Surya FM a foundation model in heliophysics

SOUTHWEST RESEARCH INSTITUTE®

Andrés Muñoz-Jaramillo and the Surya FM team

andres.munoz@swri.org

Funded by NASA's Office of the Chief Science Data Officer (OCSDO)

Vision and leadership: Lika Guhathakurta

Team

Program Scientist: Madhulika Guhathakurta

Project Scientist:
Andrés Muñoz-Jaramillo (SwRI)

Lead Engineer at IMPACT:
Sujit Roy

Lead Engineer at IBM:
Johannes Schmude

Downstream applications:

- Shah Bahauddin (EUV; LASP)
- Daniel da Silva (3D Sun; GSFC)
- Dinesh Hegde (Solar Wind; UoA)
- Jinsu Hong (Flares; GSU)
- Spiridon Kasapis (ARs; PU)
- Ryan McGranaghan (Geo; JPL)
- Chetraj Pandey (Flares; GSU)
- Vishal Upendran (SW; SETI)
- Kang Yang (Segmentation; GSU)

Development:

- Ata A. Asanjan (model training)
- Vishal Gaur (model training)
- Rohit Lal (model training)
- Amy Lin (data processing)
- Kshitiz Mandal (model training)

Development:

- Marcus Freitag (model training)
- Johannes Jakubik (model training)
- Julian Kuehnert (model optimization)
- Theodore van Kessel (model training)

Institutional Management (alphabetical): Berkay Aydin (GSU), Juan Bernabe-Moreno (IBM), Nikolai Pogorielov (UoA), Rahul Ramachandran (IMPACT), Talwinder Singh (GSU), Campbell Watson (IBM).

Why Surya FM?

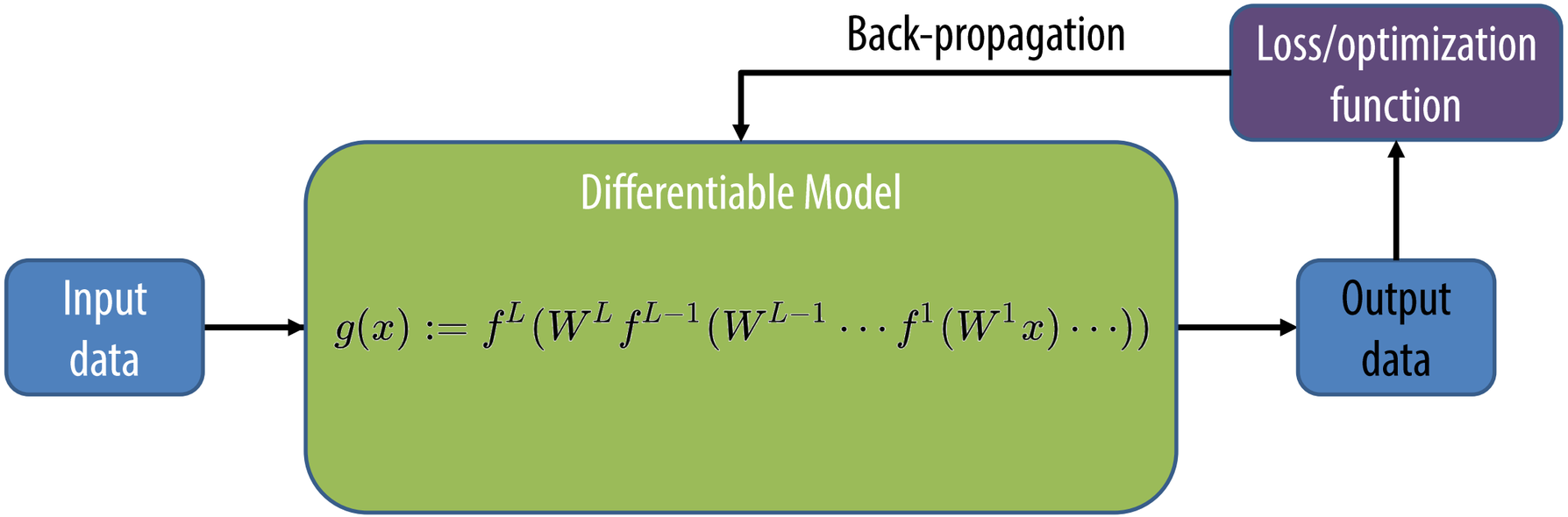
- It is impossible for the community to create complete duplicates of heliophysics data.
- AI, coupled with big data, is expensive and out of reach of most of the heliophysics community.
- Surya FM is an experiment to assess the viability and usefulness of paying for the training of big complex AI models that can be reused extensively by any of SMD's communities.
- Surya FM is an opportunity to build common AI infrastructure that enables SMD communities to take advantage of AI (regardless of AI expertise).

Take home message

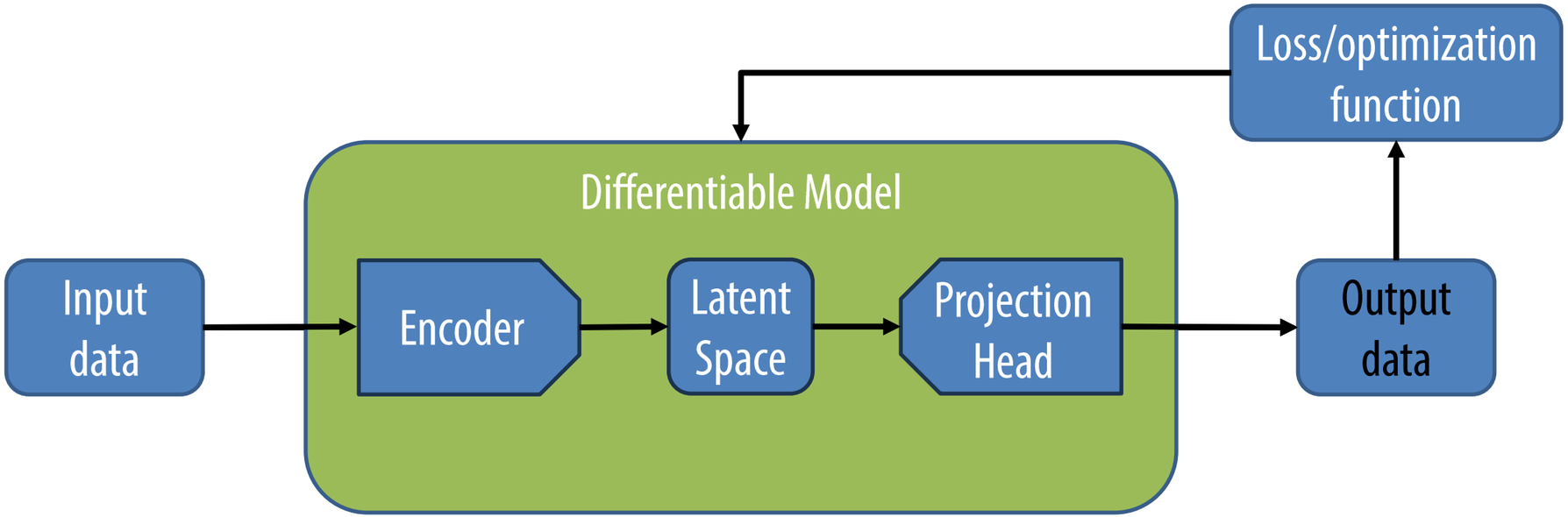
- AI Foundation models (FMs) are models that are trained on vast datasets so that they can be applied across a wide range of use cases.
- Surya FM is pre-trained using SDO data, a transformer architecture, and validated using a range of downstream applications.
- Preliminary results suggest that Surya FM is competitive when finetuned.
- For a scientific community, the value of a FM goes beyond the ability to train a model with lots of data: it's an opportunity to build a shared AI infrastructure.
- We are looking for help shaping and make this model and shared ecosystem the best in can possibly be.

How AI approximates mathematical functions

$$\frac{dC}{da^L} \cdot \frac{da^L}{dz^L} \cdot \frac{dz^L}{da^{L-1}} \cdot \frac{da^{L-1}}{dz^{L-1}} \cdot \frac{dz^{L-1}}{da^{L-2}} \cdot \ldots \cdot \frac{da^1}{dz^1} \cdot \frac{\partial z^1}{\partial x}$$

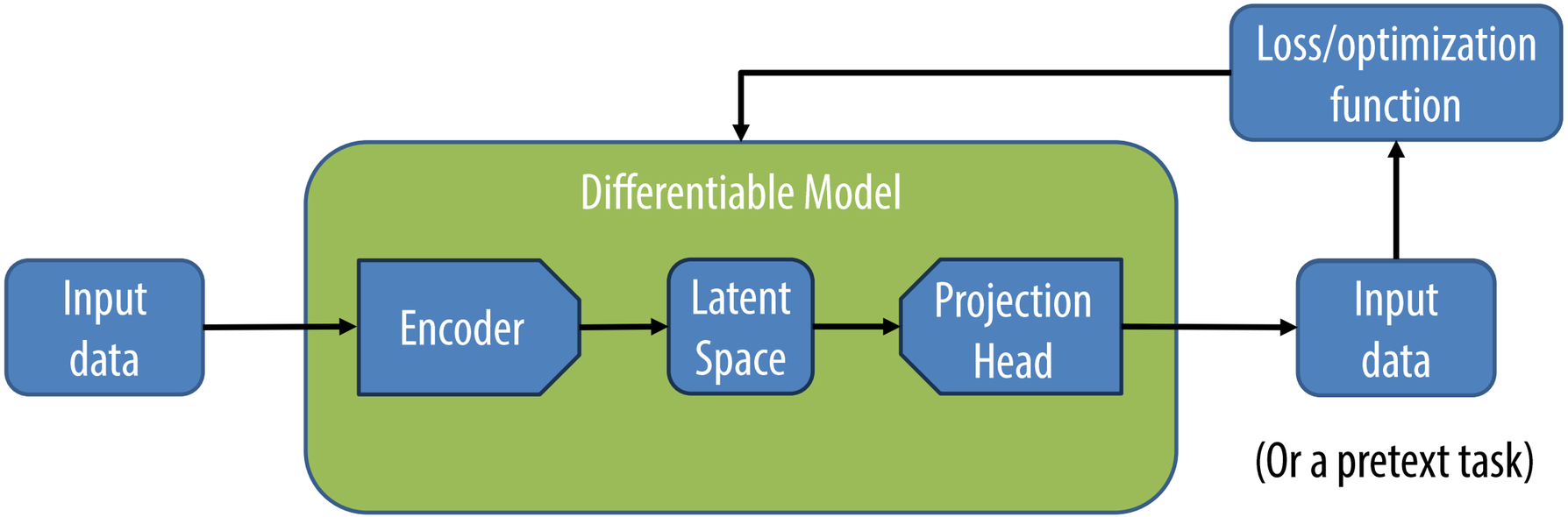


How AI approximates mathematical functions



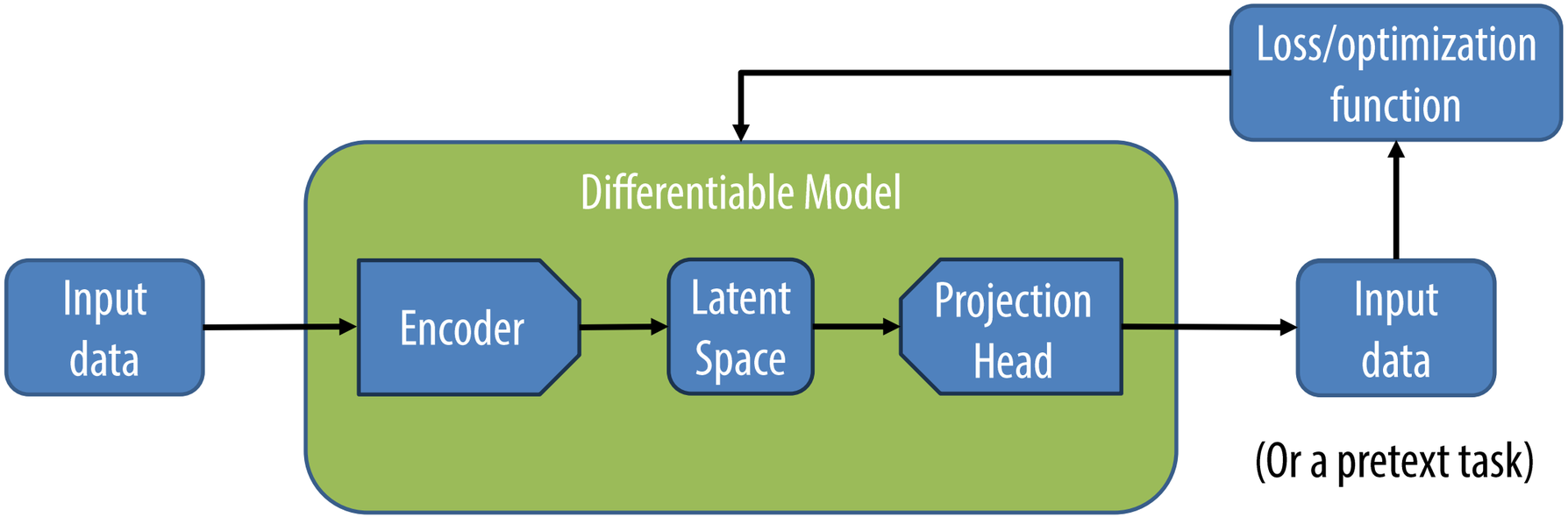
The challenge of a supervised training loop is the need for data annotation

Self-supervision solves data annotation limitations



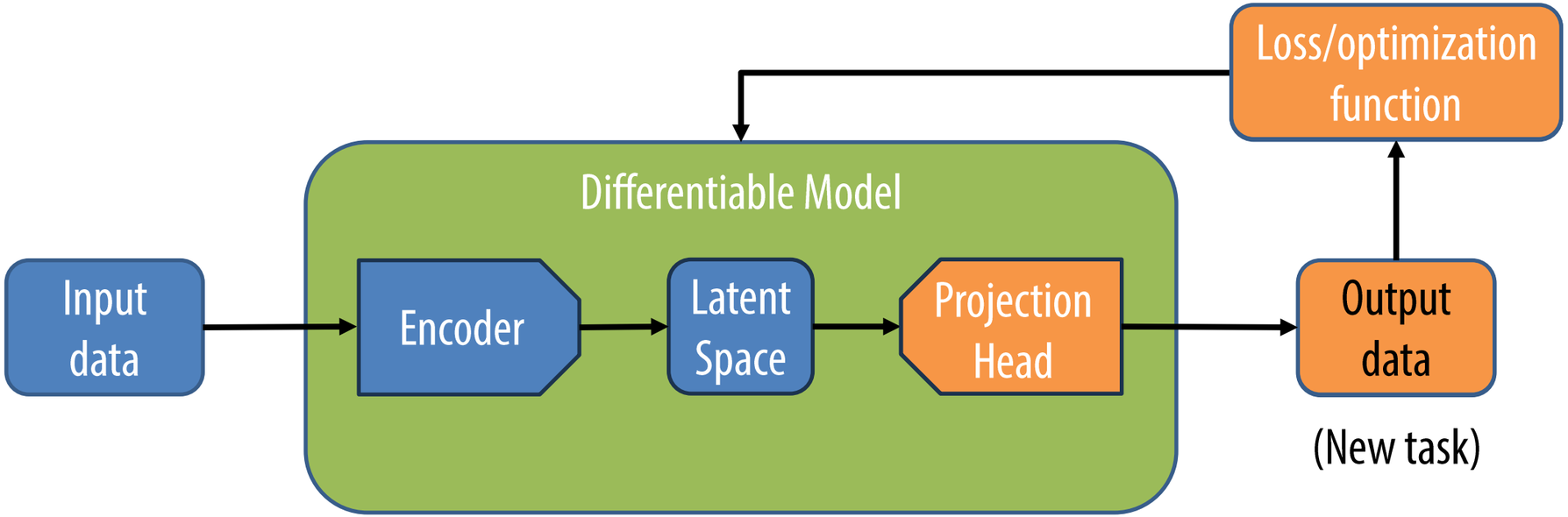
This enables training models on vast amounts of data
(with lots of computational resources)

Pretrained models can be used for many downstream tasks



This enables training models on vast amounts of data
(with lots of computational resources)

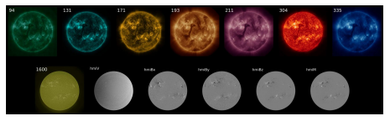
Pretrained models can be used for many downstream tasks



Only the projection head needs to be trained

SuryaBench

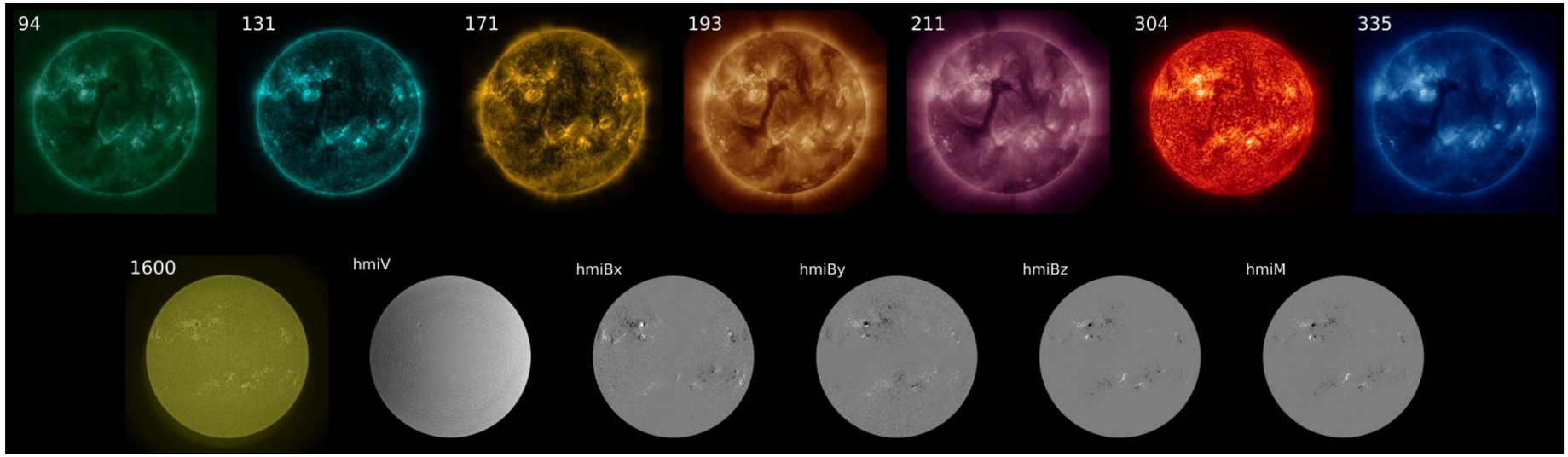
Roy et al. (2025a) https://arxiv.org/abs/2508.14107
Talk by Dinesha Hegde https://vimeo.com/1153957755?fl=pl&fe=ti
Access info https://huggingface.co/collections/nasa-ibm-ai4science/suryabench



Input data

- 8 AIA Channels
- HMI Doppler
- HMI Vector
- 4k x 4k
- 12 minutes cadence
- Full mission

UV – Diagnostics of atmospheric temperature and density

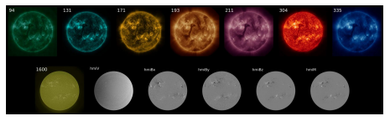


Near-surface plasma velocity

Surface magnetic field components

SuryaBench

Roy et al. (2025a) https://arxiv.org/abs/2508.14107
Talk by Dinesha Hegde https://vimeo.com/1153957755?fl=pl&fe=ti
Access info https://huggingface.co/collections/nasa-ibm-ai4science/suryabench



Input data

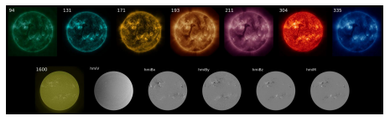
- 8 AIA Channels
- HMI Doppler
- HMI Vector
- 4k x 4k
- 12 minutes cadence
- Full mission

- AI/ML Ready.
- Residing in a public AWS that can be accessed in real time (streaming) to perform analysis and/or train models.

The Surya Foundation Model

Roy et al. (2025b) https://arxiv.org/abs/2508.14112
HAO Seminar https://youtu.be/J27Cxy47n4g?si=m2cBUvDlb5R6flm6
Access info https://github.com/NASA-IMPACT/Surya

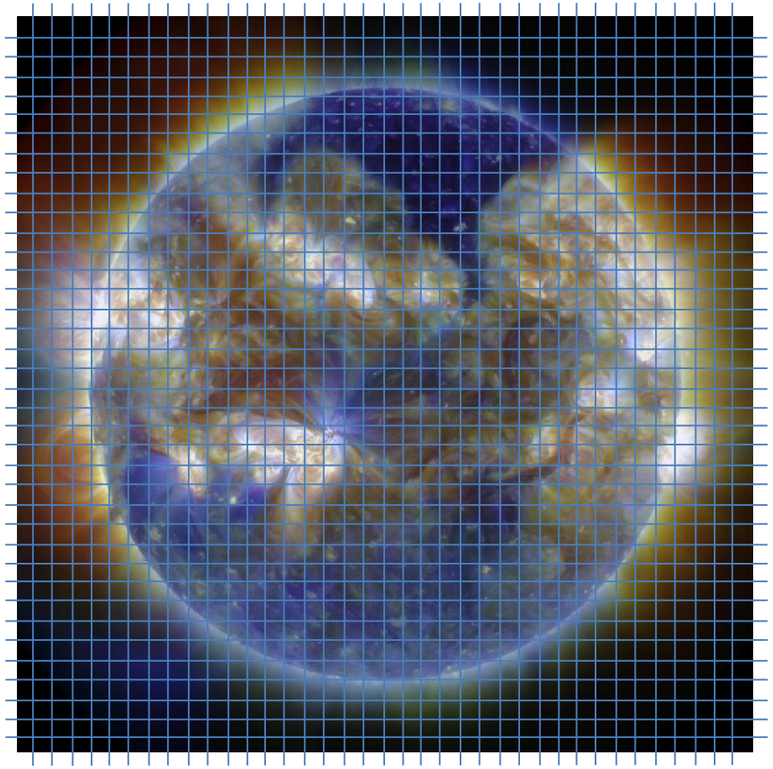
- Large, multi-purpose model pre-trained with vast quantities of data.
- Powered by a visual transformer architecture.



Input data → Encoder → Latent Space → Projection Head

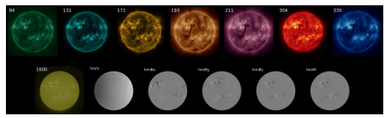
- 8 AIA Channels
- HMI Doppler
- HMI Vector
- 4k x 4k
- 12 minutes cadence
- Full mission

Builds a solar visual language

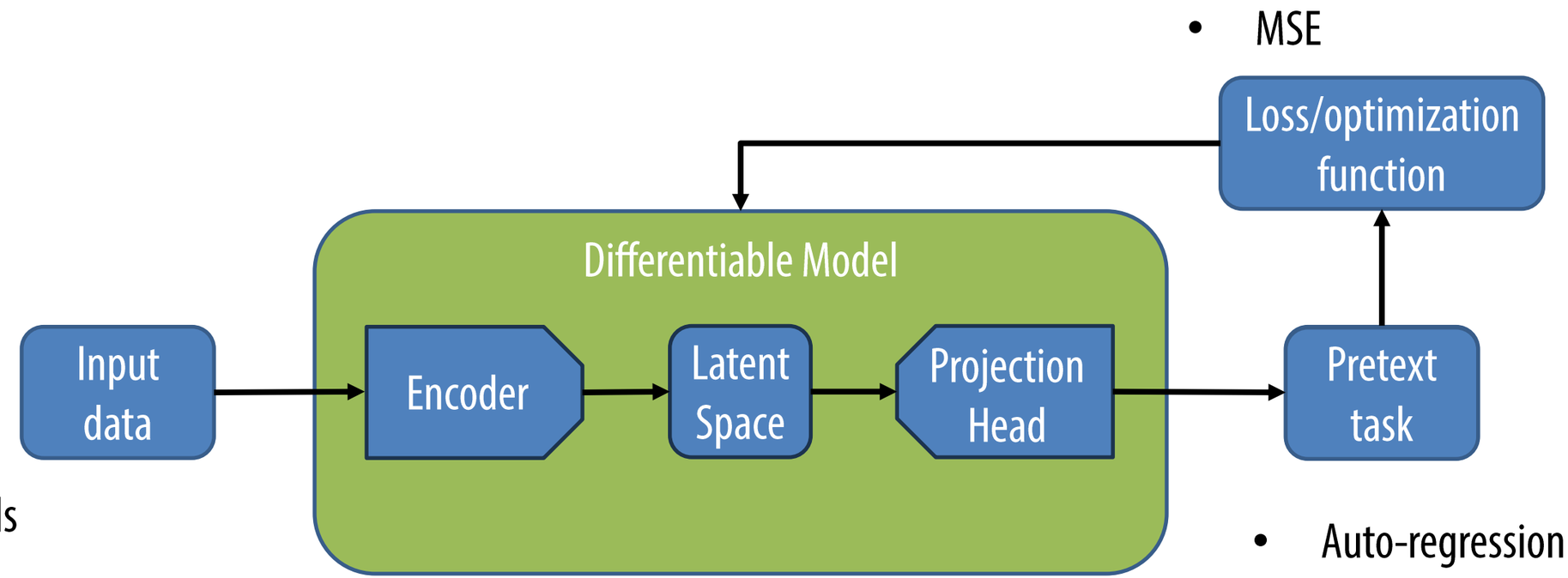


The Surya Foundation Model

Roy et al. (2025b) https://arxiv.org/abs/2508.14112
HAO Seminar https://youtu.be/J27Cxy47n4g?si=m2cBUvDlb5R6flm6
Access info https://github.com/NASA-IMPACT/Surya



- 8 AIA Channels
- HMI Doppler
- HMI Vector
- 4k x 4k
- 12 minutes cadence
- Full mission



- MSE
- Visual Transformer
- Auto-regression (Time advancement)

Highly experimental!

The good news: Downstream application validation

- Downstream applications were also chosen to represent fully the data types used in the heliophysics community:

1D Classification:
- Flare forecast

1D Regression:
- L1 Solar wind forecast
- EUV Spectrum
- K-index forecast

2D Classification:
- Segmentation

2D Regression:
- Time advancement
- Active region forecast

3D Regression
- Coronal field extrapolation
- Magnetosphere perturbation

Blue indicate downstream applications that are directly relevant to geospace.

The good news: Downstream application validation

1D Classification:
- Flare forecast

2D Classification:
- Segmentation

1D Regression:
- L1 Solar wind forecast
- EUV Spectrum

2D Regression:
- Time advancement (zero shot)

Blue indicate downstream applications that are directly relevant to geospace.

The good news: Downstream application validation

Surya is competitive in every downstream application we have tested so far!

Ask me if you are curious about any of them.

1D Classification:

- Flare forecast

2D Classification:

- Segmentation

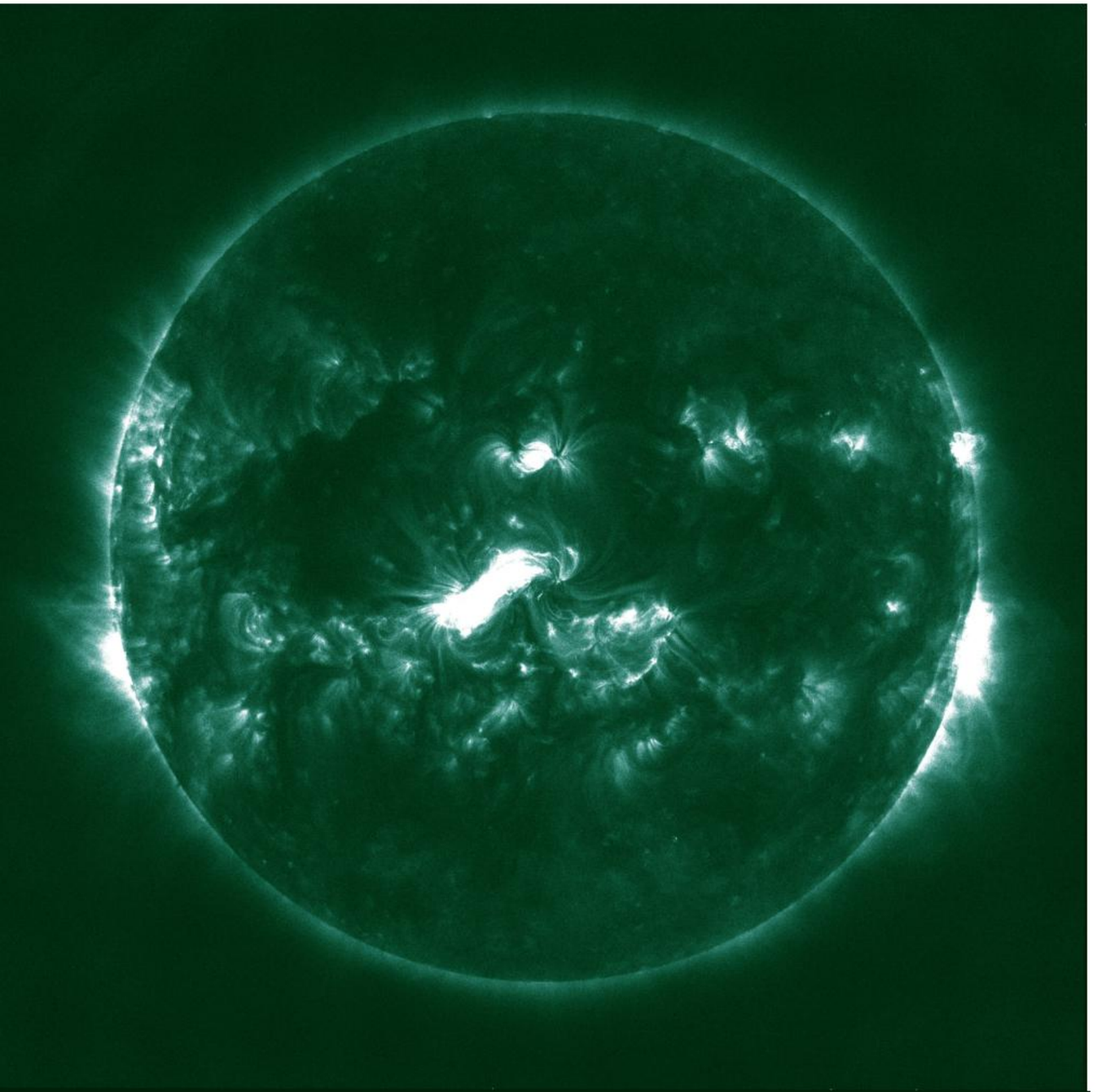
1D Regression:

- L1 Solar wind forecast
- EUV Spectrum

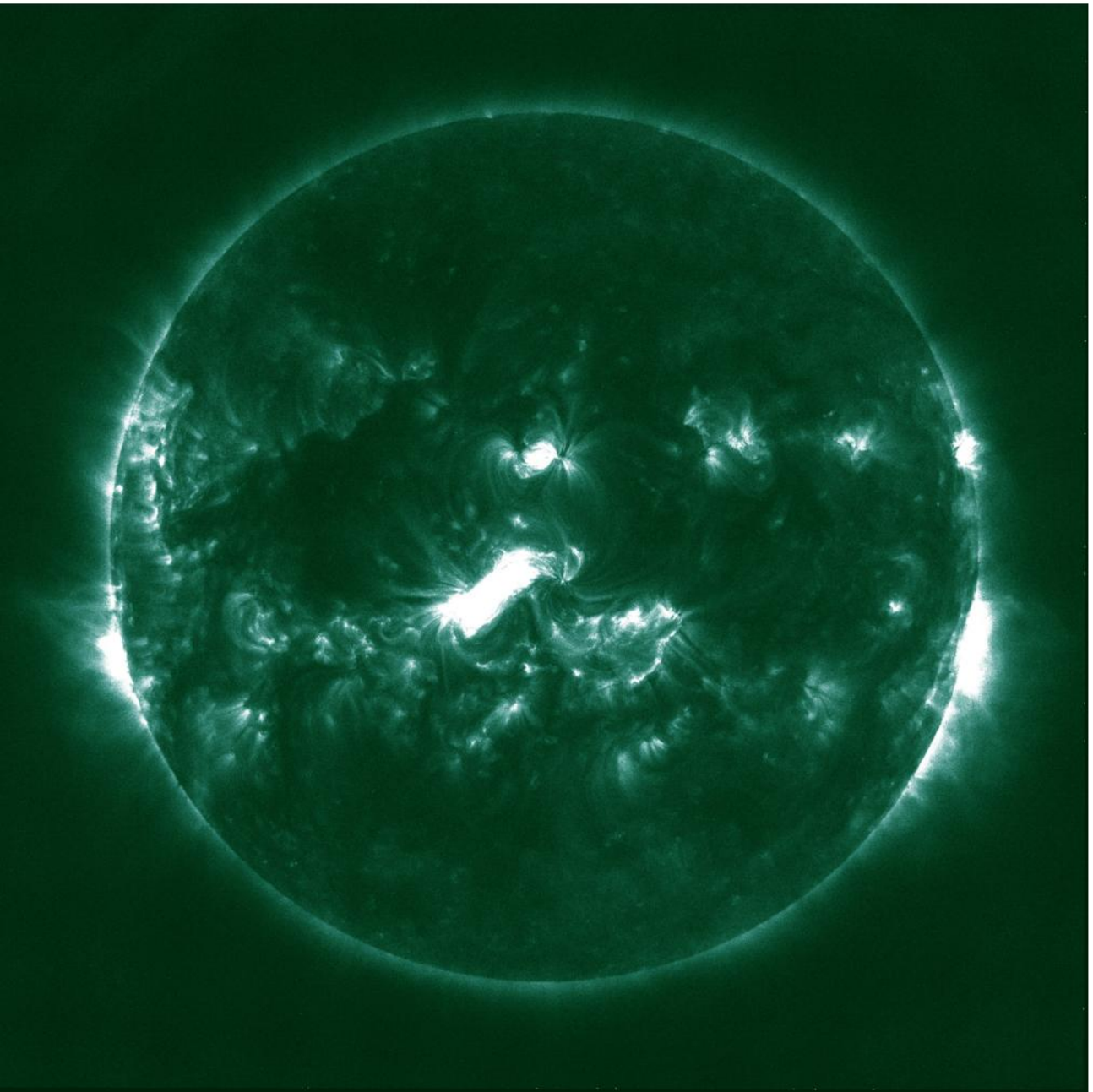
2D Regression:

- Time advancement (zero shot)

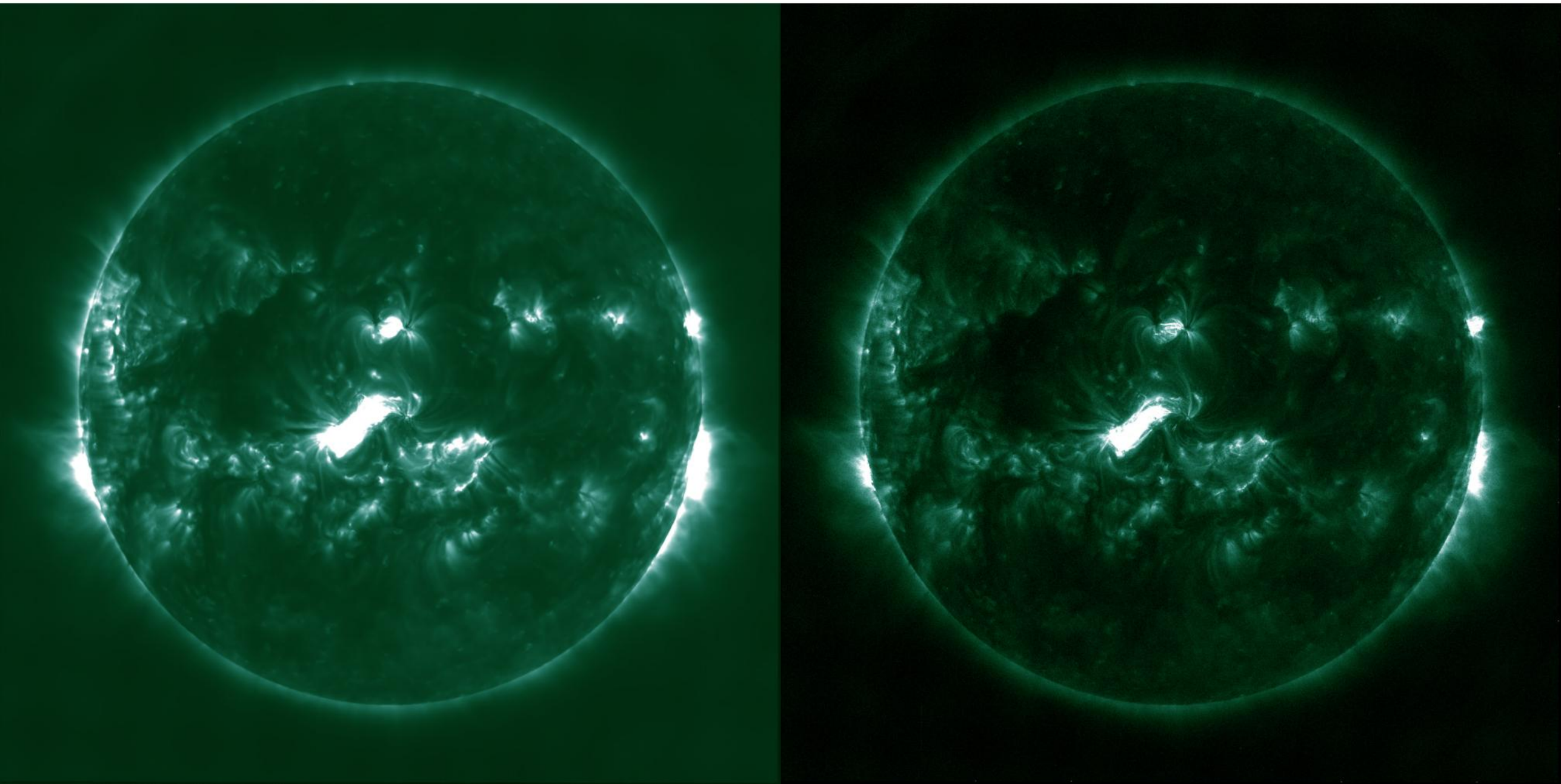
A tantalizing, but brittle result



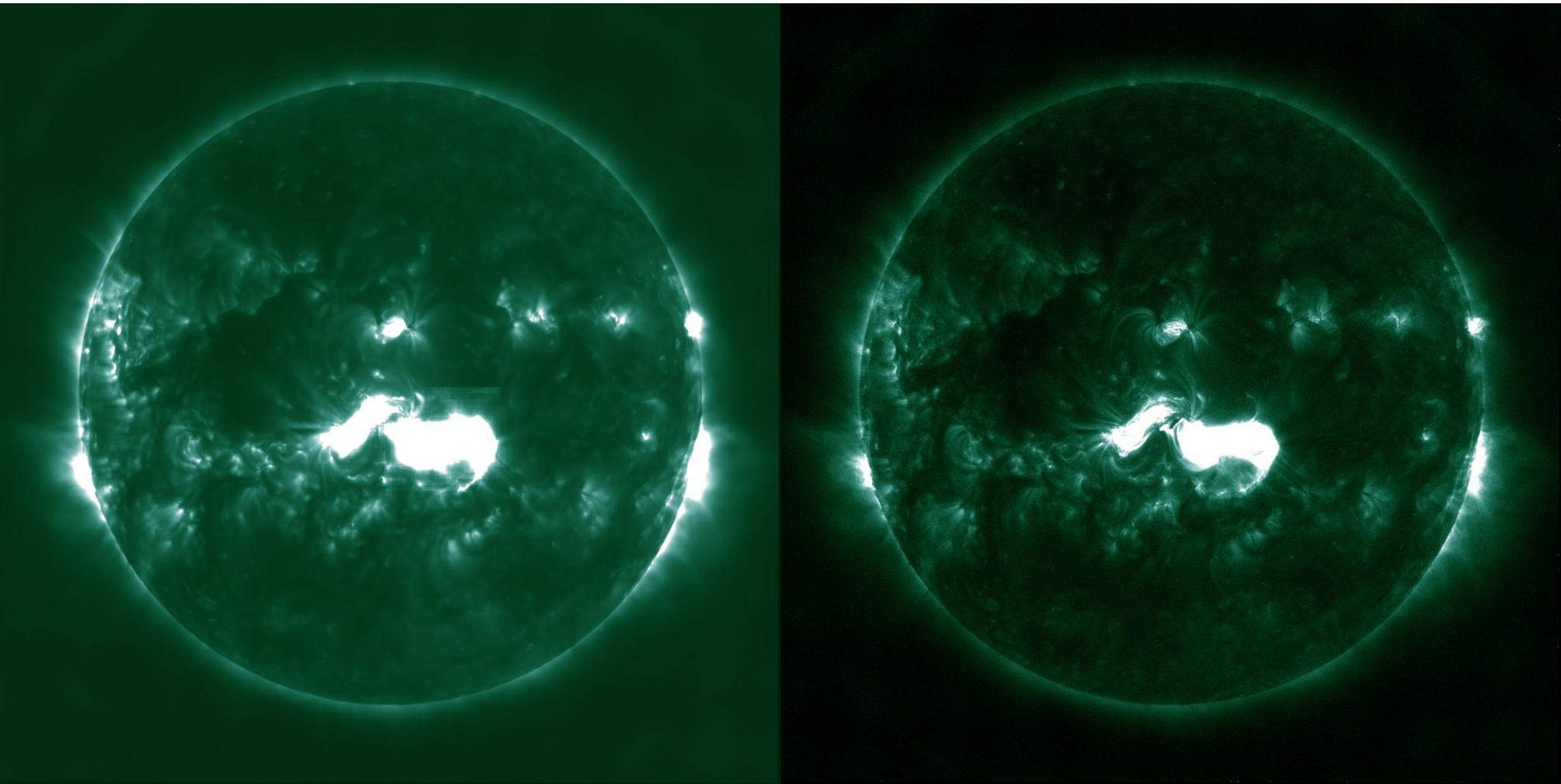
A tantalizing, but brittle result



A tantalizing, but brittle result



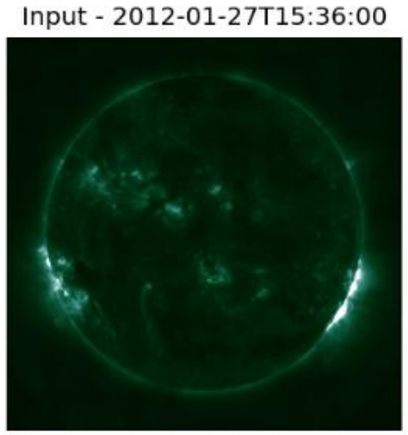
A tantalizing, but brittle result

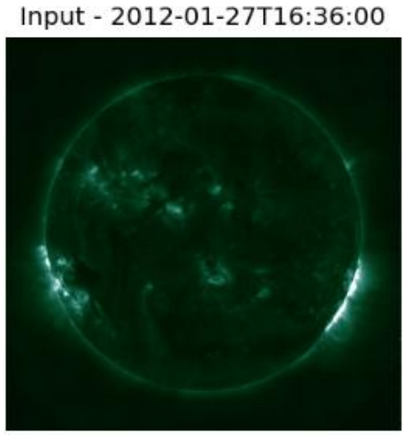


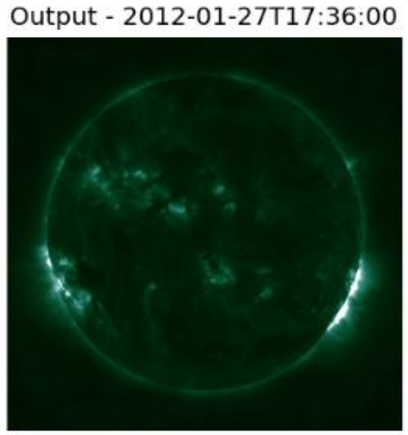
A tantalizing, but brittle result

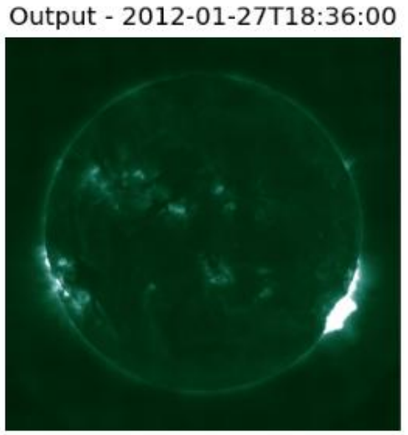
Highly experimental!

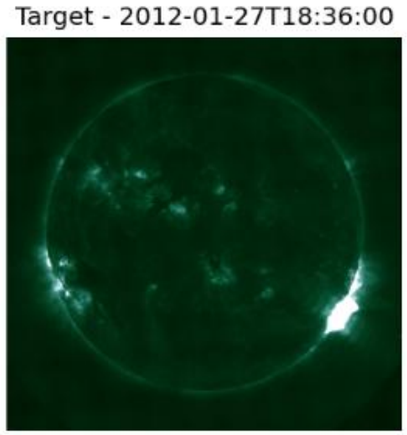
24



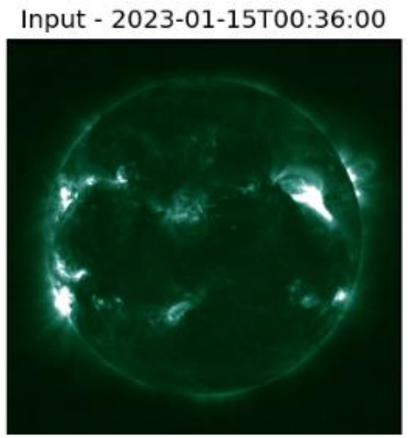


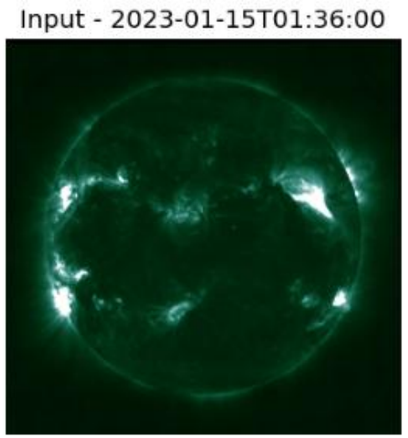


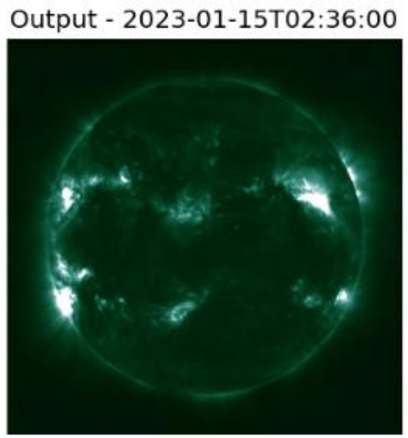


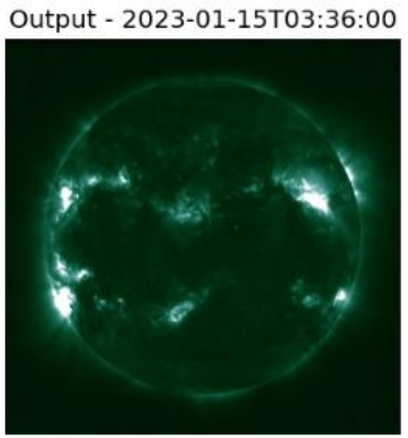


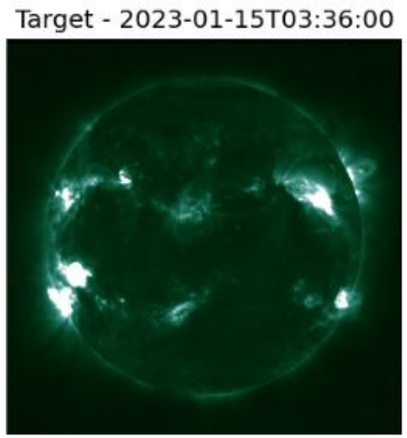
25



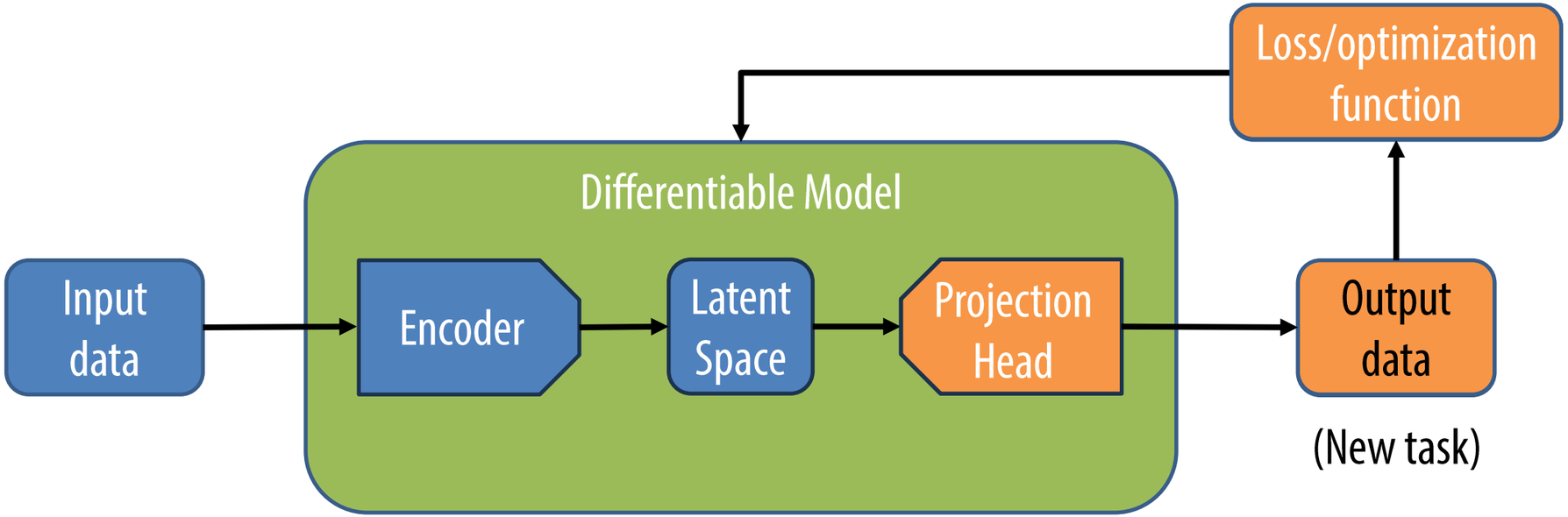








What it takes to reuse a multi-purpose model?

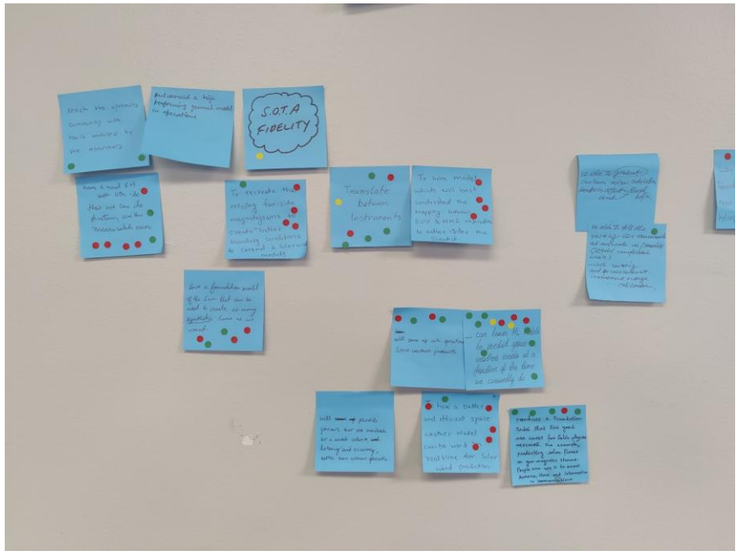


Only the projection head needs to be trained

What it takes to reuse a multi-purpose model?

- Most community members do not have the skills and understanding required to implement and validate AI training pipelines.
- Most community members already fully committed to projects and have very limited time to learn and adopt new technologies.
- Most community members are generally skeptical of the ability of AI to address scientific problems.

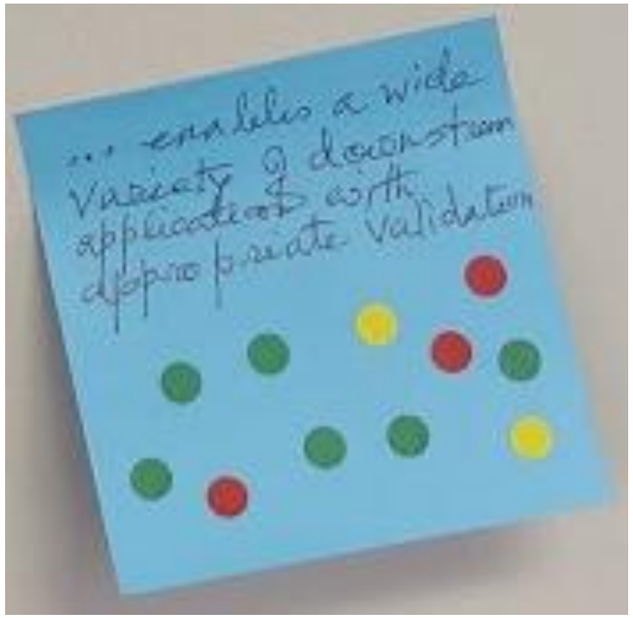
Foundational means more to us than a pretrained model

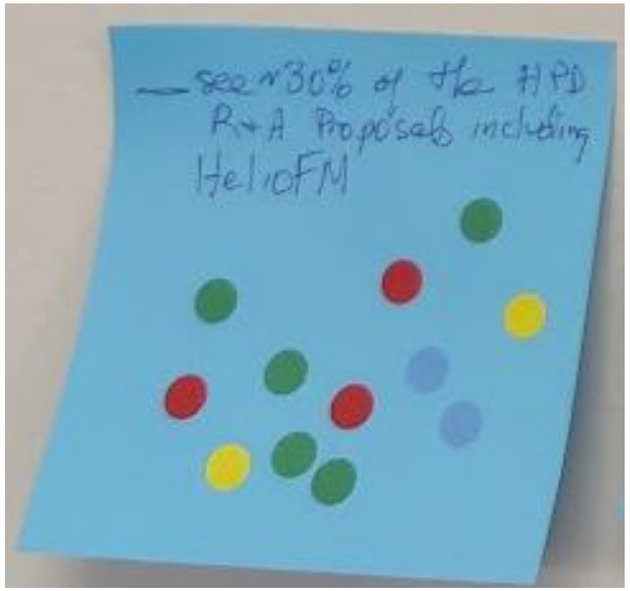


Foundational means more to us than a pretrained model

Technical

Social





To us this means meeting the community where they are.

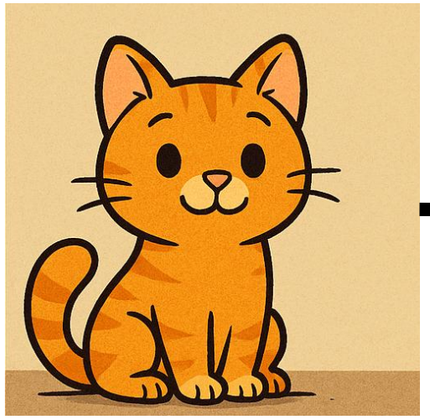
Foundational means more to us than a pretrained model

- Provide educational opportunities to learn to design, implement, and validate scientific of AI pipelines (any career stage and proficiency).
- Help design and implement Surya finetuning applications in a way that directly advances each person's scientific goals.
- Create a wide range of application examples that other scientists can use as inspiration
- Help test the limitations and capabilities of the Surya foundation model beyond the boundaries explored by the Surya team.

Supporting the AI development cycle

- Structure is critical to train all the workforce (all career stages) in the use of AI.
  - AI/ML Concepts and best practices.
  - IDEs and logging tools.
  - AI assistants.
  - Cloud computing.
- The Surya Workshop was an experiment in the creation of an educational experience open to all. 15 projects! 3 days! from zero to hero!

From zero to hero – The Surya Workshop

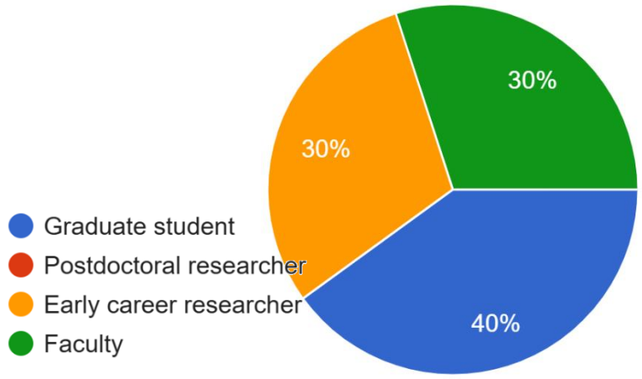


Pytorch Datasets Metrics Lightning modules

Baselines PyTorch Models

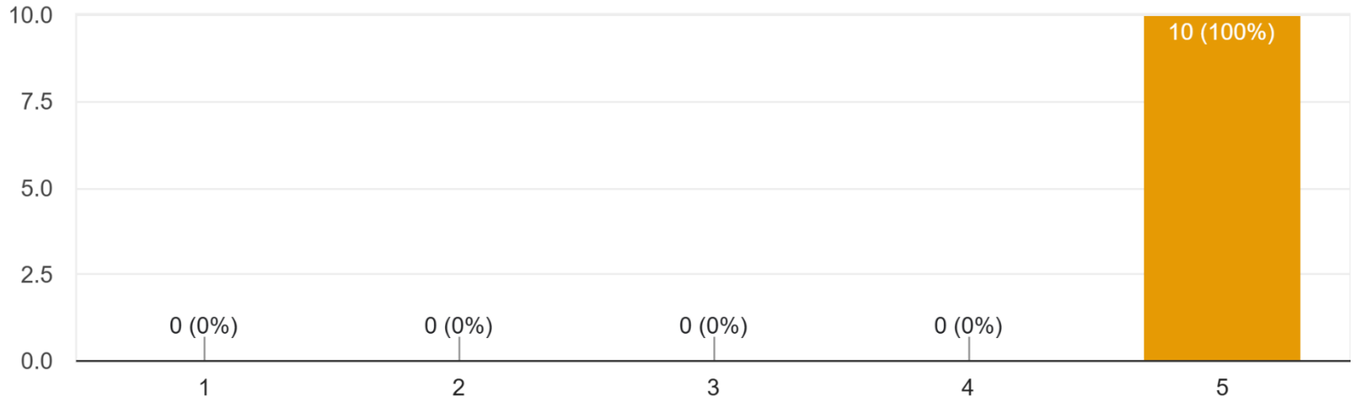
What career stage are you?

10 responses



How likely are you to recommend this workshop to future attendees?

10 responses



From zero to hero – The Surya Workshop

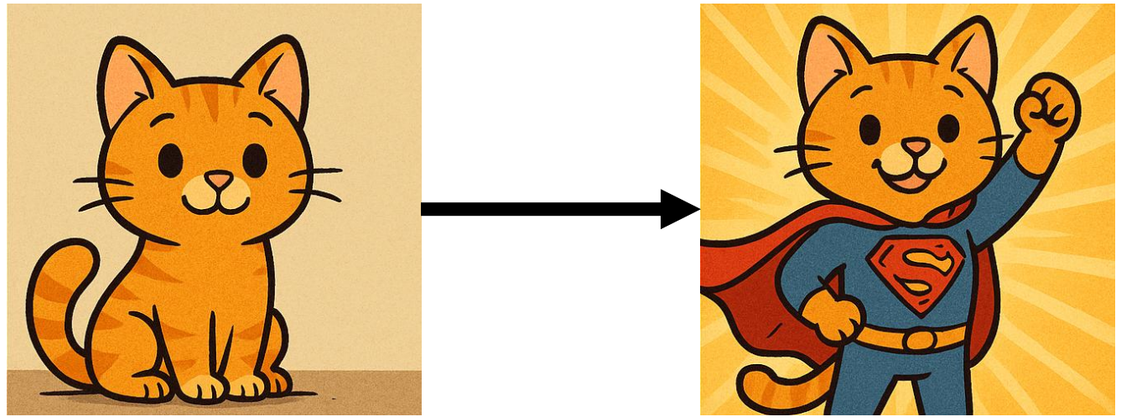
- More than 100 scientists joined the informational component and 30+ scientists joined the hands-on component.
- 13 different applications - 9 different participating institutions.
- Successful hybrid environment.
- Cloud-based infrastructure. Streaming during finetuning works!
- The Surya workshop included researchers that had little to no prior experience with GitHub, PyTorch, AI training pipelines, and online reporting tools.

Joining the process

- Chat with me.

- Join our zero-to-hero workshop in September!
  - Powered by HelioCloud.
  - Better and more streamlined!
  - Hands on fine-tuning exercise (limited availability).
  - Fine tuning presentations.

www.heliofm.org



Take home message

- AI Foundation models (FMs) are models that are trained on vast datasets so that they can be applied across a wide range of use cases.
- Surya FM is pre-trained using SDO data, a transformer architecture, and validated using a range of downstream applications.
- Preliminary results suggest that Surya FM is competitive when finetuned.
- For a scientific community, the value of a FM goes beyond the ability to train a model with lots of data: it's an opportunity to build a shared AI infrastructure.
- We are looking for help shaping and make this model and shared ecosystem the best in can possibly be.

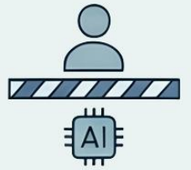
The VOCAL Framework: Elevating Human-AI Collaboration through Domain Vocabulary

MAIN THESIS

Conceptual vocabulary **(domain-specific terms, relationships, and semantic structures), NOT prompting skill, is the primary bottleneck in productive human-AI collaboration.**

Tier 1 (Everyday)
Tier 2 (Academic/Analytical)
Tier 3 (Domain-Specific)

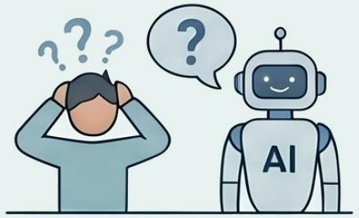
The Three-Tier Challenge.
Productive dialogue requires command of Tier 1, Tier 2, and Tier 3 vocabulary.



The Lexical Interaction Threshold.
A minimum vocabulary floor is required. Below this floor, AI is incomprehensible and learner cannot give direction.

THE 5-LEVEL AI COLLABORATION SCALE

Creative Subordination

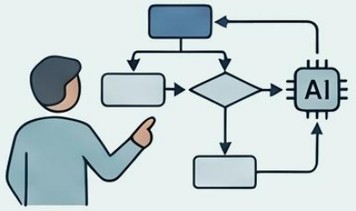


Lacks vocabulary to specify goals or recognize errors. Cognitive surrender; blindly accepts AI output.

Surface Checking

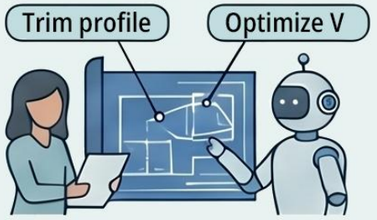
Can spot gross errors (e.g., wrong chart type) but lacks terminology to evaluate deeper choices.

Active Direction



Can decompose tasks and identify wrong approaches, but relies on AI to fill significant knowledge gaps.

Critical Collaboration



Acts as technical peer, catching non-obvious errors and pushing back with specific, domain-justified alternatives.

Equal Partnership

Has full epistemic agency; distinguishes between "better" and "just different," could complete task independently.

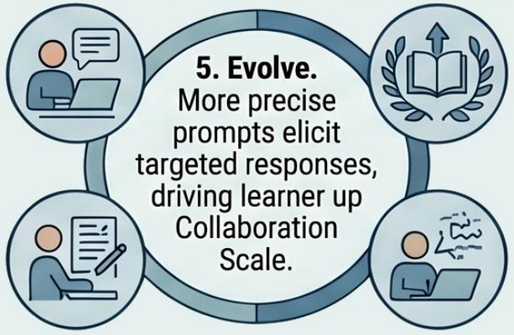
THE DIALOGIC BOOTSTRAPPING CYCLE

1. Prompt.
Learner engages AI using current vocabulary.

2. Introduce.
AI response introduces new terms at "Zone of Proximal Development."

3. Acquire.
Learner identifies and internalizes subset of new terms through context.

4. Refine.
Expanded vocabulary enables learner to write more precise, domain-appropriate prompts.



DESIGN PRINCIPLES FOR CURRICULA

Principle 1: Threshold Priming.
Explicitly teach minimum domain vocabulary BEFORE AI engagement.

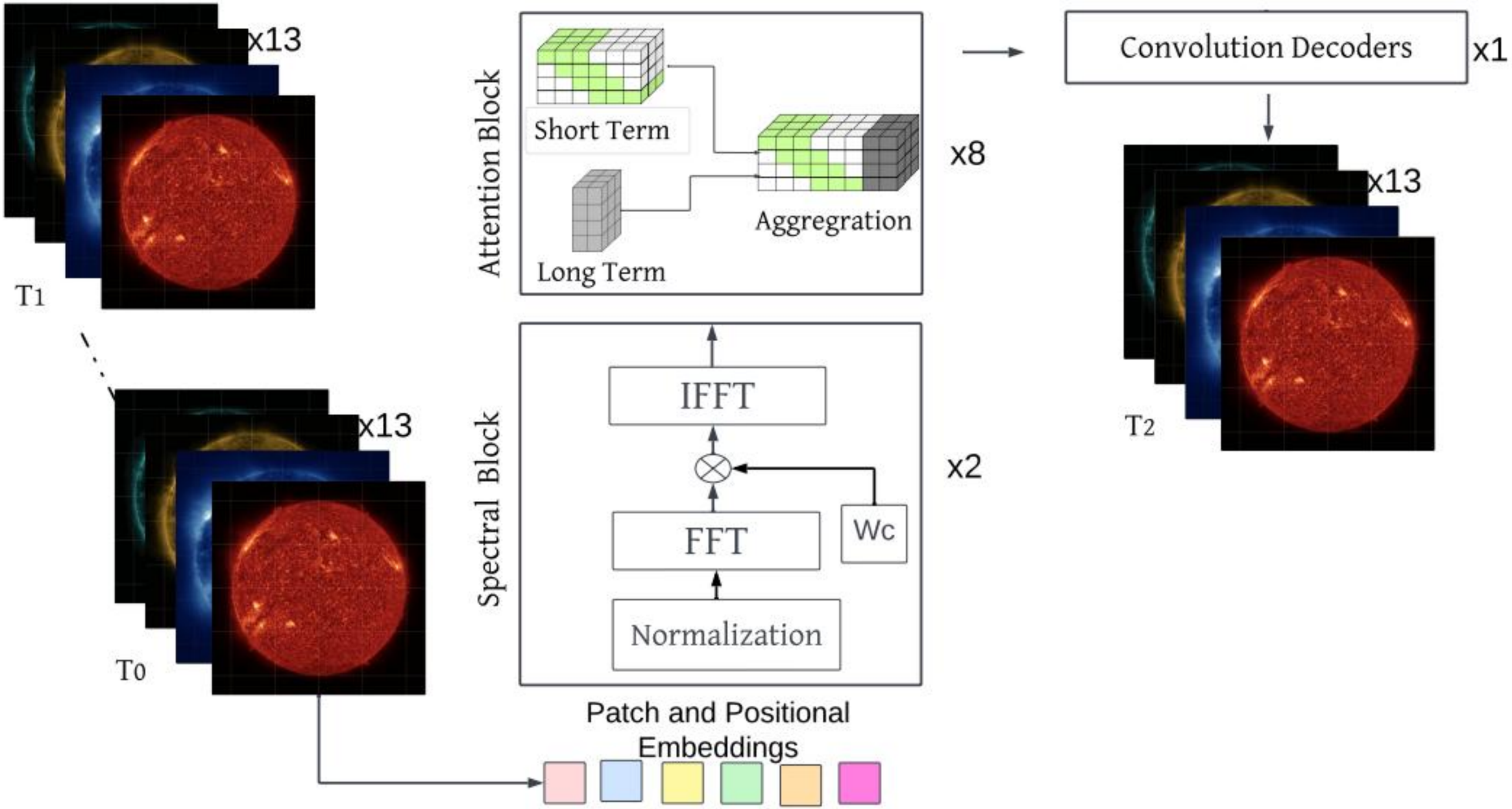
Principle 2: Tiered Sequencing.
Structure tasks to build Tier 2 academic language alongside Tier 3 domain terms.

Principle 3: Embedded Vocabulary.
Design prompts that REQUIRE specific domain terms for task completion.

Principle 4: Evaluative Output.
Require students to produce an artifact proving productive use of new vocabulary.

Principle 5: Explicit Metacognition.
Help students track their own movement on the Collaboration Scale.

Surya's Architecture



Flare forecast

$$f(\mathbf{x}_t) \rightarrow y_t \in \{0, 1\}, \qquad y_t = \begin{cases} 1, & \text{if a strong flare occurs in } [t, t + 24h) \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{TSS} = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}.$$

$$\text{F1} = \frac{2TP}{2TP + FP + FN}.$$

$$\text{HSS} = \frac{2(TP \cdot TN - FP \cdot FN)}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)}$$

TSS: True Skill Score

HSS: Heidke Skill Score

Model	TSS	HSS	F1
AlexNet	0.358	0.398	0.454
ResNet50	0.018	0.028	0.055
Surya	**0.436**	**0.522**	**0.561**

L1 Solar wind forecast

$$y_{t+\Delta t} = f(\mathbf{x}_t), \qquad \Delta t = 4 \text{ days}$$

Table 6: Solar Wind Prediction results comparing baseline models with Surya

Model	**RMSE** ($\mathrm{km\ s^{-1}}$)	**MAE**($\mathrm{km\ s^{-1}}$)	**Validation MSE** ($\mathrm{km^2 s^{-2}}$)
AlexNet	118.6	95.7 km	13839.49
ResNet50	93.76	74.65	8547.924
Surya	**75.92**	**58.06**	**5698.62**

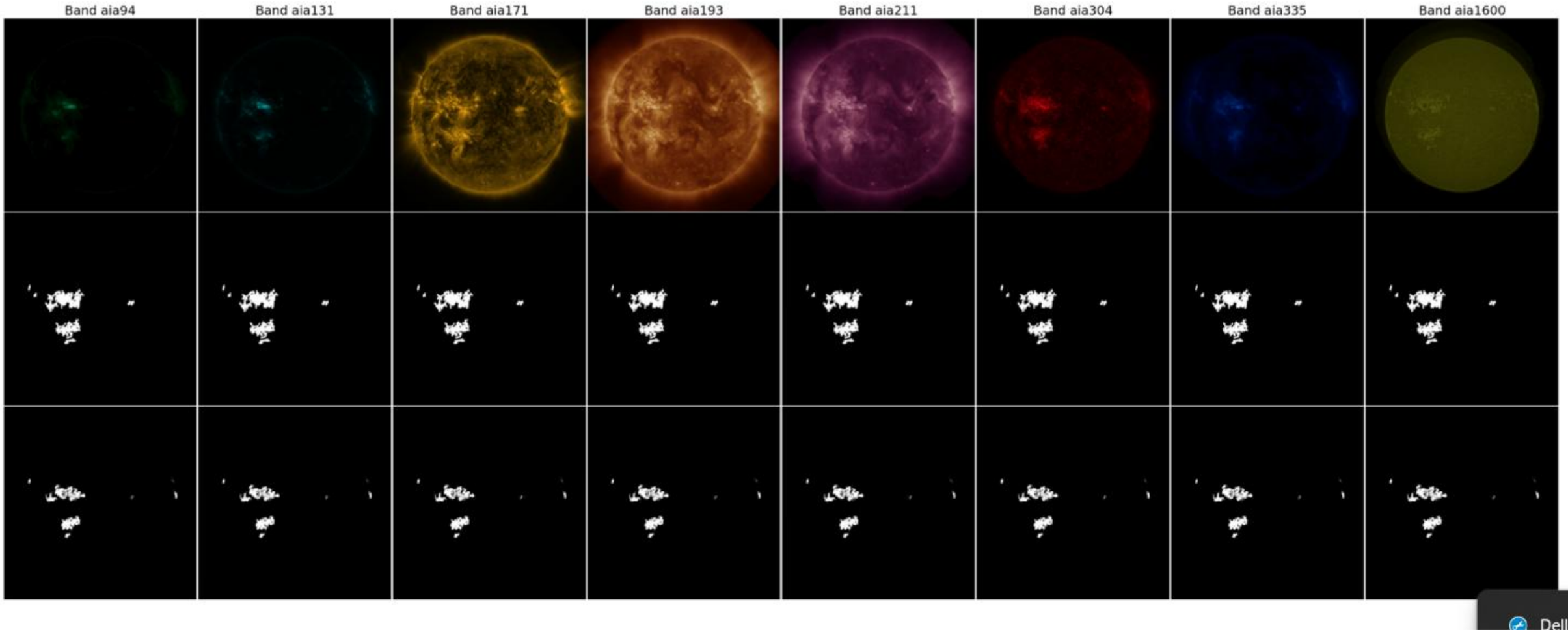
EUV Spectrum

$$y_t = f(\mathbf{x}_t). \qquad y_t \in \mathbb{R}^{1343}$$

Table 5: EUV spectra modeling results comparing baseline models with Surya

Model	MSE	MAE	MAPE
Alexnet	0.0001311409	0.0061858603	1.6834715604
ResNet50	0.0030414662	0.0529725812	8.8635520935
Surya	**0.0001260741**	**0.0045114677**	**1.4792510271**

Active Region Segmentation



Active Region Segmentation

Table 3: AR Segmentation results comparing baseline models with Surya

Model	Params	IoU	Dice Coeff
Unet	9.2 M	0.688	0.801
Surya	**4.1 M**	**0.768**	**0.853**

IoU: Intersection over Union

The power of attention

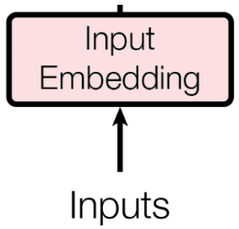
France is my home. I speak fluent . . .

Inputs

The power of attention

153 3000 23999 42 356 88 478 . . .

France is my home. I speak fluent . . .

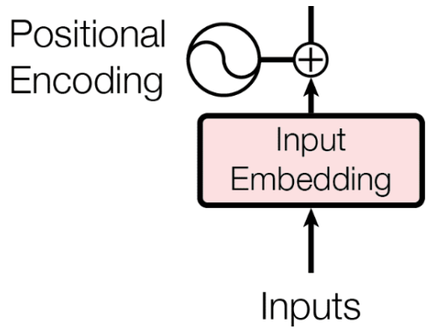


The power of attention

0 1 2 3 4 5 6

153 3000 23999 42 356 88 478 . . .

France is my home. I speak fluent . . .



The power of attention

France is my home. I speak fluent . . .

France is my home. I speak fluent . . .

France is my home. I speak fluent . . .

Ownership

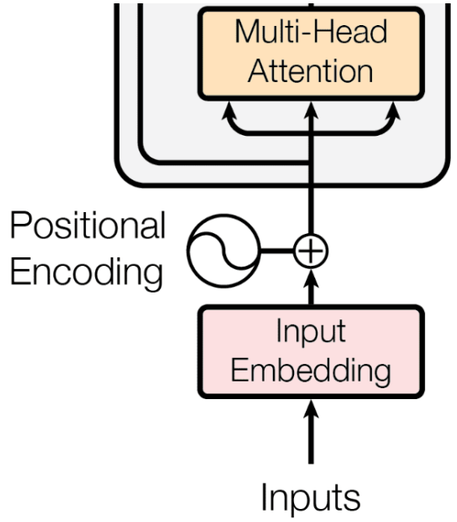
Location

Topic

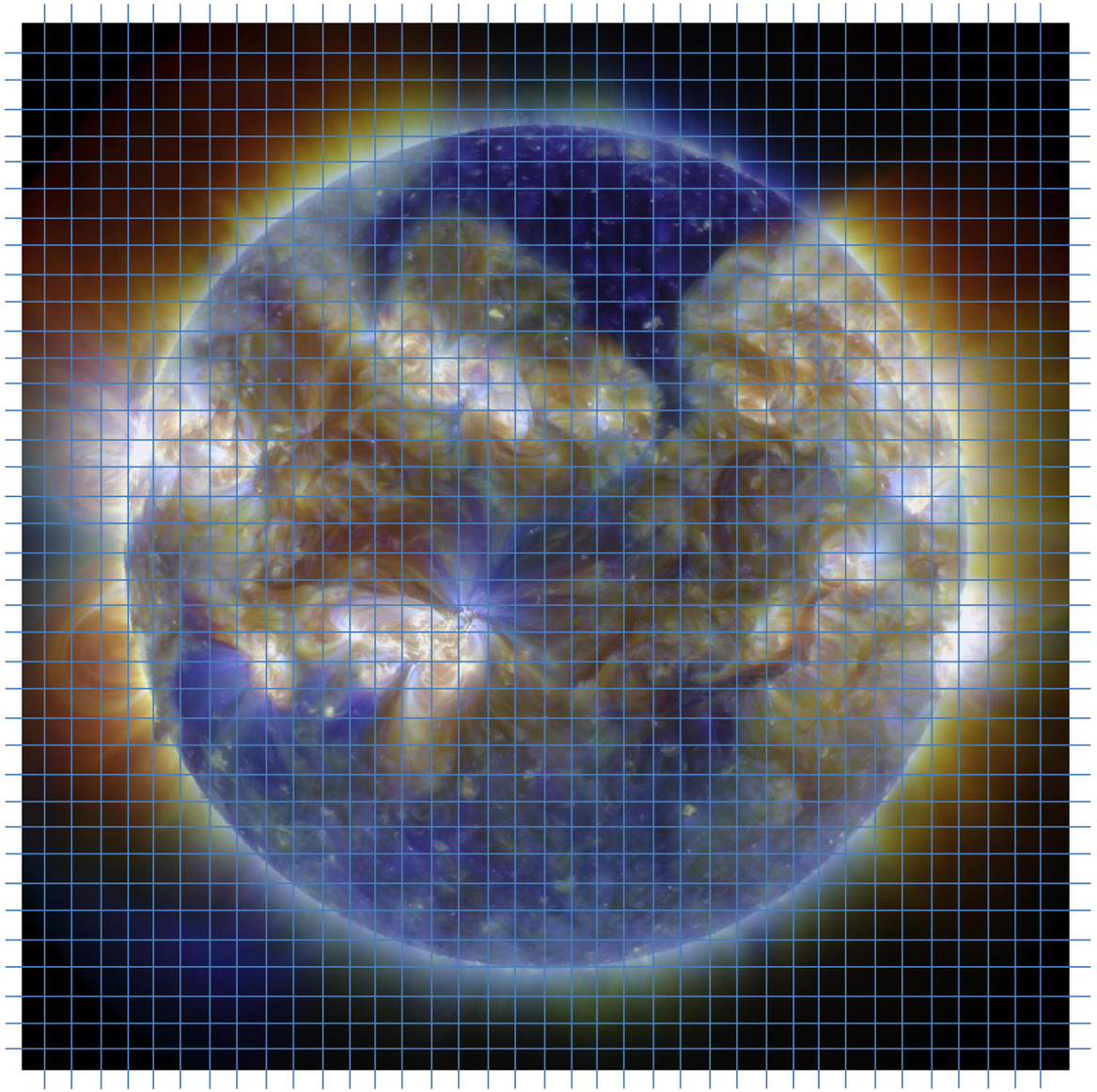
0 1 2 3 4 5 6

153 3000 23999 42 356 88 478 . . .

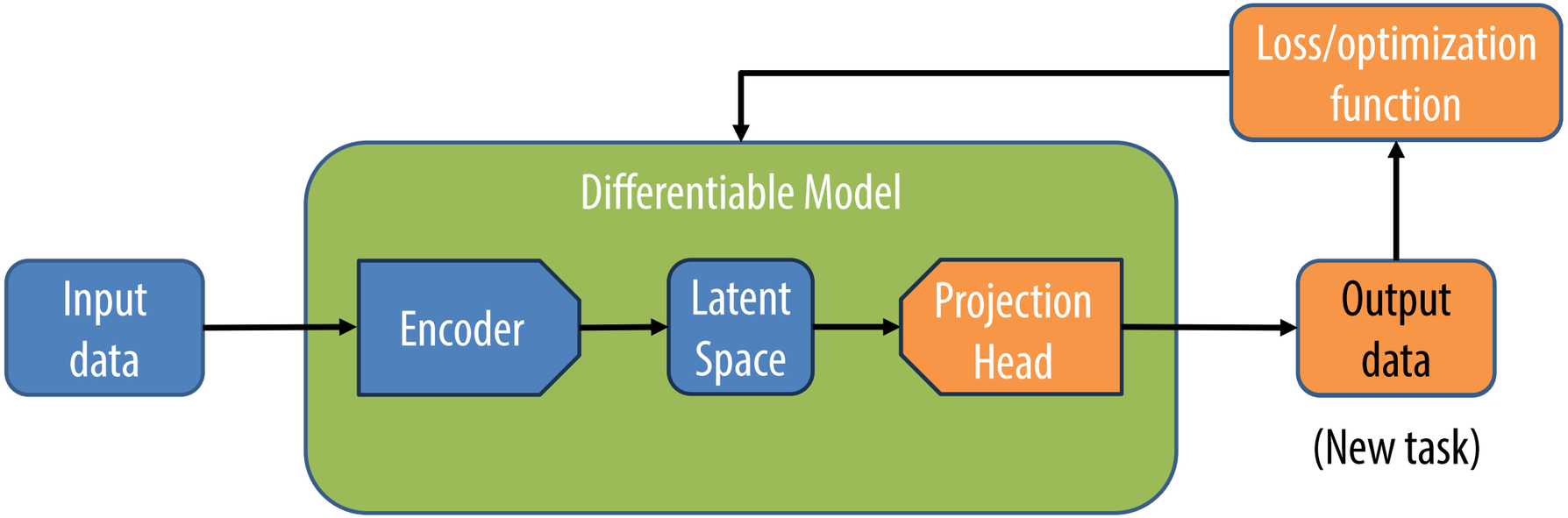
France is my home. I speak fluent . . .



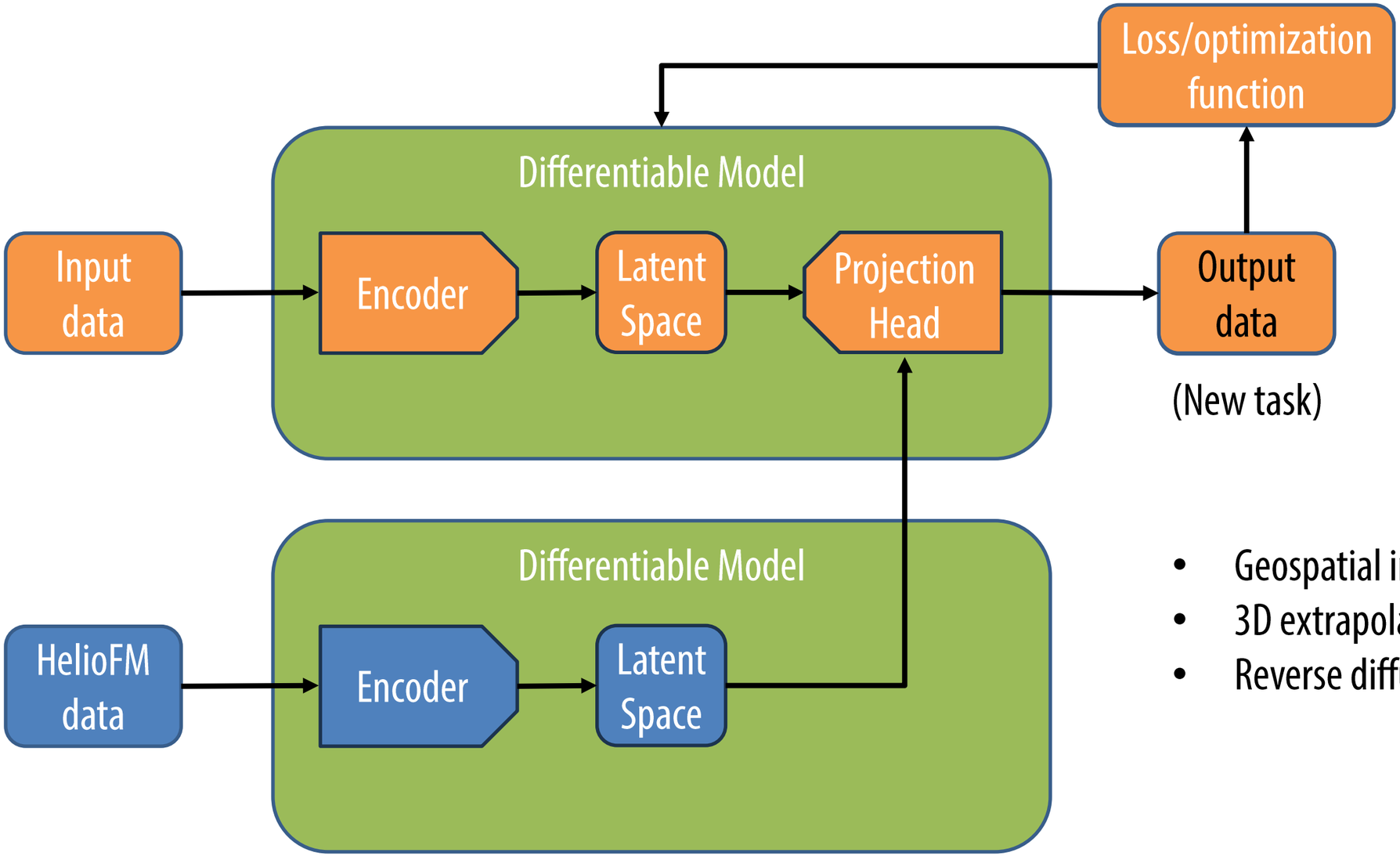
Visual Transformer



Conditioning

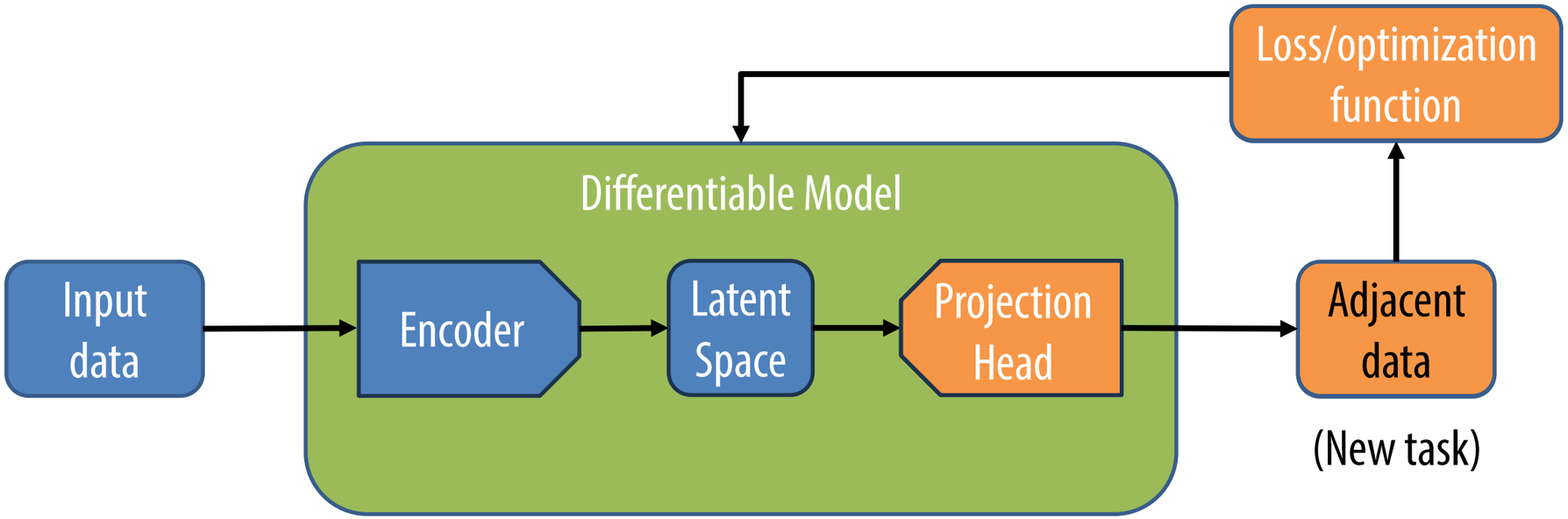


Conditioning

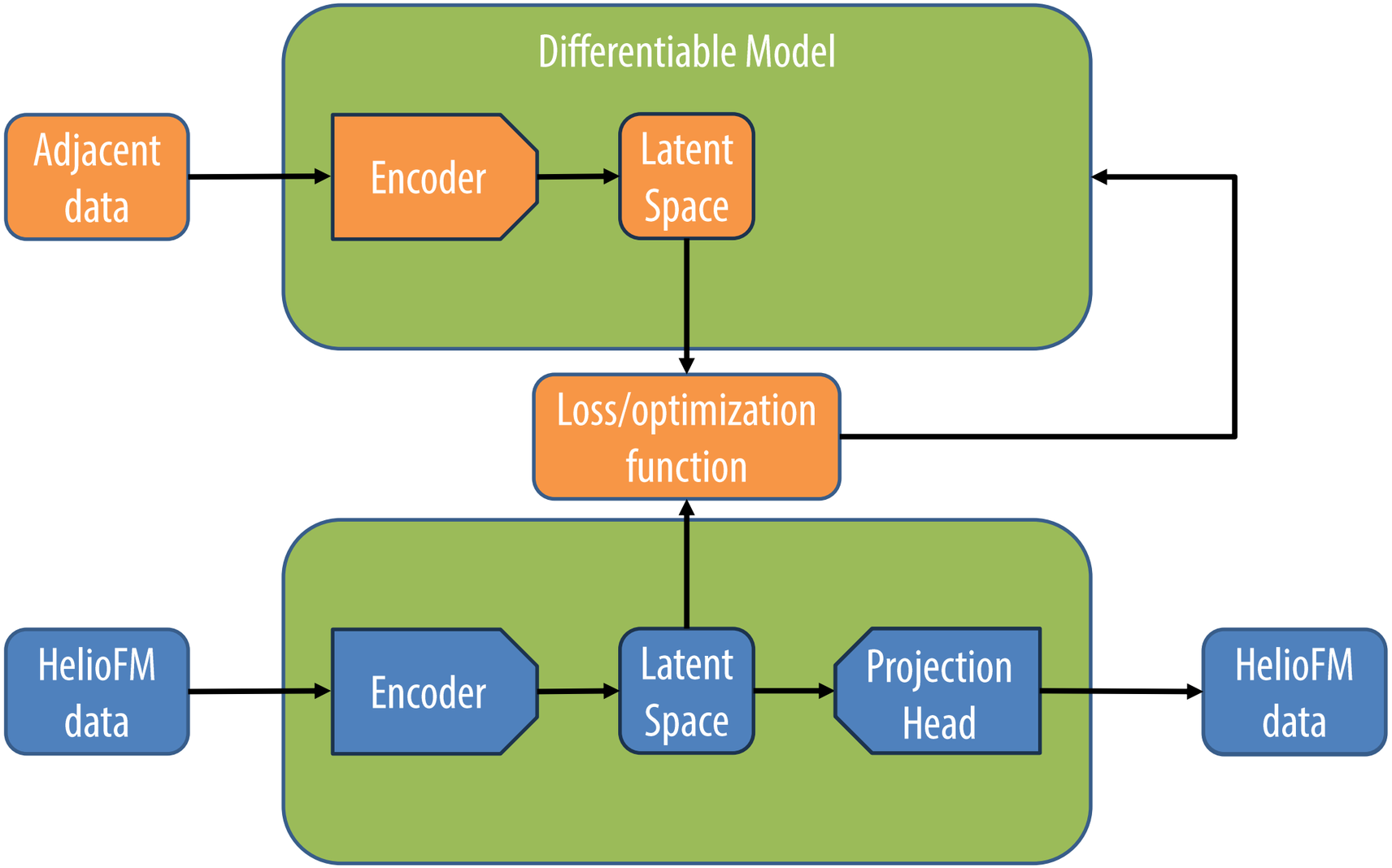


- Geospatial inference.
- 3D extrapolation.
- Reverse diffusion.

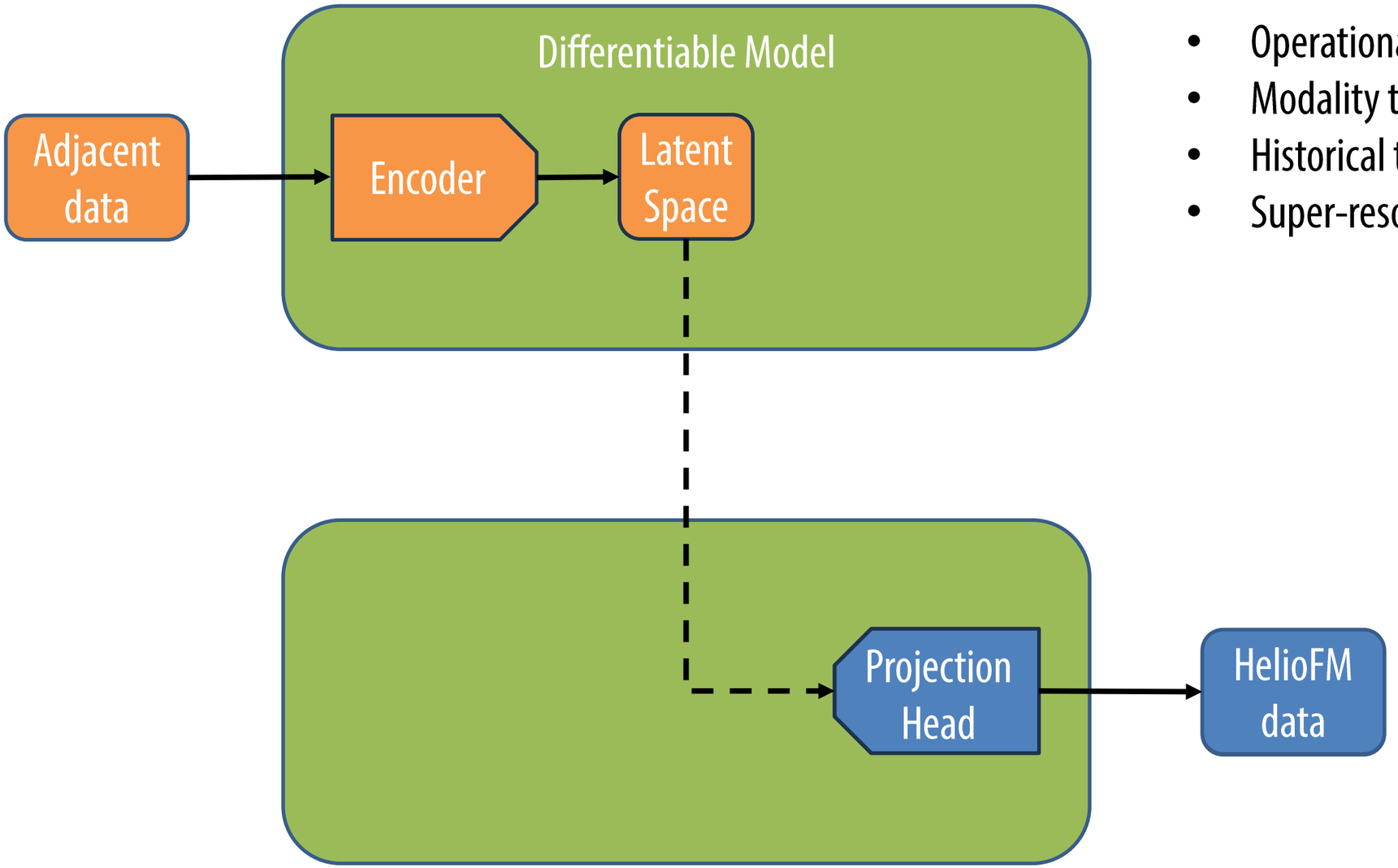
Mapping



Mapping



Mapping



- Operational to scientific.
- Modality translation.
- Historical to modern.
- Super-resolution