OCSDO
OFFICE OF THE CHIEF SCIENCE DATA OFFICER

ARTIFICIAL INTELLIGENCE WORKSHOP REPORT

NASA Science Mission Directorate

March 2024

TABLE OF CONTENTS

Workshop agenda, recordings, and presentations can be found here.

EXECUTIVE SUMMARY

The 4th NASA Science Mission Directorate (SMD) Artificial Intelligence (AI) Workshop, held during March 25-27, 2024, in Huntsville, AL, highlighted the significant potential of AI and machine learning (ML) in scientific research and processes. The workshop, supported by the NASA Office of Chief Science Data Officer (OCSDO), emphasized the critical role of foundation models (FMs) and large language models (LLMs) in advancing scientific disciplines. The event brought together domain scientists, computer scientists, AI experts, program managers, program scientists, and industry partners to address key challenges and explore opportunities in applying these advanced technologies.

The primary goals of the workshop were to:

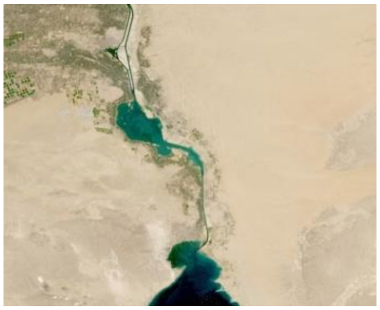
- Identify the most pressing scientific problems suitable for AI, particularly FMs;
- Discuss the essential collaborative strategies and resources needed for developing and applying FMs in science; and
- Share best practices and insights for effectively developing and using FMs to enhance scientific research processes and operational efficiency.

Several key outcomes emerged from the workshop:

- The workshop identified how foundation models have accelerated the application of AI to SMD disciplines and their use cases.
- The workshop’s breakout sessions helped develop a list of critical challenges in scientific research that could be addressed using FMs.
- Participants developed skills in utilizing FMs and LLMs for various scientific applications through hands-on sessions and expert-led presentations.
- The workshop facilitated the building of a robust community of AI and science professionals, promoting partnerships and collaborative projects.
- The workshop uncovered several emerging research opportunities, offering insights and potential focus areas for prioritization.

There was a consensus that science foundation models will significantly advance NASA science by addressing the challenges of large data volumes, enabling efficient analysis, and reducing the resource demands associated with developing AI applications, thus benefiting the entire community.

FOUNDATION MODELS



Harmonized Landsat Sentinel-2 (HLS) image of irrigated agricultural field near Sadat City. Prithvi-Geospatial Foundation Model was built using HLS dataset. Credit NASA IMPACT.

Artificial intelligence (AI) models can learn from millions of examples to solve complex problems, which is key to supervised learning approach. Traditionally, building these systems has required significant time and data. However, the foundation model approach is changing this landscape. Foundation models are trained on vast amounts of unlabeled data and can be adapted for various tasks with minimal fine-tuning, making AI more accessible and efficient.

What Are Foundation Models?

Foundation models are a new class of AI models designed to handle multiple tasks. Instead of being trained for a specific task with a labeled dataset, they learn from a broad range of data without explicit labeling. This allows them to apply general knowledge to specific tasks with minimal additional training.

How Do They Work?

Foundation models use two main techniques:

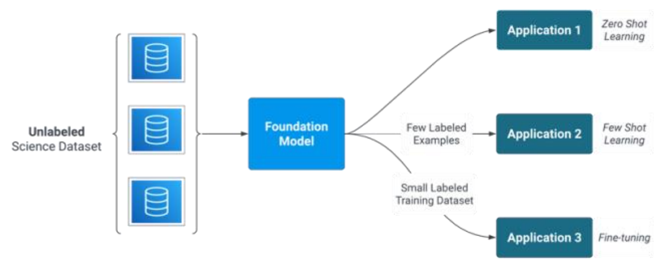


Figure 1. Foundation model approach

1. **Self-Supervised Learning:**
   This involves training the model on large datasets without labeled examples. The model learns to predict parts of the data from other parts, effectively teaching itself.
2. **Transfer Learning:**
   Once trained, the model can transfer its knowledge to new tasks. For example, a model trained on general text can be fine-tuned to understand medical documents with a smaller, specific dataset.

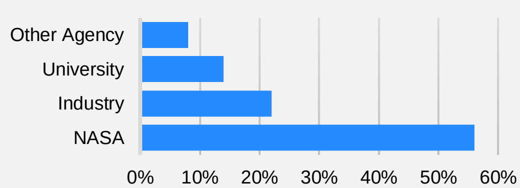
WORKSHOP STRUCTURE

The workshop was structured over three days, each focusing on different aspects of AI and FMs in scientific research.

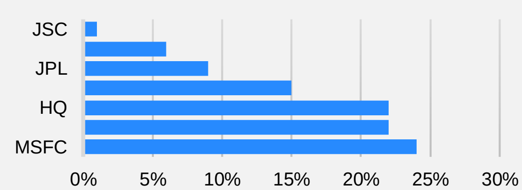
- Day 1 provided an overview of the current AI landscape, including keynote addresses, panel discussions on AI ethics and governance, and hands-on sessions on the applications and best practices of using LLMs for science.
- Day 2 focused on a deep dive into FMs for science, explored new areas and opportunities, hands-on exercises in fine-tuning FMs for scientific applications, and consensus-driven breakout sessions.
- Day 3 focused on summarizing breakout group findings, planning future directions, and concluding remarks.

152	20	06	05
PARTICIPANTS	PRESENTATIONS	BREAKOUT SESSIONS	HANDS-ON ACTIVITIES

ORGANIZATIONAL AFFILIATION TYPE



NASA CENTER REPRESENTATION



Keynote Addresses & Panel Discussions

The first day provided an extensive overview of the current AI landscape, featuring keynote addresses, and AI ethics and governance panel discussion. Full descriptions of each discussion can be found in the Appendix.

Panel Insights

There was a consensus on the panel that best practices must be considered from both a human perspective (e.g., fairness and equity) and a scientific perspective (e.g., robustness, reproducibility). A convergence of open science and open AI was discussed as open science was seen as supporting transparent and ethical AI. Other insights were that:

- Innovation should move at the speed of trust;
- Foundation models should enhance our understanding of scientific physical processes, not just serve to accelerate tasks;
- Ethics also encompasses education of the users as to exactly what foundation models are;
- Speed of AI development outpaces governance and policy development;
- Increasing system complexity can lead to difficulty in evaluating the ethical performance of the system; and
- There is a need to modify scientific ethical standards for the use of FMs in scientific research.

The broader conversation considered concerns about the trustworthiness of outputs and biases and that addressing biases can lead to challenges in trustworthiness. Several questions were raised for further consideration.

Group photo of workshop participants

- What are methods of questioning a FM and its output?
- How can ethical questions be defined as quantitative measures?
- What does trustworthiness in FMs mean?
- What guardrails should we put in place to keep FM trustworthy?

Hands-on Activities

Two hands-on activities were conducted during the workshop: (i) Day 1: best practices on using the LLMs for scientific applications and processes and (ii) Day 2: fine-tune the Prithvi 100M geospatial foundation model for down-stream tasks. The goal of the hands-on activities was to build capacity to properly use large AI models while adhering to open science principles.

The first hands-on activities centered on using LLMs for three scenarios: prompt patterns for science (for any audience), application development and deployment (for app developers), and fine-tuning (for ML developers). Participants explored technologies like LangFlow, retrieval-augmented generation, and ReAct. They applied these to search open science repositories, create chatbots with minimal configuration, query Earth science or astronomical datasets, curate datasets, and enhance access to resources such as environmental justice data.

The second hands-on session focused on fine-tuning Prithvi for burn scar detection. Participants utilized Geocroissant, a geospatial version of Croissant that provides a standardized metadata format for ML-ready data. They followed the Prithvi fine-tuning demonstration on their devices and deployed their fine-tuned models.

Foundation Models for Scientific Applications

The second day was dedicated to exploring FMs in-depth for scientific applications. Several key presentations provided insights into the development and application of these models. Full descriptions of each discussion can be found in the Appendix.

Breakout Sessions

Workshop participants engaged in five different breakout sessions. The session topics were selected based on participants' interests. Participants engaged in comprehensive discussions, sharing insights and formulating strategies for the future.

Earth and remote sensing applications session highlighted FMs role in data assimilation, observational analysis, and productivity for Earth scientists. The findings included emphasizing AI model architectures for multi-modal data and defining benchmarks for physics-aware models.

Enhancing data infrastructure and uncertainty quantification session focused on open data infrastructure, standardized preprocessing, and effective uncertainty quantification. The session discussed challenges related to data accessibility, documentation, quality, and maintenance; and emphasized the needs of establishing working groups for standardization, exploring public-private partnerships for data sharing, and researching methods to incorporate uncertainty.

The space applications session considered several focus questions, including adapting FMs for diverse spectral and image data, implementing federated learning, and leveraging digital twin technologies. The applications identified included Earth, Moon, and Mars FMs, space radiation prediction, and digital twin astronaut models.

The downstream tasks session reiterated that FM is in its early stages, requiring new tools, infrastructure, and resources to support diverse user groups, including FM developers, scientist FM users, and applied science users. For scientists, key workflow elements are discovery, selection, and usability. Several concepts were suggested guided by FAIR principles, benchmarking, research results, and comprehensive documentation to aid in the appropriate use of FMs. Incentives for FM producers to prioritize accessibility, user support, and community building were emphasized.

The architecture and systems session explored advancements to enhance FMs. Key points included expanding the FM ecosystem by involving universities, especially MSIs, and fostering collaboration with interagency partners. Suggestions also included initiating pilot projects and implementing federated learning.

CONCLUSIONS AND FUTURE DIRECTIONS

The workshop concluded with a consensus on prioritizing foundation model development for NASA science. In addition, participants advocated for community-led governance of open-source models to ensure effective and immediate sharing and access. The necessity of standardization in model development and integrating scientific questions with appropriate models was emphasized. Standards for interoperability were also brought up as critical for reusing models for different scientific applications.

Several opportunities for foundation models and an emphasis on addressing existing challenges in advancing AI for NASA science were identified. These include:

- Prioritizing future foundation models across science domains;
- Continuing to evaluate and build applications using LLMs to enhance science research and processes;
- Developing new tools and infrastructure built around scientists' workflows to support the use of FMs;
- Establishing and demonstrating measurable improvements in the application of foundation models for NASA science use cases, including benchmarking;
- Creating a centralized platform for discovering, comparing, and evaluating FMs;
- Building general tools to simplify the use of FMs in scientific research, focusing on low-code/no-code tools for model finetuning and deployment;
- Incentivizing FM producers to make their models accessible and user-friendly, encouraging adoption and proper use;
- Addressing best practices by considering both human aspects, such as fairness and equity, and scientific aspects, such as robustness and reproducibility; and
- Developing guidelines for the appropriate use of foundation models, drawing on lessons learned, to serve as a resource for the community.

The workshop was well-received, with over 76% of respondents indicating it met their expectations, highlighting its effectiveness in fostering collaboration and advancing the integration of AI and ML in scientific research. This success also reflects the group's openness to diverse perspectives, as the presence of all SMD divisions created an inclusive and open-minded community where participants felt comfortable sharing their opinions.

Moving forward, the insights and findings from the workshop will inform the development of strategies for the responsible and effective integration of AI into SMD. Continued collaboration, innovation, and standardization will be key to unlocking the full potential of AI and FMs in advancing scientific research.

ACKNOWLEDGMENTS

Workshop Lead:

Manil Maskey (MSFC/HQ/CSDO)

OCSDO Data Science Innovation Board Members

Kevin Murphy
Megan Ansdell
Sylvain Costes
Joel Scott
Roopesh Ojha
Robin Fergason
Quang-viet Nguyn
Marc Kuchner
Tsengdar Lee
Carolyn Mercer
Chelle Gentemann
Alessandra Aloisi
Alan Zide
Rebekah Dawson-Rigas
Rachel Paseka
Sanaz Vahidinia
Steve Crawford
Andrew Mitchell

Organizing Committee:

Manil Maskey (MSFC/HQ/CSDO)
Mark Carroll (GSFC)
Dan Crichton (JPL)
Suman Sinha Ray (GRC)
Ed Mclarney (LaRC/HQ)
Nikunj Oza (ARC)
Kenton Fisher (JSC)
Elizabeth Fancher (MSFC)

Steering Committee:

Rahul Ramachandran (MSFC)
Sylvain Costes (ARC)
Daniel Duffy (GSFC)
Megan Ansdell (HQ)
Madhulika Guhathakurta (HQ)
Roopesh Ojha (HQ)
Sanaz Vahidinia (HQ)

Workshop Sponsor: Office of the Chief Science Data Officer, Science Mission Directorate, NASA

Report Formatting:

Lauren Perkins (MSFC)

APPENDIX

Descriptions of Keynote Addresses & Panel Discussions

Opening Keynote Address: NASA's Chief Science Data Officer, Kevin Murphy, delivered a keynote presentation on Data and Open Science Strategy that emphasized the transition from AI for experts to AI for everyone. He highlighted the potential of AI to transform every step of the scientific research lifecycle, emphasizing the importance of maximizing AI's benefits while managing associated risks. Murphy pointed out that AI could accelerate the scientific discovery process and enhance the data lifecycle, thus significantly benefiting scientific research.

Collaboration and Innovation: Saleem Hussain from IBM Research discussed the AI Alliance's role in fostering collaboration. The AI Alliance brings together over 80 institutions, including leading universities, startups, big enterprises, and scientific institutions, to accelerate open innovation and technology development. Hussain emphasized that collaboration is key to driving forward AI advancements.

AI Advancements in Geospatial Science: Dr. Dalton Lunga from Department of Energy's Oak Ridge National Lab highlighted the Center for AI Security Research's focus on risks, threats, control, robustness, and reliability. He stressed the importance of understanding performance and risk-benefit trade-offs in the context of human-AI alignment, particularly in geospatial science and human security.

National AI Research Resource (NAIRR): National Science Foundation's Katie Antypas presented on the NAIRR, which aims to connect research and education communities to AI advancement resources. She noted that many communities lack access to AI research resources, which are often concentrated in the largest companies and institutions. Antypas emphasized the need to increase diversity in AI talent and coordinate access to AI resources, focusing on trustworthy, transparent, and responsible AI.

Foundation Models for Fundamental Physics: Dr. Mariel Pettee (Yale University) discussed the development of an FM for fundamental physics, addressing challenges such as erratic LLM embeddings of numbers. She proposed creating a new numerical encoding that uses a single token stretched by the number's magnitude, providing continuity, interpolation, and efficiency.

Large Models for Understanding Complex Systems: Dr. Georgia Gkioxari's (California Institute of Technology) presentation explored leveraging large models to understand complex systems through data. She proposed innovative approaches to time series representation, such as tokenizing continuous time series into a discrete vocabulary that improves model performance across various domains, including climate, neuroscience, and earthquake analysis.

Air Quality Forecasting with FMs: Dr. Jennifer Sleeman (Johns Hopkins University) advanced a path towards using FMs for air quality forecasting through the next-generation ensemble prediction (NGEP) system. This system utilizes deep learning to speed up ensembling and generates 10-day forecasts from small training sets, significantly reducing the time required to train deep learning models.

Generalist Medical AI Systems: Dr. Michael Moor (Stanford University) presented on the potential of generalist medical AI systems, highlighting their flexibility, reusability, and integration of domain knowledge. These systems address challenges such as narrow applicability, lack of domain knowledge, and fixed data modalities by being adaptable and reasoning with medical knowledge across multiple modalities.

GraphCast: Dr. Remi Lam (Google) presented GraphCast, an ML model for medium-range global weather forecasting, which predicts Earth's surface and atmospheric weather up to 10 days ahead at high spatial resolution. GraphCast achieves about one day of improved accuracy compared to high-resolution models, and approximately nine days of accuracy gain in predicting tropical cyclones.

Panel on AI Ethics: A panel on AI ethics in science featured Charles Haley (NASA MSFC), Dr. Douglas Rao (NC State University and NOAA), and Dr. Barbara Thompson (NASA GSFC). The discussion covered ethical considerations on the use of generative AI for science and the role of data stewards in providing guardrails for ethics. This conversation examined the challenges in governing large pre-trained models and recognized that the human use of generative AI tools introduces uncharted territory.

Descriptions of Foundation Models for Scientific Applications

Geospatial FM - Prithvi: Dr. Juan Bernabé-Moreno (IBM Research) discussed Prithvi as an example of transitioning from AI as a tool to AI as a platform. Developed using Harmonized Landsat and Sentinel-2 data, Prithvi employs a masked autoencoder model adapted for time dilation and multispectral bands. Its applications include flood detection and wildfire analysis, demonstrating the model's versatility and effectiveness.

SatVision: Dr. Jordan Caraballo-Vega (NASA GSFC) discussed how science drives deep learning, focusing on improving scientific outcomes efficiently and making FMs more adaptable and accessible. He highlighted that FMs could provide experimental answers more quickly, reduce the data and compute time required, and enhance understanding of input-output relationships. The SatVision model suite supports faster scientific discovery while saving costs.

Application of Foundation Models to Astronomy and Biology: Dr. Ashish Mahabal (California Institute of Technology) discussed the application of foundation models (FMs) in astronomy and biology, highlighting the Segment Anything Model (SAM). This model has learned a general concept of objects, allowing it to generalize to unfamiliar objects and images without additional training. SAM offers versatility and can be deployed “out of the box” for various tasks across multiple datasets.

Does Where your Pre-training Data Come from Matter for Geospatial Foundation Model Performance?: Mirali Purohit (Arizona State University) addressed the impact of pre-training data sources on geospatial foundation model performance. The presentation explored how the spatial distribution of pre-training data affects self-supervised geospatial models, emphasizing the importance of data quantity, diversity, and distribution.

Large Mars Model: Umaa Rebbapragada (NASA JPL) discussed the research goals of large Mars models, which apply machine learning to Martian datasets with the aim to eliminate hours of labeling and the need for custom classifiers and workflows, enabling users to directly upload images for fine-tuning. The project seeks to determine if a custom foundation model can improve performance over current benchmarks.

An AI Manager’s Journey into the Era of Foundation Models: James Parr (Trillium Technologies) discussed an AI manager’s experience in the era of foundation models, highlighting challenges such as team balancing, compute management, scaling, and prioritizing use cases. Key lessons included that task generalization doesn’t equate to regional generalization and that not all model architectures suit every downstream task. He emphasized that building FMs requires learning from multiple modalities and noted that scientific data is more complex than text, as it is statistically dynamic.

Foundation Model Playbook: Dr. Rahul Ramachandran (NASA MSFC) shared lessons learned from working on building three FMs as a playbook on building FMs for science. Key points included working with SMEs and collecting use cases and benchmarks to validate model and methodology, deriving requirements for FMs from use cases, carefully selecting which dataset(s) to use for pretraining, and evaluating FMs relative to the relevant use cases. The presentation also stressed the need for a diverse group to support FM development process that includes AI researchers, ML engineers, data engineers, infrastructure engineers, and science experts.