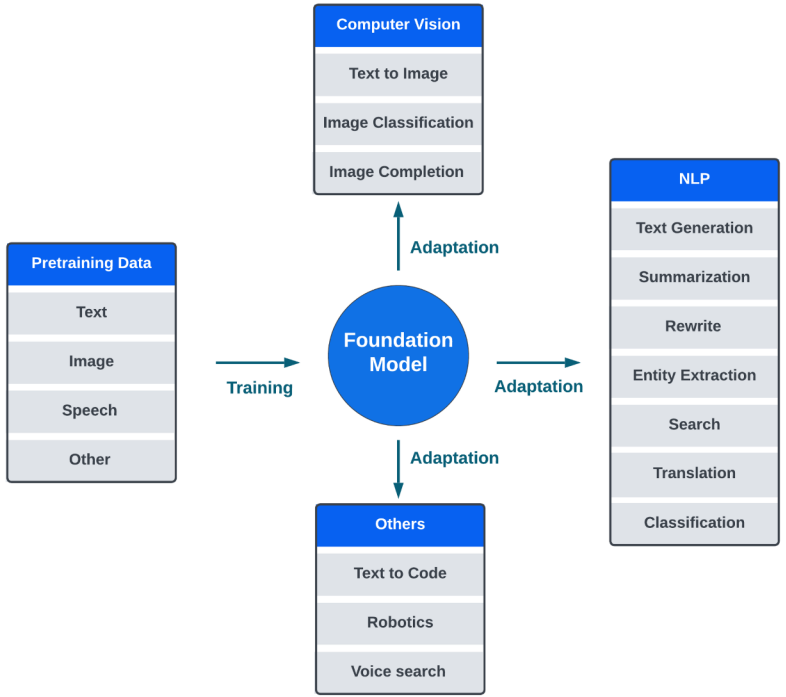
AI Foundation Models for Science: An Open Collaborative Initiative

Michael M Little

NASA IMPACT
NASA Marshall Space Flight Center
NASA Goddard Spaceflight Center
13 March 2024

What are Foundation Models?

- Large-scale AI models pre-trained on vast amounts of data, serving as a starting point for fine-tuning on specific tasks.
- Unlike traditional models, FMs are pre-trained on general data and then adapted to specialized tasks.
- Pre-training captures broad knowledge, allowing for versatility across multiple domains.



Foundation models: training and adaptability

Advantages of Foundation Models

- Reduces the need for extensive task-specific data and training time
- A single foundation model can be fine-tuned for a wide range of generally related applications
- Foundation models often achieve state-of-the-art performance on various tasks, even with limited labeled data
- Fast inference on a laptop in the field

Need for Collaboration in AI for Science

Complexity of the Problem and Vastness of Scientific Data

- Complex science problems by nature require interdisciplinary teams
- Volume and diversity of data in scientific fields require diverse expertise

Limitations of Individual Research Groups or Institutions

- Resources
- Not realistic to possess diverse skill sets and perspectives
- Expertise in various AI subfields

Pooling downstream use cases (labeled datasets/benchmarks) helps develop FMs that has been validated by different groups using variety of use cases

Our Open Collaborative Approach

- Encourage participation from diverse groups, ensuring a wide range of perspectives in research.
- Engage Key Stakeholders:
  - Science experts dedicated to advancing knowledge in their respective fields.
  - Universities, research labs, and organizations that provide the infrastructure and support for research.
  - Tech Companies that offer technological solutions, platforms, and resources essential for modern research.
- Grounded in Open Science Principles:
  - Ensure that research is conducted transparently and that findings are shared openly with the community.
  - Promote reproducibility by making methodologies and data accessible.
  - Promote data sharing, reducing redundancy in data collection efforts.

Foundation Models in Disaster Management

- Planning
  - Land Surface change detection – detect precursors or enablers
- Active
  - Provide low latency model output in the field
- Post-event evaluation
  - Burn scars
  - Flood extent

Prithvi100M
Earth Surface Foundation Model

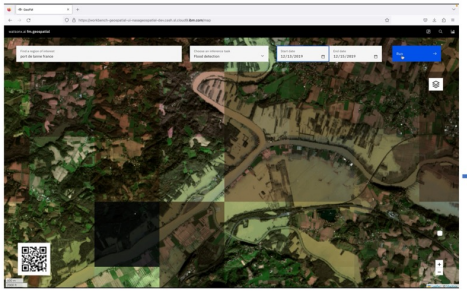
Objective:
- Segment/Classify Earth Surface

Strategy:
- IBM/NASA Partnership
- Redhat OpenShift tooling
- Data: Harmonized Landsat-Sentinel
- Transformer/Encoder only

Status:
- Prithvi100M published in huggingface for CONUS
- Independent Validation in progress
- Partner: FZ Jülich for Global Validation

- Compare downstream apps
- Fidelity: Flood map and burn scar comparisons

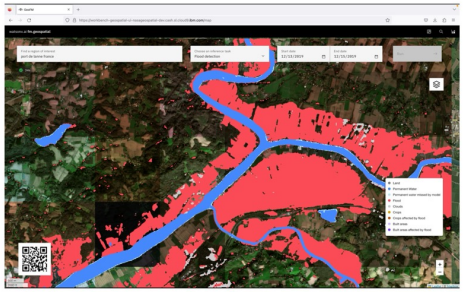
Inference insights by Prithvi – Flood Mapping



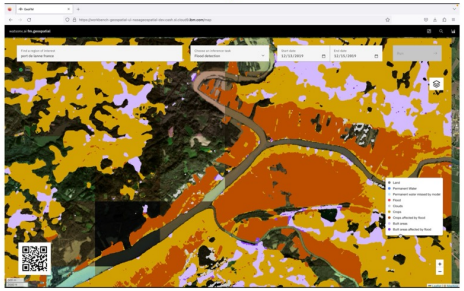
"Prompt": Image(s) (spatial + temporal domains)

Insights: Flood detection

<< Inference>>
(e.g., flood task)



Insights: Flood impact



	IoU (water class)	F1 (water class)	IoU	F1 score	Accuracy
Baseline [44]	24.21	–	–	–	–
U-Net-based SOTA [43]	69.12	81.74	93.85	96.65	96.44
ViT-base [19]	66.52	79.89	90.92	94.97	94.97
Swin [46]	74.75	85.55	92.38	95.90	94.73
Prithvi (not pretrained)	79.23	88.41	94.52	97.09	97.07
Prithvi (pretrained)	**80.10**	**88.95**	**94.78**	**97.23**	**97.23**

Team

NASA MSFC/IMPACT

- Sujit Roy, Kumar Ankur, Christopher Phillips, Iksha Gurung, Muthukumaran Ramasubramanian
- Rahul Ramachandran, Manil Maskey, Pontus Olofsson, Elizabeth Fancher

NASA HQ

- Tsengdar Lee, Kevin Murphy

NASA GSFC

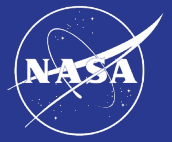
- Dan Duffy, Mike Little

IBM Research

- Johannes Jakubik, Linsong Chu, Paolo Fraccaro, Ranjini Bangalore, Kamal Das, Daiki Kimura, Naomi Simumba, Daniela Szwarcman, Michal Muszynski, Carlos Gomes, Dario Oliveira, Karthik Mukkavilli, Campbell Watson , Kommy Weldemariam, Bianca Zadrozny, Raghu Ganti, Carlos Costa

Clark University

- Hamed Alemohammad, Steve Li, Michael Cecil, Sam Khallaghi, Denys Godwin, Maryam Ahmadi, Fatemeh Kordi



Thank you.

m.m.little@nasa.gov