Multimodal LLMs: A New Way to Classify Astronomical Transient Images

Dr Fiorenzo Stoppa

YITP long-term workshop, Kyoto, 18-02-2026

Multi-Messenger Astrophysics in the Dynamic Universe

UNIVERSITY OF OXFORD

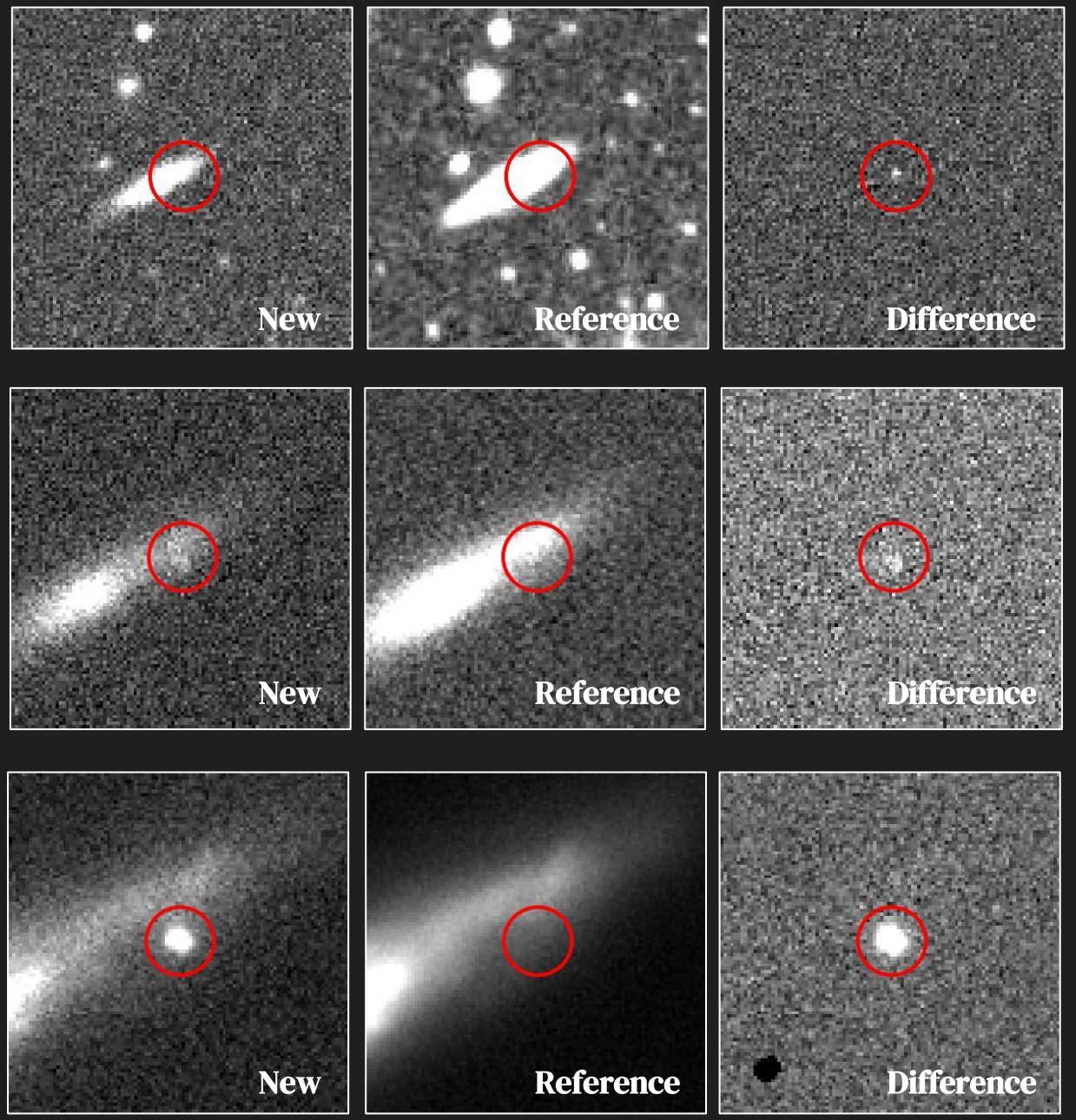
LLMs for Transients Detection

Using Gemini we want to not only classify transients, but also have an explanation of what is in the image.

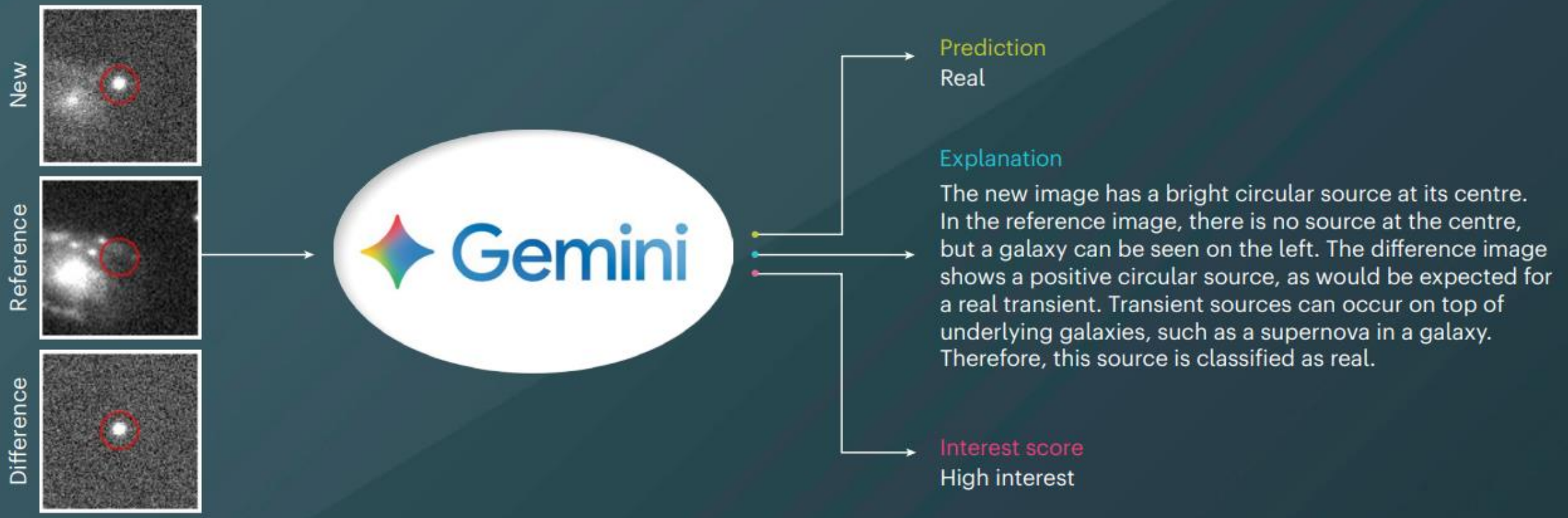
ATLAS
(1.86 arcsec/pixel)

MeerLICHT
(0.56 arcsec/pixel)

Pan-STARRS
(0.258 arcsec/pixel)



F. Stoppa & T. Bulmus et al., 2025, Nature Astronomy

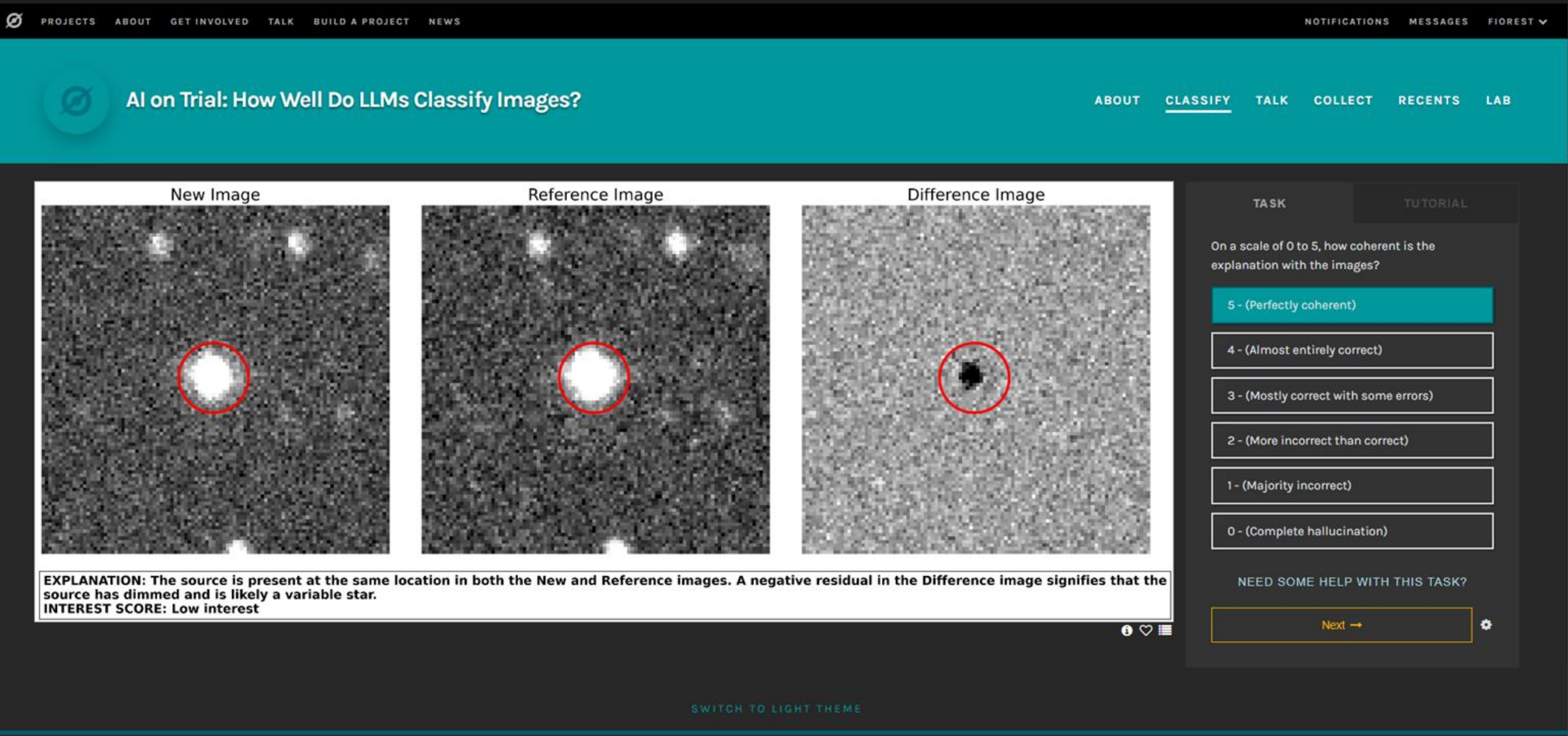


It is not trained!

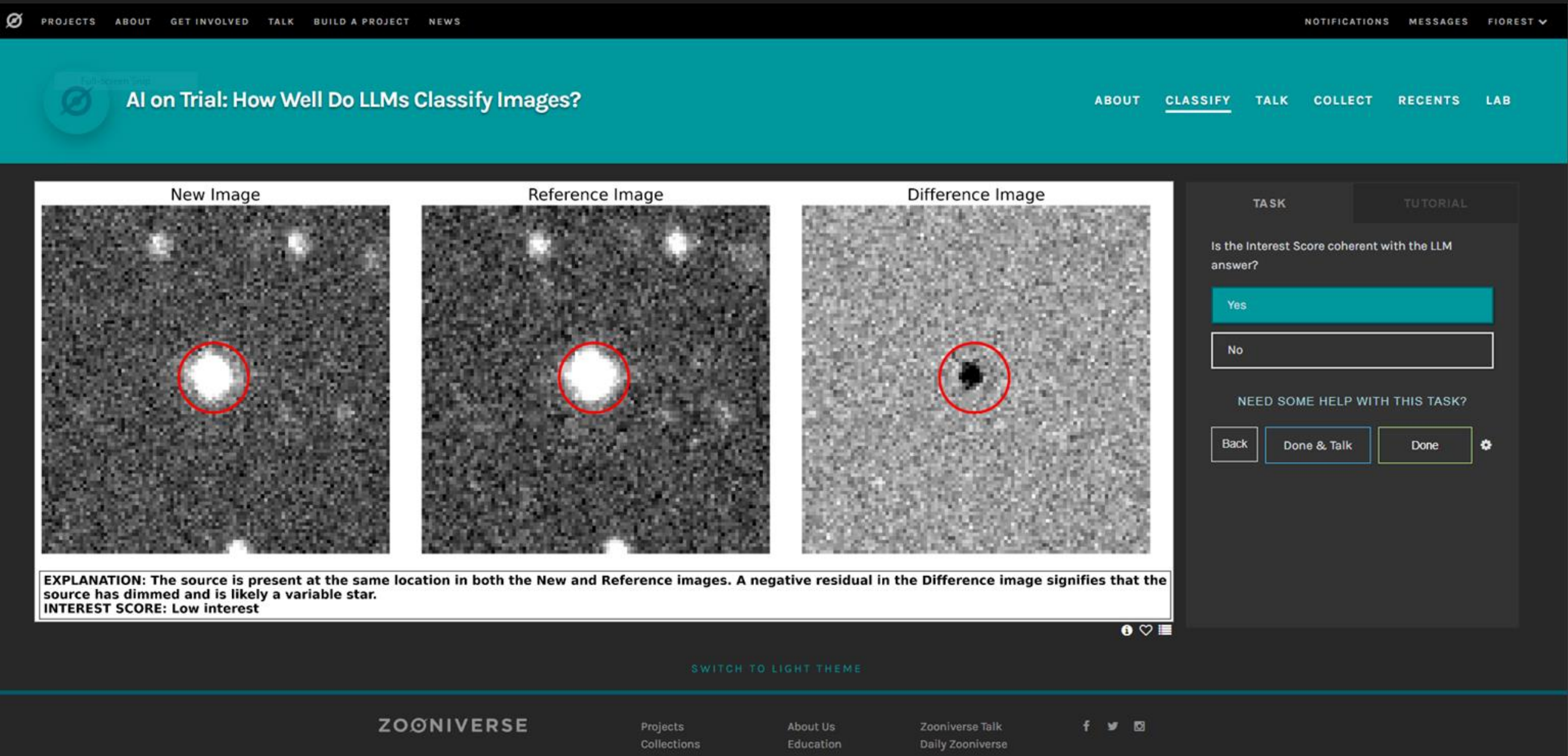
You only pass a set of instructions and 15 annotated triples like the one above.

Telescope	Accuracy (%)	Precision (%)	Recall (%)
ATLAS	91.9	88.5	94.5
MeerLICHT	93.4	87.7	98.7
Pan-STARRS	94.1	95.4	93.1

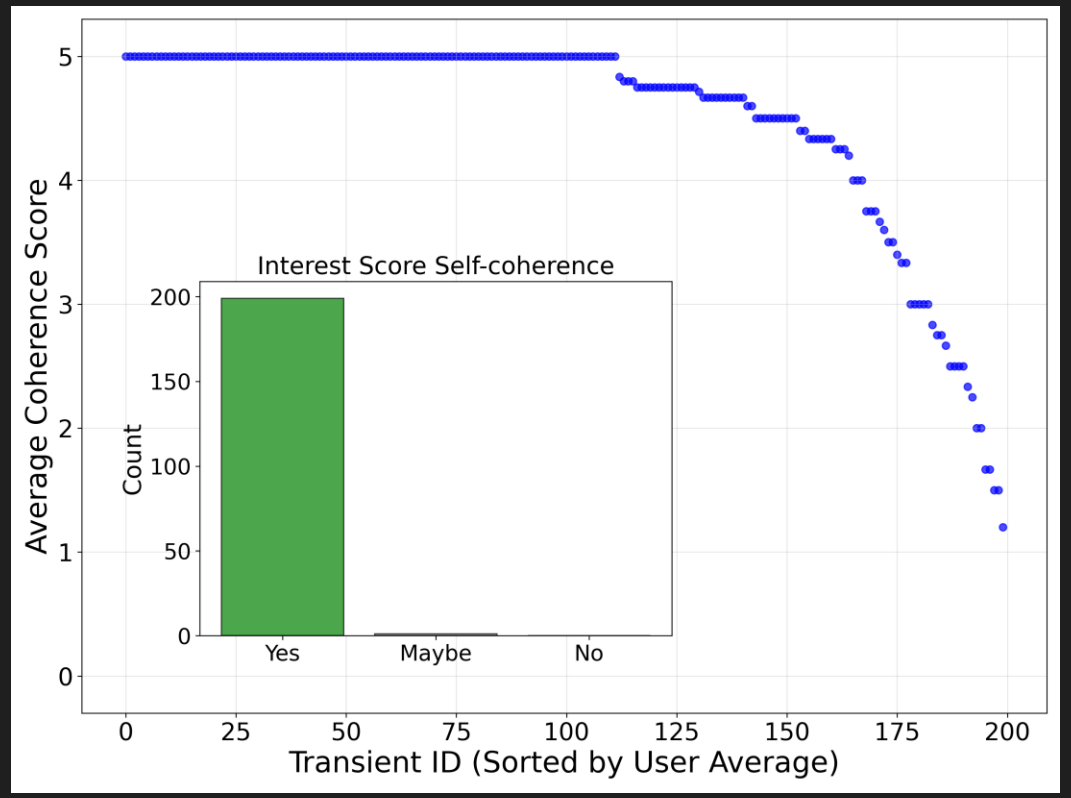
LLMs answers coherence

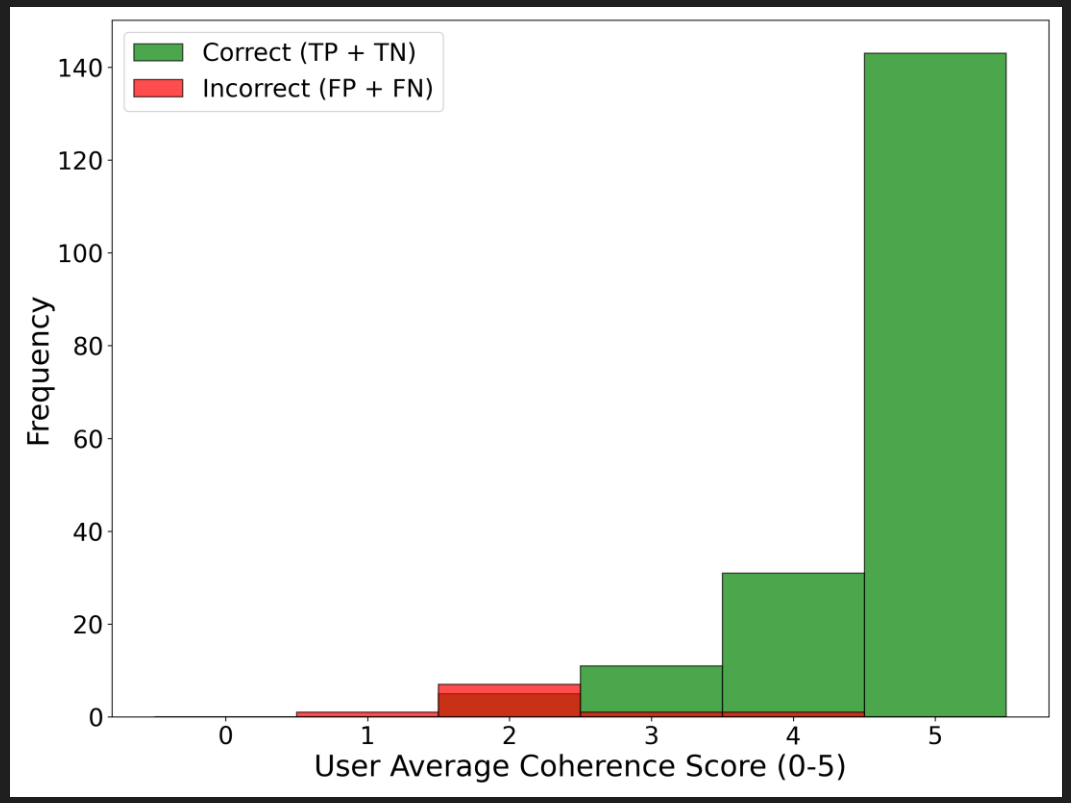


LLMs answers coherence



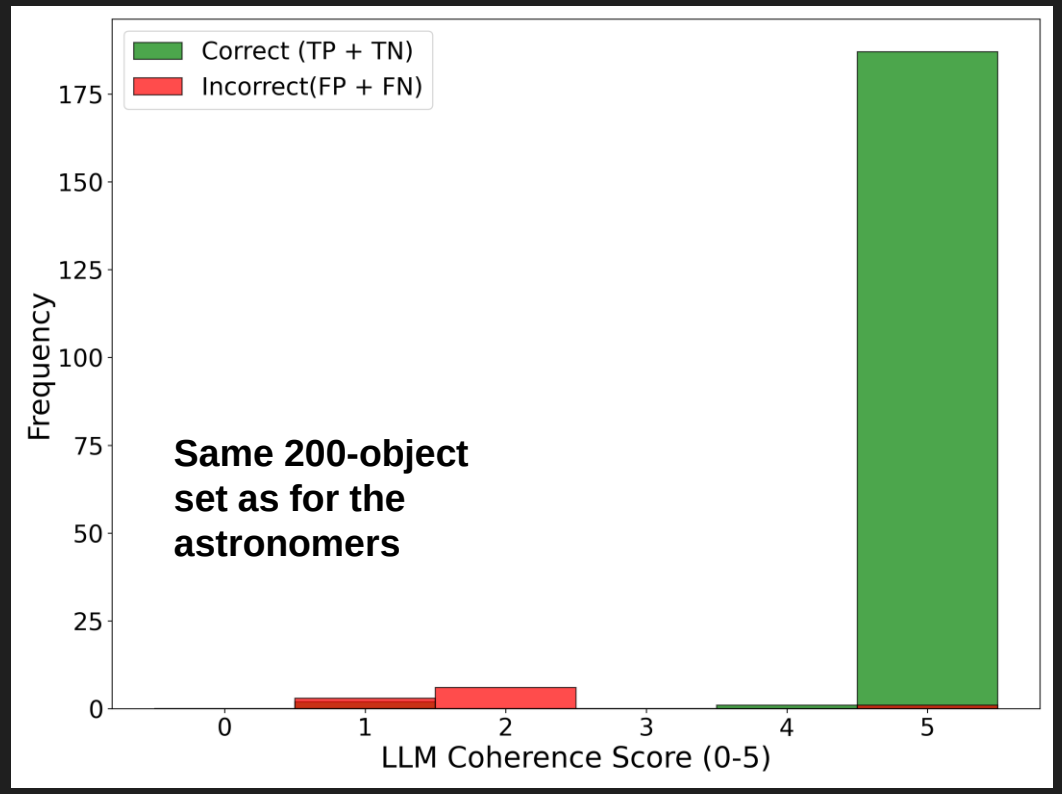
Astronomers judging LLMs

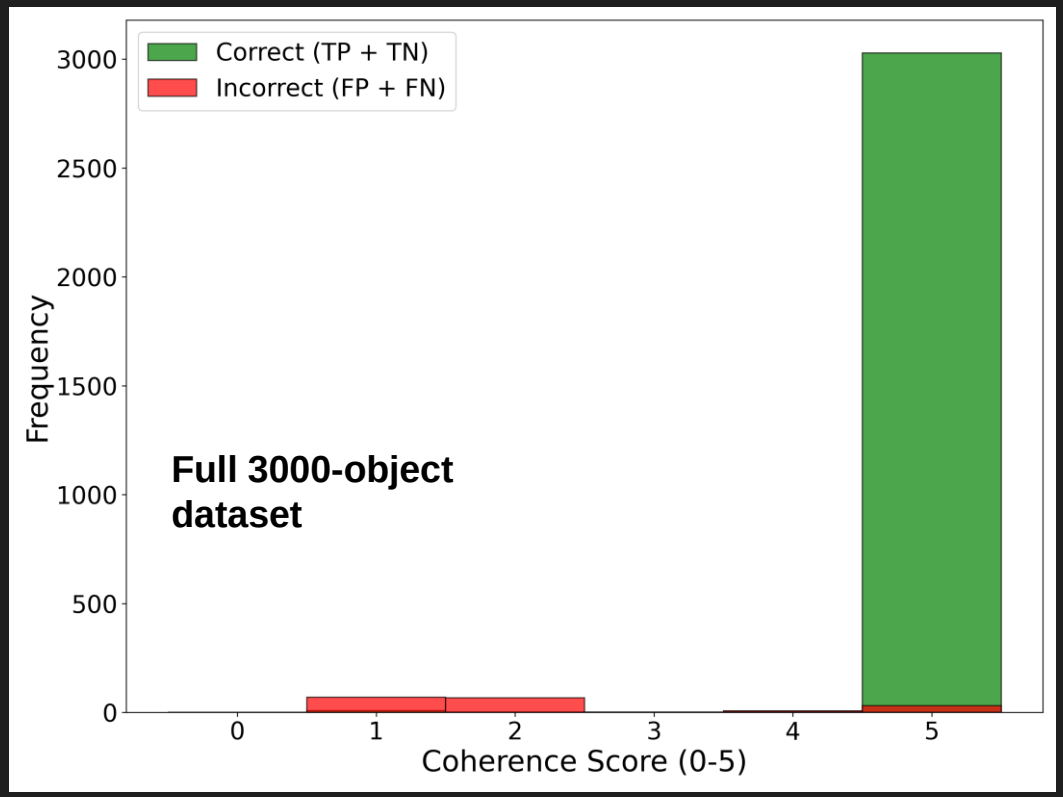




Expert ratings show that the explanations provided by Gemini are generally consistent and coherent (left plot, most transients get 5 or 4). The ones that get a lower coherence score are also the ones that were misclassified (right plot).

LLMs judging LLMs



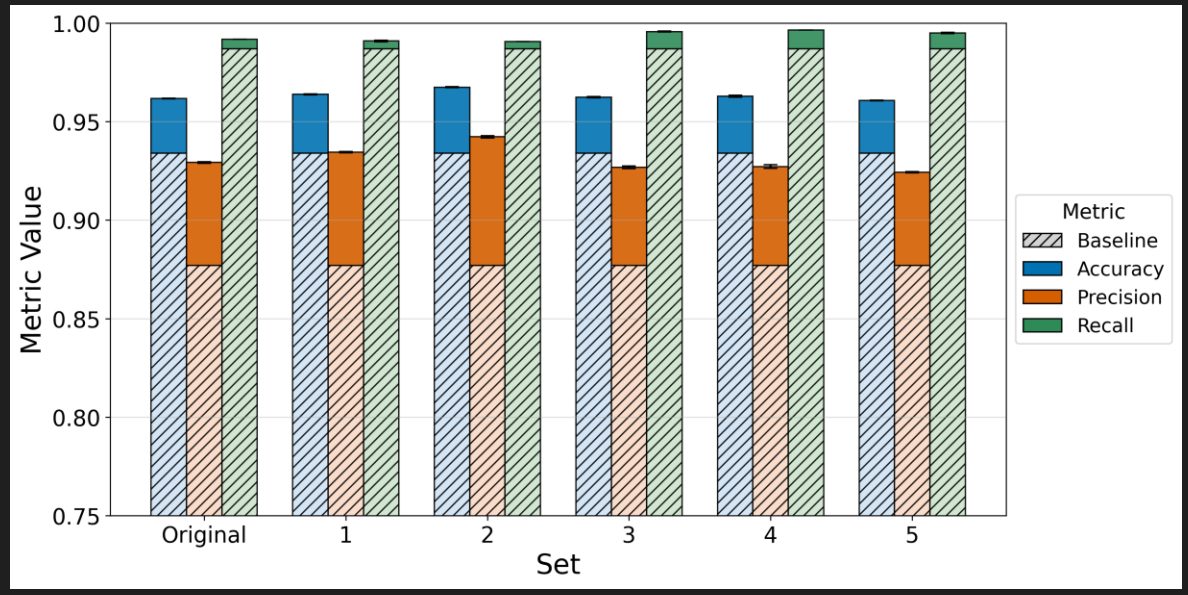


Coherence scores self-assigned by Gemini are more extreme, but similarly correlate with classification correctness.

LLMs Self-Improvement

Low coherence scores reliably flag Gemini's misclassifications, enabling a targeted, iterative data-augmentation loop.

Adding just 2 triplets to the original 15 it boosted accuracy from ~93.4% to ~96.7%.



Repeated results for different sets of inputs (baseline). Improved results with the same sets due to internal Gemini upgrade (no control over the model being used).

Pros/Cons of LLMs for Astro

Pros

1) No need for training!

2) Good and easy to improve

3) Accessible to non ML experts

Cons

1) Way too expensive

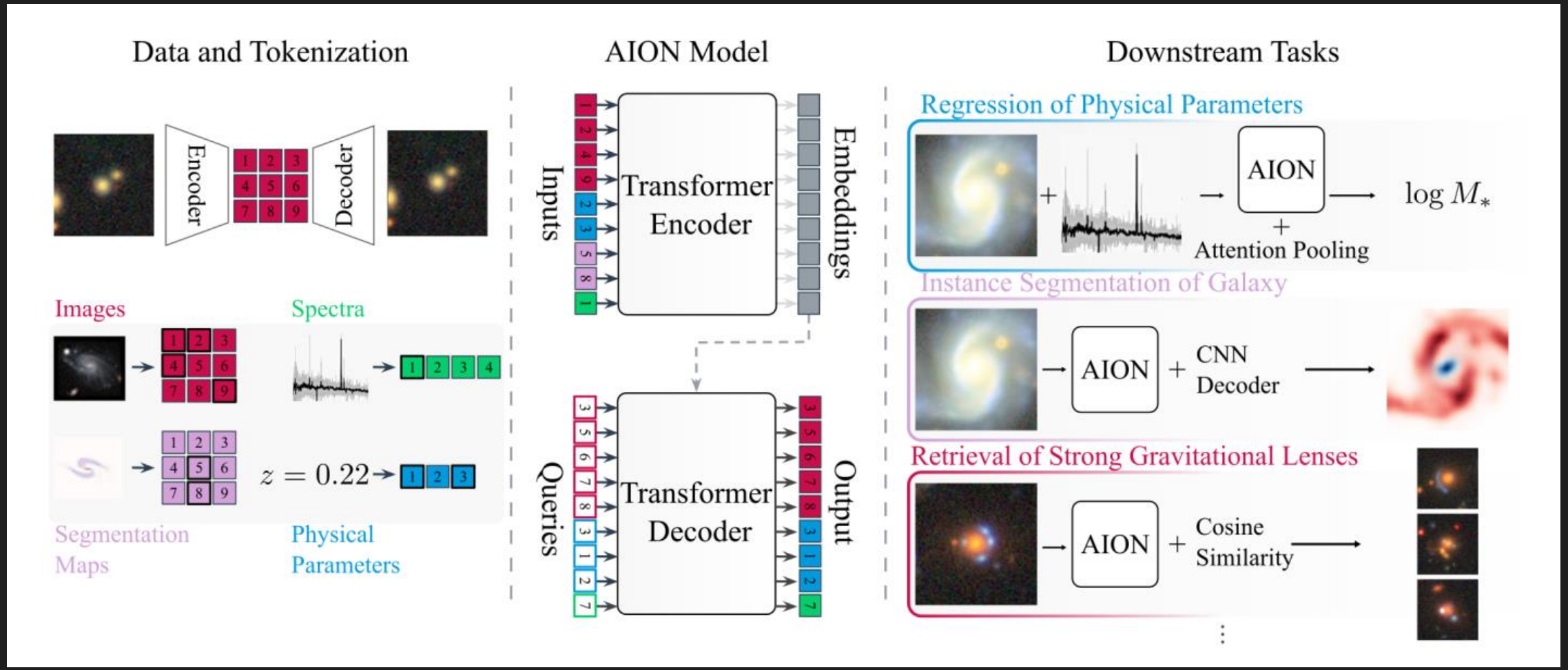
2) Slow

3) At the mercy of the provider

Overall, this shows that a general foundation model already encodes meaningful visual structure relevant to astronomy.

As foundation models strengthen their visual representations, astronomical data become increasingly aligned with their training paradigm.

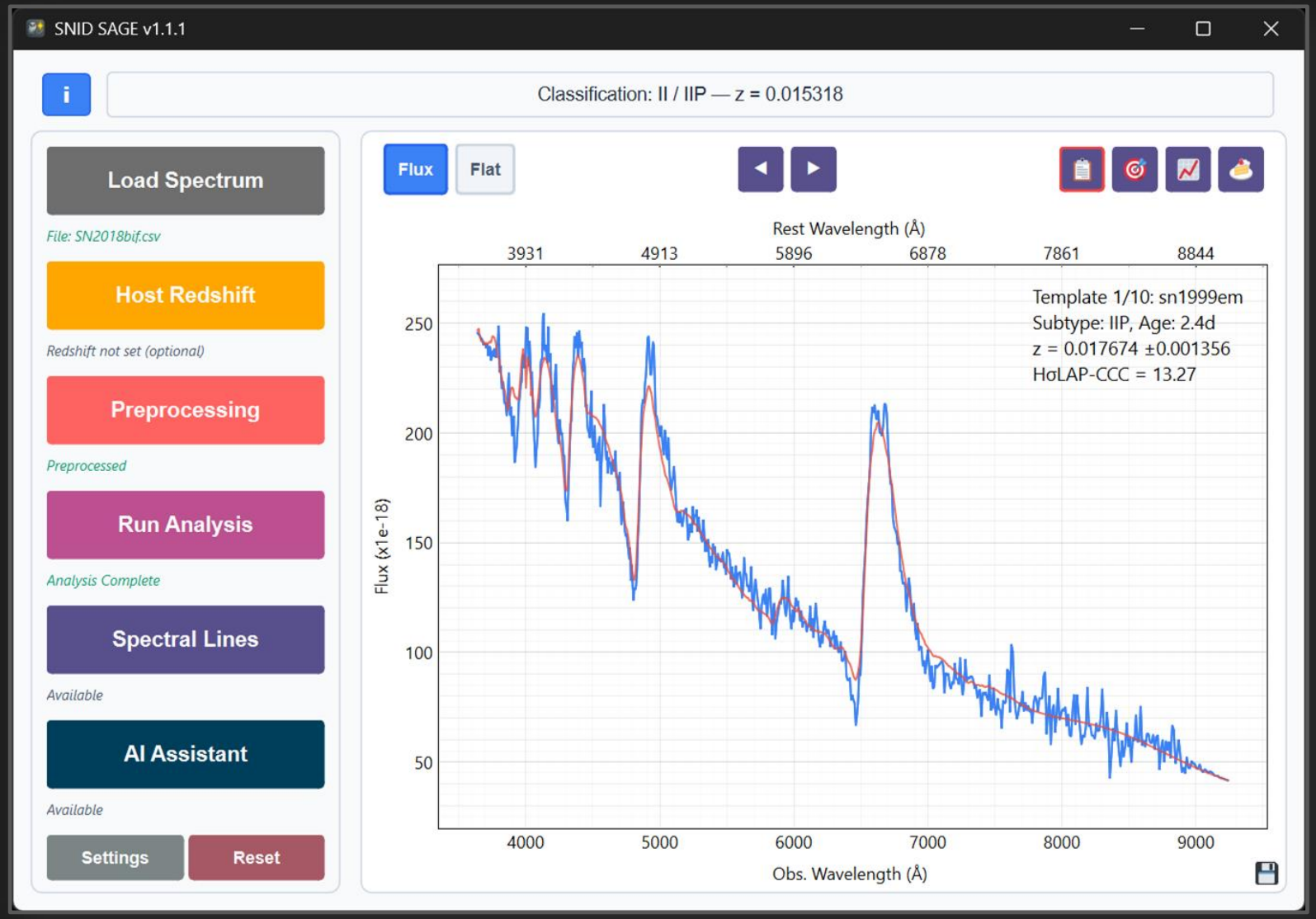
Foundation Models for Astronomy



Parker, L. et al., 2025, "AION-1: Omnimodal Foundation Model for Astronomical Sciences"

SNID-SAGE For Spectral Classification

- Python-native reimplementation of SNID
- Improved metrics and aggregation of final matches
- Some other nice optional features



SNID-SAGE For Spectral Classification

snid-sage-templates GUI

Type	Templates	Spectra
Ia	268	2883
II	108	1215
Ib	54	509
Ic	68	627
Ibn	10	68
Icn	1	2
SLSN	53	252
TDE	16	103
AGN	12	16
Galaxy	27	27
CV	9	52
GAP	10	54
LFBOT	3	21
KN	1	10
Star	3	13
Total	643	5852