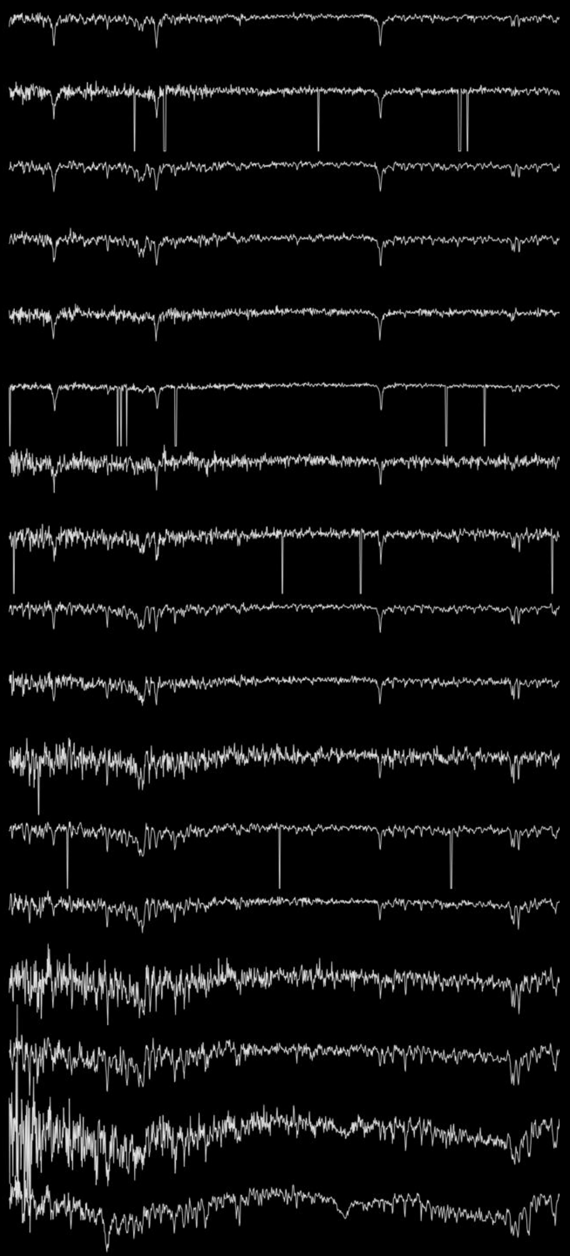
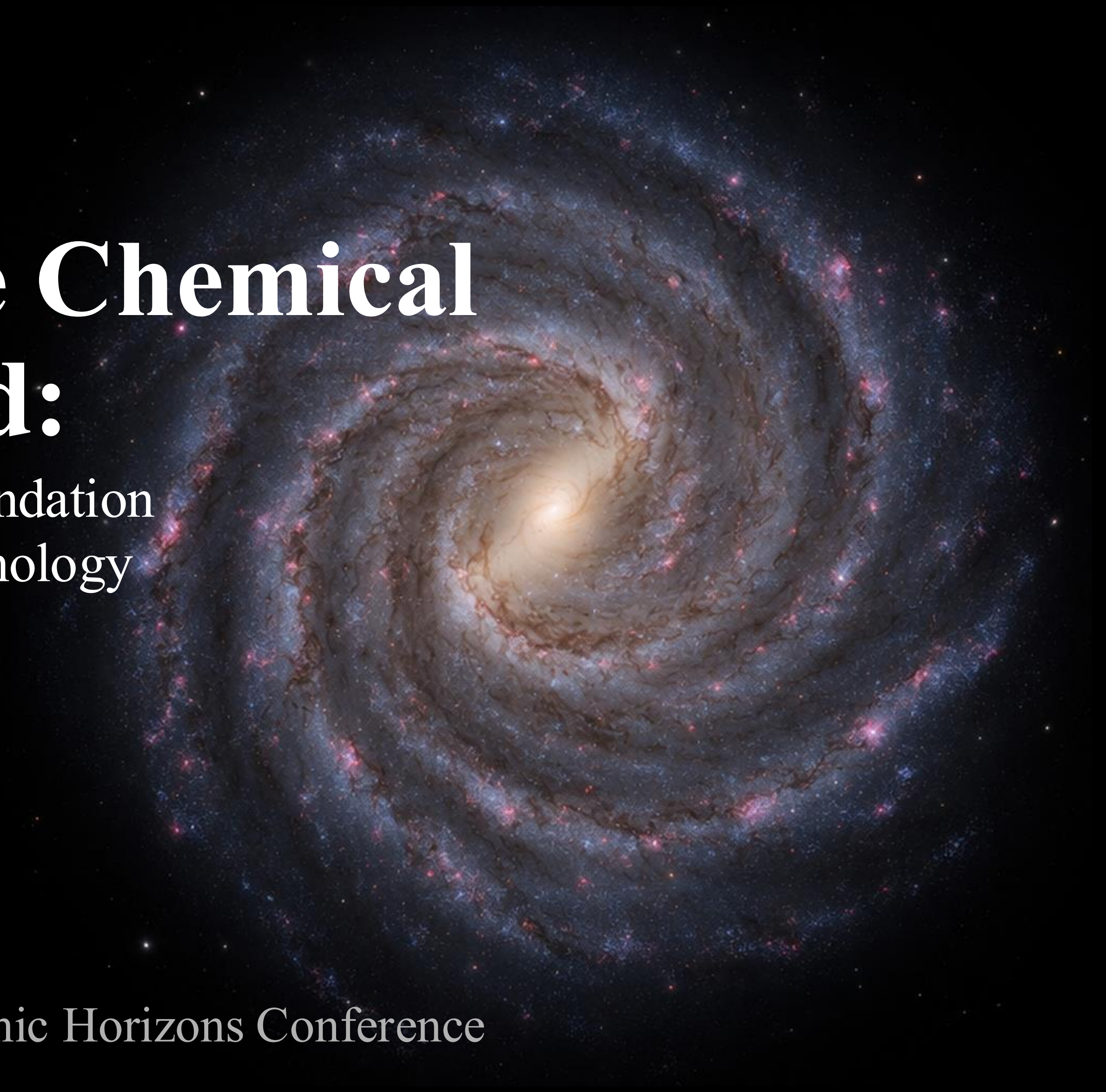


Decoding the Chemical Fossil Record:

Machine Learning and Foundation Models in Near-Field Cosmology

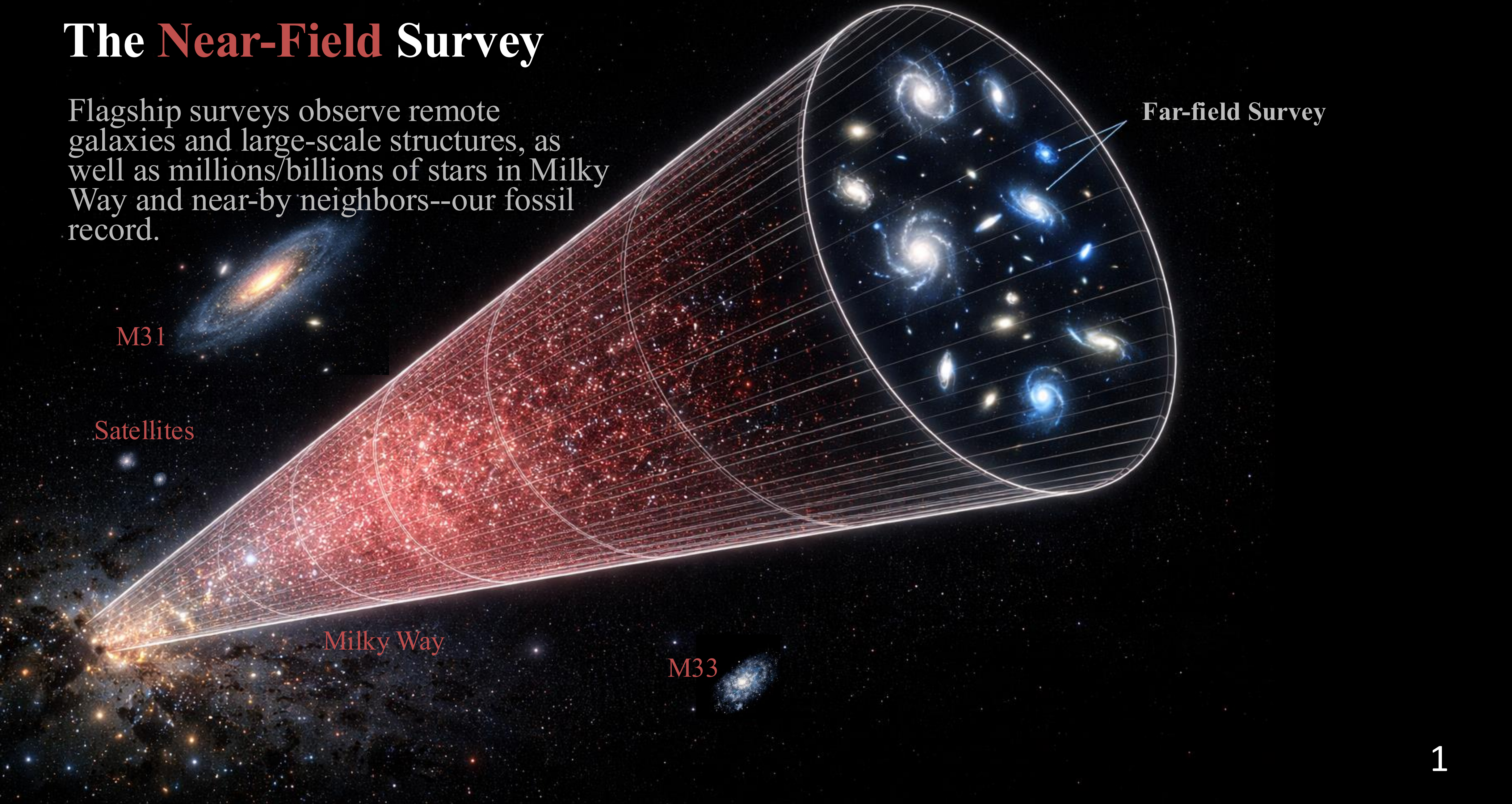


Xiaosheng Zhao (JHU)

07/14/2026, 2nd Annual Cosmic Horizons Conference

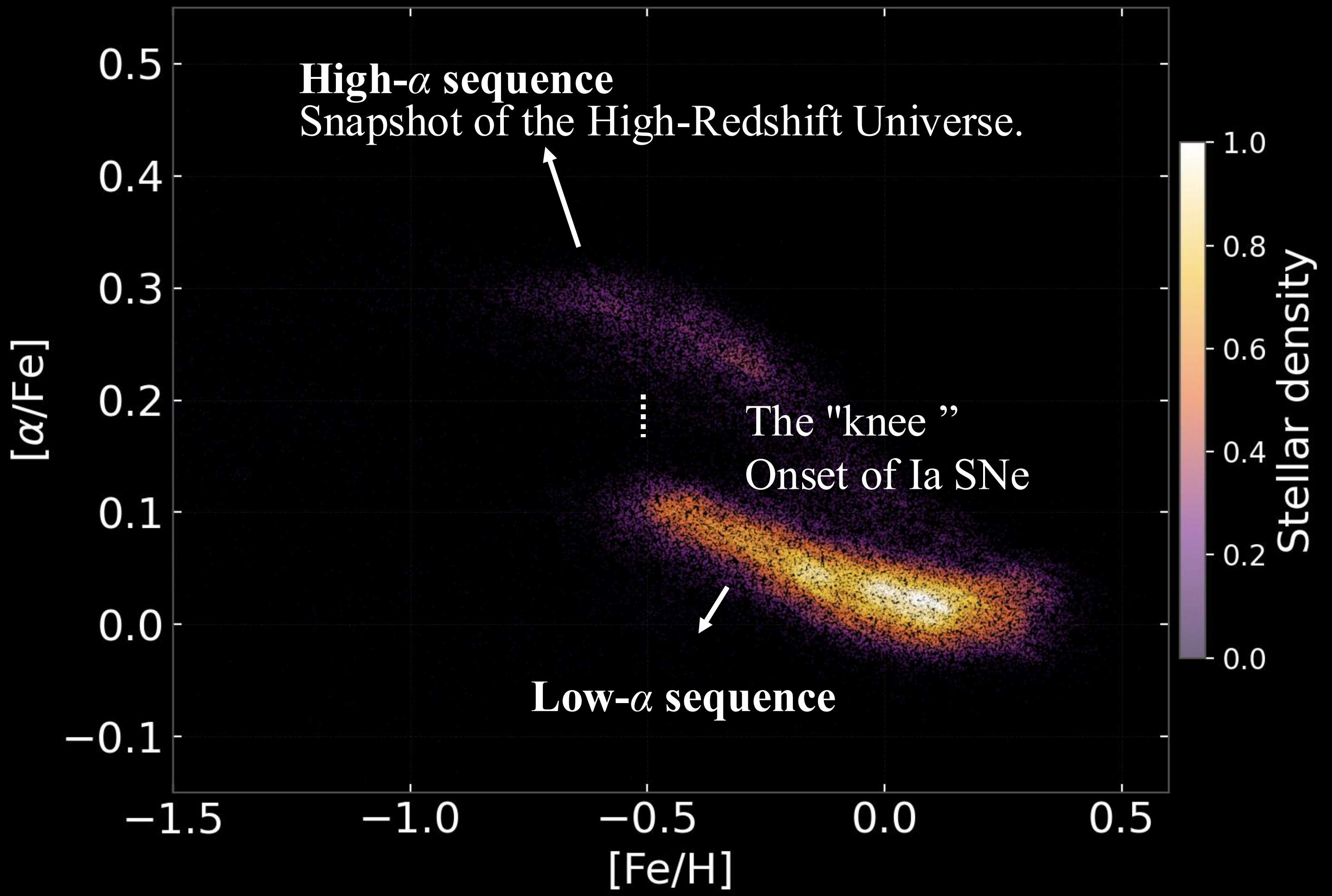
The Near-Field Survey

Flagship surveys observe remote galaxies and large-scale structures, as well as millions/billions of stars in Milky Way and near-by neighbors--our fossil record.



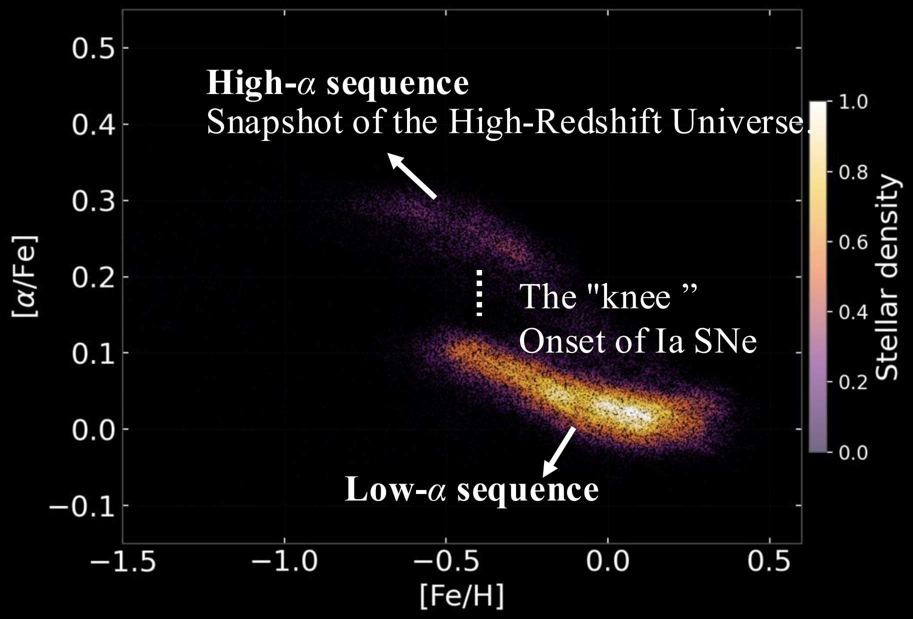
Chemical Tagging: Data-Mining the Galaxy using Stellar 'DNA'

Alpha-element Ratio
(Proxy for Star
Formation Speed)



Overall Metallicity (Proxy for Time/Generations)

Chemical Tagging: Data-Mining the Galaxy using Stellar 'DNA'



The Chemical Clock (Tracing Star Formation)

- **High-α:** rapid early burst; dominated by Type II SNe
- **The "knee"**— onset of delayed Type Ia enrichment; defines the transition epoch
- **Low-α:** slower, extended disk formation

COSMOLOGICAL LEVERAGE

Galaxy assembly & merger history

Accreted populations (e.g. GSE, Sequoia) carry distinct chemical fingerprints — a direct test of hierarchical clustering predictions.

Gas physics & feedback

The bimodality places direct constraints on gas inflow rates and supernova-driven outflow efficiency

Nucleosynthesis & early IMF

Metal-poor stars constrain the integrated yields of the first stellar generations, probing the early Universe IMF and Pop III enrichment channels.

Classical pipelines are not enough

WHERE CLASSICAL METHODS FAIL

Empirical library limits
Reference libraries have finite parameter coverage — the metal-poor regime is simply unreachable.

Domain mismatch across surveys
A pipeline trained on one survey fails when directly applied to another — different resolution, wavelength coverage, and stellar populations break the model.

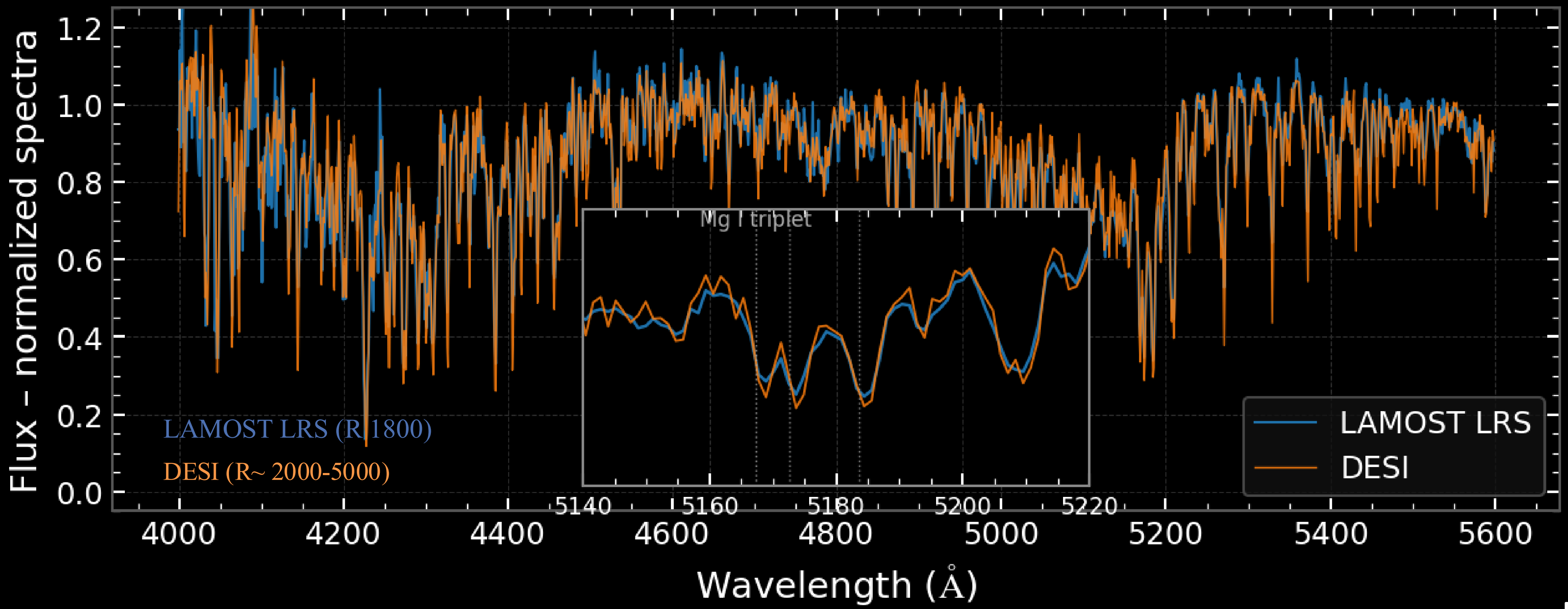
THE ML RESPONSE

Learn from data and high-res libraries
Train directly on millions of observed spectra — no reference library ceiling, except high-res library for labels (from, e.g., APOGEE)

Pre-train → fine-tune
Pre-train on a large, label-rich survey once. A small label set in the target survey is enough to adapt — zero-shot or few-shot transfer.

The Challenge of Heterogeneous Surveys

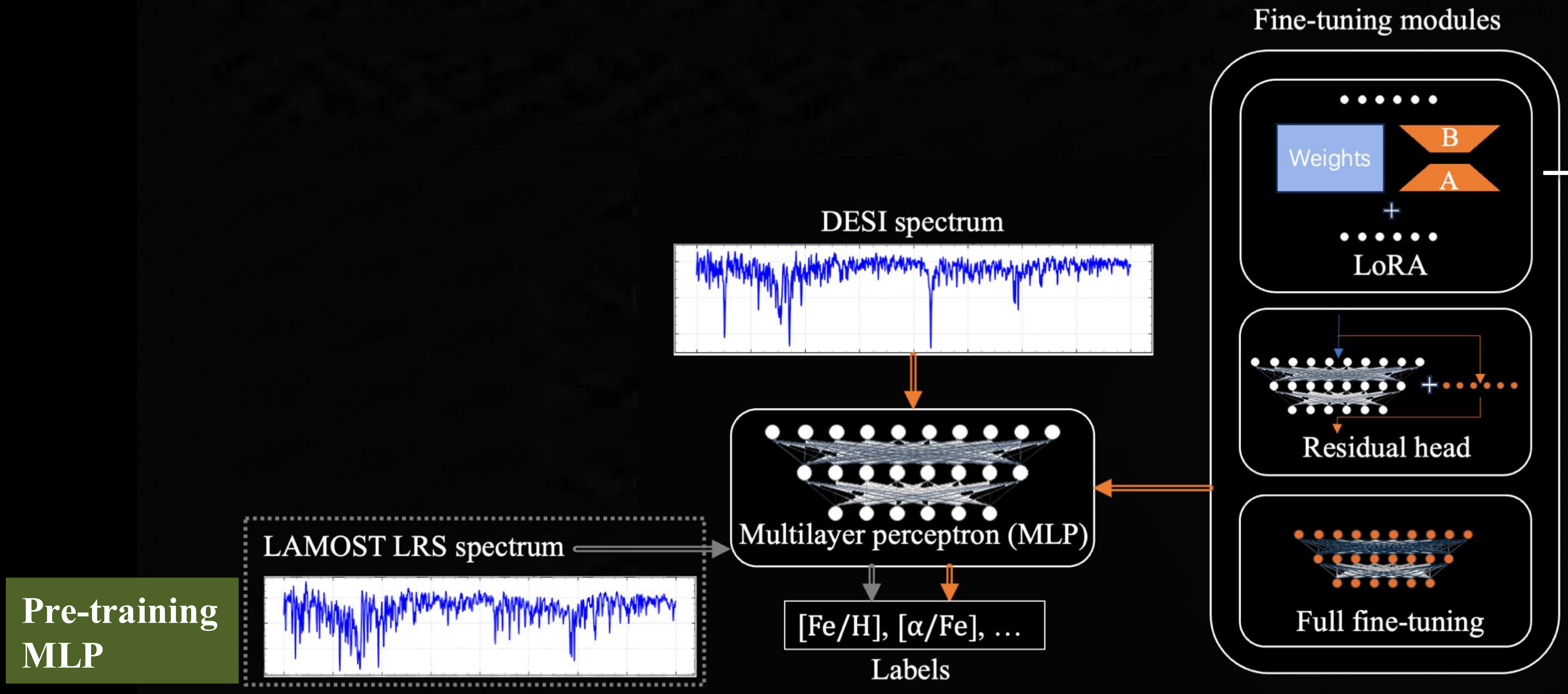
The Challenge: How do we transfer models measuring physical labels ([Fe/H], [alpha/Fe]) from a well-studied, low-res survey to a new, massive, medium-res survey without re-measuring everything?



Pre-trained on LAMOST LRS → Fine-tuned on DESI

Tool A: The Effective Simple MLP with Proper Finetuning

2 Million Parameters. Pre-trained on LAMOST LRS (90k spectra). Fine-tuned on DESI.

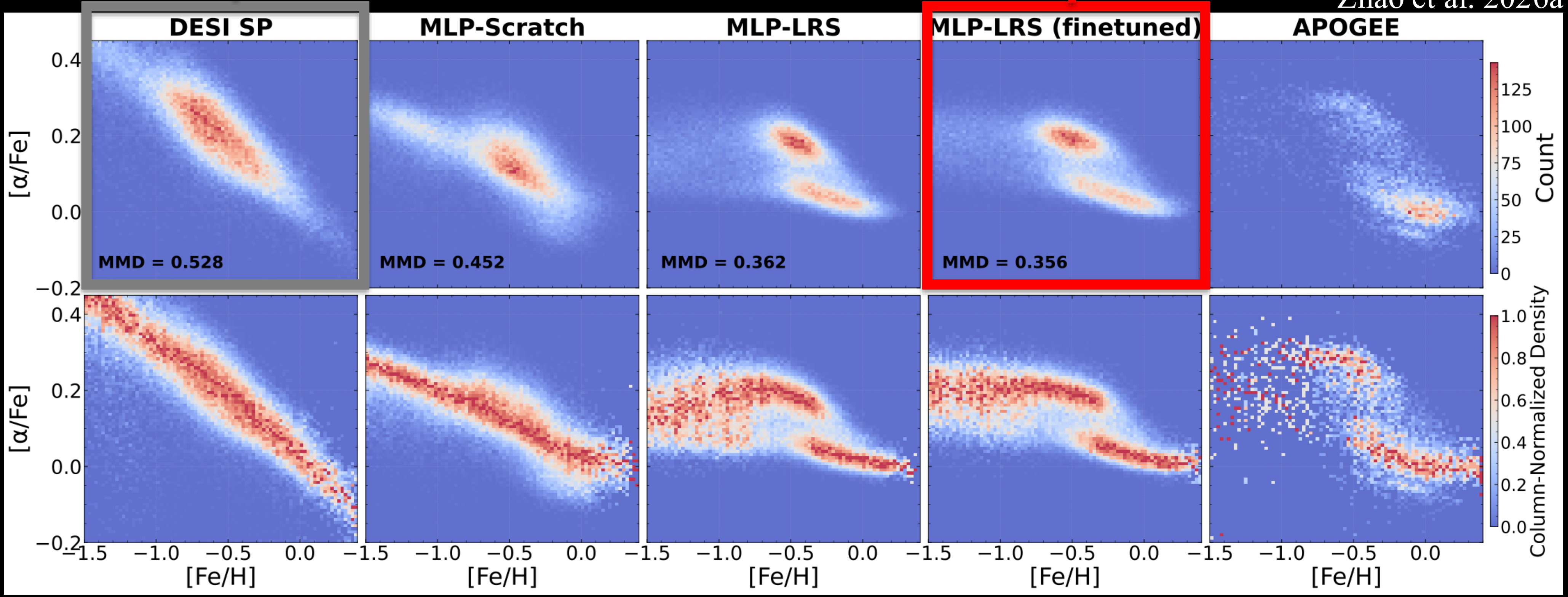


See also:
Zhao et al. 2025
ICML ML4Astro

Recovering the Thin & Thick Disk Bimodality

The standard pipeline smeared the data. The MLP recovered the galactic structure. MMD (distance to APOGEE) improved.

Zhao et al. 2026a



Beyond MLPs: Foundation Models for Spectra

Motivation of a foundation model

Why we explore foundation models?

MLP transfers — but may shallowly

Pre-trained MLPs transfer across surveys, but treat each spectrum as a flat vector. Survey-specific systematics remain entangled with physical parameters.

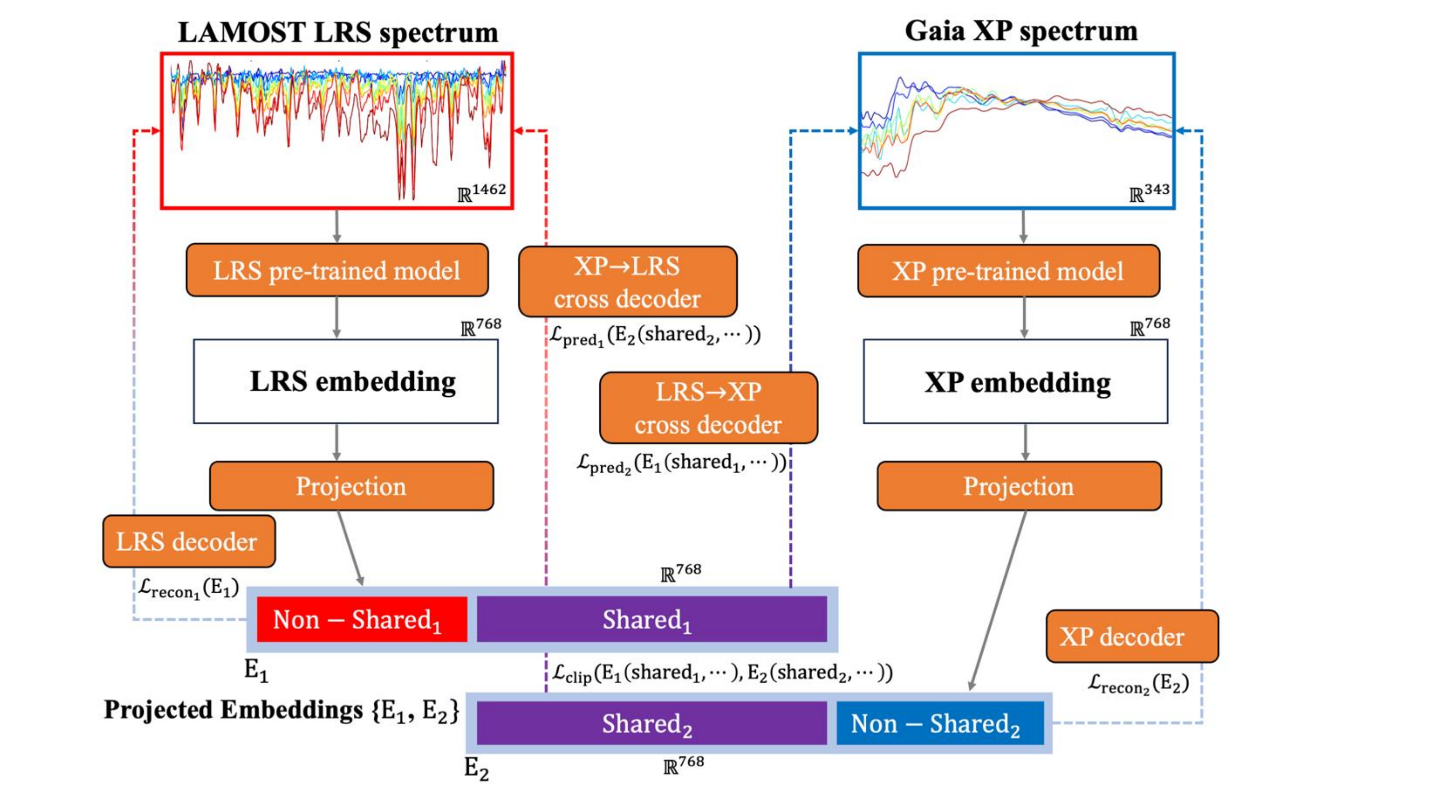
Foundation models learn structure

Pre-trained on massive, diverse data, they learn the underlying spectral language — representations that are physically meaningful, not survey-specific.

Beyond MLPs: Foundation Models for Spectra

SpecCLIP: Aligning LAMOST LRS and Gaia XP. Using Contrastive Learning to push spectra of the same star together in vector space, with decoders to cross-predict and retain spectra-specific information

Zhao et al. 2026b



MLP transfers — but may shallowly

Pre-trained MLPs transfer across surveys, but treat each spectrum as a flat vector. Survey-specific systematics remain entangled with physical parameters.

Foundation models learn structure

Pre-trained on massive, diverse data, they learn the underlying spectral language — representations that are physically meaningful, not survey-specific.

Downstream tasks*: Stellar parameters, chemical abundances, extinction, RV estimation; Cross-modal prediction; In-modal & Cross-modal Retrieval*

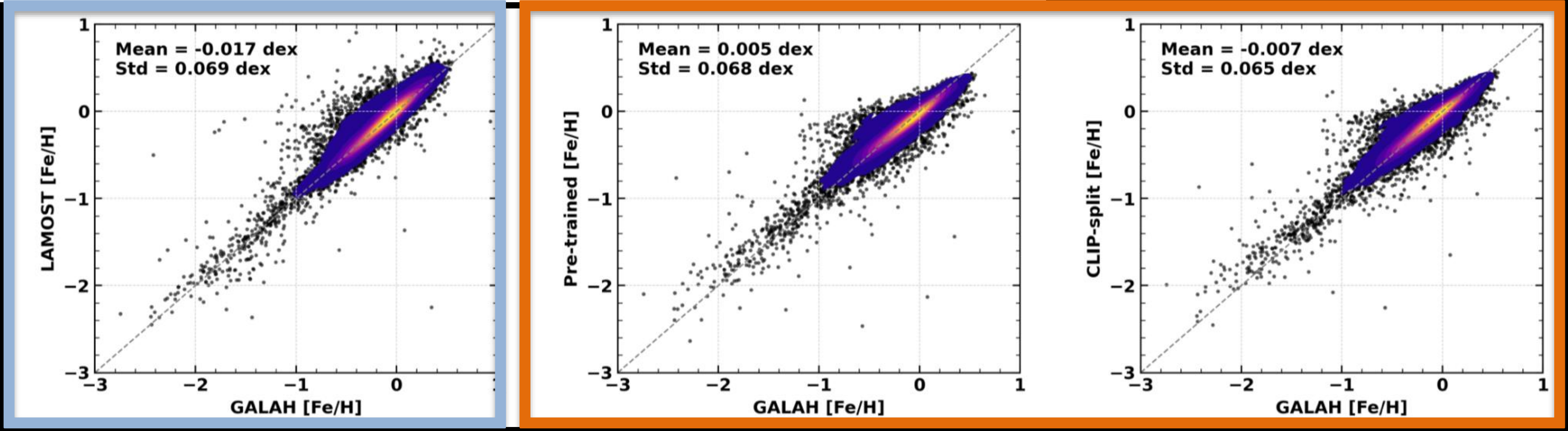
Model Evaluation

Parameter	LRS Models						
	Raw Spectra σ / R^2	Pre-trained σ / R^2	CLIP σ / R^2	CLIP-r σ / R^2	CLIP-p σ / R^2	CLIP-pr σ / R^2	CLIP-split σ / R^2
Atmospheric Parameters							
[Fe/H]	0.070 / −0.882	0.066 / 0.939	0.058 / 0.949	0.057/0.949	0.058 / 0.949	0.057 / 0.949	**0.056 / 0.954**
[α/Fe]	0.023 / 0.872	0.021 / 0.906	**0.020** / 0.912	**0.020**/0.913	**0.020** / 0.911	**0.020 / 0.916**	**0.020** / 0.911

Raw spectra < Embeddings

Full table in
Zhao et al. 2026b

LAMOST LRS
Comparable or
better than
official pipeline



LAMOST official

Ours

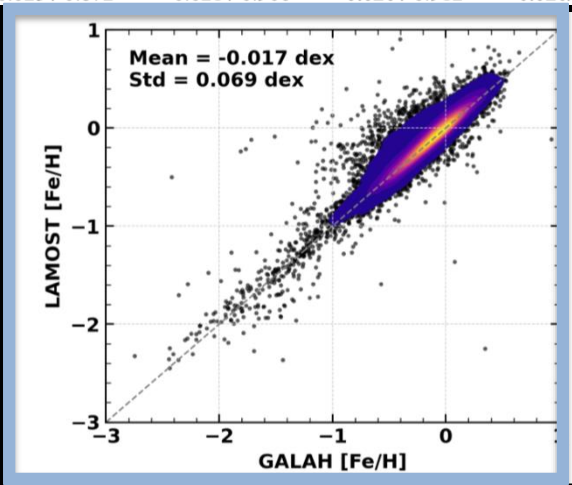
Model Evaluation

			LRS Models				
Parameter	Raw Spectra σ / R^2	Pre-trained σ / R^2	CLIP σ / R^2	CLIP-r σ / R^2	CLIP-p σ / R^2	CLIP-pr σ / R^2	CLIP-split σ / R^2
Atmospheric Parameters							
[Fe/H]	0.070 / −0.882	0.066 / 0.939	0.058 / 0.949	0.057/0.949	0.058 / 0.949	0.057 / 0.949	**0.056 / 0.954**
[α/Fe]	0.023 / 0.872	0.021 / 0.906	**0.020** / 0.912	**0.020**/0.913	**0.020** / 0.911	**0.020 / 0.916**	**0.020** / 0.911

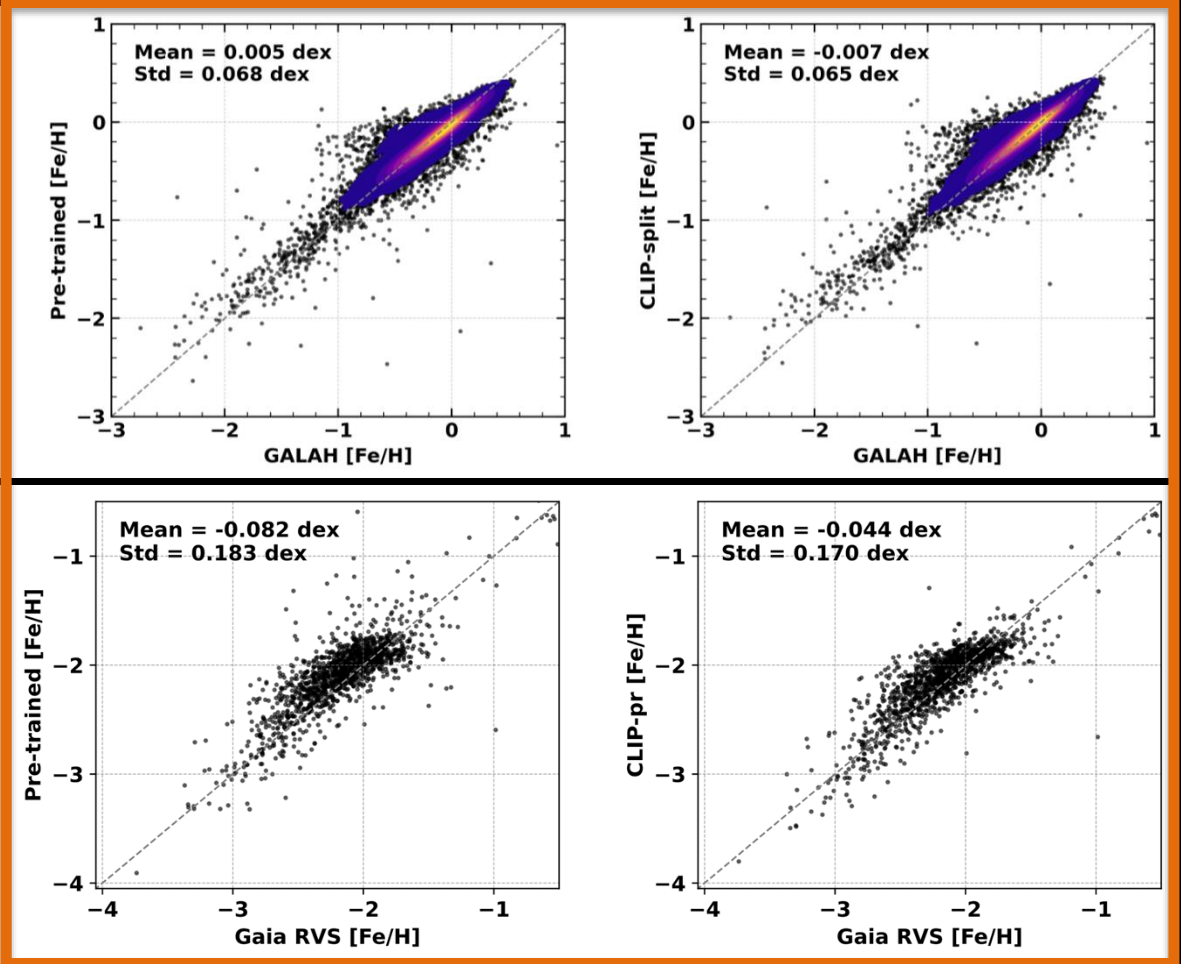
Raw spectra < Embeddings

Full table in
Zhao et al. 2026b

LAMOST LRS
Comparable or
better than
official pipeline



LAMOST official



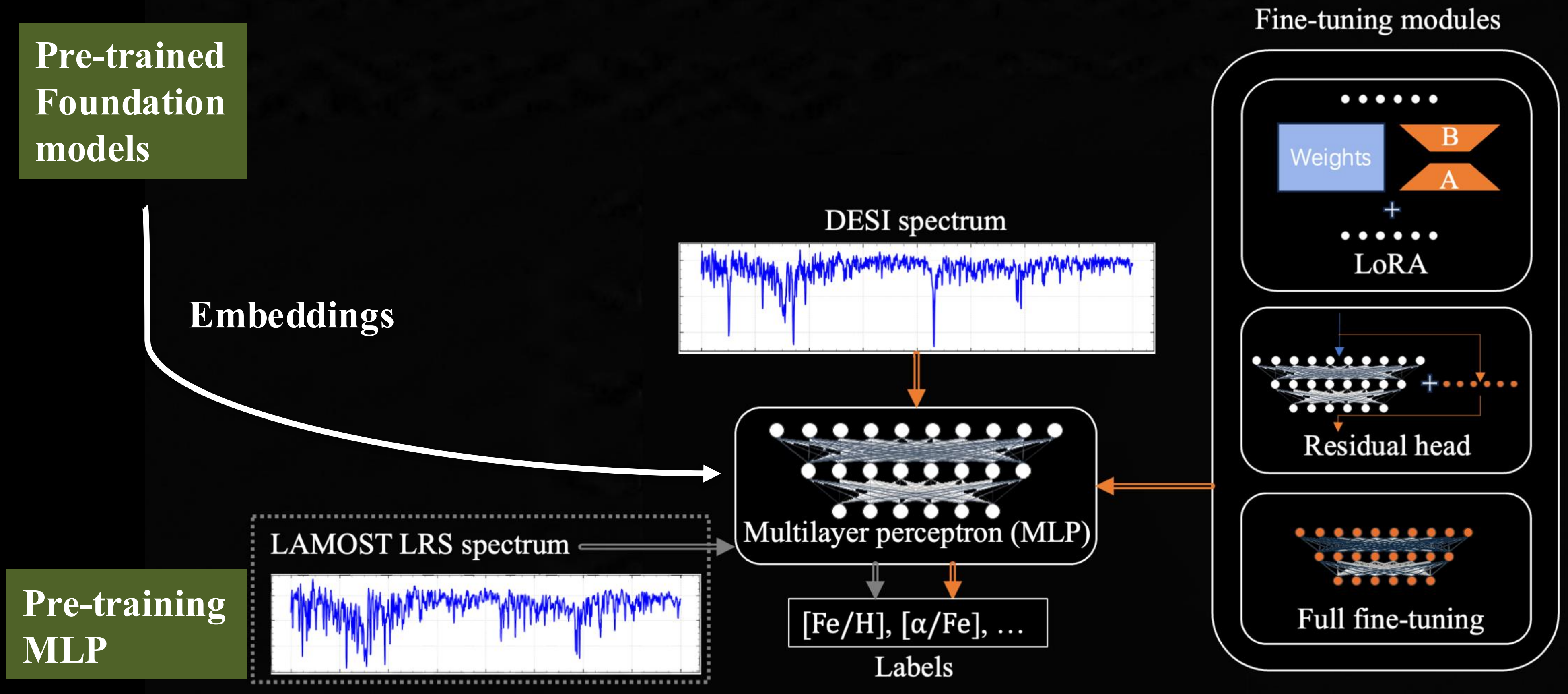
Ours

Gaia XP
Effectiveness
extending to metal-poor

How do foundation models help to generalize from one survey to a **brand-new** one?

TOOL B: Explore the Embedding from a Foundation Model

How could embeddings from a pre-trained foundation model perform?



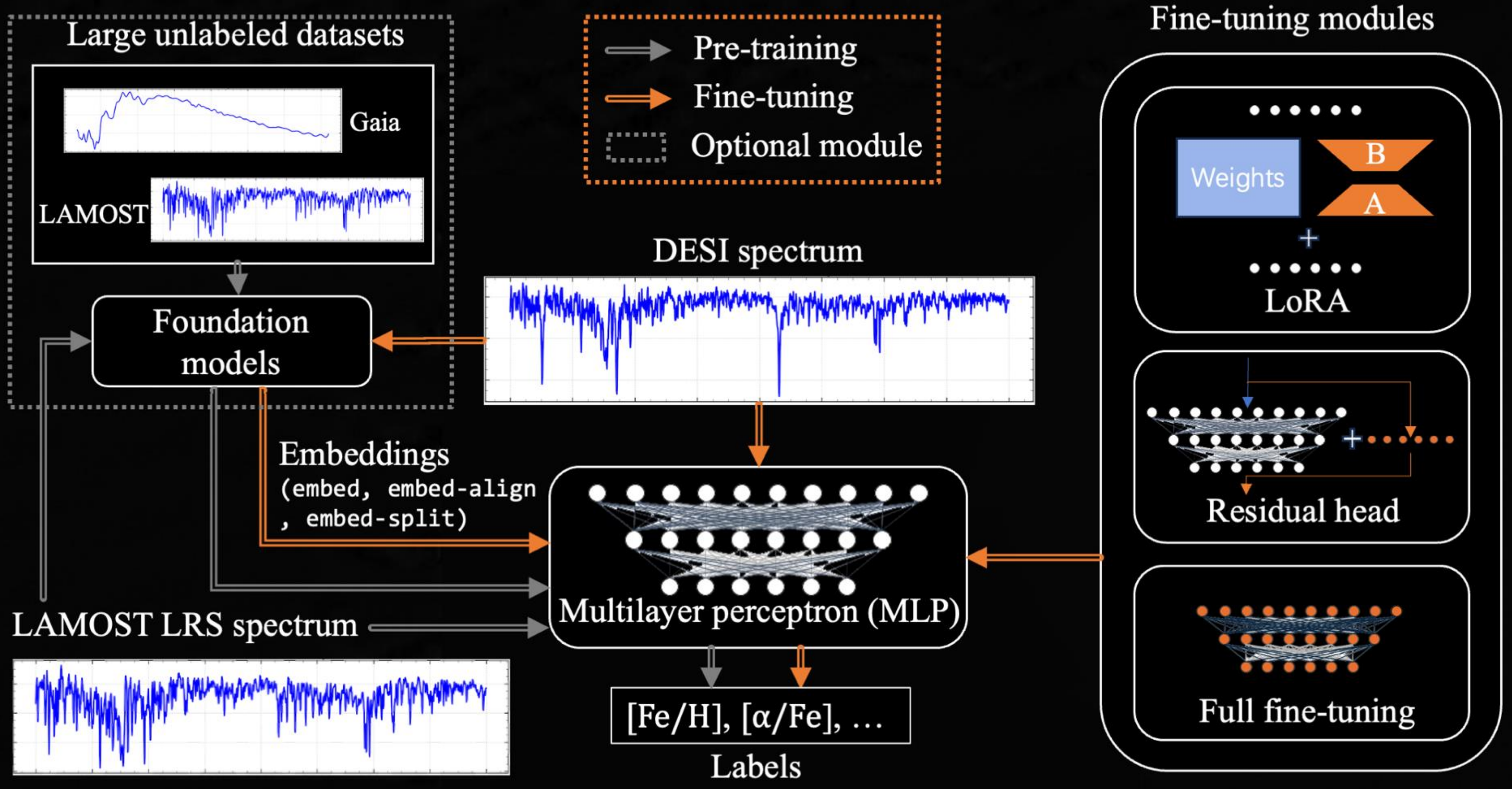
TOOL B: Explore the Embedding from a Foundation Model

How could embeddings from a pre-trained foundation model perform?

Zhao et al. 2026a

Pre-trained Foundation models

Pre-training MLP



The Model Showdown: Where Foundation Model Wins

METAL-POOR ([Fe/H] < - 1.0)

Winner: Simple MLP

Direct training captures faint features better.

METAL-RICH ([Fe/H] > -1.0)

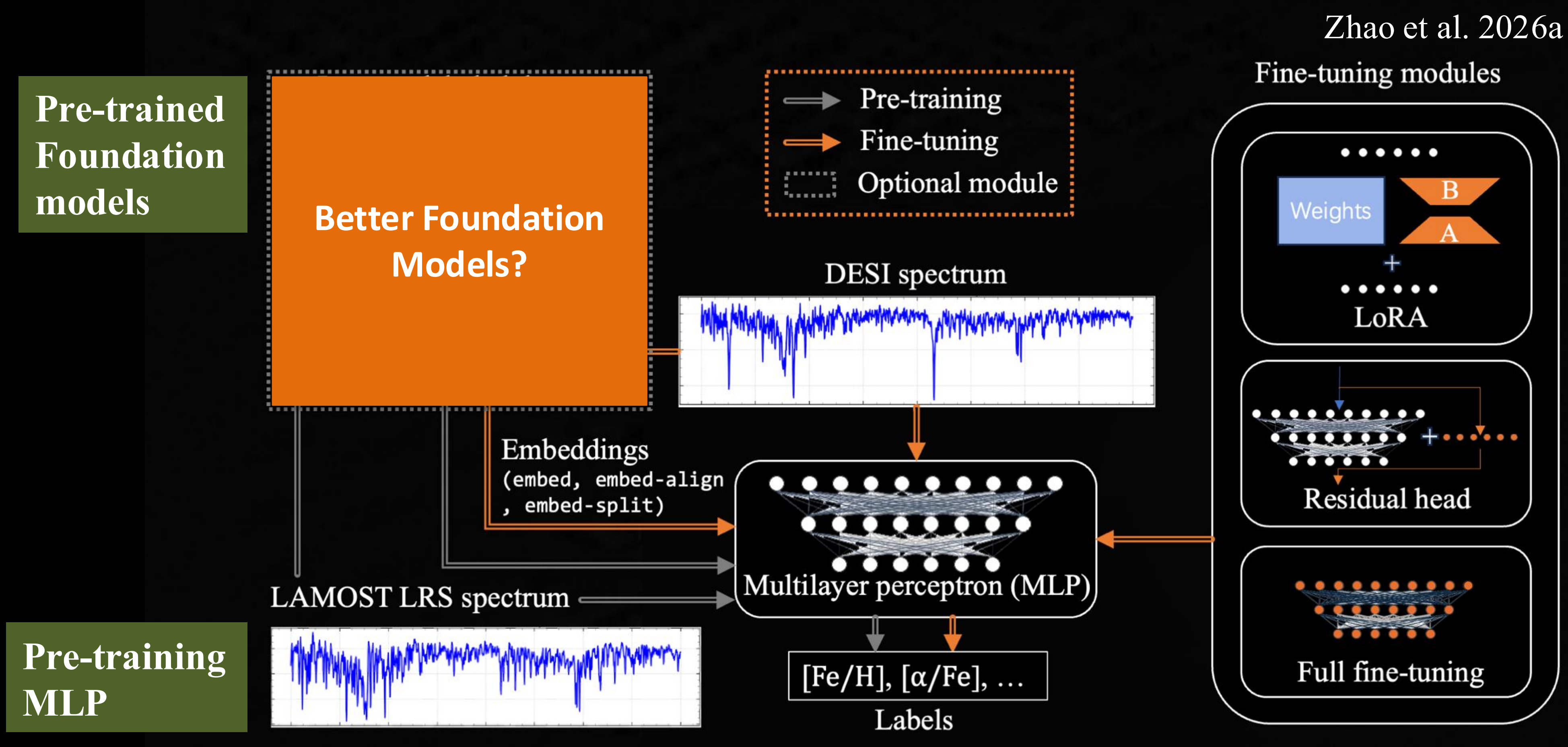
Winner: SpecCLIP

Embeddings capture complex correlations best.

Nuance matters. Bigger isn't always better for every task.

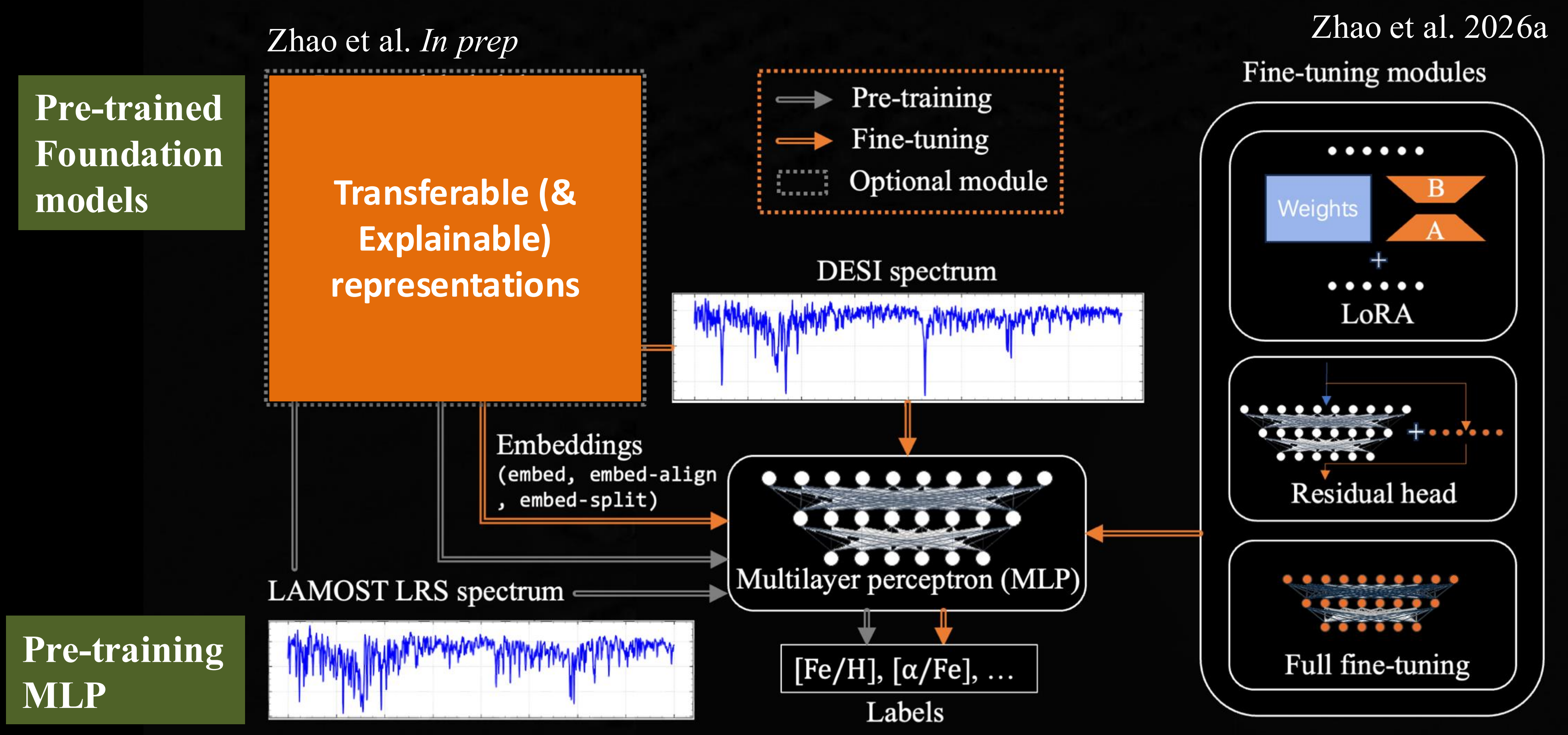
TOOL B: Explore the Embedding from a Foundation Model

How could embeddings from a pre-trained foundation model perform?



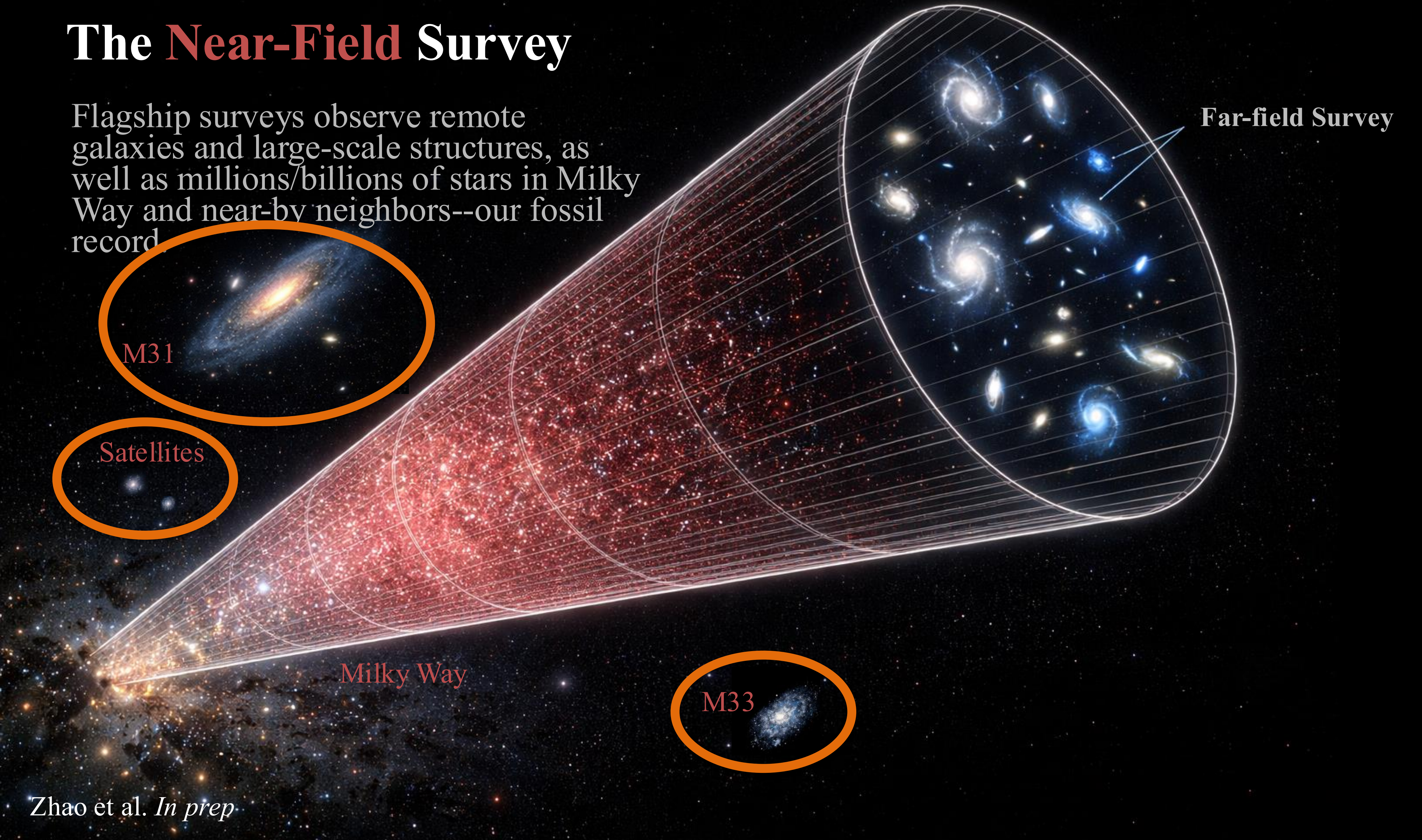
TOOL B: Explore the Embedding from a Foundation Model

How could embeddings from a pre-trained foundation model perform?



The Near-Field Survey

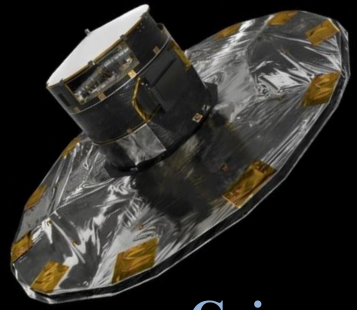
Flagship surveys observe remote galaxies and large-scale structures, as well as millions/billions of stars in Milky Way and near-by neighbors--our fossil record.



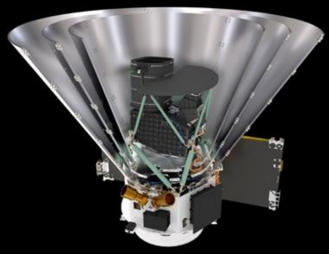
Zhao et al. *In prep*

The Blueprint for the Upcoming Years…

Building multi-modal pipeline. As data volume grows, ML/foundation model-driven analysis/discovery may be a key complement.



Gaia



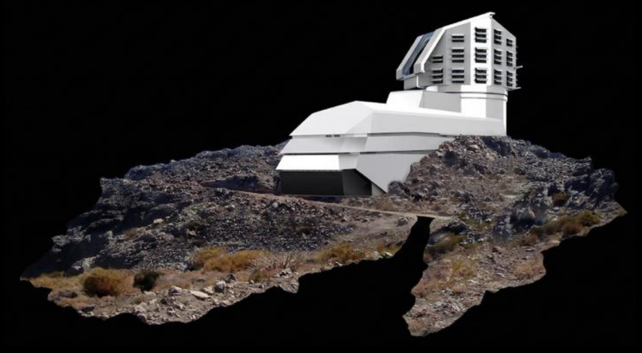
+ LAMOST, Via, Roman…

SPHEREx

DESI

PFS

SDSS



LSST

Summary

The Method:

- Simple MLPs bridge survey heterogeneity and show transferrability.

The Scale:

- Foundation model bring robust embeddings at some parameter regions and unlocks cross-modal search and prediction, benefiting parameter estimation and anomaly detections at large scale.

The Yield:

- Reveal the thin-thick disk, providing observational constraints on early galaxy assembly and chemical enrichment history.
- Provides broader implications for multi-modal transfer learning in any domain where instruments are mismatched and labels are scarce.

Limitation:

- The embeddings from SpecCLIP struggle at some physical contents and some physical parameter regime for cross-survey generalization to a brand-new domain
- “Foundation model” should prove its transferability across instruments, populations, and tasks

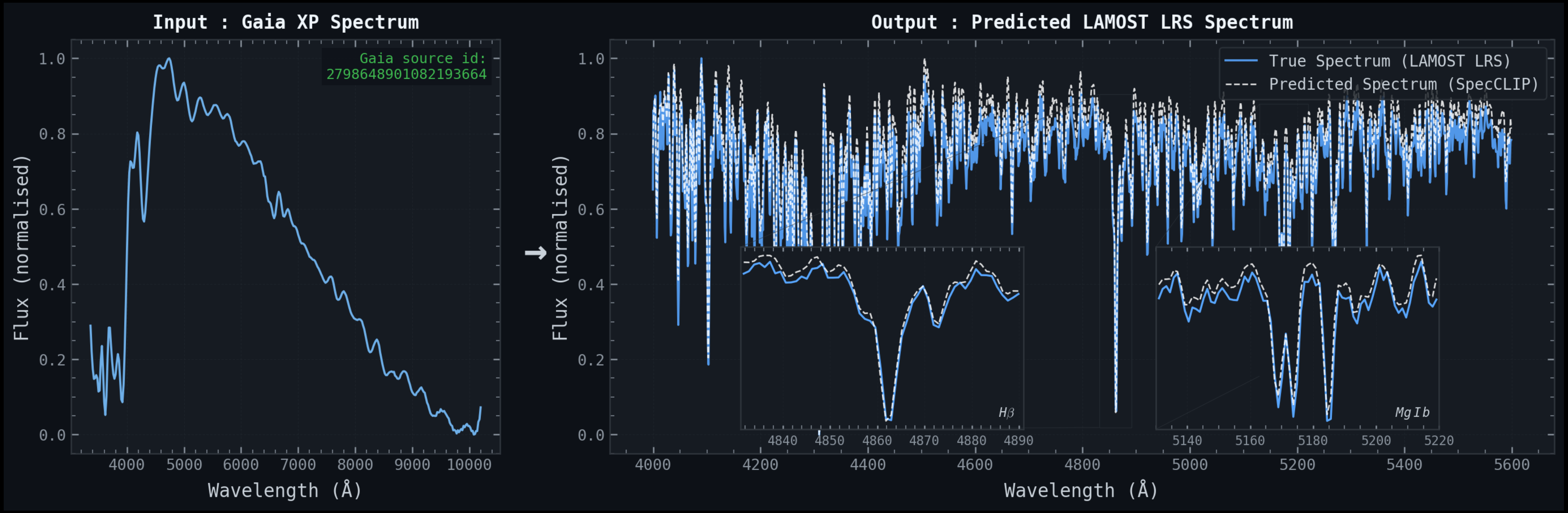
Back Up

Cross-Modal Retrieval

Retrieve a similar star in another survey

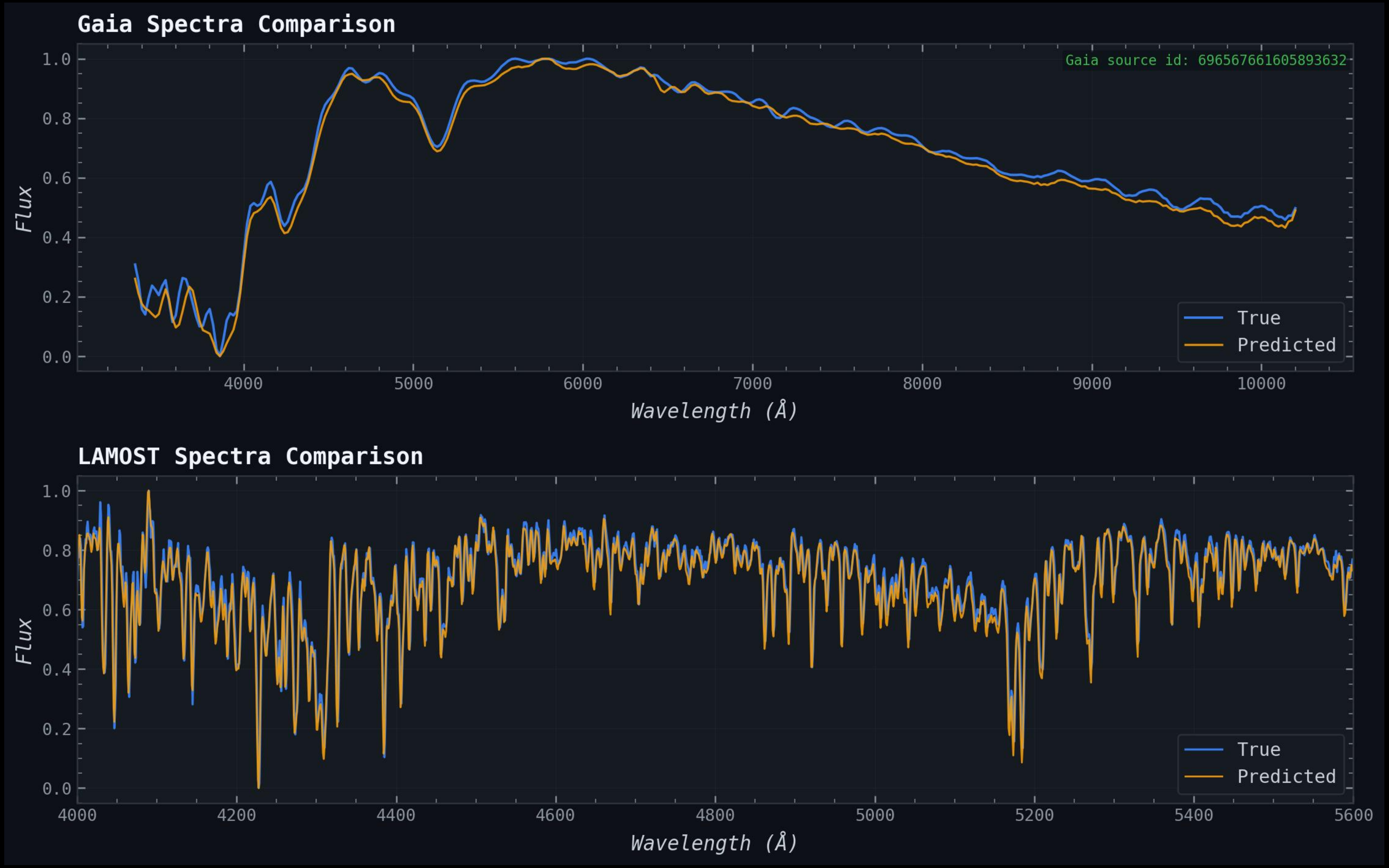
Cross-Modal Spectral Translation

SpecCLIP predict a higher-res spectrum from the spectrophotometry (via shared physical information)



Cross-Modal Translation

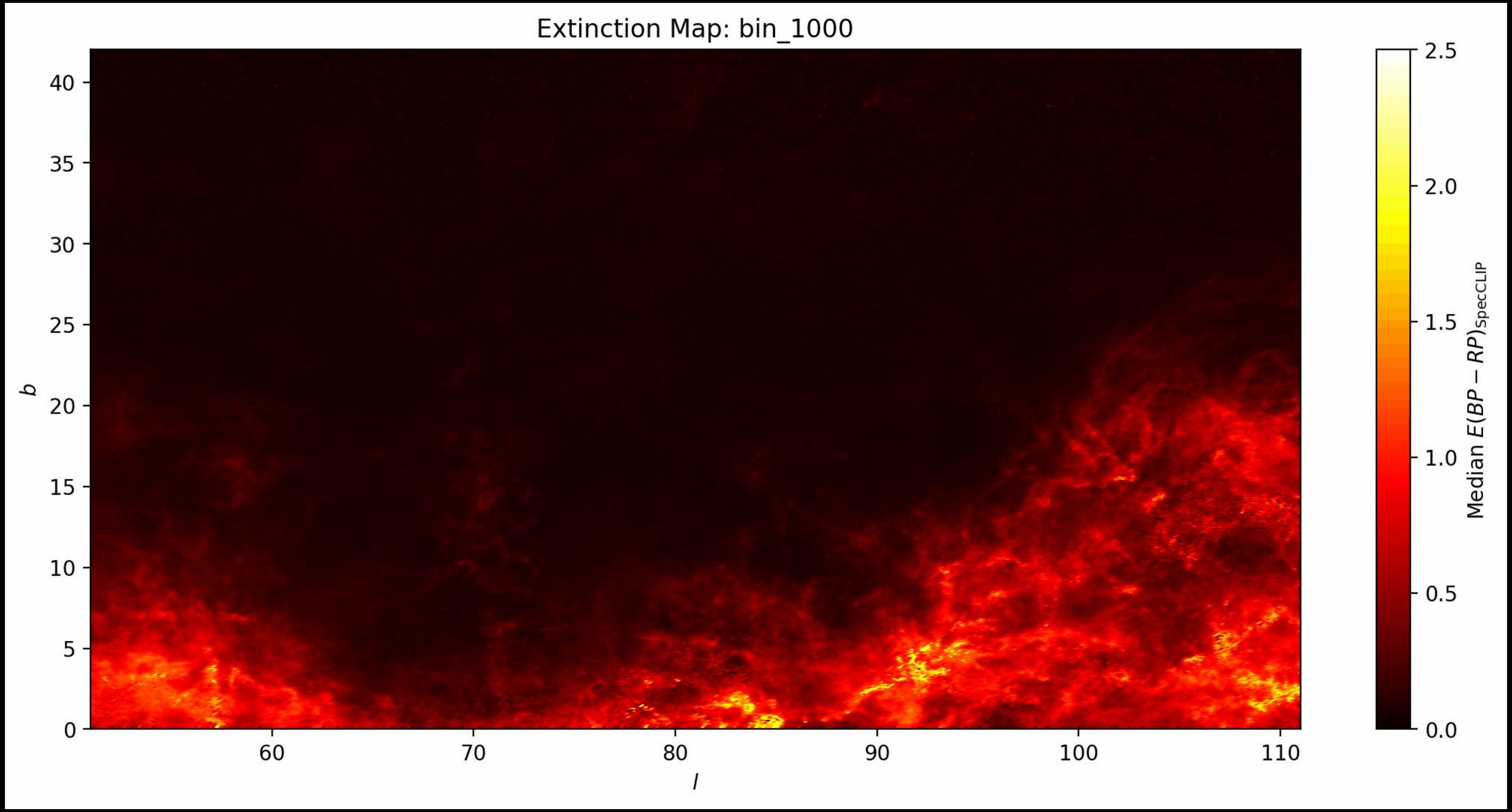
Cross-Modal Translation



Cross-Modal Translation

Mapping the Milky Way's Extinction

Preliminary



Details revealed when increasing bins

The Model Showdown

Zhao et al. 2026a

(Full/Metal-poor/Metal-rich)

Table 1. Coefficient of determination (R^2) for MLPs with different fine-tuning strategies and input types in [Fe/H] and [α/Fe] estimation on DESI spectra.

(Full/MP/MR)	lrs	embed	embed-align	embed-split
[Fe/H]				
residual	0.923 / 0.636 / 0.904	0.922 / 0.196 / 0.913	0.922 / 0.003 / 0.918	0.933 / 0.219 / 0.927
LoRA	0.922 / 0.571 / 0.904	0.921 / 0.031 / 0.916	0.928 / 0.131 / 0.922	0.933 / 0.133 / 0.930
full fine-tune	0.923 / 0.329 / 0.911	0.926 / 0.173 / 0.918	0.927 / 0.085 / 0.922	0.933 / 0.079 / 0.931
zero-shot	0.873 / 0.678 / 0.833	0.833 / 0.694 / 0.779	0.851 / 0.465 / 0.810	0.839 / 0.537 / 0.791
from scratch	0.880 / -0.736 / 0.878	0.877 / -1.246 / 0.886	0.647 / -0.072 / 0.542	0.794 / -0.649 / 0.759
[α/Fe]				
residual	0.799 / 0.366 / 0.802	0.790 / 0.396 / 0.792	0.796 / 0.232 / 0.801	0.796 / 0.335 / 0.800
LoRA	0.829 / 0.418 / 0.831	0.807 / 0.370 / 0.810	0.769 / 0.290 / 0.772	0.733 / 0.205 / 0.736
full-finetune	0.816 / 0.381 / 0.819	0.810 / 0.315 / 0.813	0.784 / 0.247 / 0.787	0.773 / 0.231 / 0.777
zero-shot	0.776 / 0.318 / 0.778	0.730 / 0.183 / 0.733	0.772 / 0.202 / 0.776	0.681 / 0.105 / 0.685
from scratch	0.761 / 0.067 / 0.766	0.729 / 0.401 / 0.730	0.611 / 0.144 / 0.612	0.749 / 0.374 / 0.751

Raw spectra as input **Embeddings as input**

Foundation models perform well at metal-rich regime, though underperform at the metal-poor regime for [Fe/H]; consistently degraded performance on [α/Fe].