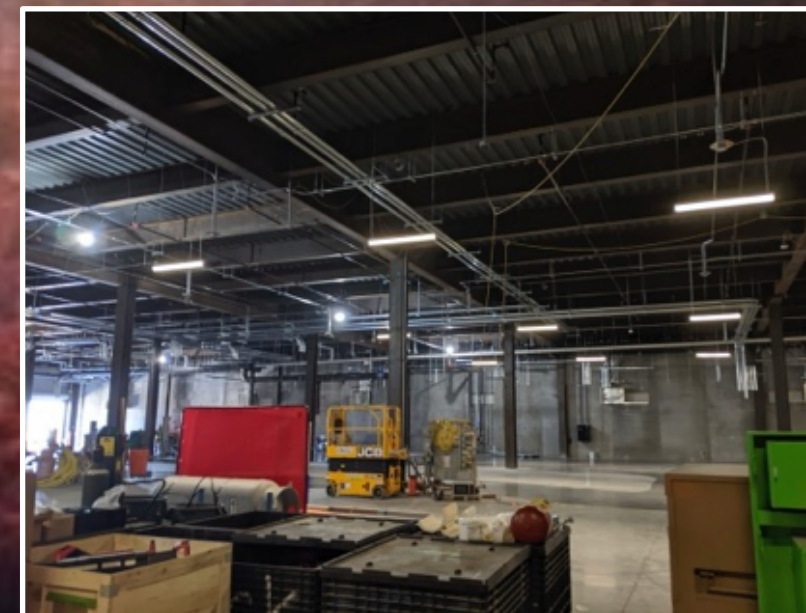
AI Certificate
Program with a
Badge in
Astrophysics

AI-Astro
Bootcamps

UT Austin

TACC resources supporting CosmicAI

- **Vista (currently) for AI compute**
 - 600 Grace Hopper (H200) nodes with 400 Gbps interconnect
 - CosmicAI is allocated **80,000 GPU hours** annually
- **Corral for data sharing and preservation**
 - 40 PB shared file system for data hosting
 - **200 TB** initial allocation
 - On prem cloud environments for service hosting
- **Expertise & Technical Support**
 - TACC personnel are **integrated into the working groups**
 - Developed and Implemented **Astro-AI Bootcamp**
 - **Data and Software Managers** help to store, curate, and manage products
- **Leadership Class Computing Facility and Horizon (coming soon!)**
 - TACC will lead the LCCF for the NSF for the next decade
 - Horizon will be the largest compute cluster for open science in the world starting in the **Fall 2025**
 - 4000 Vera Rubin CPU nodes, 4000 Blackwell GPUs
 - 400 PB all flash file system

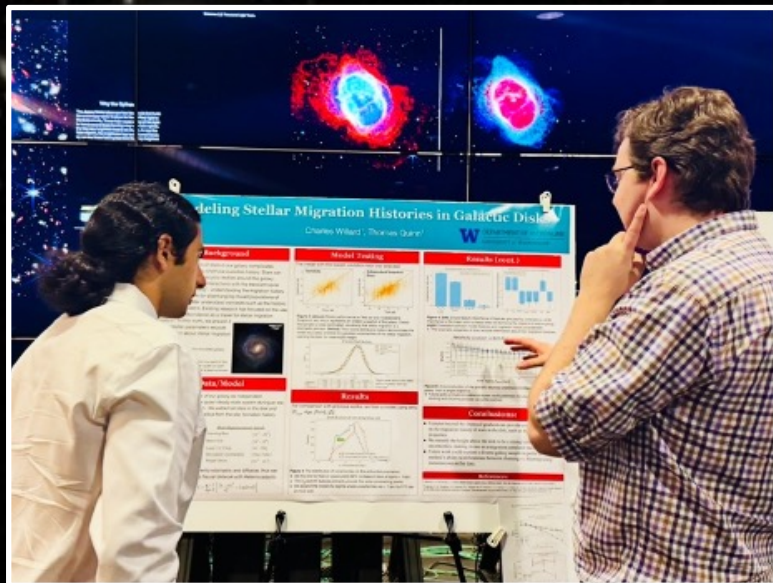


Credit: NASA, ESA and the Hubble Heritage Team (STScI/AURA).
Acknowledgment: J. Gallagher (University of Wisconsin), M. Mountain (STScI) and P. Puxley (NSF).

Year 1 Research Highlights



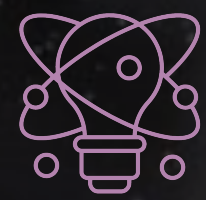
Cosmic Horizons
Conference (UT)



Formed and Launched 4 Research Groups



Focus on Building Astro Benchmarks +
Training Sets for Models in all Groups



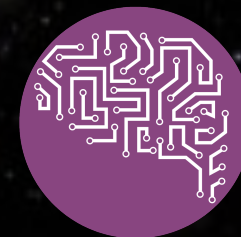
1st 100% CosmicAI project completed +
accepted to NeurIPS (AstroVisBench)



18 CosmicAI-Related Publications



CosmicAI Research Featured in
talks at 17 Venues



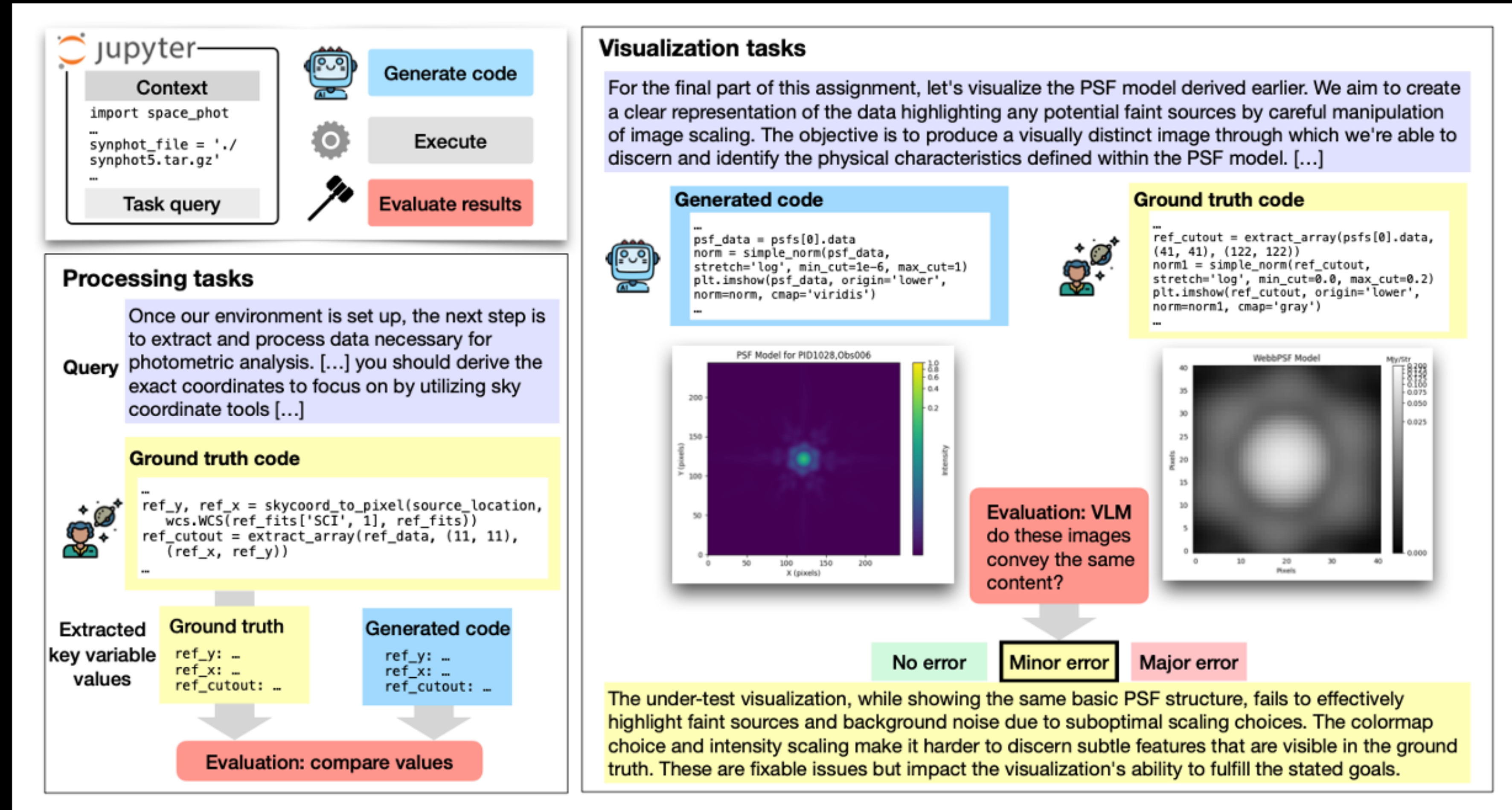
8 Seed Research Proposals Awarded



Achievement Highlights 1: AstroVisBench

A language model benchmark to evaluate astronomy workflow tasks that lead to a visualization

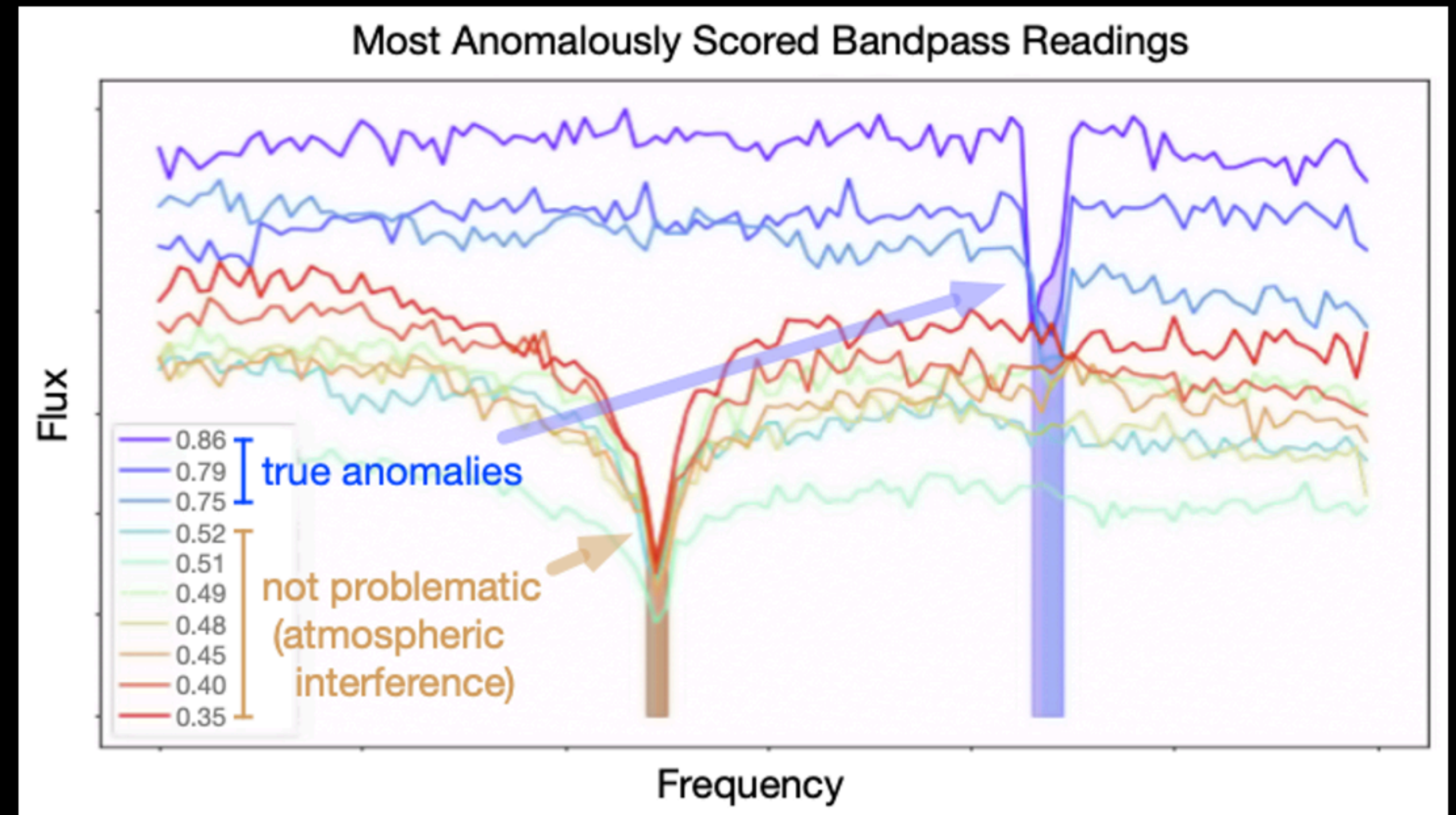
- Uses a novel LLM-as-a-judge workflow, validated by five professional astronomers
- Includes 432 coding tasks for visualization sourced from astronomy research python notebooks
- Evaluations of 8 state-of-the-art language models indicate significant deficits in their ability to be useful research assistants



Achievement Highlights 2: Data Cleaning

Goal: Use machine learning to automate the processing of radio telescope data

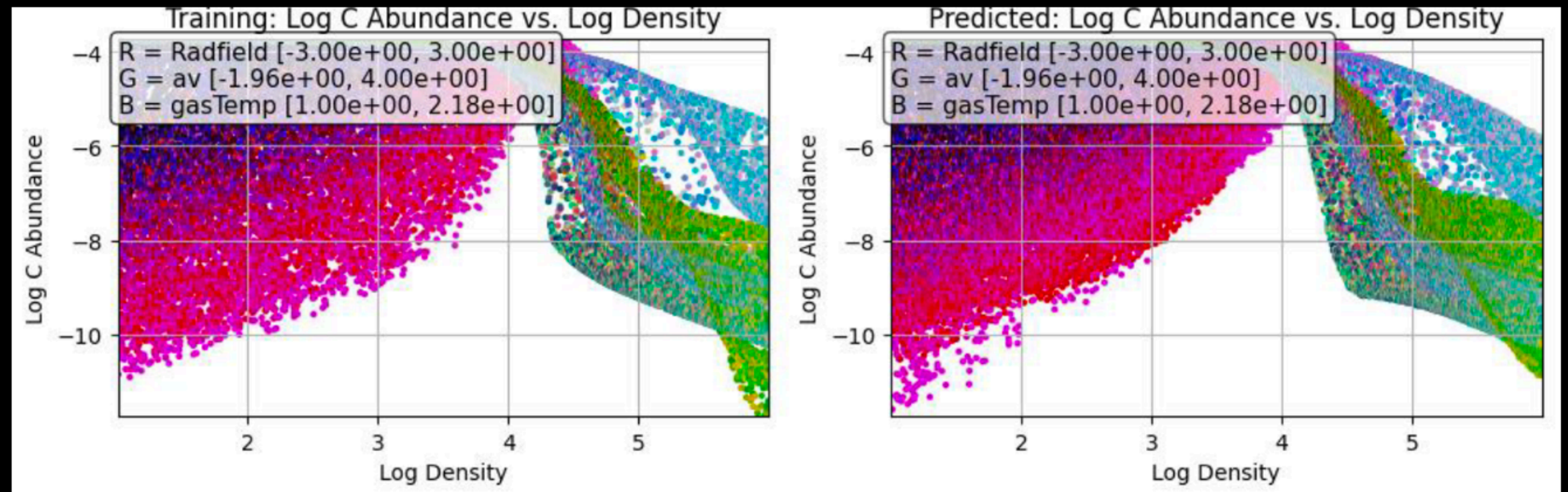
- Use scan statistics to identify frequency regions with measurement interference
- Method accounts for:
 - Low signal to noise
 - Non-uniform signal baseline
 - Atmospheric interference
- Designed new efficient algorithms, using a kernel regression base model
- Identified anomalies validated by expert astronomers



Achievement Highlights 3: Chemistry Surrogate

A deep learning model that predicts the evolution of chemical abundances

- Predicts abundances for an astrochemistry network relevant for star and planet formation that contains 333 species and 3,187 reactions
- Conserves the species to 1% accuracy
- The model is 1,300 / 3,500 times faster than directly solving the network equations using CPUs / GPUs



Data input to the model during training

Abundances predicted by the model

Achievement Highlights 4: Dark Matter Decay

Goal: Use machine learning to predict the properties of dark matter

Using the [DREAMS Simulations](#) (Rose, Torrey et al. 2025) we aim to answer how the dark matter decay or annihilation signal changes as a function of:

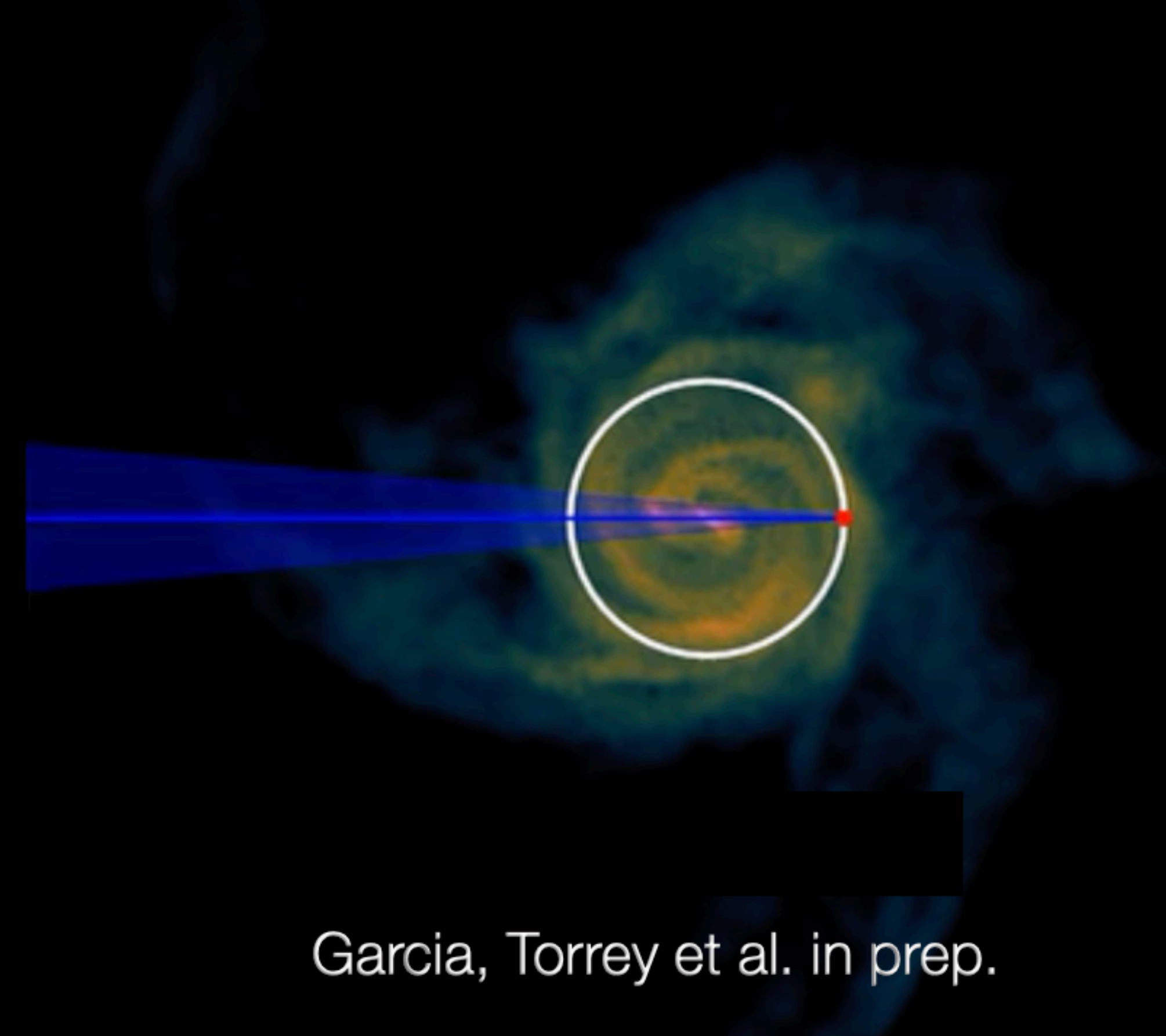
- Dark matter physics
- Uncertainty in the cosmology
- Uncertainty in the astrophysics modeling of simulations

We use surrogate models trained on simulations to efficiently learn the signal's dependence on all model parameters.

AI enables rapid prediction and analysis that would be too computationally expensive to sample directly.

Gas Density

Example configuration of a Milky Way galaxy



Education & Outreach Highlights



Astro-AI Bootcamp (UT)



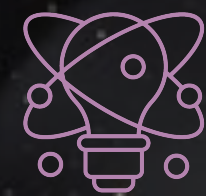
AI Teacher Training (UVA)



CosmicAI Seminar: Alternating weeks (142 people)



CosmicAI-UVA K12 Teacher Training — AI in the Classroom (38 teachers)



Cosmic Horizons Conference (106 registered from 43 institutions)



CosmicAI AstroAI Bootcamp (25 students, 12 institutions)



Research Training (24 students & postdocs)



AI-Astro Certificate Program (Application closed Sept 1, 203 apps, 119 for the Astro-AI Badge)



AI-Astro Advocacy



COSMIC HORIZONS CONFERENCE

AI-Powered Insights into the Universe

Bringing together researchers who are actively developing and applying AI/ML methods in astronomy

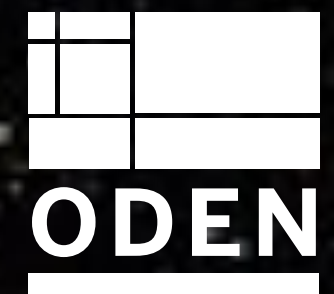
Location

The University of
Texas at Austin

Dates

May 6th-8th
2025

Learn more



AI BOOTCAMP FOR ASTRONOMERS

PRESENTED BY THE NSF-SIMONS AI INSTITUTE FOR COSMIC ORIGINS
AND THE TEXAS ADVANCED COMPUTING CENTER

Overview of the fundamentals and theory of machine learning & deep learning techniques

Tutorials on reinforcement learning, large language models, & surrogate models

Practice applying AI techniques in a high performance computing environment

Demonstrations of public AI/ astronomy software packages



Location
The University of Texas at Austin

Dates
May 5th-May 9th 2025

AI Foundations Certificate with a Concentration in Astrophysics

New Graduate Certificate Program

Equips professionals with AI skills for data analysis, predictive modeling and research, fostering career advancement and long-term growth in AI-driven fields

Advanced AI Techniques

Astrophysics Specialization

12 credit hours
4 courses

Asynchronous, instructor-paced

Accessed from anywhere

Transcripted and Accredited

Sign up for our interest form



Applications Open June 1st, 2025

Applications Close : September 15th, 2025
Program Start : Spring 2026



CosmicAI Team:

88

Senior Personnel: 28

Postdocs: 8

Students: 17

Collaborators: 14

External Advisory Board: 5

Staff and BI Program Managers: 16

Year 2 Plans

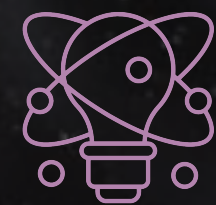
Research



Explorable: AI Assistant training focusing on multi-turn interaction



Observable: Develop an AI system to automate the detection and handling of a wide variety of anomalies for data calibration.



Explainable: Generating Training Data [ongoing], Developing Inference Models Across Scales



Accelerated: Release benchmarks and astrochem solver. Continue surrogate evaluation and architecture exploration.



Data Platform: Astro dataset curation & transfer to the data platform



Year 2 Plans

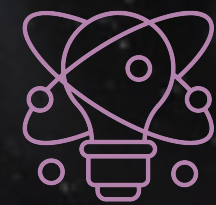
Education & Outreach



1st semester of Certificate Program (Spring 2026)



1st Winter & Summer Schools @NRAO (2026)



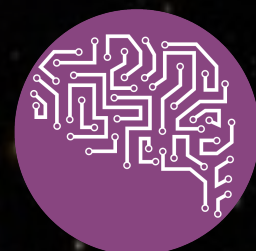
2nd AI K12 Teaching Training @UVA, maybe @UT



2nd AI-Astro Bootcamp @UT



Continue research seminar series + conference(s)



Continue to grow collaborations, esp. student & postdoc members



Year 2 Plans

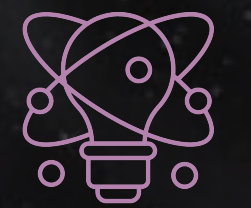
Research



Explorable: AI Assistant training focusing on multi-turn interaction



Observable: Develop an AI system to automate the detection and handling of a wide variety of anomalies for data calibration



Explainable: Generating Training Data [ongoing], Developing Inference Models Across Scales



Accelerated: Release benchmarks and astrochemistry solver. Continue surrogate evaluation and architecture exploration.

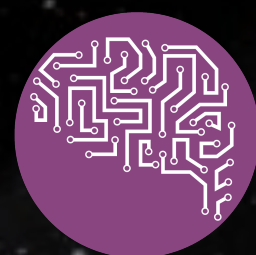


Data Platform: Astronomy dataset curation & transfer to the data platform.



Year 2 Plans

Education & Outreach



1st semester of Certificate Program (Spring 2026)



1st Winter & Summer Schools @NRAO (2026)



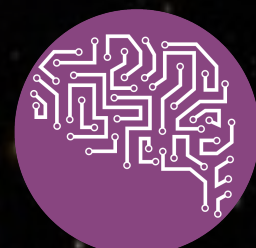
2nd AI K12 Teaching Training @UVA, maybe @UT



2nd AI-Astro Bootcamp @UT



Continue research seminar series + conference(s)



Continue to grow collaborations, esp. student & postdoc members



Questions and Discussion

Community Engagement: Ways to Join

Attend

CosmicAI Seminars
Cosmic Horizon Conferences
Astro-AI Bootcamps
Astro-AI Certificate

Subscribe

Email
Slack
Social Media

Collaborate

Research:
Explorable
Observable
Explainable
Accelerated +
Education Initiatives

Apply

Seed funds for new
Astro-AI projects

Contribute

Data & code to our
data platform, open
source
infrastructure

Participate

Use data, code,
models and tools
developed by
CosmicAI

NSF Annual Review

Site Visit in July

Rating: Evolving / On Track

“CosmicAI has had a successful first year and is exactly on track with where it needs to be at this stage. The strong management and governance structure that are in place, the well-designed interdisciplinary research thrusts, and the energy that the team is devoting to all its activities, have positioned the institute to have significant impact as it ramps up over the next few years.”

External Advisory Board

April 2025 review

“The Board was extremely impressed by the fast start of the CosmicAI Institute, in terms of administrative structure, organization of research, ambitious goals, and strategic planning. The buy-in and engagements of the partners and stakeholders is remarkable and sets up the Institute for success.”



David Hogg (Chair)
NYU
Astro/AI



Viviana Acquaviva
CUNY
Astro/AI/Education



Josh Peek
Stsci
Astro/Data



Yusu Wang
UC San Diego
AI (TILOS director)



Michael Brenner
Harvard / Google
AI

Year 1 Interactions:

- EAB + Directors meet & greet (Feb)
- 1-day virtual review + report (April)
- In-person meeting, Chair + Director (May)

