Causally-Motivated Foundation Models

Disentangling Physics from Instruments

Daniel Muthukrishna
Research Scientist — MIT Kavli Institute for Astrophysics
AstroAI Fellow — Harvard & Smithsonian CfA

Coauthors:

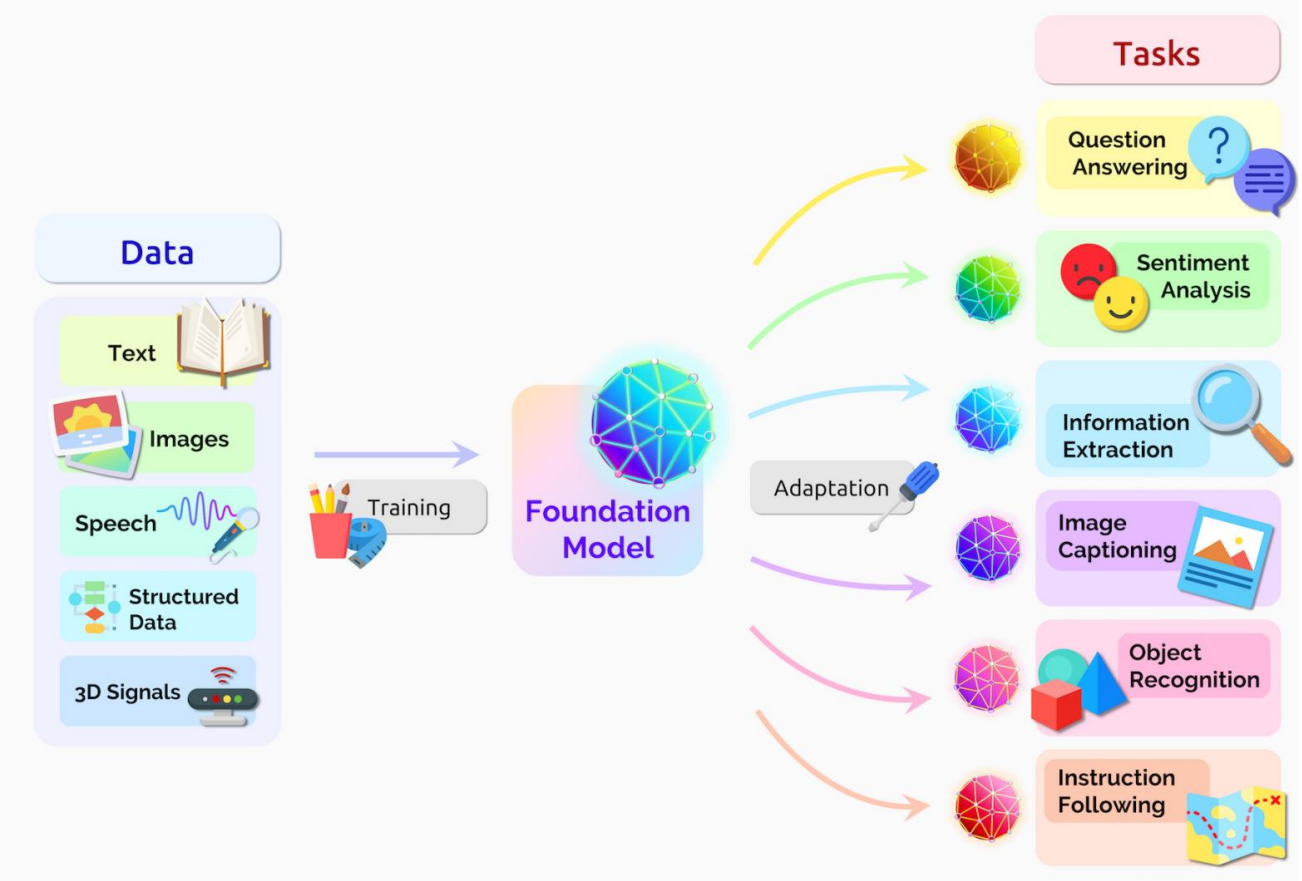
Pablo Mercader-Perez, Carolina Cuesta-Lazaro, Jeroen Audenaert, David Hogg, Ashley Villar

What is a Foundation Model?

- Foundation models are the base pre-trained networks for LLMs like ChatGPT or Claude, vision models like DALL-E, and appear to have emergent capabilities

- The goal is to find **general representations** that are useful for many (any?) downstream task.

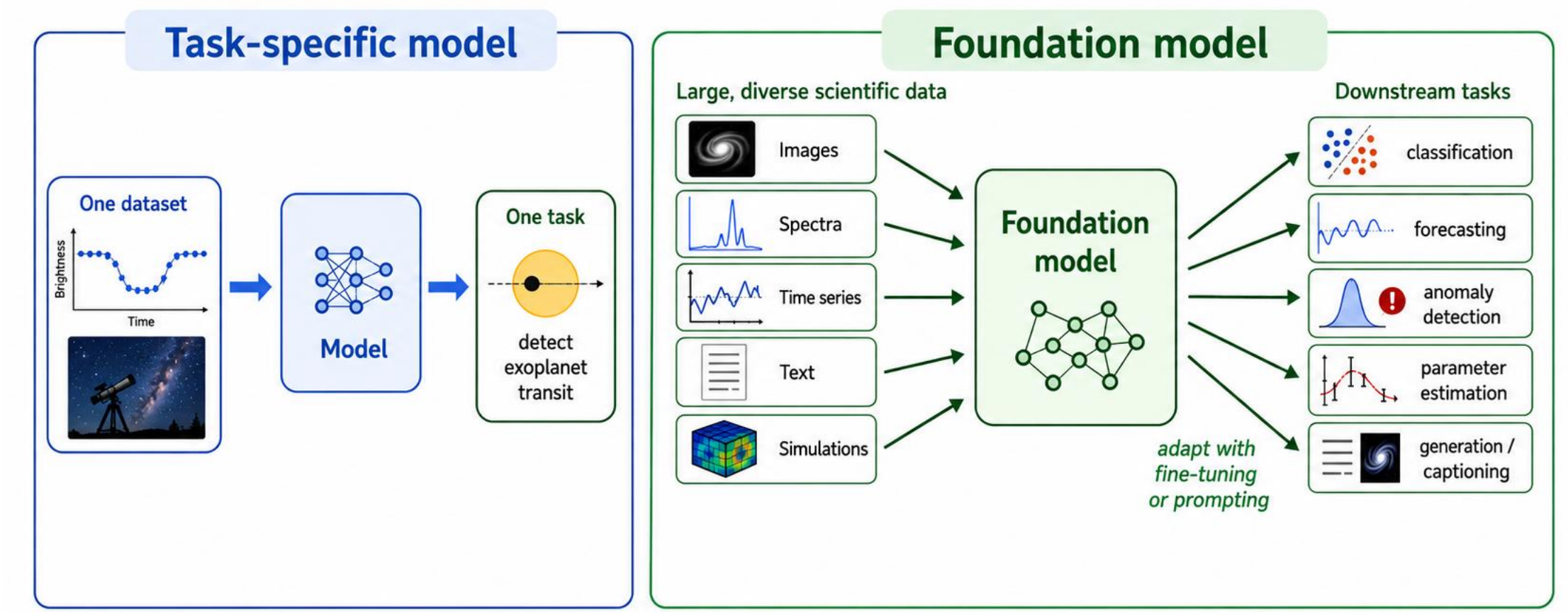
- They're not trained in the usual supervised fashion
  - Instead, FMs typically undergo *self-supervised* learning by optimizing a contrastive or generative objective.



Bossmani et al. 2021

What is a Foundation Model?

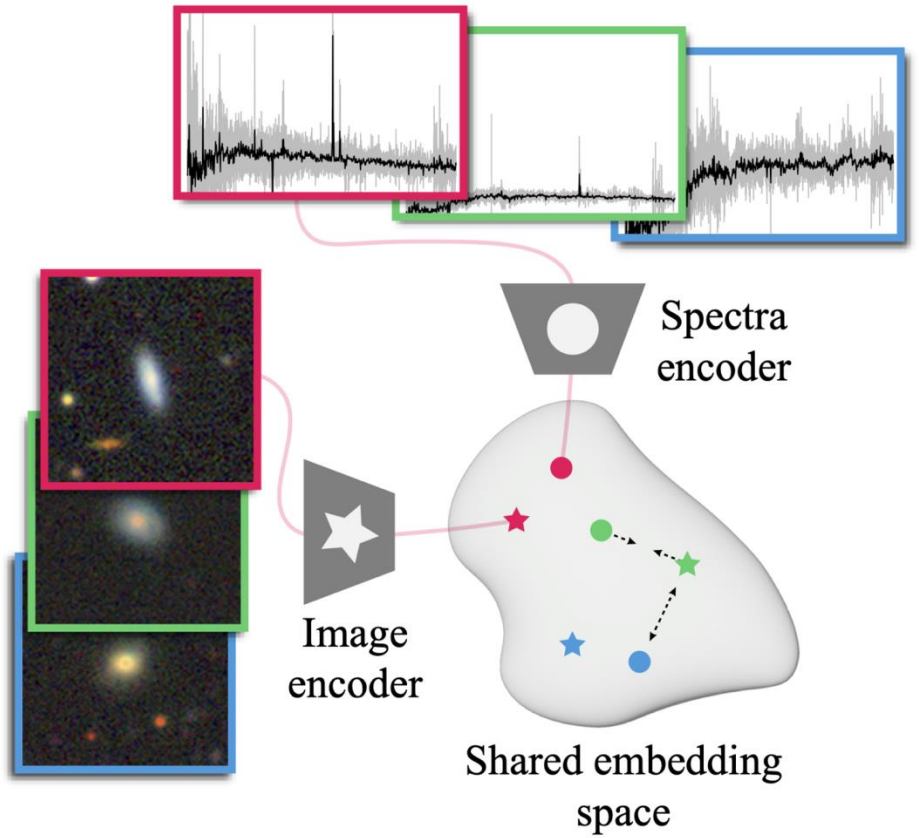
- FMs learn **general representations** that are useful for many (any?) downstream task.
- They're not trained in the usual supervised fashion
  - Instead, they undergo *self-supervised* learning by optimizing a contrastive or generative objective.



I prompted ChatGPT for an illustration…

Foundation Models in Astronomy

Foundation Model	Domain	Method
AstroCLIP (Parker et al. 2023; Github)	Multi-modal (images and spectra)	Contrastive
Maven (Zhang et al. 2024)	Multi-modal time series (photometry and spectra)	Contrastive
AstroM³ (Rizhko & Bloom 2024)	Multi-modal time series (photometry, spectra, and metadata)	Contrastive
*AstroPT-Euclid (Siudek et al. 2025)	Multi-modal (images and photometry)	Generative
FALCO (Zuo et al. 2025)	Kepler time-series	Generative
SpectraFM (Koblischke & Bovy 2024)	Stellar spectra (synthetic & real)	Generative
*Gaia spectra (Buck & Schwarz 2024)	Stellar spectra (Gaia XP and RVS)	Contrastive
SSL for DESI Legacy Survey (Stein et al. 2021; Github)	DESI Legacy Survey galaxy images	Contrastive
GZ-Evo (Walmsley et al. 2022; Github)	Galaxy images (multiple observatories)	Constrastive
AstroPT (Smith et al. 2024; Github)	DESI Legacy Survey Galaxy images	Generative
Radio Galaxy Zoo (Slijepcevic et al. 2022)	Radio-wavelength galaxy images	Constrastive
SSL for Radio Interferometric Images (Cecconello et al. 2024; Github)	Radio interferometric images	Constrastive
SSL for LOFAR (Baron Perez et al. 2025)	Radio galaxy images (LoTSS-DR2)	Constrastive

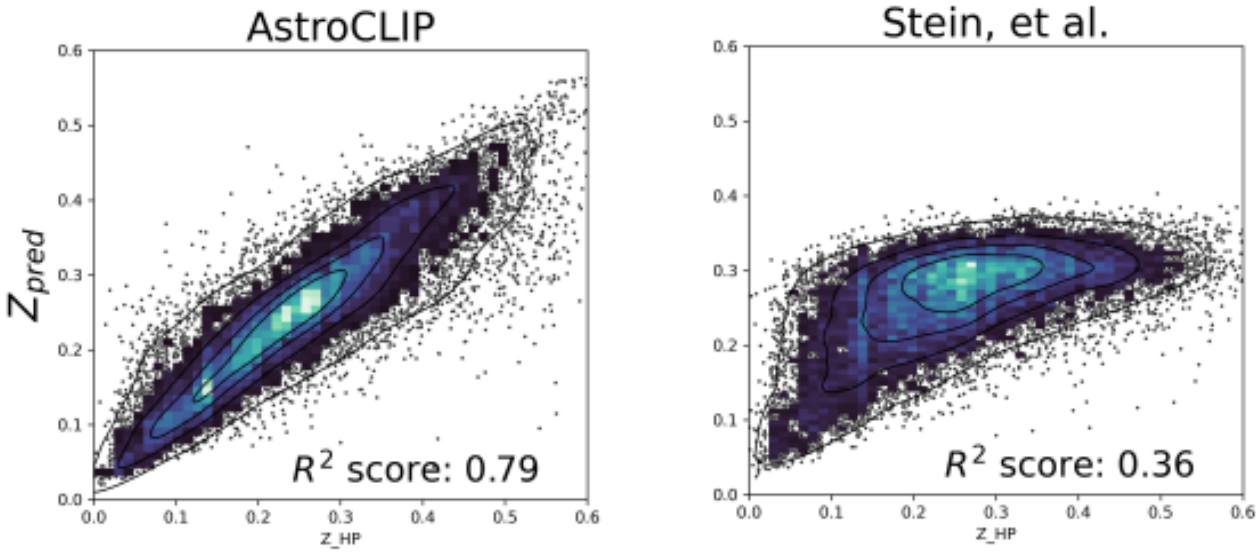


AstroCLIP, Parker et al. 2024

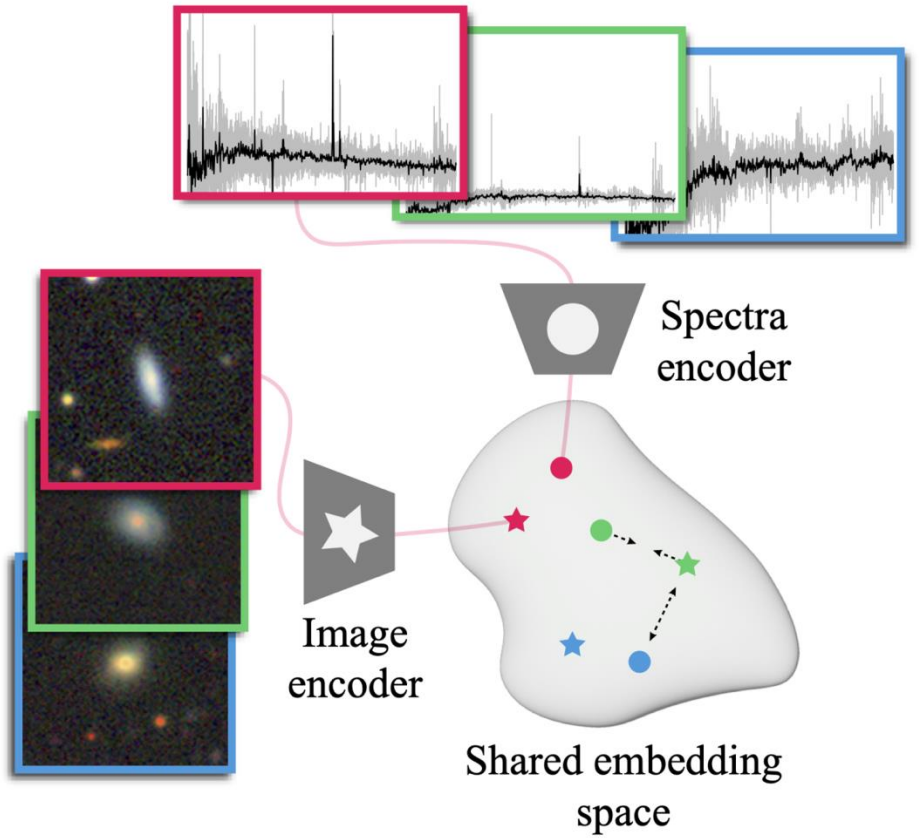
Blog by John Wu about Foundation Models: https://jwuphysics.github.io/blog/2025/05/foundation-models-in-astronomy/

Foundation Models in Astronomy: AstroCLIP

Downstream Task: Redshift Prediction

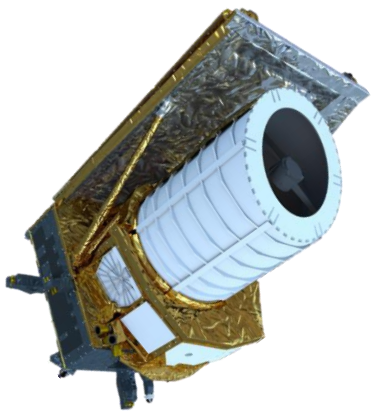


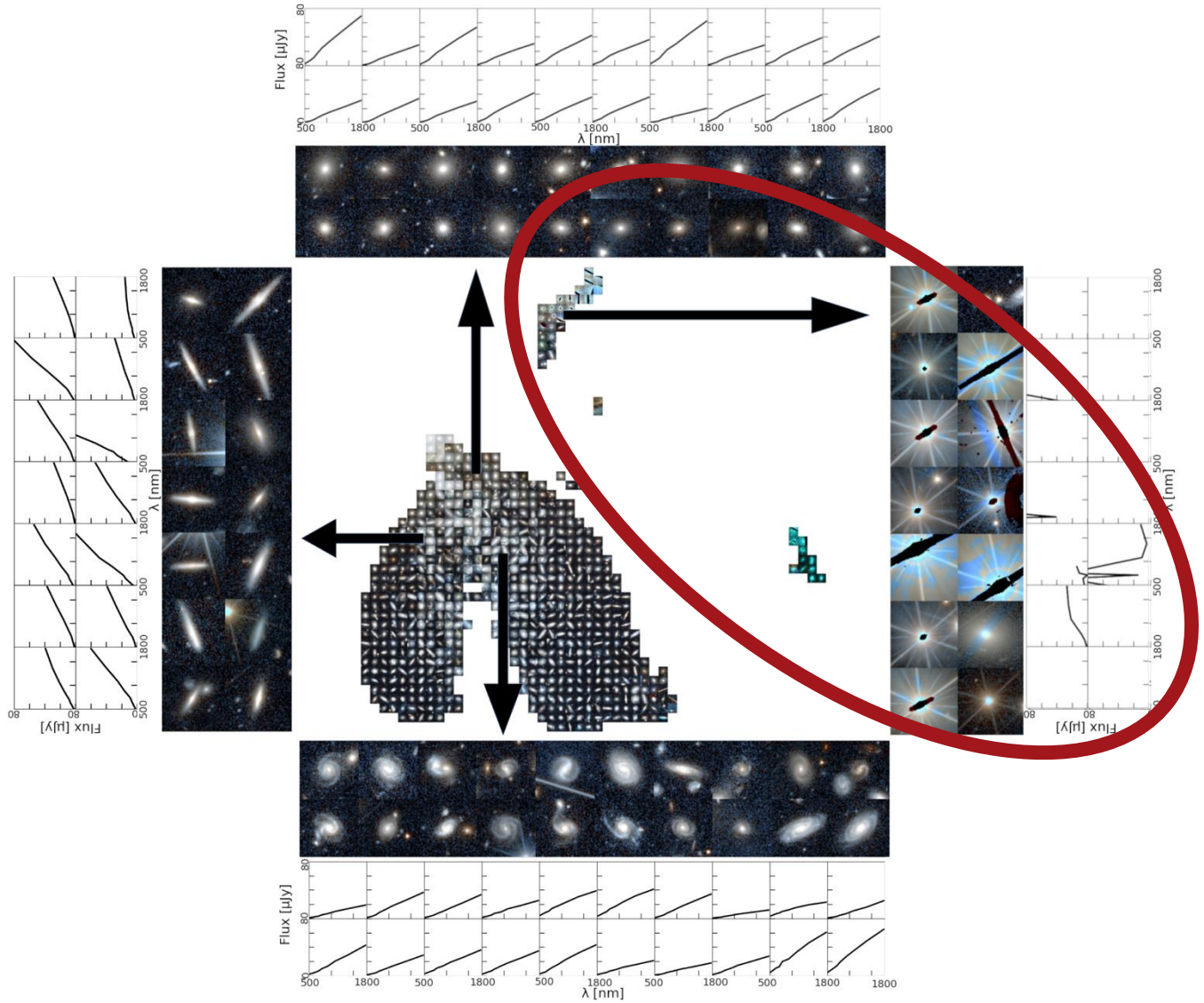
AstroCLIP had better redshift prediction
Without ever knowing about redshift! (zero-shot)



AstroCLIP, Parker et al. 2024

The anomalies from standard FMs are often instrument systematics – not discoveries

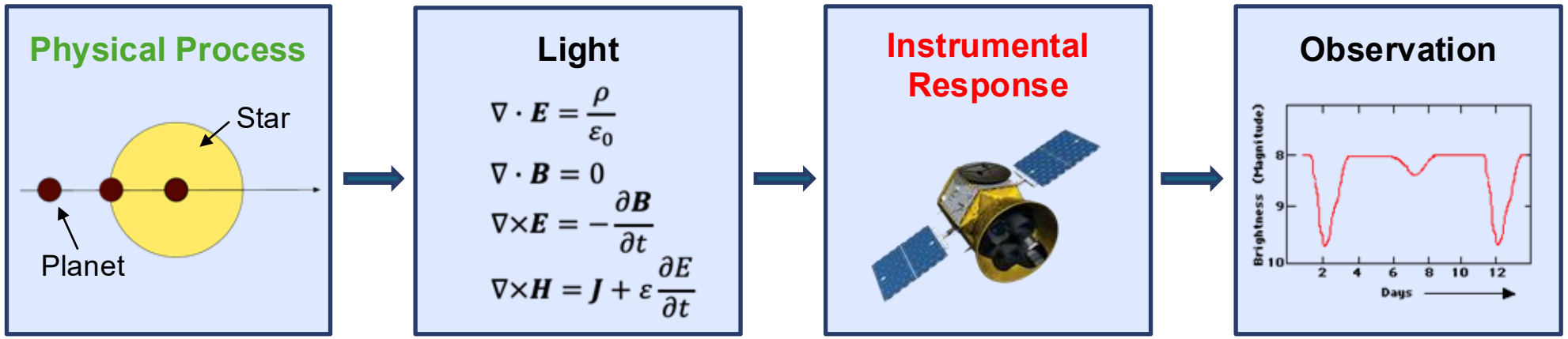




Euclid Collaboration, Siudek et al. 2025

Instrument systematics distort our signal

An observation is the downstream result of a combined causal process:



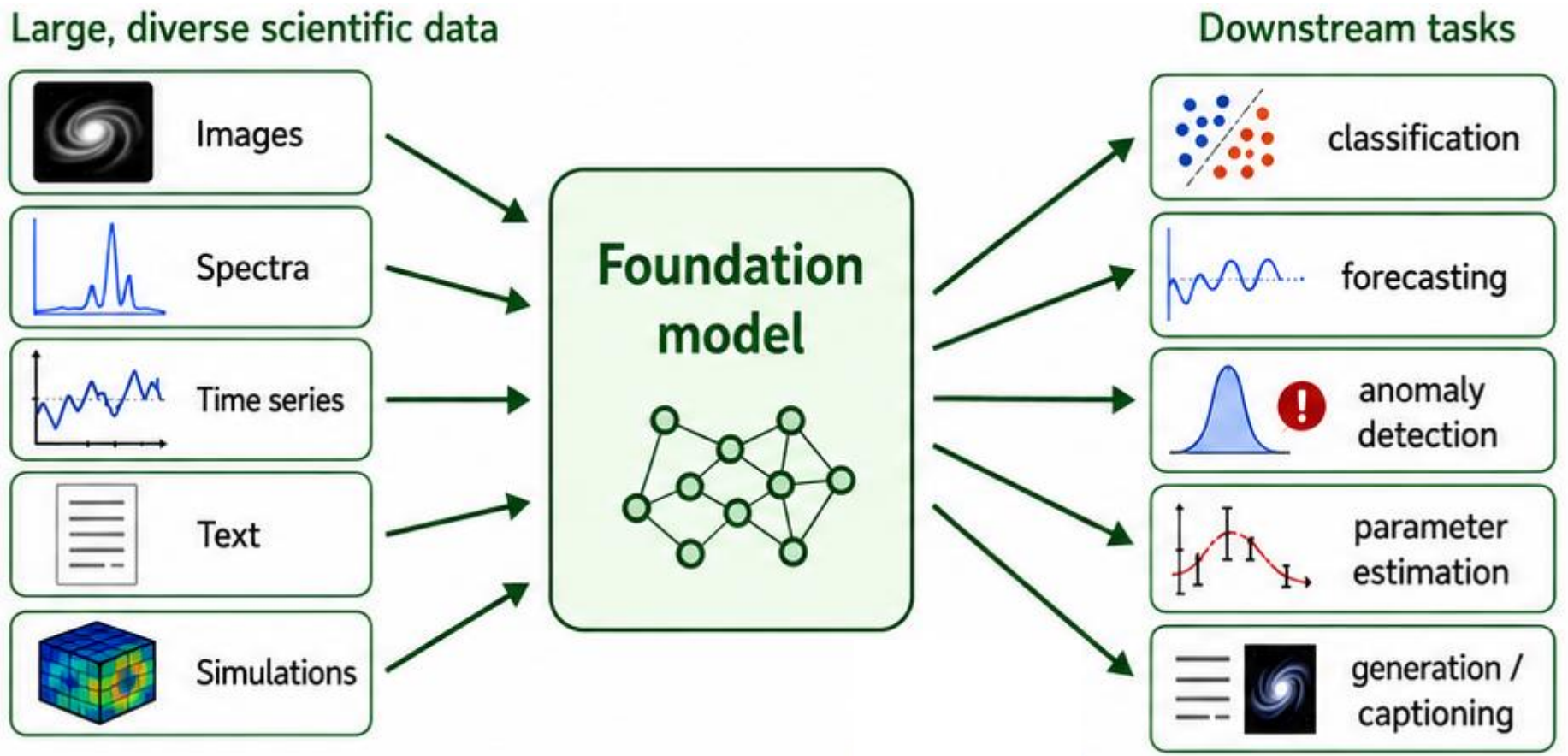
Observation = f(Physics, Instrument Systematics)

Originates somewhere in the universe and is independent of the measurement
Science variables

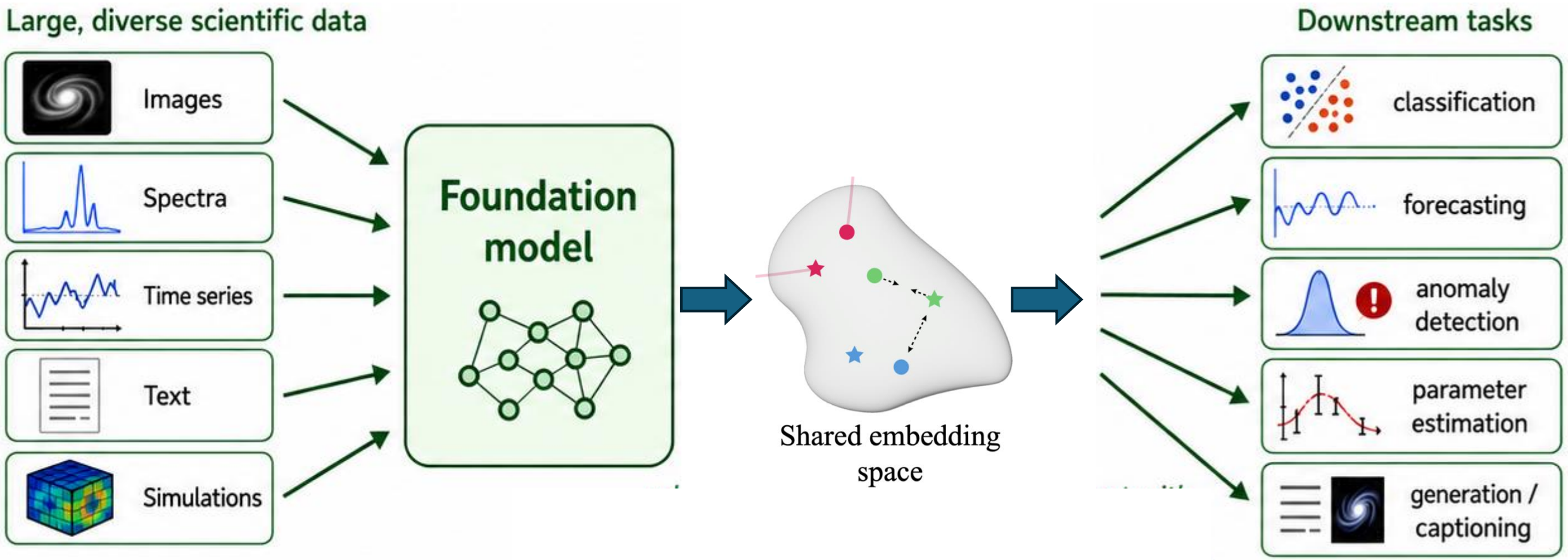
Distortions introduced by the measurement
Confounding variables

Slide credit: Pablo Mercader-Perez

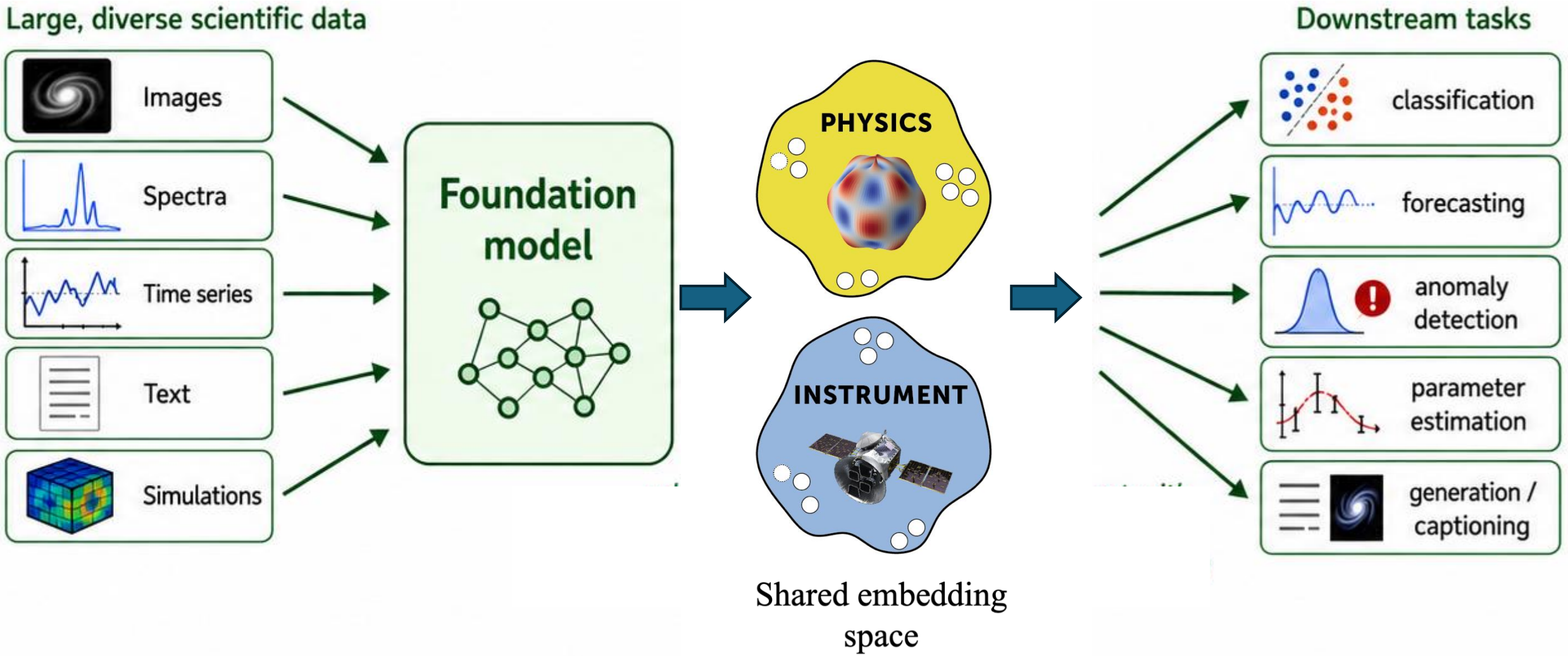
Typical Foundation Model



Typical Foundation Model



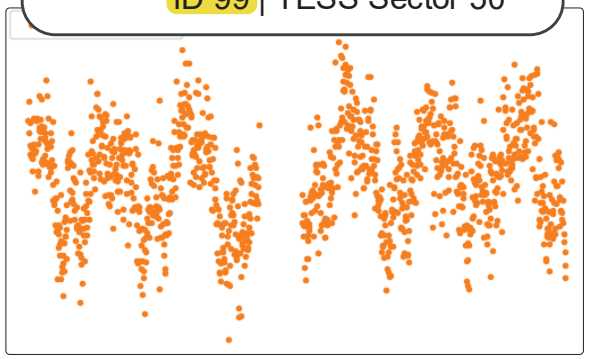
Causal Foundation Model



Making Data “Triplets”

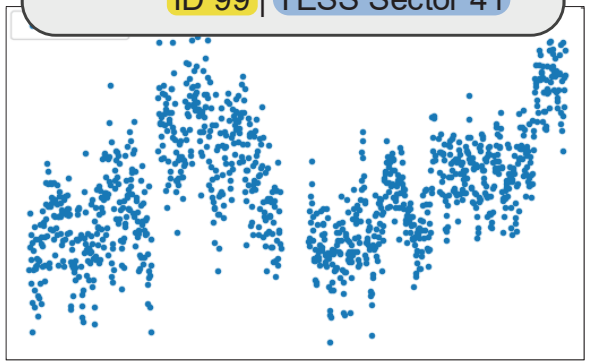
Same star

STAR ID 99 | TESS Sector 50



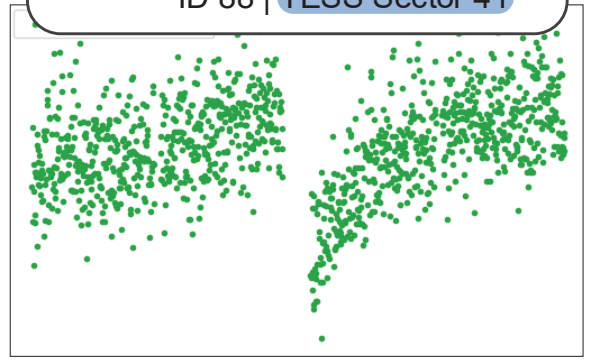
Original star

STAR ID 99 | TESS Sector 41



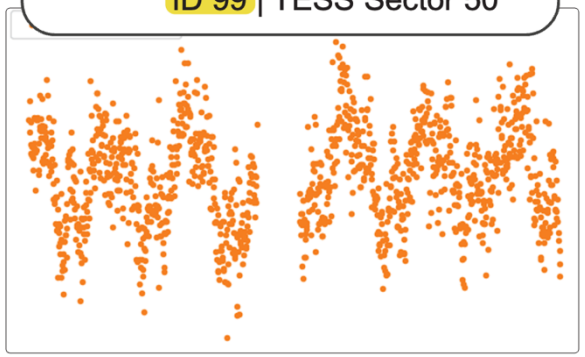
Same instrument

STAR ID 88 | TESS Sector 41

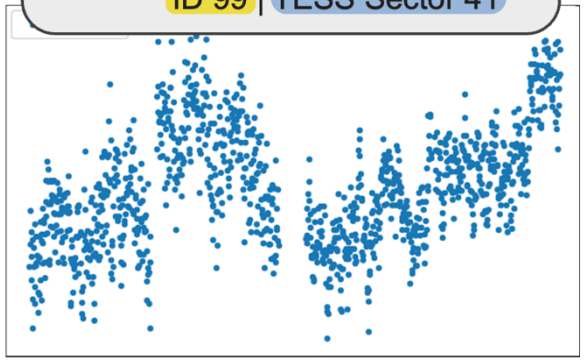


Making Data “Triplets”

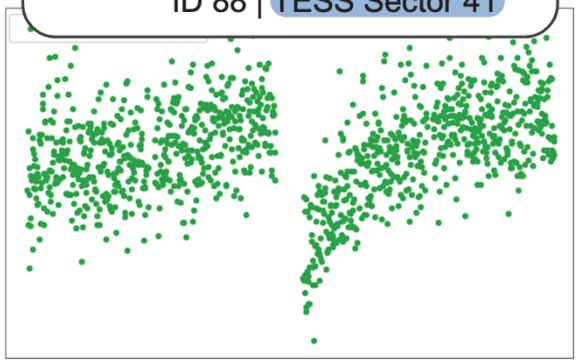
Same star
STAR ID 99 | TESS Sector 50



Original star
STAR ID 99 | TESS Sector 41

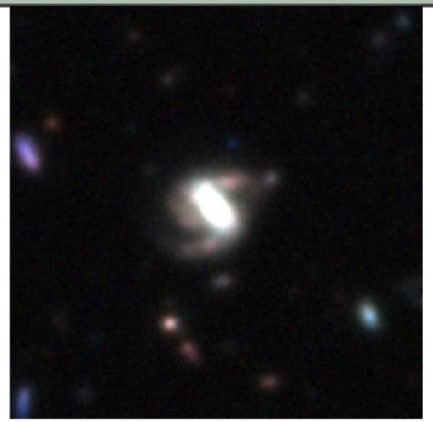


Same instrument
STAR ID 88 | TESS Sector 41

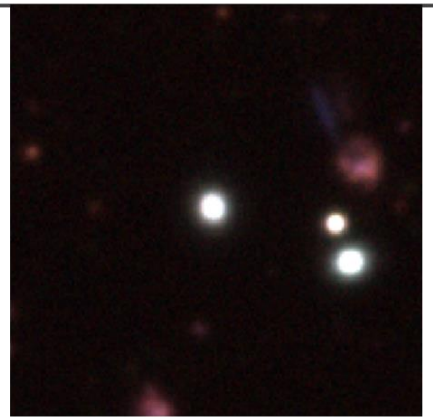


Same galaxy
GALAXY ID 99 | Legacy Survey

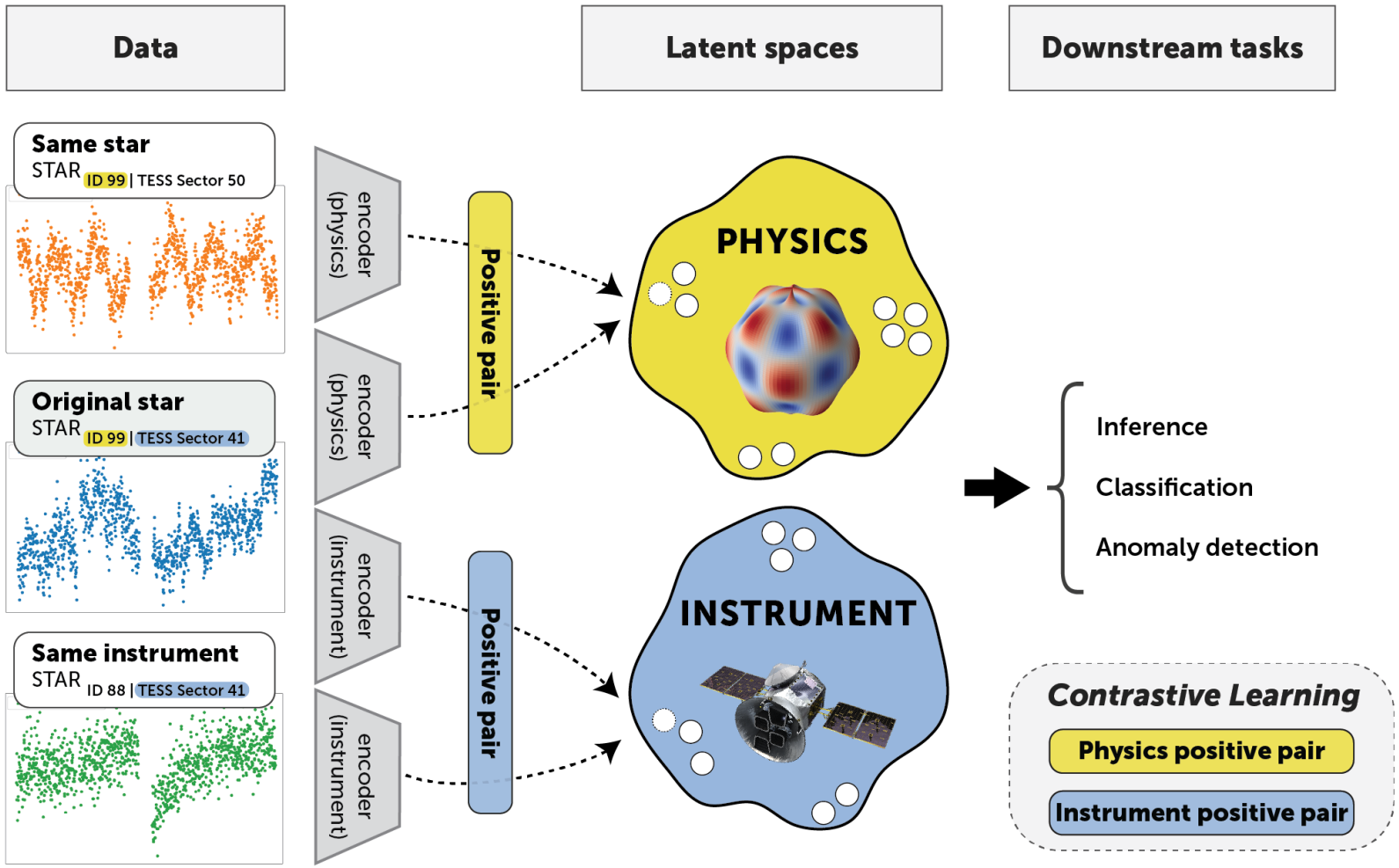
Anchor image
GALAXY ID 99 | HSC



Same instrument
GALAXY ID 88 | HSC

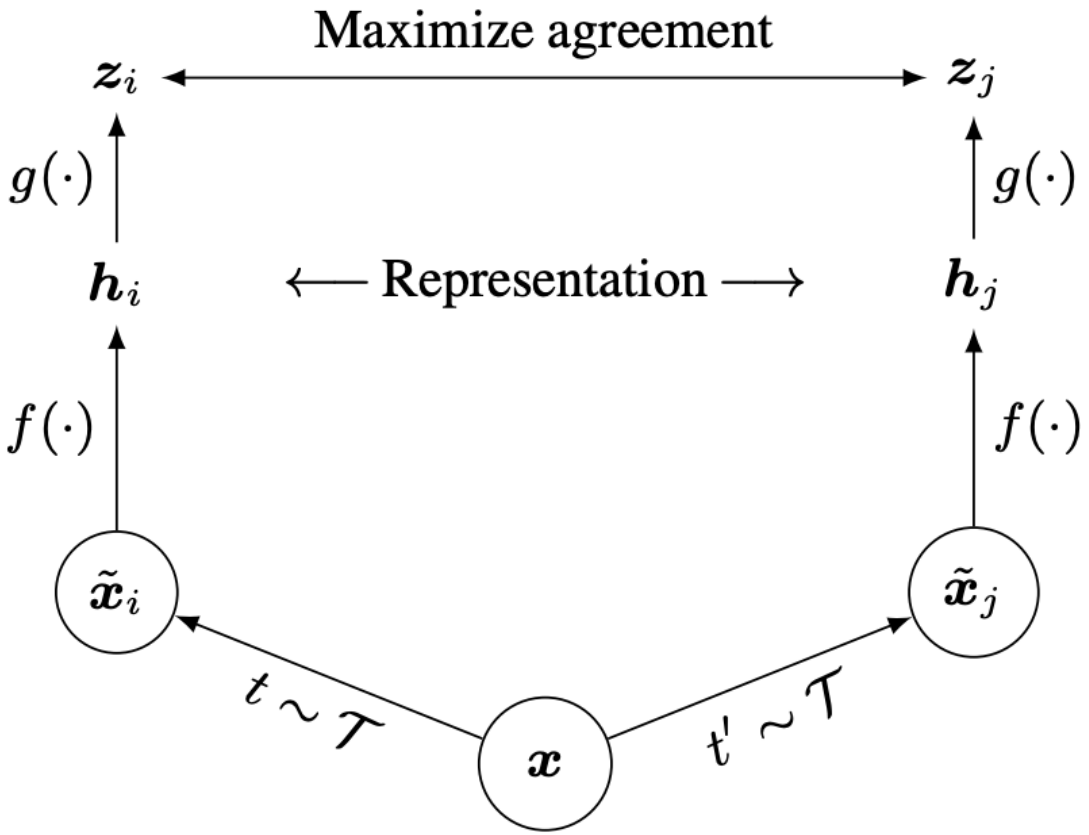


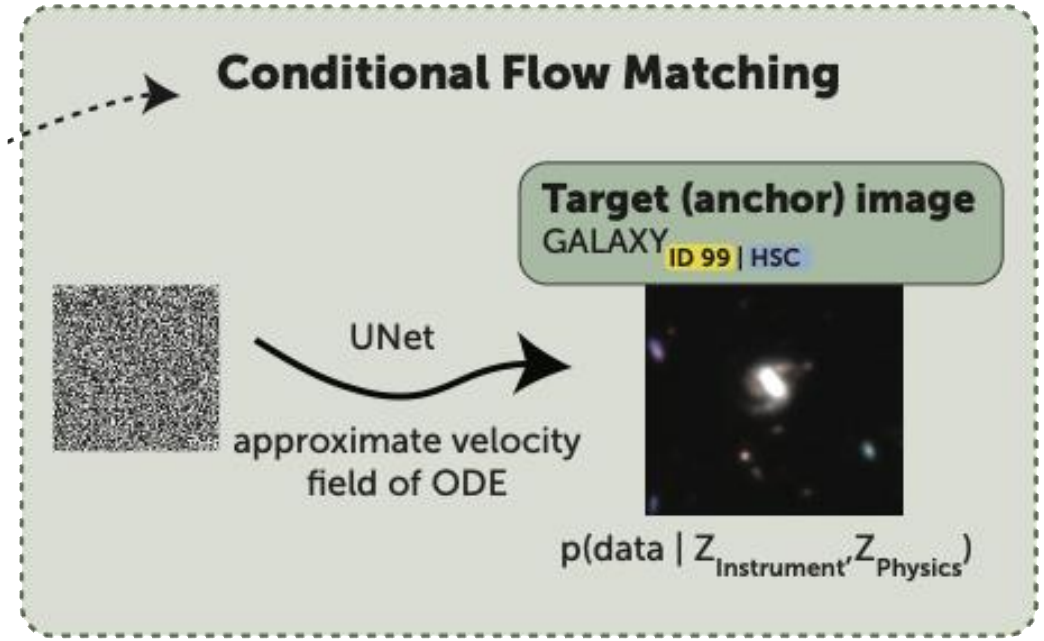
Causal foundation model



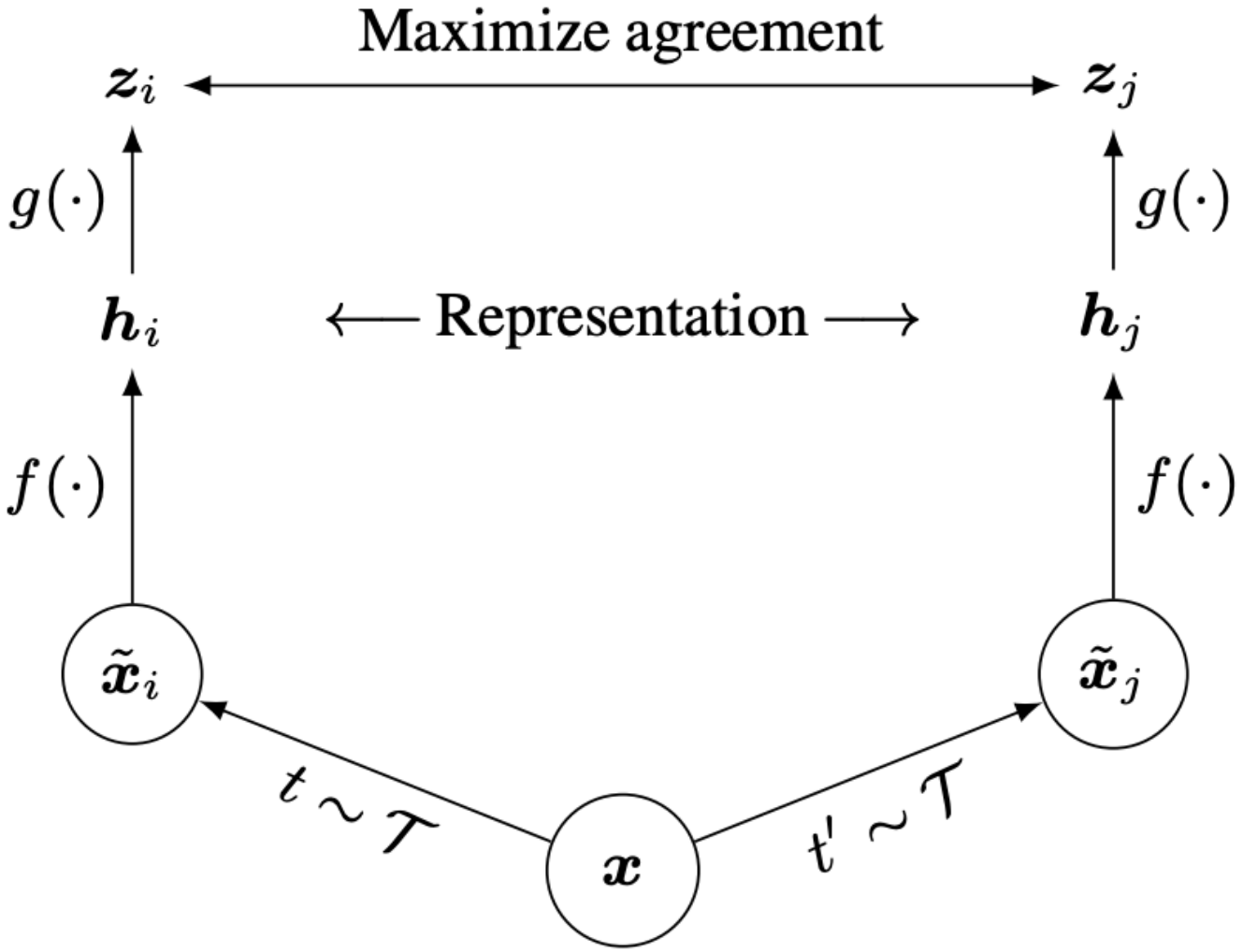
Audenaert, Muthukrishna, Gregory, Hogg, Villar (2025)

Contrastive vs Generative objective



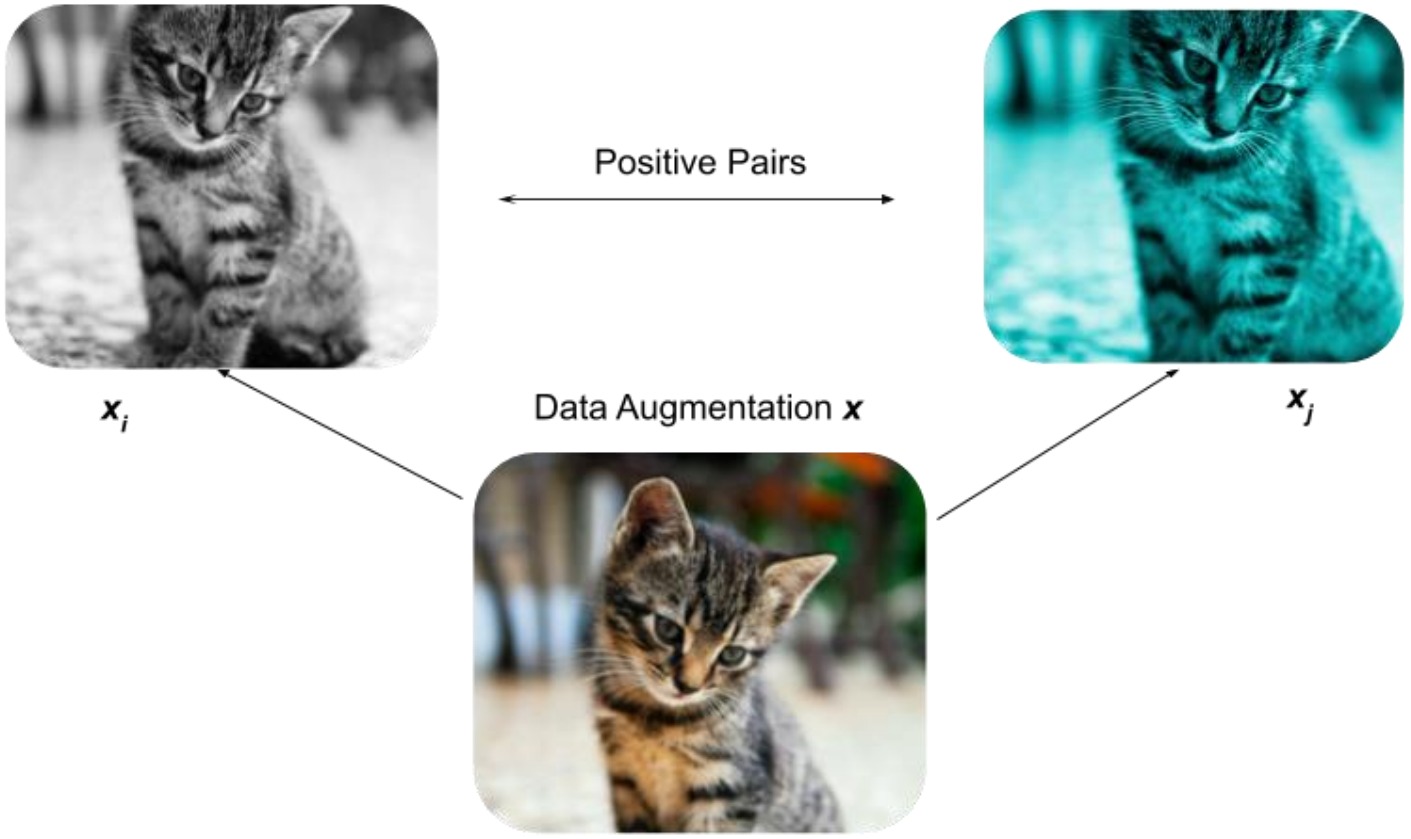


SimCLR – Contrastive Learning



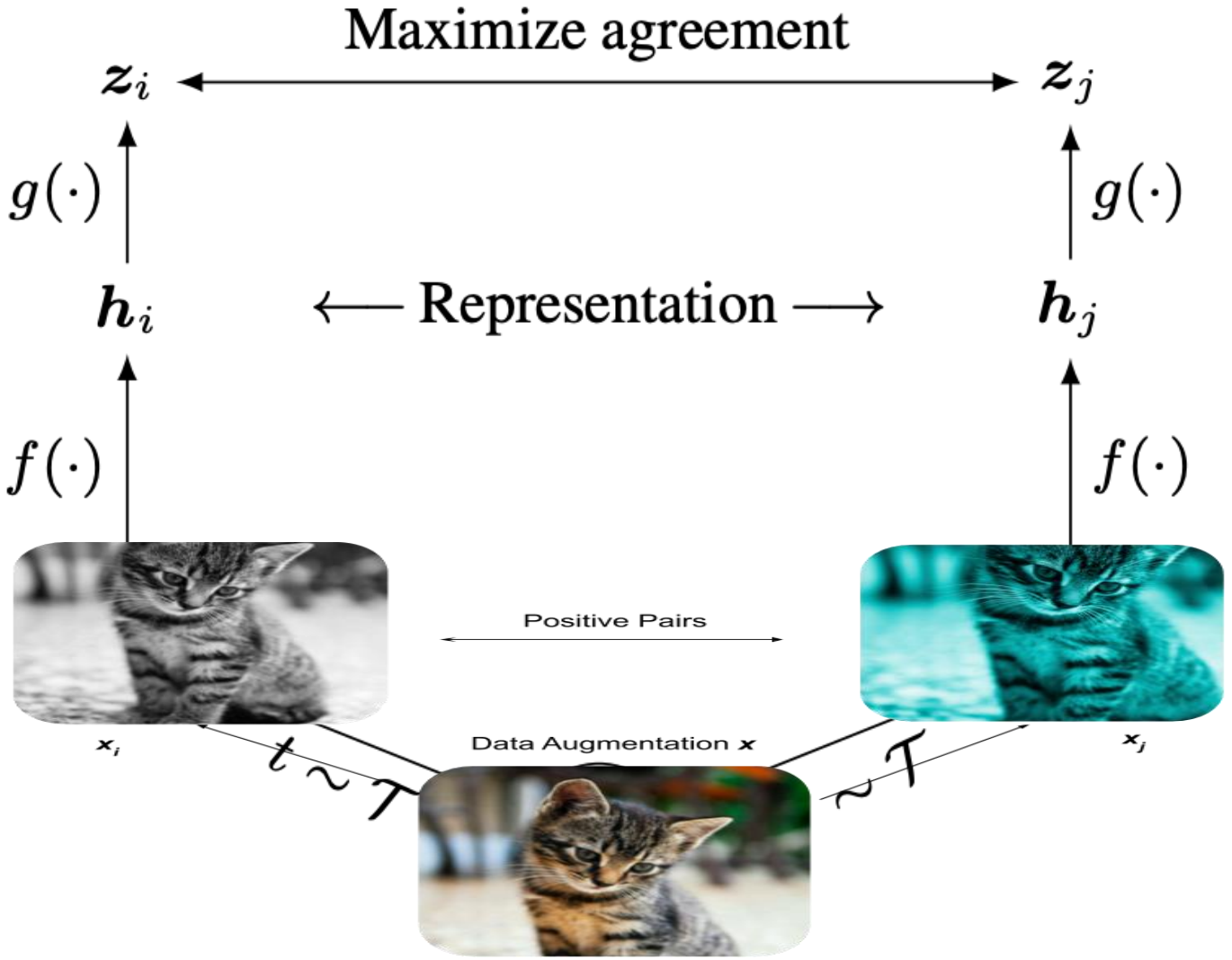
Chen et al. (2020)

SimCLR – Contrastive Learning



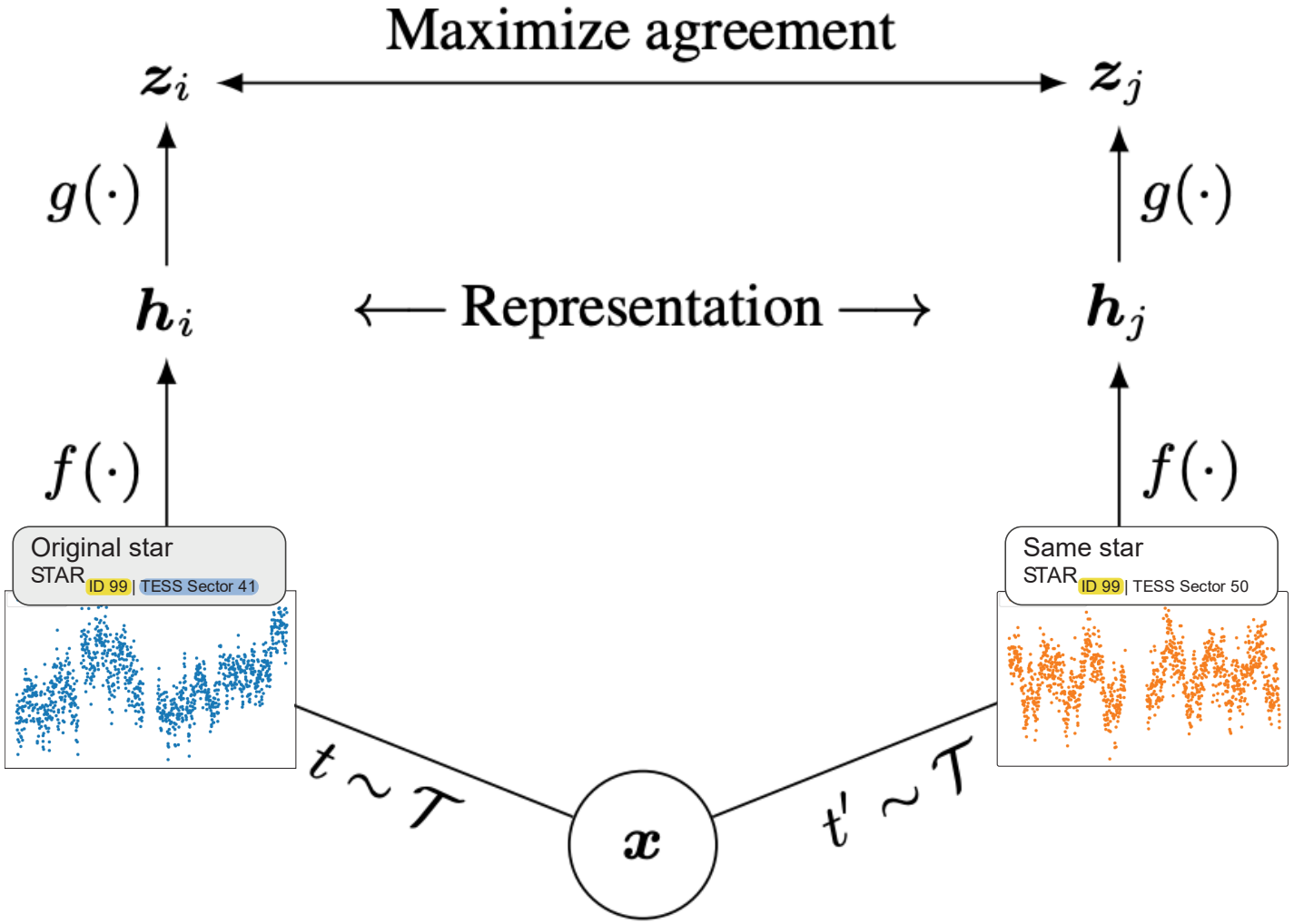
Chen et al. (2020)

SimCLR – Contrastive Learning



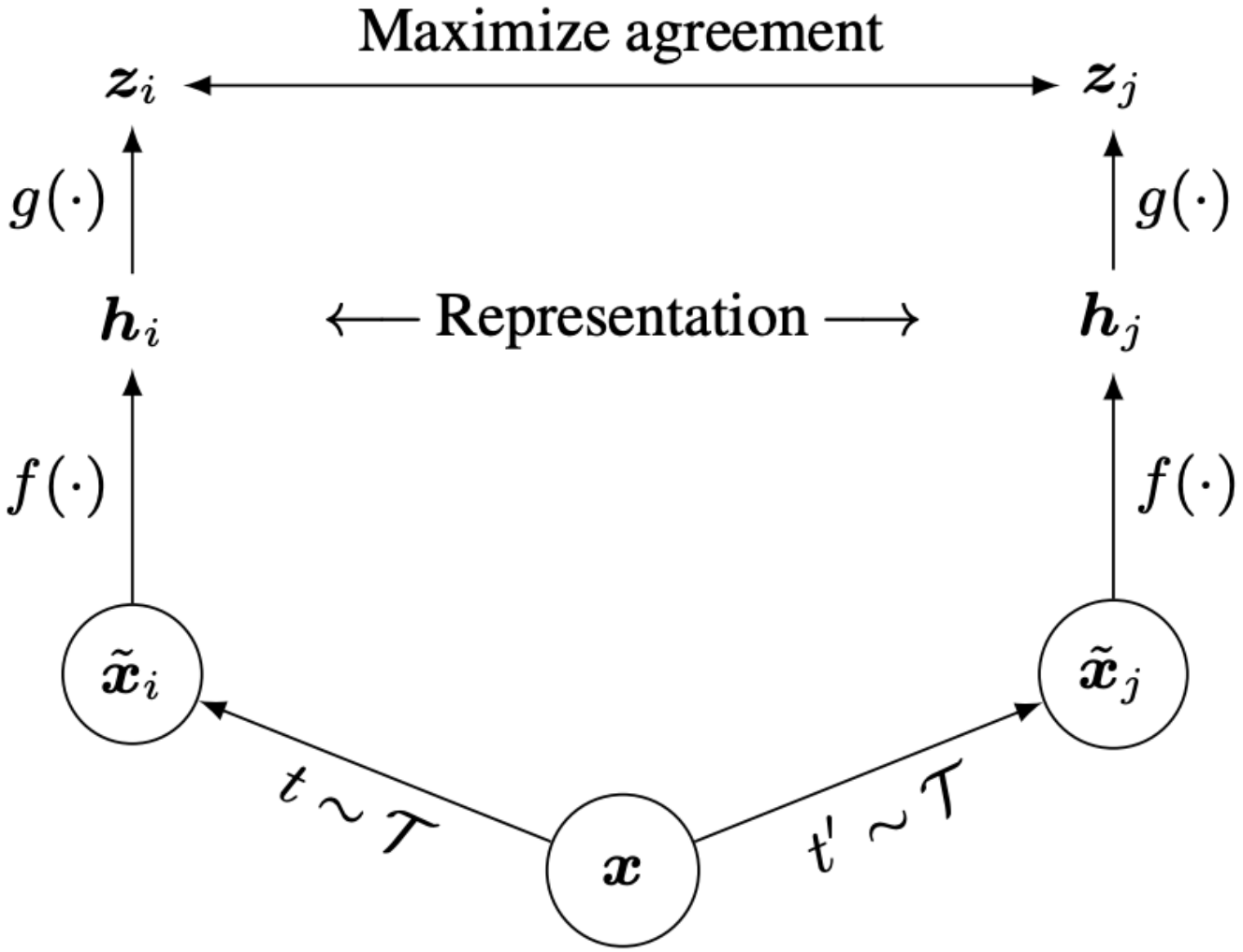
Chen et al. (2020)

SimCLR – Contrastive Learning



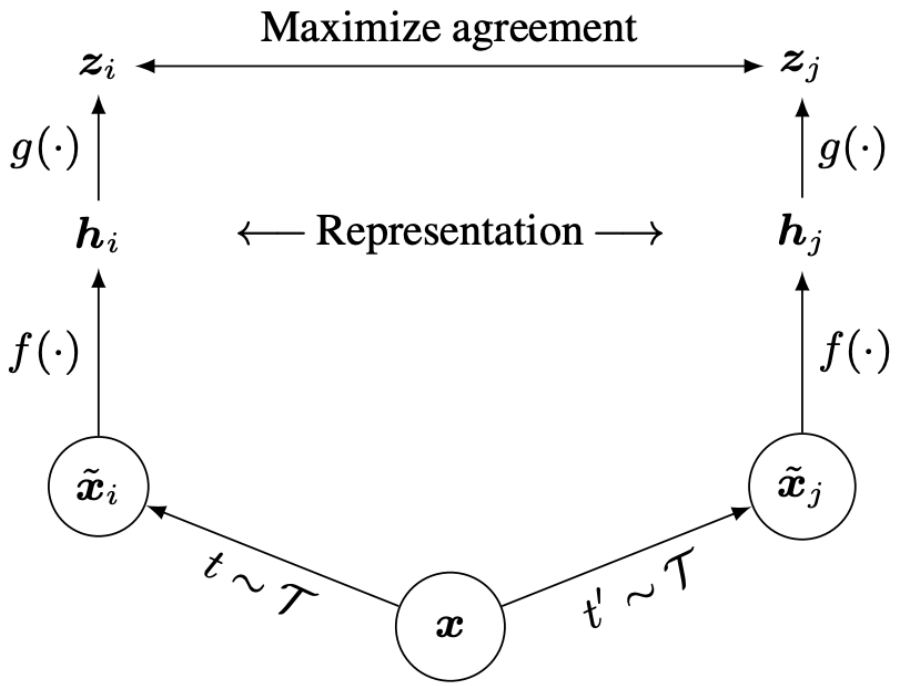
Chen et al. (2020)

SimCLR – Contrastive Learning

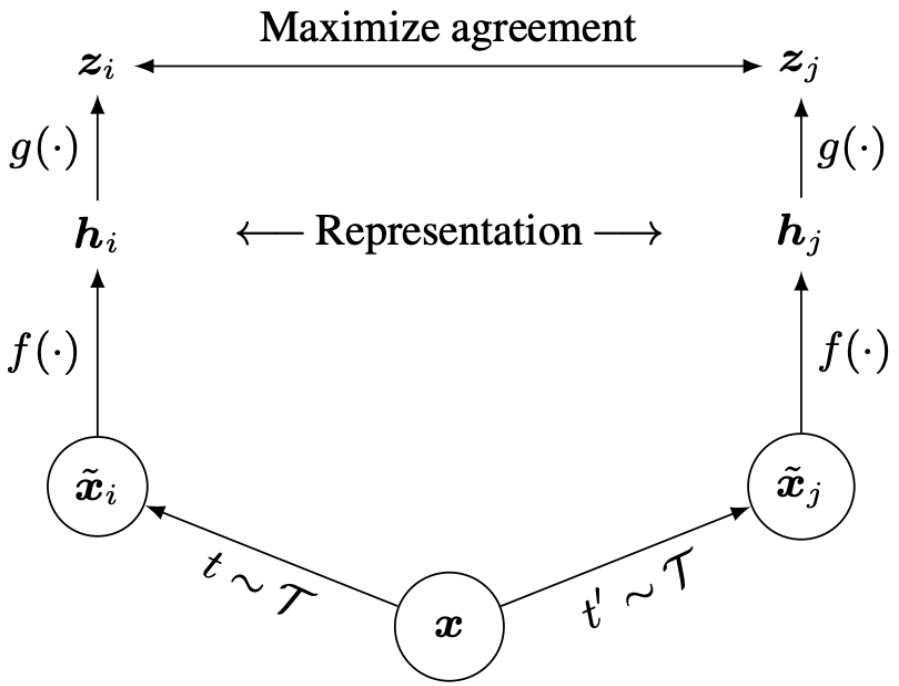


Chen et al. (2020)

SimCLR – Contrastive Learning



Physics Space



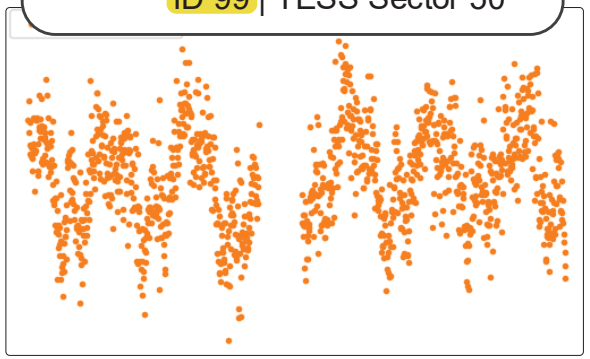
Instrument Space

Massachusetts Institute of Technology

Making Triplets

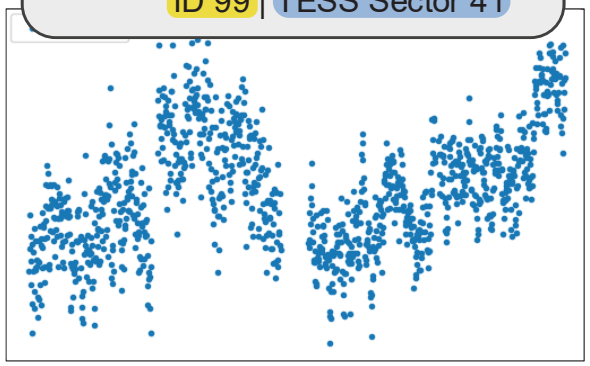
Same star

STAR ID 99 | TESS Sector 50



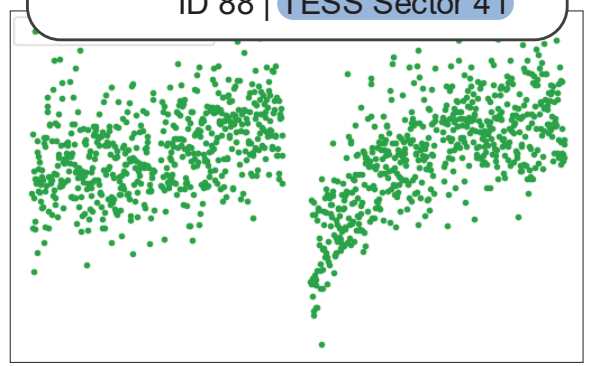
Original star

STAR ID 99 | TESS Sector 41

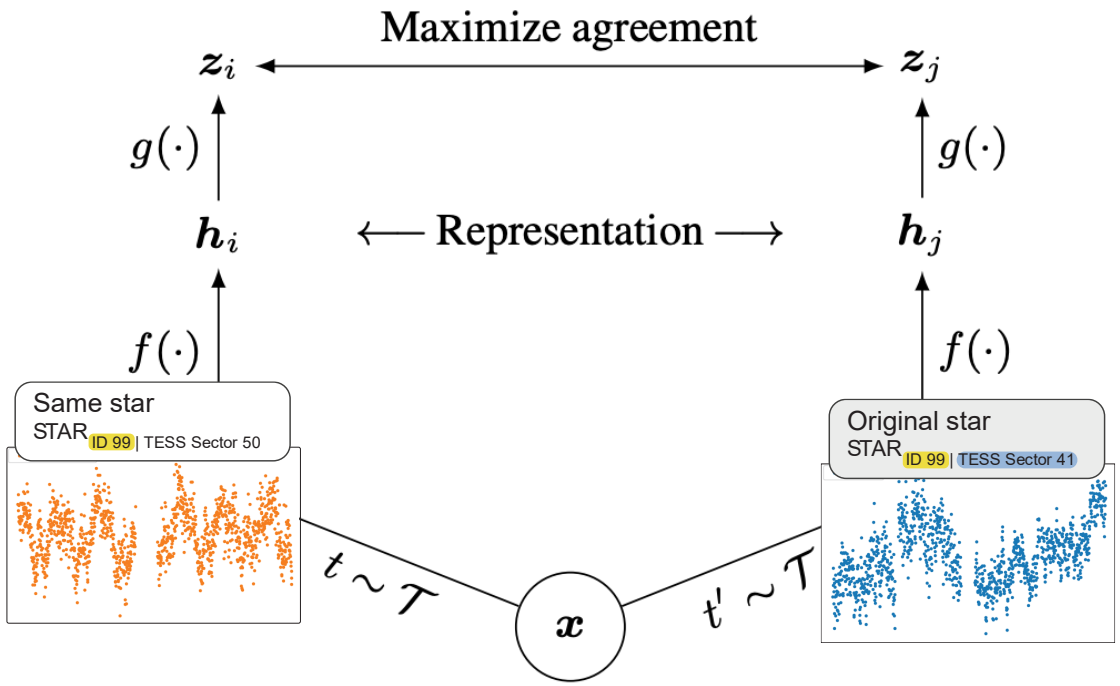


Same instrument

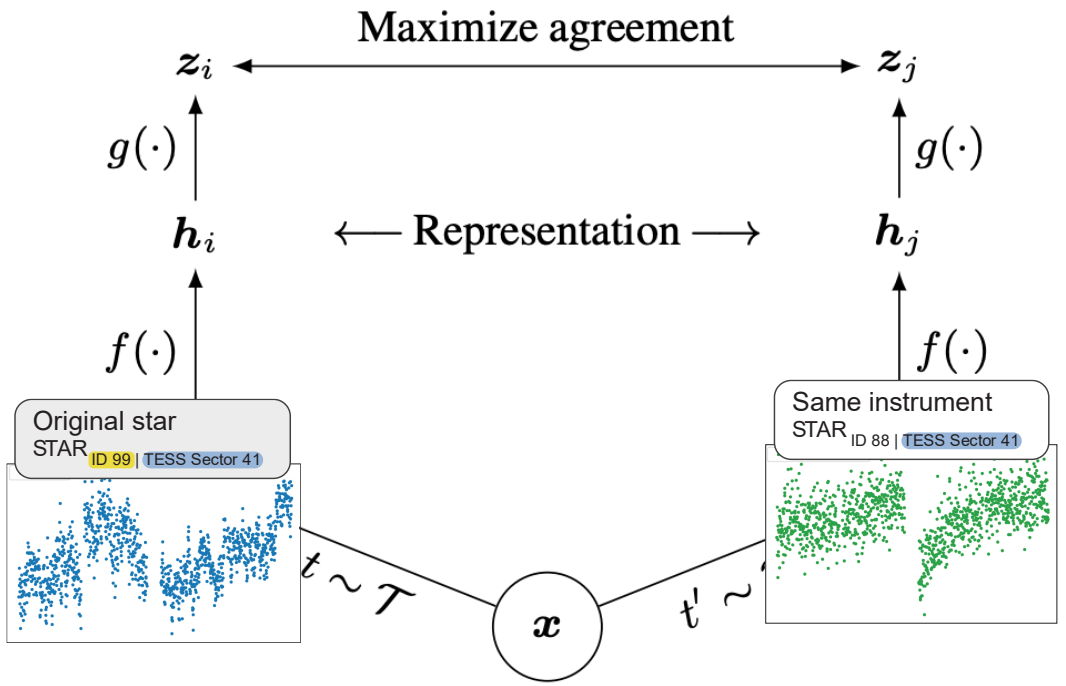
STAR ID 88 | TESS Sector 41



SimCLR – Contrastive Learning

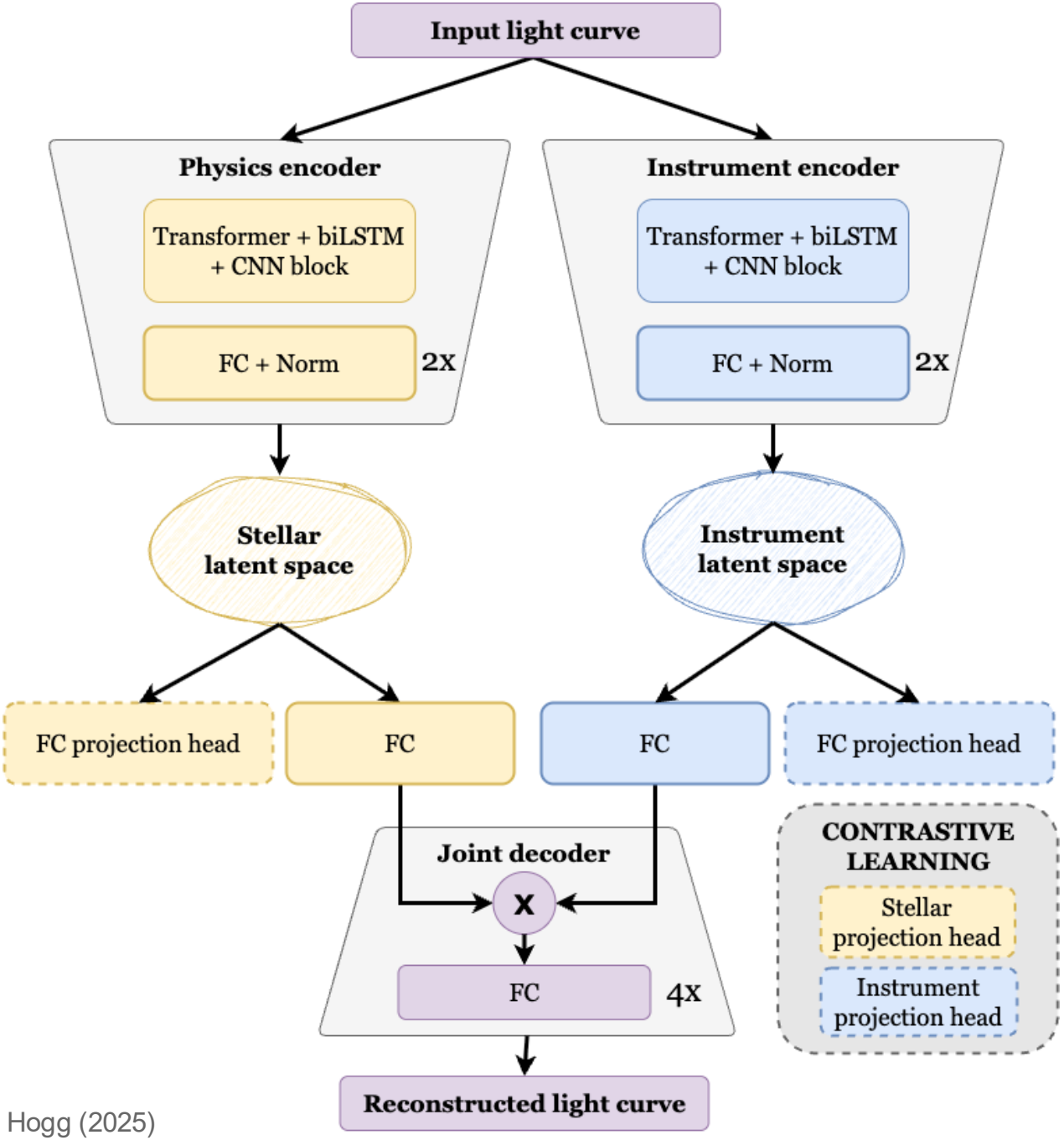


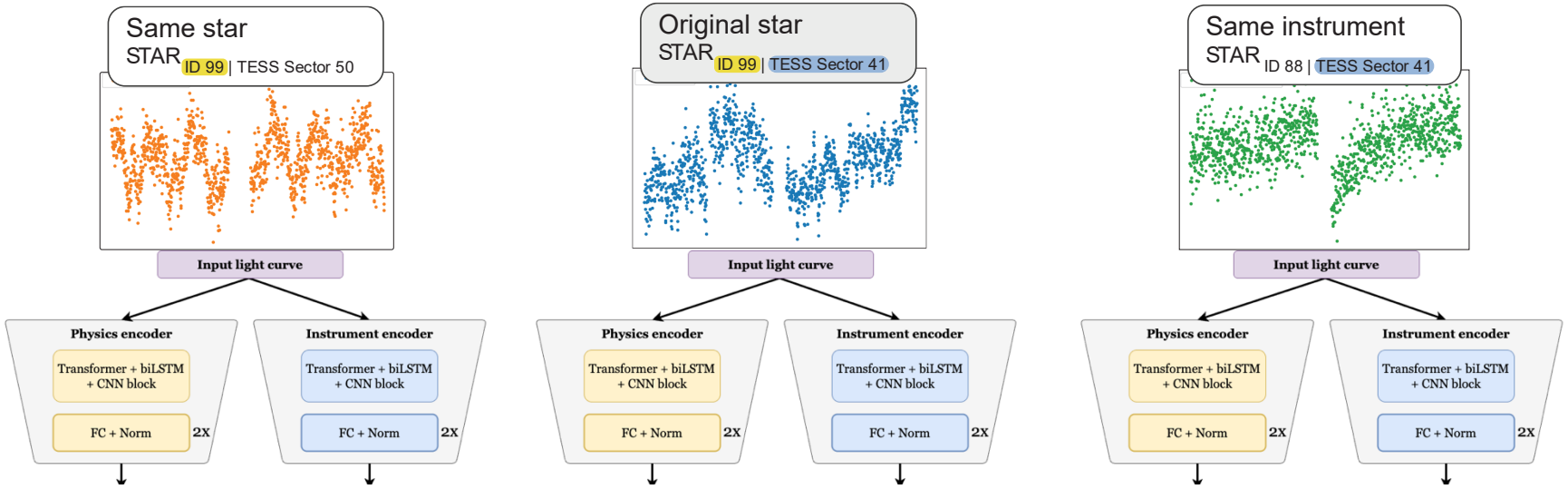
Physics Space

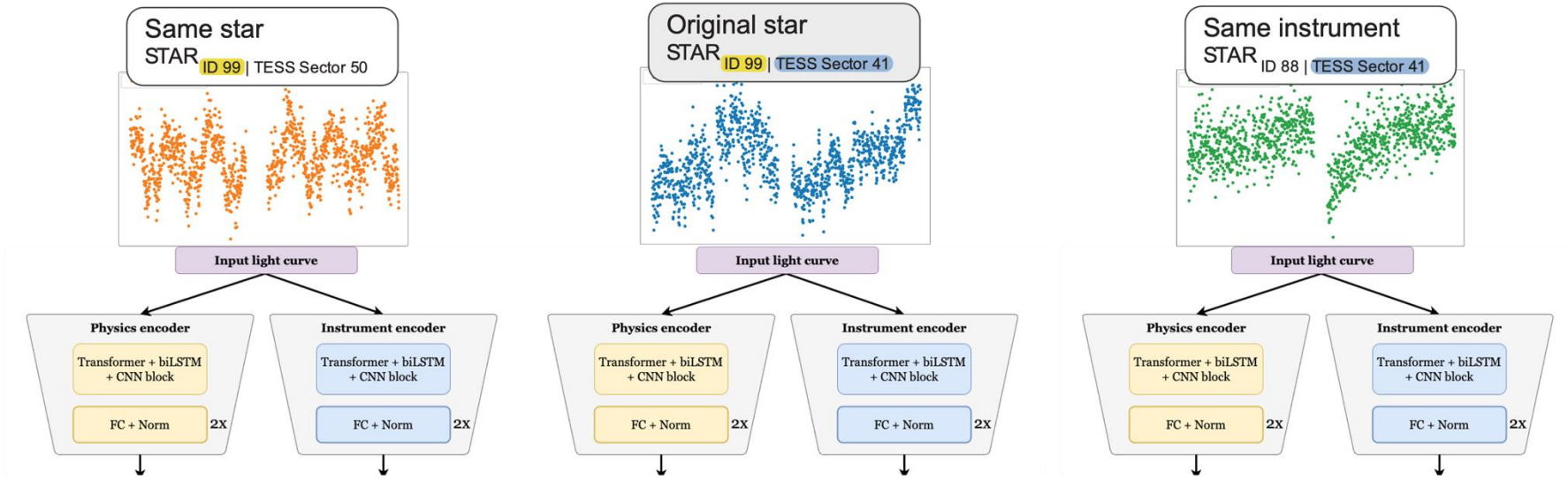


Instrument Space

Massachusetts Institute of Technology

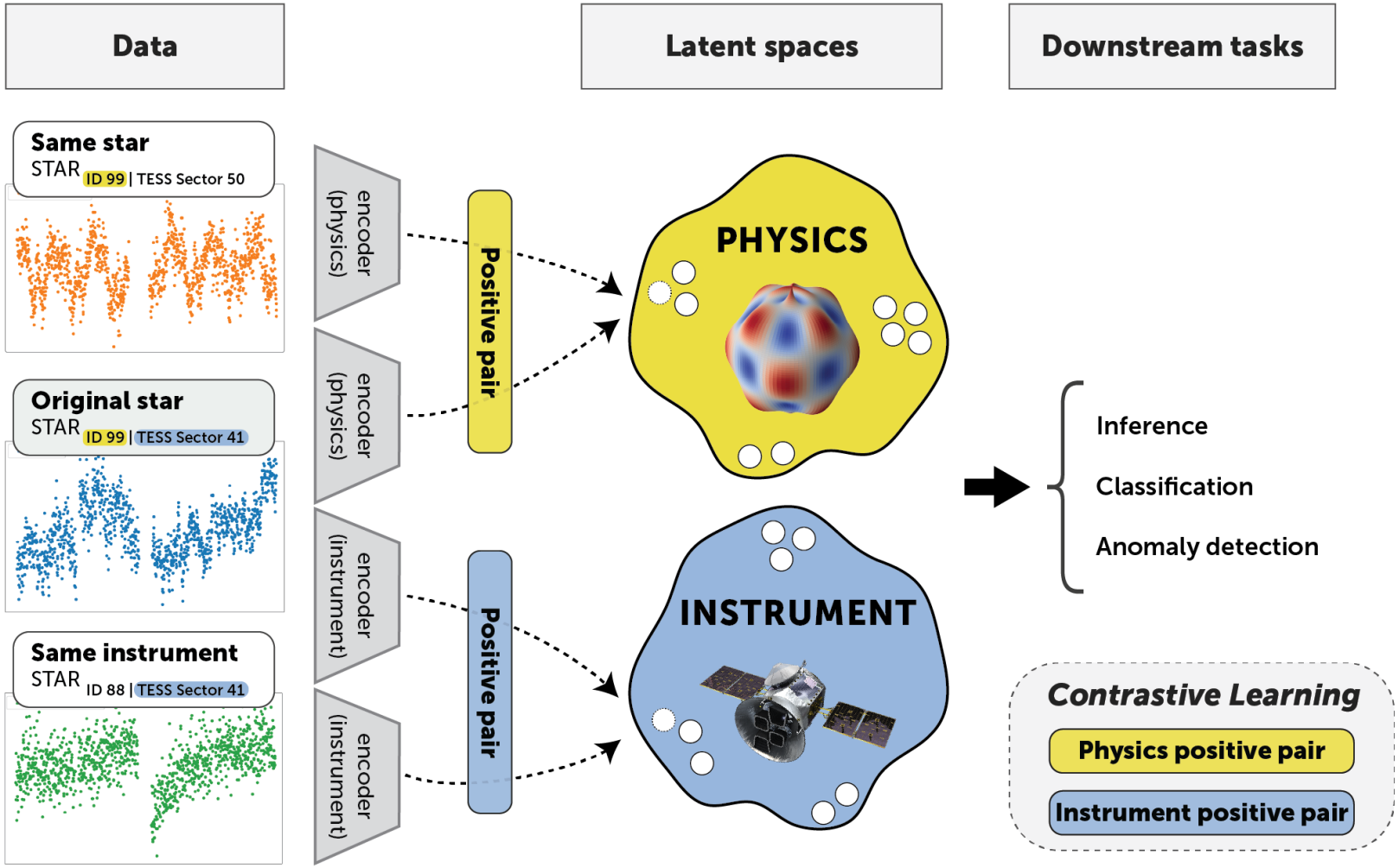

Input light curve
Physics encoder
Transformer + biLSTM + CNN block
FC + Norm
2X
Instrument encoder
Transformer + biLSTM + CNN block
FC + Norm
2X
Stellar latent space
Instrument latent space
FC projection head
FC
FC
FC projection head
CONTRASTIVE LEARNING
Stellar projection head
Instrument projection head
Joint decoder
X
FC
4X
Reconstructed light curve


Same star
STAR ID 99 | TESS Sector 50
Original star
STAR ID 99 | TESS Sector 41
Same instrument
STAR ID 88 | TESS Sector 41
Input light curve
Physics encoder
Instrument encoder
Transformer + biLSTM + CNN block
FC + Norm
2X


Same star
STAR ID 99 | TESS Sector 50
Original star
STAR ID 99 | TESS Sector 41
Same instrument
STAR ID 88 | TESS Sector 41
Input light curve
Physics encoder
Instrument encoder
Transformer + biLSTM + CNN block
FC + Norm
2x

Causal Foundation Model

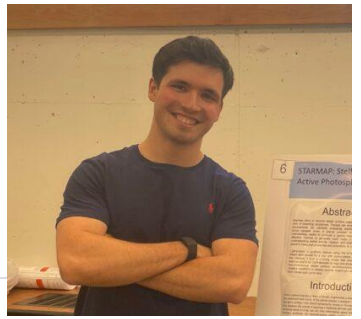
arXiv:2507.05333



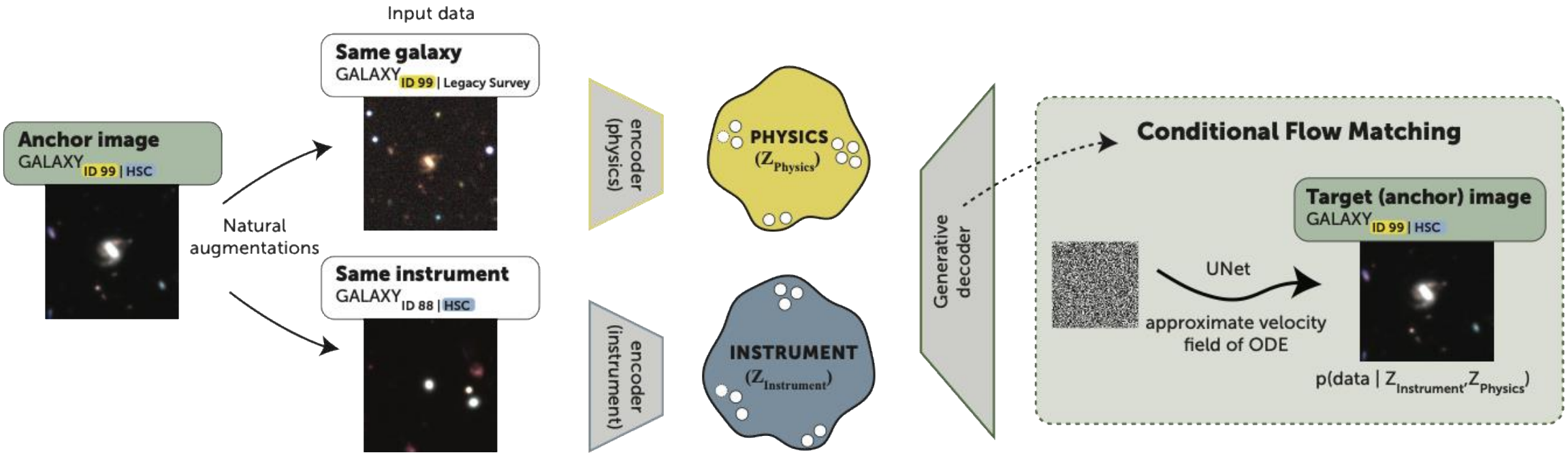
Muthukrishna*, Audenaert*, Gregory, Villar, Hogg (2025)

Massachusetts Institute of Technology

CausalFM — disentangling physics from instrument
Preliminary work on real data

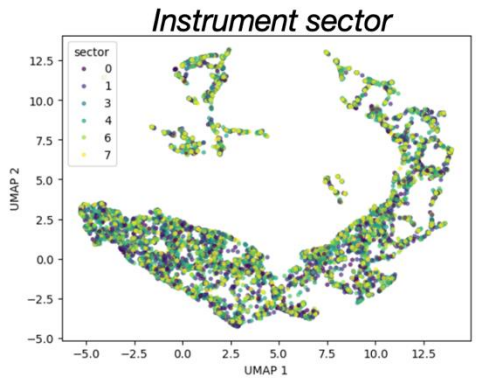


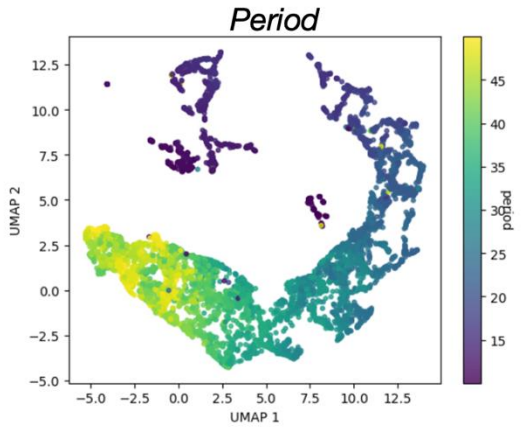
Pablo Mercader-Perez
MIT PhD Student

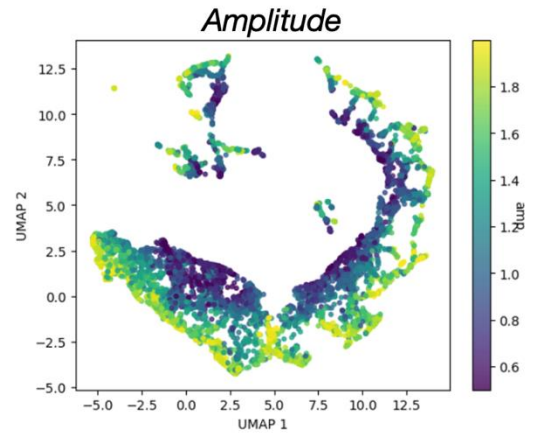


Muthukrishna, Audenaert et al. 2025 · https://arxiv.org/abs/2507.05333
Mercader-Perez, Cuesta, Muthukrishna et al. 2026 · https://arxiv.org/abs/2604.09787

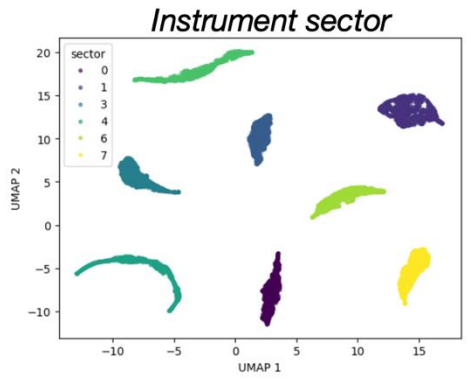
Physics Latent Space

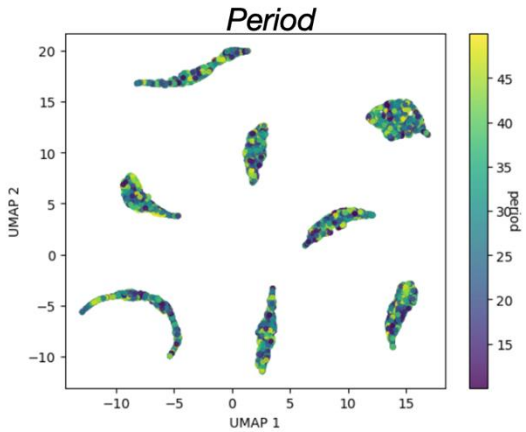


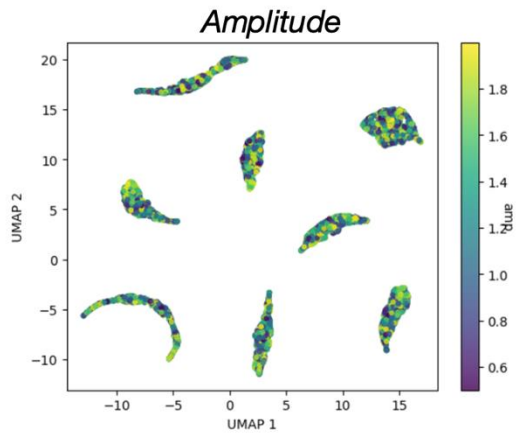




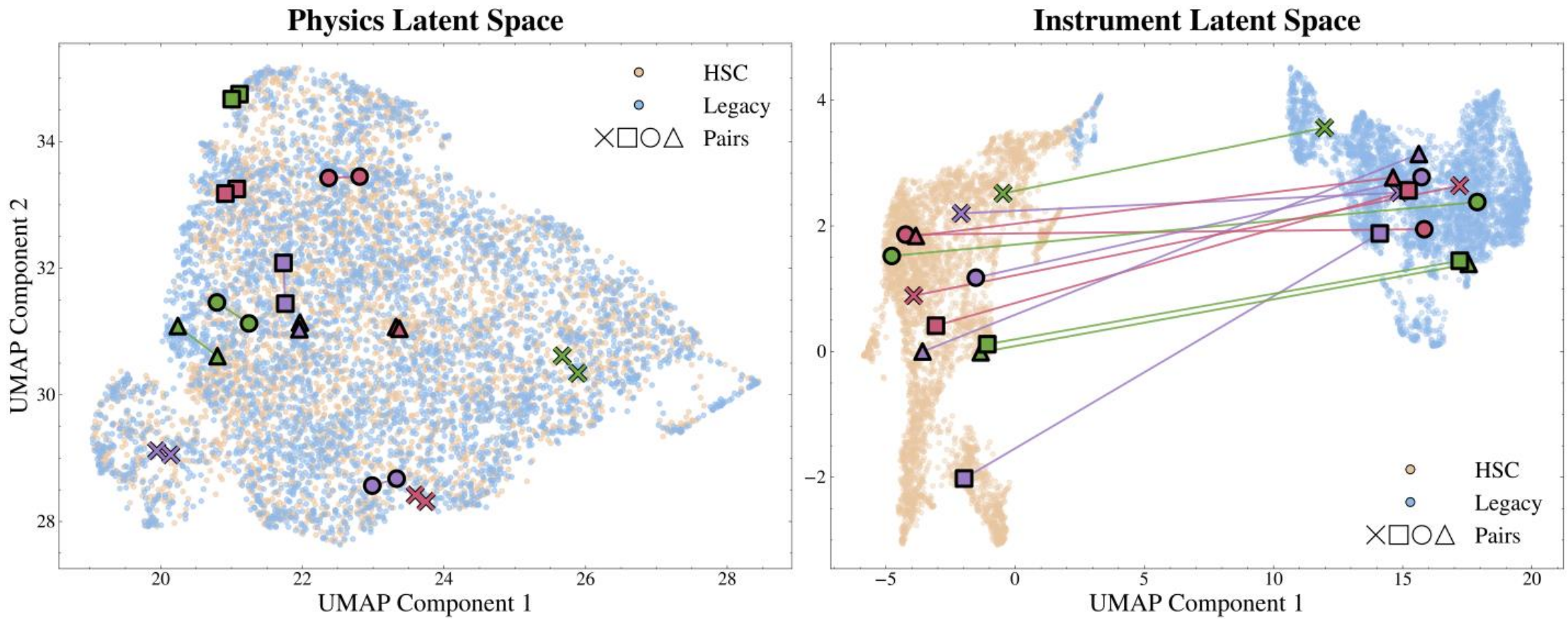
Instrument Latent Space



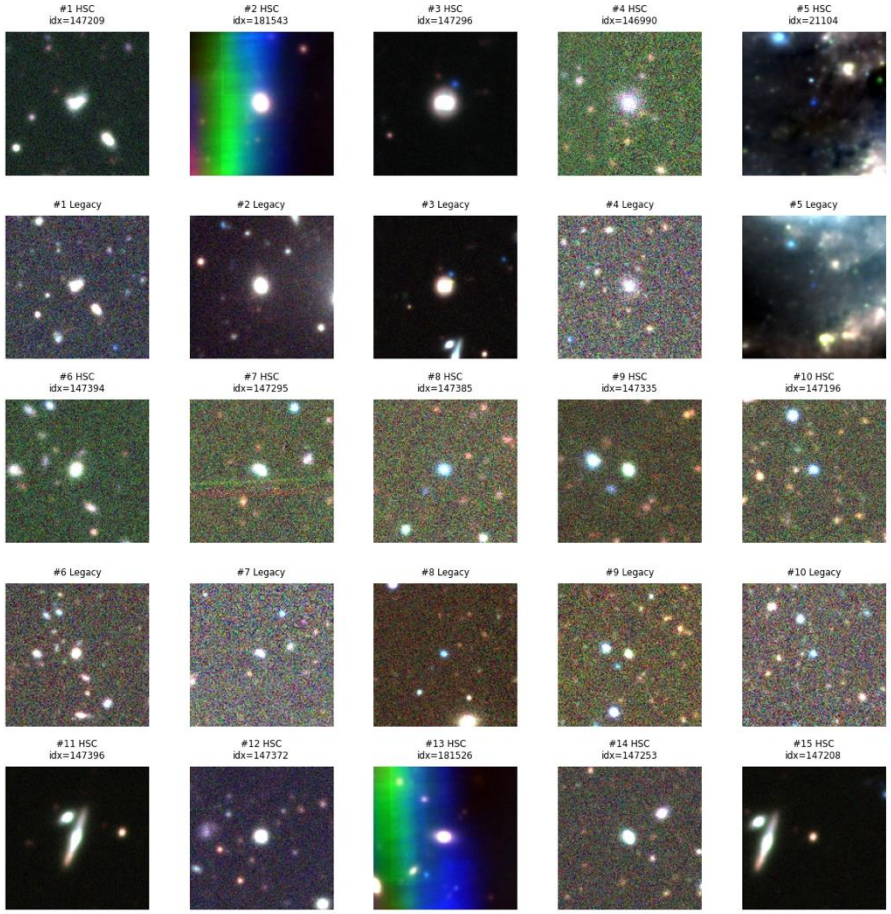




Latent Space

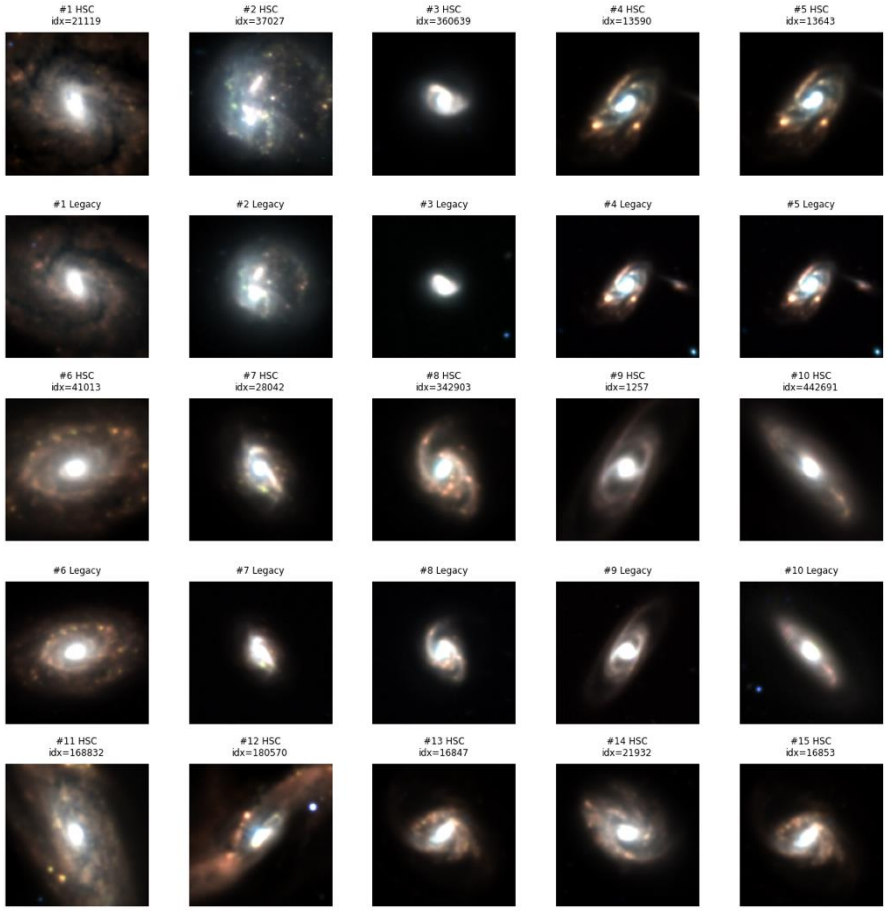


AION's (state of the art) top 25 anomalies



AION, Polymathic Team, Parker+ 2025, https://arxiv.org/abs/2510.17960

CausalFM – Physics latent top 25 anomalies

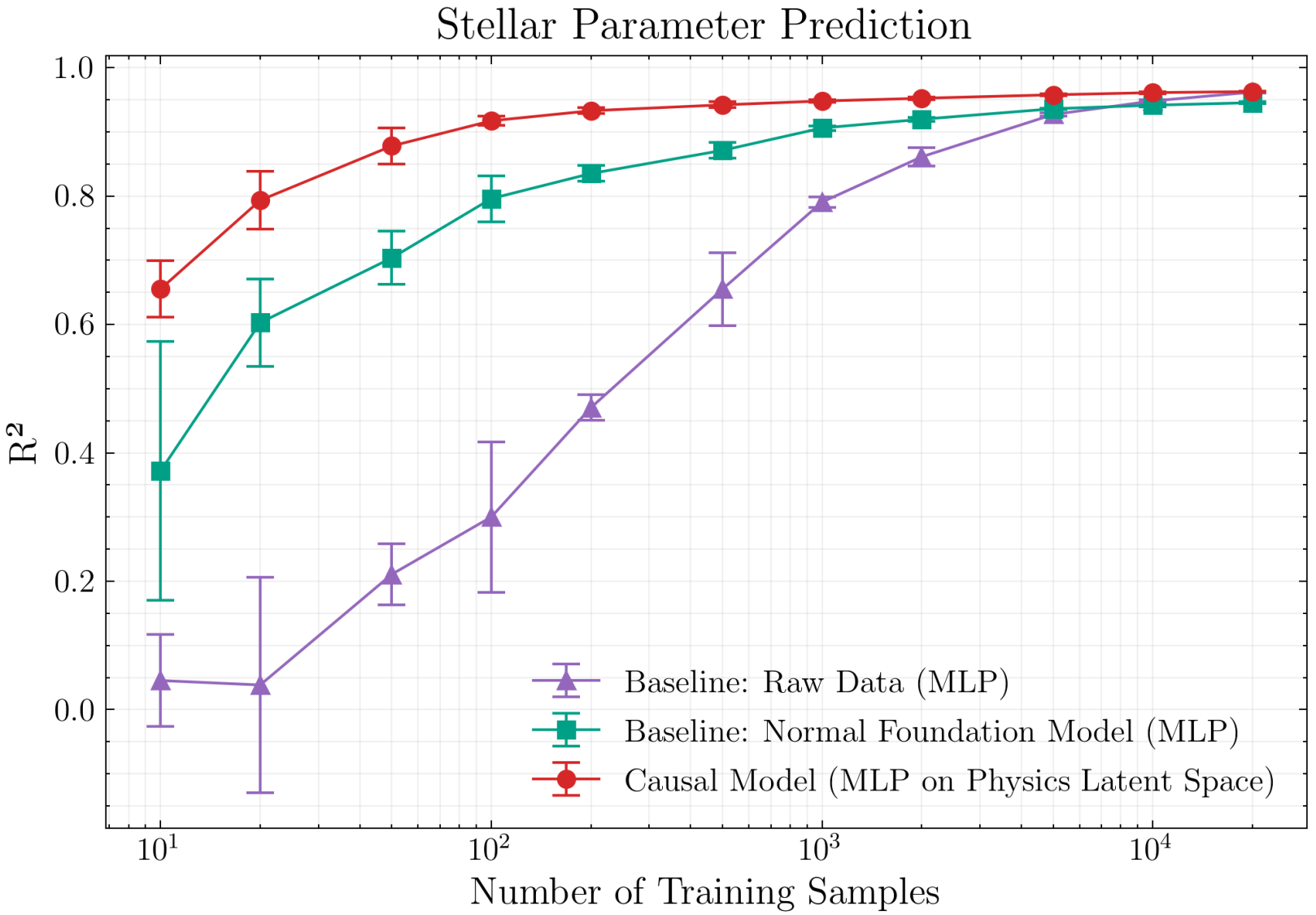


Pablo Mercader-Perez
MIT PhD Student

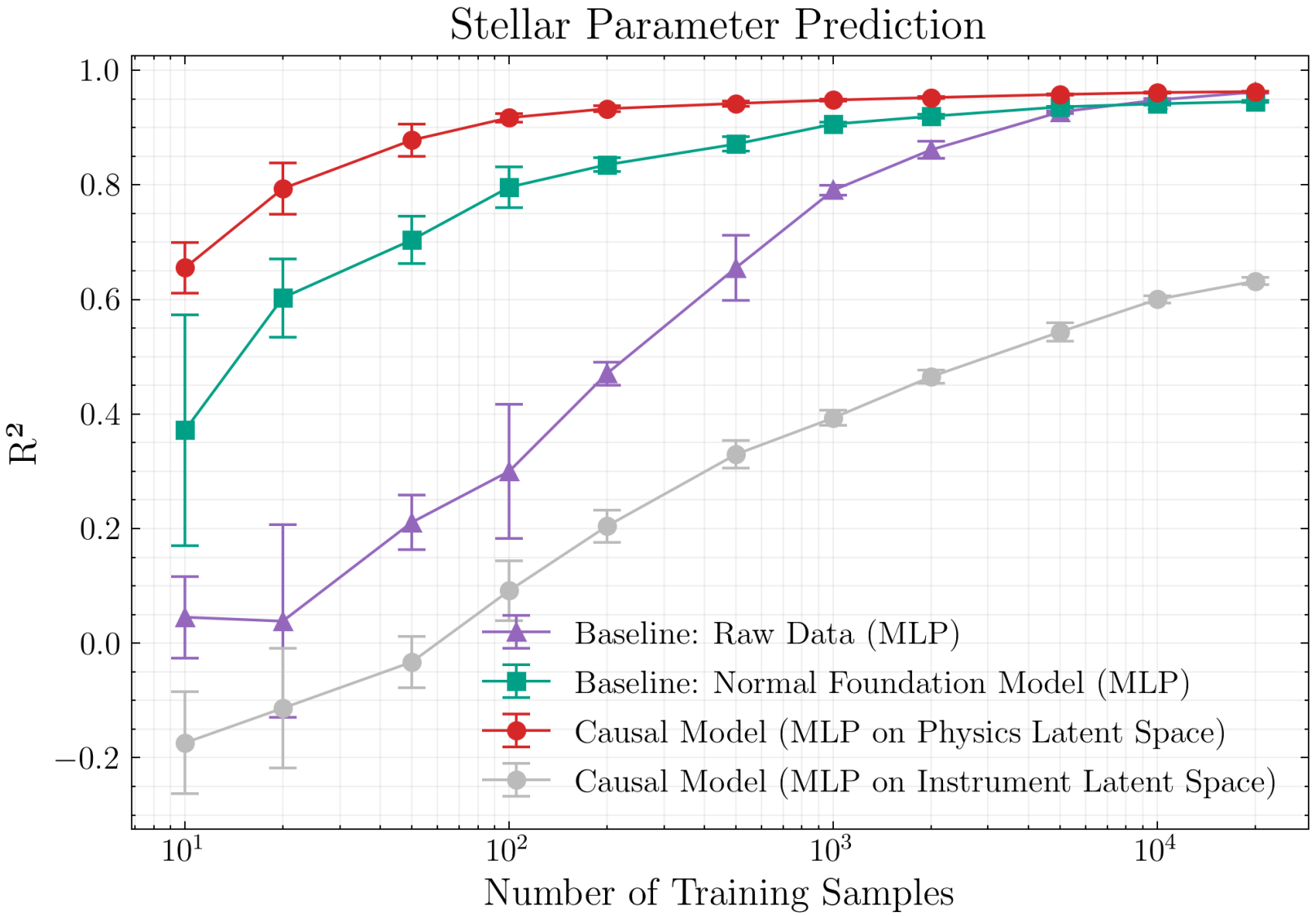
Paper at ICLR applying novel CausalFM to galaxy images
*Mercader-Perez, Cuesta, **Muthukrishna** et al. 2026 · https://arxiv.org/abs/2604.09787*

→ Standard methods' anomalies are often instrument artifacts. ClearSkyAI will find **real discoveries**

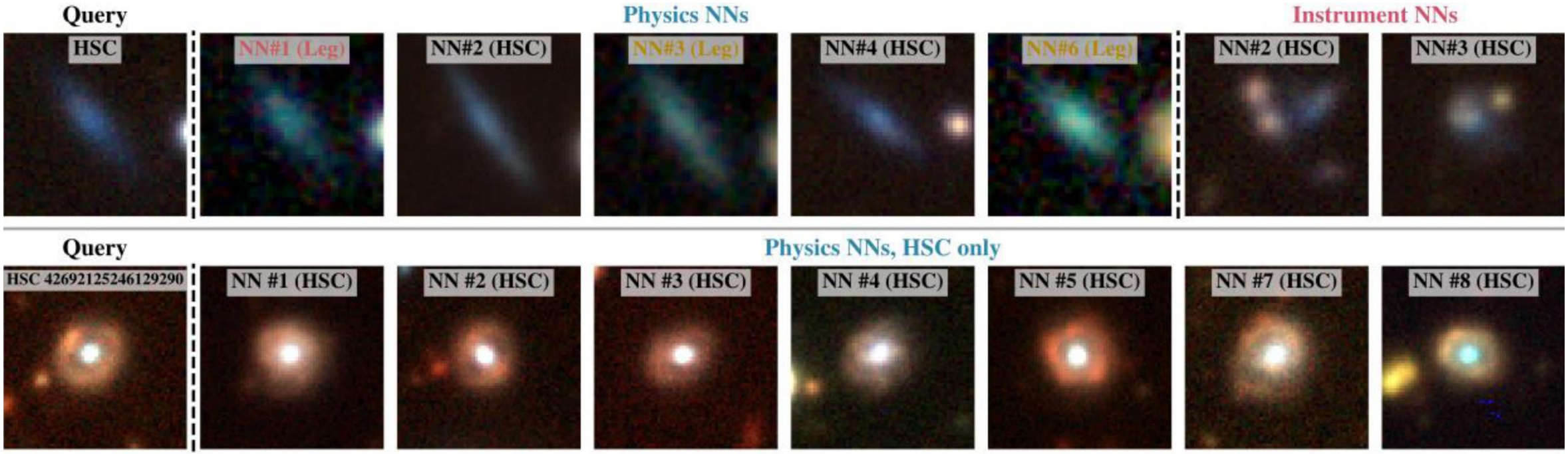
Downstream Tasks Comparison



Downstream Tasks Comparison



Instrument-Invariant Search and Rare Object Discovery



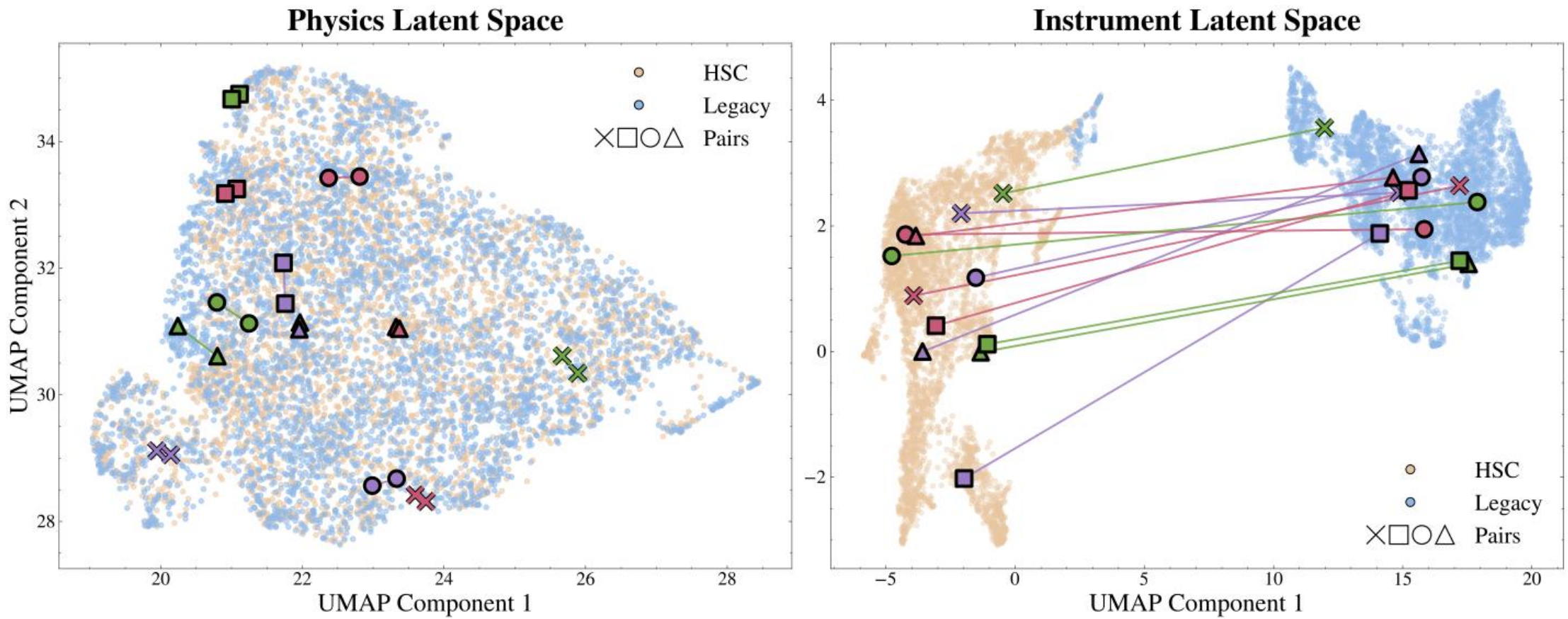
Nearest neighbours in the physics latent space across Legacy and HSC, enabling cross-survey object search.

Searching for known strong gravitational lens as query yields interesting examples.

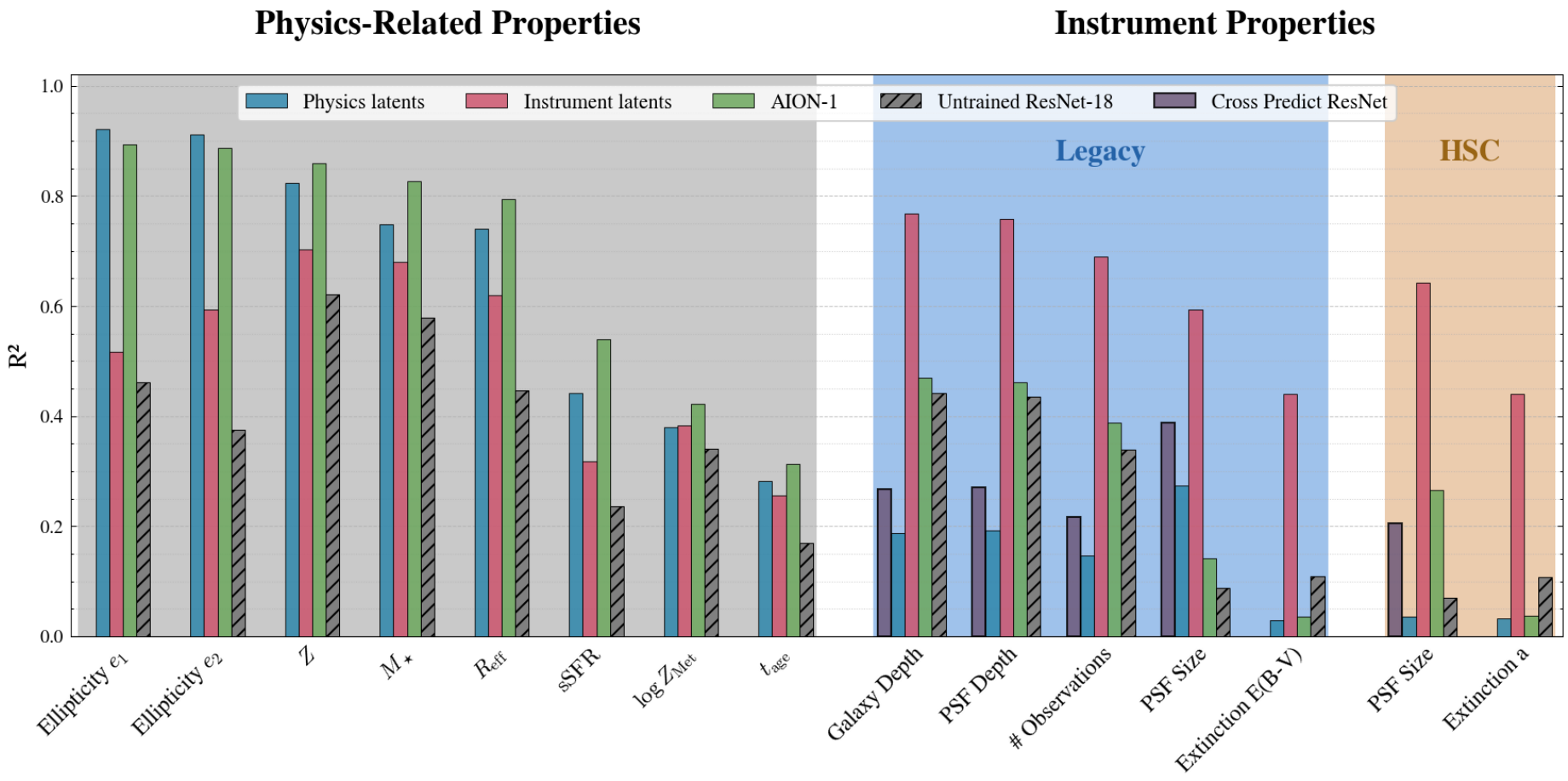
Conclusions

- We know our instruments cause distortions, let's build this into the model!
- Building foundation models using a causally-motivated framework may lead to **better performance than standard single-latent space FMs** and can enable a **domain-agnostic (survey-independent)** model
- Our Causal FM consists of:
  - **Two latent spaces** for the Physics and Instruments
- A generative approach can condition on multiple telescopes at once and enable cross-survey generation
  - i.e. what would this galaxy look like if observed by a different instrument?

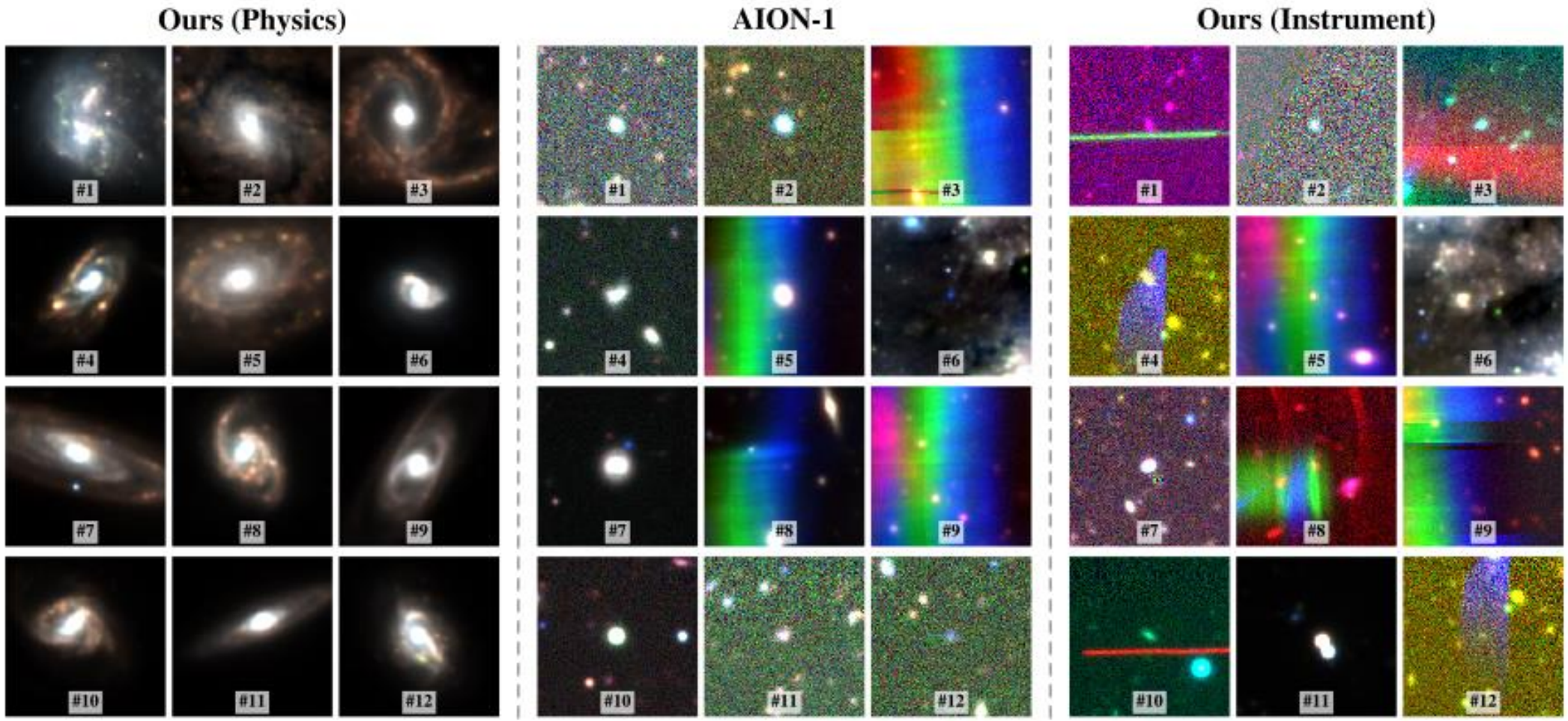
Latent Space



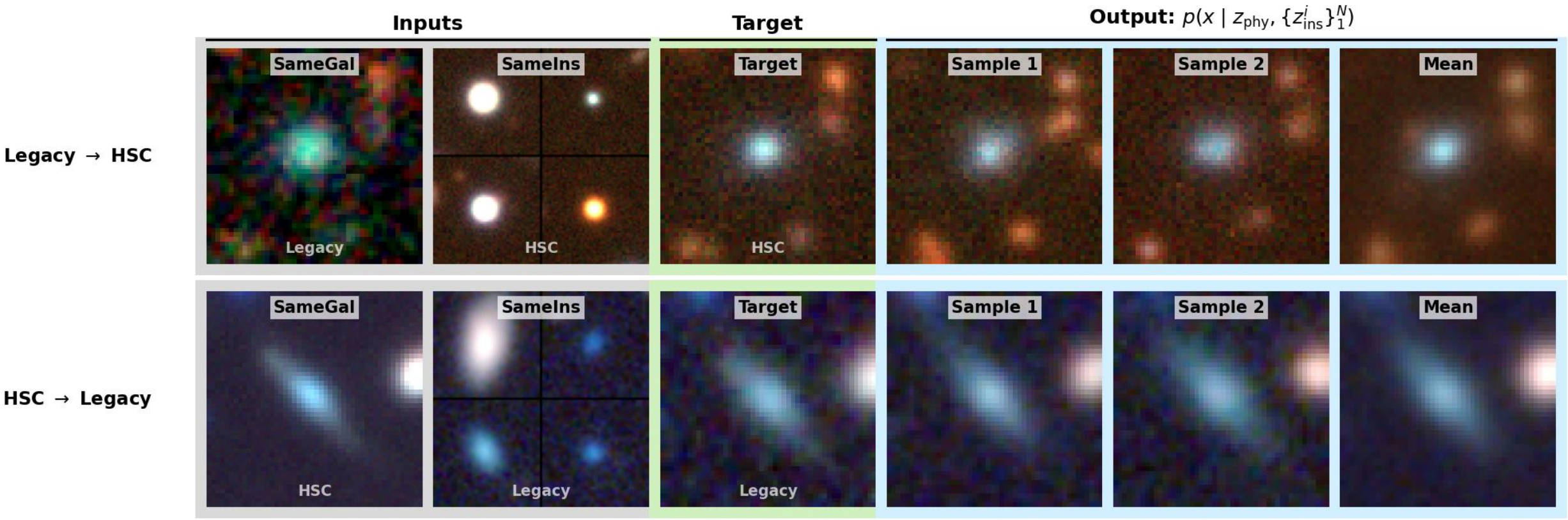
Downstream Tasks



Anomaly Detection



Counterfactual Reconstruction



Nearest Neighbours Search