The Platonic Universe: Do Foundation Models See the Same Sky?

UniverseTBD

Trinidad Borrell [1 2] **Steven Dillmann** [3] **Kshitij Duraphe** [4] **Furkan Eris** [5] **Ashod Khederlarian** [6] **Aman Kumar** [7]
Giovanni Marraffini [8 9] **Michael J. Smith** [10 11] **Shashwat Sourav** [12 13 14] **Rocco Di Tella** [10 11] **John F. Wu** [15 16]

Abstract

We test the Platonic Representation Hypothesis (PRH) and its Aristotelian refinement (ARH) by using diverse astronomical data to measure representational convergence across foundation models. We propose that astronomy is a natural testbed for this: the historical success of astrophysics is itself evidence that a compact, modality-invariant description of galaxy observables exists, and so representation convergence toward reality should be measurable against the physical parameters astronomers already use. Given this framework, we evaluate eleven foundation model families (spanning supervised classification, self-distillation, joint-embedding prediction, masked autoencoding, vision-language pre-training, and astronomy-specific architectures from $\mathcal{O}$(10M)$\rightarrow\mathcal{O}$(10B) parameters) on cross-matched JWST, HSC, and Legacy Survey imagery, and DESI spectroscopy. All models are evaluated frozen, with no astronomy-specific finetuning. We probe redshift, stellar mass, and specific star formation rate via linear probes, and local (MKNN) and global (CKA) embedding geometry within families, between modalities, and across architectures. We find that physics performance scales predictably with capacity; probe directions align consistently with expected astrophysical correlations and selection effects; and local (but not global) embedding alignment tracks physics performance, including between DESI spectra and HSC imagery—modalities that share essentially no low-level statistics. Our results support the ARH over the strict PRH, and suggest that astro-foundation models can build on general-purpose pre-trained architectures, capitalizing on the broader open machine learning community's already-spent computational investment.

1. Astronomy and the Platonic and Aristotelian Representation Hypotheses

Three historical waves of increasingly automated connectionism have swept the shores of astronomy. The late 1980s brought MLPs tuned for astronomical applications on manually selected inputs (e.g. Adorf & Johnston, 1988; Angel et al., 1990; Odewahn et al., 1992). With the advent of CNNs, RNNs, and deep learning, these MLP models gave way to raw data ingestion (e.g. Dieleman et al., 2015; Charnock et al., 2018; Wu & Peek, 2020; Khederlarian et al., 2026). And the third wave of unsupervised and self-supervised learning largely removed the need for task-specific human-generated labelling, with connectionist methods inferring astronomical knowledge directly from the raw data (e.g. Sarmiento et al., 2021; Smith et al., 2022). A fourth wave has recently been seeded by the discovery of predictable neural scaling laws (Kaplan et al., 2020; Hoffmann et al., 2022)—the application of foundation models to astronomical observations, publications, and survey data (Smith & Geach, 2023). This fourth wave has brought with it diverse approaches in the search for a viable path towards a single, canonical, astro-foundation model: from contrastive methods (e.g. Slijepcevic et al., 2024; Parker et al., 2024; Mishra-Sharma et al., 2024; Zhao et al., 2025), to generative architectures (e.g. Leung & Bovy, 2023; Koblischke & Bovy, 2024; Ore et al., 2024), to autoregressive modelling (e.g. Smith et al., 2024; Pan et al., 2024; Euclid Collaboration et al., 2025; Zuo et al., 2025; Heneka et al., 2025; Moriwaki et al., 2025), to finetuning of large language models on astronomical text (e.g. Nguyen et al., 2023; Perkowski et al., 2024; de Haan et al., 2024; Zaman et al., 2025). The

[1]Paris Brain Institute, Paris, France [2]Sorbonne Université, Paris, France [3]Stanford University, Stanford, CA, USA [4]Independent Researcher [5]Prolific [6]University of Pittsburgh, Pittsburgh, PA, USA [7]IUCAA, Pune, India [8]INRIA, France [9]Sigma Nova [10]AstroAI [11]Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA [12]Washington University in St. Louis, St. Louis, MO, USA [13]Oak Ridge National Lab, Oak Ridge, TN, USA [14]Lawrence Berkeley National Lab, Berkeley, CA, USA [15]Space Telescope Science Institute, Baltimore, MD, USA [16]Johns Hopkins University, Baltimore, MD, USA. Correspondence to: Trinidad Borrell <trinidad.borrell@icm-institute.org>.

Mechanistic Interpretability Workshop at the 43^{rd} International Conference on Machine Learning, Seoul, South Korea, 2026.

sheer diversity of these approaches raises a natural question: does the choice of architecture, training regime, and data modality matter, or do sufficiently performant models converge to equivalent representations?

The Platonic Representation Hypothesis (PRH; Huh et al. (2024)) offers one answer: neural networks trained on different modalities are converging toward a shared global statistical model of reality in their representation spaces. The PRH attributes this convergence to three mechanisms—*task generality* (models trained on more diverse tasks require representations that capture more information about underlying reality), *model capacity* (larger models are more likely to find optimal representations), and *simplicity bias* (neural networks naturally favour simpler solutions that generalise better)—and draws inspiration from Plato's Allegory of the Cave (Plato, c. 375 BCE), in which the cave-dwellers mistake shadows on a wall for reality itself. In this analogy, our training data are the shadowy projections of an underlying reality, and our models learn to recover representations (or 'Forms') of the reality that generates them. Under the PRH, larger models trained on more diverse tasks should converge toward a Platonic Ideal: a perfect, lossless Form of underlying reality. A recent refinement, the Aristotelian Representation Hypothesis (ARH; Gröger et al. (2026)), tempers this strong claim. Rather than agreeing on a shared global coordinate system, the ARH proposes that converging models agree on local neighborhood structure rather than global embedding geometry. This distinction has direct empirical consequences in that *local* rank statistics are predicted to track convergence, while *global* similarity measures are not.

Astronomical observations provide a natural testbed for both hypotheses due to the observations' fundamental nature as different projections of the same underlying cosmic reality. The historical success of astrophysics is itself evidence that a compact, modality-invariant description of galaxy observables exists, in that distinct instruments observing the same galaxies can be reconciled into a single physical account. These observations are inherently linked through shared physical processes; a galaxy's morphology (captured in images), chemical composition (revealed through spectroscopy), and integrated properties (measured via photometry) all emerge from the same stellar populations, gas and dust dynamics, and underlying matter distributions (Conroy, 2013). Given this, foundation models should converge toward representations that capture the underlying fundamental physics governing these phenomena. All the pieces are in place to test the PRH in astronomy: the scale and diversity of modern surveys provide the data volume necessary to test convergence across multiple model architectures and training objectives, and recent work (The Multimodal Universe Collaboration, 2024; Caplar et al., 2025) has standardized crossmatching across astronomical modes[1].

In this paper we test both the PRH and the ARH across a basket of eleven foundation model families spanning supervised classification, self-distillation, joint-embedding prediction, masked autoencoding, vision–language pre-training, and astronomy-specific architectures from $\mathcal{O}$(10M)→$\mathcal{O}$(10B) parameters, evaluated on crossmatched JWST, HSC, and Legacy Survey imaging, and DESI spectroscopy. [2] We organise our analysis around three questions:

1. Do larger models encode more galaxy physics? We train linear probes on frozen embeddings to predict redshift, stellar mass, and specific star formation rate given galaxy images, and ask whether probe performance scales with model capacity (§3.1 ¶1).

2. Do different architectures converge to the same physical representation? We probe crossmodal performance, and measure the distance between probe weight vectors across models and ask whether they organise physics similarly (§3.1 ¶2-3).

3. Does geometric alignment scale with performance—locally, globally, or both? We measure MKNN and CKA within model families, between astronomical modalities, and across architectures, and regress alignment against probe performance (§3.2).

Conflict of Interest Disclosure. The authors declare no financial conflicts of interest. All foundation models evaluated in this study (ViT, ConvNeXtV2, DINOv3, I-JEPA, V-JEPA, ViT-MAE, CLIP, LLaVA-1.5, PaliGemma 2, AstroPTv2, and Specformer) are publicly released, third-party models that were not developed by the authors or by the organisations that employ them; all were used frozen, without fine-tuning or modification. No author has a financial relationship with the entities that produced these models that could constitute a conflict of interest.

2. Datasets, models, and metrics

We briefly describe and motivate our chosen data, model architectures, and metrics below.

[1] We define an astronomical 'mode' or 'modality' as the information captured by a specific instrument so that (e.g.) JWST and HSC imaging are separate modes.

[2] All code used to produce the results in this paper (data preprocessing and crossmatching, embedding extraction, linear probing, MKNN/CKA alignment, and the statistical analyses and figures) is available at https://github.com/UniverseTBD/platonic-universe.

2.1. Datasets used

We test across four crossmatched astronomical datasets that capture fundamentally different projections of galaxy Forms. We use the MMU to crossmatch between data modes at a 1" radius (The Multimodal Universe Collaboration, 2024). **Hyper Suprime-Cam (HSC)** (Miyazaki et al., 2018) provides our reference baseline. HSC is an 8.2m ground-based optical survey with ∼0.7" seeing and 5-band (*grizy*) coverage. We use HSC as it provides our largest overlap sample across all other modalities, making it a natural anchor for pairwise comparison. **DESI Legacy Imaging Survey** (Dey et al., 2019) is a complementary ground-based optical survey at ∼1" seeing, its inclusion allows us to test representational alignment across different ground-based imaging survey strategies. Our crossmatched Legacy-HSC sample contains 102k galaxies. **JWST NIRCam imaging** (Gardner et al., 2023) provides our most extreme imaging comparison: space-based near-infrared and infrared observations that can reveal dust emission and dust-obscured stellar populations invisible to our HSC and Legacy optical surveys. The JWST-HSC crossmatched sample contains 1.67k galaxies. **DESI spectra** (DESI Collaboration et al., 2024) provide our cross-modal non-imaging test. Spectroscopy probes integrated galaxy properties (e.g. stellar properties, kinematics) that are complementary to the spatial information captured by imaging. The DESI-HSC crossmatched sample contains 18.6k galaxies.

To test the convergence of representations with astronomical parameters, we use HSC and JWST images from **COSMOS-Web** (Casey et al., 2023). The COSMOS field is one of the most observed areas of the night sky; due to imaging from many different telescopes across wide wavelength ranges, galaxies in COSMOS have very accurate measurements of their physical parameters (Shuntov et al., 2025). To ensure that our physical labels are high quality, we select 45 000 galaxies with both HSC i and JWST F150W band flux signal-to-noise ratios greater than 25, and use their redshifts (many-band photometric redshifts), stellar masses, and specific star formation rates (sSFRs) obtained from the template-fitting method LePHARE (Arnouts et al., 1999; Ilbert et al., 2006). These parameters are canonical physical descriptors of galaxies; they capture a galaxy's size, intrinsic luminosity, current evolutionary state, and distance from Earth. To obtain image cutouts, we use the DAWN JWST Archive (DJA)[3] and the image cutout tool from the HSC release website[4].

For our MMU-crossmatched surveys we assemble three-channel RGB composites from the raw flux cubes: g, r, z for HSC and Legacy Survey, and F090W, F277W, F444W for JWST, ensuring maximum wavelength coverage while remaining suitable for our foundation models trained on RGB natural images. Pixel intensities are normalised per band with an arcsinh stretch anchored to precomputed 1st and 99th percentile flux values (estimated per survey over batches of up to 10 000 images): $\bar{x} = \text{arcsinh}(\alpha(x - p_1)/(p_{99} - p_1))/\text{arcsinh}(\alpha)$, with stretch parameter $\alpha = 20$ and the output clipped to $[0, 1]$. Arcsinh stretching preserves faint low-surface-brightness structure while compressing the dynamic range of bright galaxy centres, we choose $\alpha = 20$ following Lupton et al. (2004). In the COSMOS-Web dataset, we use g, r, z for HSC and $F115W$, $F150W$, and $F277W$ for JWST. We crop HSC and Legacy Survey galaxies and bilinearly rescale the image cutouts to match the $3.84'' \times 3.84''$ angular size of the JWST cutouts (96×96 pixels at $0.04''/\text{pixel}$). Finally, before passing data into a model, any model-specific preprocessing is applied according to each model's published processing pipeline.

[3] https://dawn-cph.github.io

[4] https://hsc-release.mtk.nao.ac.jp/doc/index.php/data-access__pdr3/

2.2. Model architectures

Our model selection (Tab. 1) is designed to test the PRH along three axes. First, we include models spanning supervised classification (ViT, ConvNeXtv2), self-supervised learning via knowledge distillation (DINOv3), joint-embedding prediction (IJEPA, VJEPA), and masked autoencoding (ViT-MAE), allowing us to test whether convergence depends on training objective. Second, we include vision-language models (CLIP, LLaVA-1.5, PaliGemma) to test whether architectures trained to bridge vision and language converge toward the same representations as unimodal vision models. Third, we include two astronomy-specific models: AstroPTv2, an autoregressive transformer pre-trained exclusively on DESI Legacy Survey imaging, and Specformer, a transformer trained on one-dimensional DESI spectra. AstroPTv2 tests whether domain-specific pretraining yields representations that diverge from or converge with those of general-purpose models. Specformer represents our most out-of-distribution test case: a fundamentally different input modality (spectra vs. images) processed by a model that has never seen natural images. We extract our embeddings from the literature-recommended embedding layer of each model, or from the penultimate layer where there is no recommendation. For each architecture family, we test multiple model sizes to measure whether representational alignment scales with capacity or performance.

2.3. Metrics used to measure representational alignment

Following the methodology established in the original PRH work, we measure representational alignment using the mutual k-nearest neighbour (MKNN) metric (Chechik et al., 2010). Given two embeddings $(\mathbf{z}_1, \mathbf{z}_2)$ corresponding to the same object as viewed by two different in-

Table 1. Foundation models used in this study, spanning supervised, self-supervised, autoregressive, and multimodal training paradigms across vision, language, and spectral modalities.

Model	Training regime	Modality
Supervised		
ViT (Dosovitskiy et al., 2020)	Supervised classification	Images
ConvNeXtv2 (Woo et al., 2023)	Supervised classification	Images
Self-supervised		
DINOv3 (Siméoni et al., 2025)	Self-distillation	Images
IJEPA (Assran et al., 2023)	Joint-embedding prediction	Images
VJEPA (Assran et al., 2025)	Joint-embedding prediction	Video
ViT-MAE (He et al., 2022)	Masked autoencoding	Images
Multimodal / Vision-Language		
LLaVA-1.5 (Liu et al., 2023)	Vision-language	Images + Text
PaliGemma (Beyer et al., 2024)	Vision-language	Images + Text
CLIP (Radford et al., 2021)	Contrastive image-text	Images + Text
Astronomy-specific		
AstroPTv2 (Smith et al., 2024)	Next-token prediction	Astro. images
Specformer (Parker et al., 2024)	Contrastive	Astro. spectra

struments or models, the MKNN score is computed as the cardinality of intersections for each object's k-nearest neighbours in the embedding space: $\text{MKNN}(\mathbf{z}_1, \mathbf{z}_2) = k^{-1}|N_k(\mathbf{z}_1) \cap N_k(\mathbf{z}_2)|$, averaged over all objects in the dataset, where N_k is the k-nearest neighbours operation under the cosine distance, and $|\cdot|$ denotes set cardinality. We follow Huh et al. (2024) and set $k = 10$ across all experiments. We complement MKNN with the Centered Kernel Alignment (CKA) metric (Kornblith et al., 2019): $\text{CKA} = (\|AB^T\|_F^2)/(\|AA^T\|_F\|BB^T\|_F)$, where A and B are mean centered embedding matrices and $\|\cdot\|_F$ denotes the Frobenius norm. We use both MKNN and CKA as MKNN captures local neighborhood agreement, whereas CKA captures global embedding similarity. To eliminate confounds from network scale we follow Gröger et al. (2026) and use permutation-calibrated versions of the MKNN and CKA metrics in all of our experiments, which subtract an empirical null obtained by shuffling sample correspondences to correct for width-driven baseline inflation.

3. Experiments and results

We structure our experiments to ask first whether foundation models encode galaxy physics, and then whether the geometry of their embeddings converges in ways that track this physical content. §3.1 probes redshift, stellar mass, and sSFR from frozen embeddings to test whether physics knowledge scales with capacity and transfers across instruments. §3.2 then measures embedding alignment across four increasingly challenging settings—within model families, between modalities, and between architectures—each relaxing one shared assumption to isolate its contribution to representational convergence.

3.1. Physical interpretation experiments

Having reliable per-galaxy measurements of redshift, stellar mass, and sSFR allows us to probe two questions directly: **(1)** do larger models encode more knowledge in their embeddings, and **(2)** do different architectures and modalities share similar internal representations?

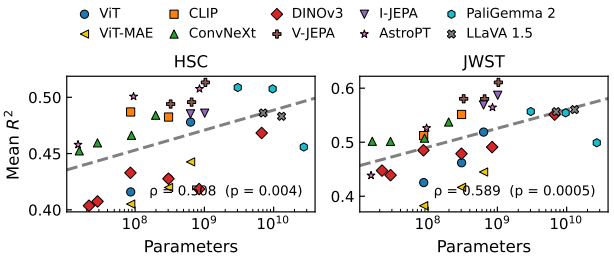


Figure 1. Linear probe R^2 as a function of model size, averaged over mass, sSFR, and redshift. We find a significant positive correlation, even though AstroPT is the only model in our basket pretrained substantially on galaxy imagery.

Larger models encode more physics knowledge. We focus on three physical properties: stellar mass ($M_\star$), specific star formation rate (sSFR), and redshift (z). A model that has internalised the physics of galaxy images should encode all three canonical descriptors. After clipping values to the 1st and 99th percentiles to remove outliers, we train linear probes on the embeddings of each model to predict these properties. For both HSC and JWST, we run 10-fold cross-validation using the 45 000 selected galaxies with matched physical parameter measurements from the COSMOS-Web catalog (40 000 train / 5 000 validation per fold). Fig. 1 shows the average R^2 score across properties versus the parameter count; full per-property results are in App. B. A Spearman's rank test gives a significant positive correlation for our JWST ($\rho = 0.508, p = 0.004$) and HSC ($\rho = 0.589, p = 0.0005$) cases, indicating that larger models recover more of the underlying astrophysics from images alone, despite AstroPT being our only tested model pretrained significantly on galaxies.

Physics performance is correlated between modalities. In Fig. 2, we show JWST R^2 scores as a function of HSC R^2 scores in a grid of $M_\star$, sSFR, and z. We find a strong cross-modal correlation ($p<<0.01$ via a Spearman's rank test) in all cases, indicating that a model's capacity to linearly encode physical information is a consistent property of the model itself rather than that of the observed wavelength range. A principal component analysis across the six probe R^2 scores confirms this with 87% of the variance captured in a single axis with uniformly positive loadings—a 'model quality' axis along which the same models that recover more astrophysics from optical imagery also recover more from near-infrared imagery, despite the two probing largely different physical processes (§2.1). One would expect this if our

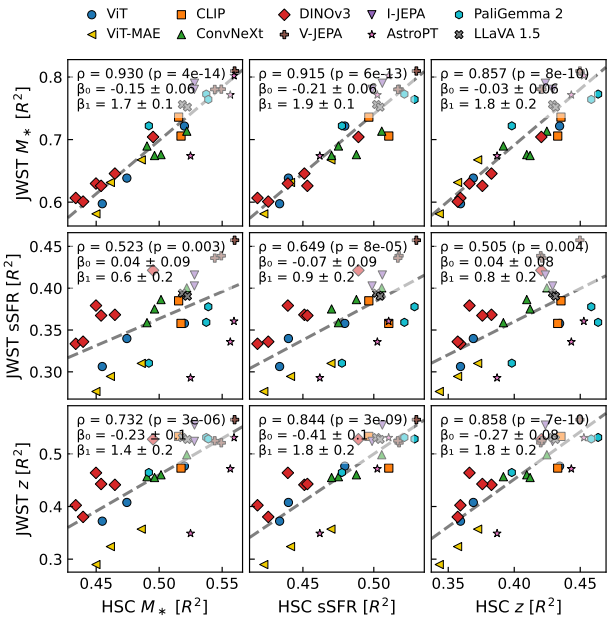


Figure 2. Linear probe R^2 scores of JWST versus HSC modalities on a grid of physical parameters. Model performance is strongly correlated between the instruments, indicating that a model's capacity to linearly encode physical information is a property of its representations rather than a function of the observed wavelength range or modality.

models are converging on representations of the underlying galaxies rather than on modality-specific image statistics.

JWST samples near-infrared stellar light, where emission is dominated by relatively low mass (long living) stars. At near-IR wavelengths, photospheric emission is also less affected by dust obscuration. JWST imagery is therefore a clean tracer of already-built stellar mass but carries little information about ongoing star formation. HSC instead samples UV/optical light, and so is expected to be the more reliable tracer of recent and ongoing star formation (Conroy, 2013). Finally, JWST's longer wavelength baseline pins down redshift more tightly than HSC's (Salvato et al., 2019), so we expect superunity linear regression slopes for $M_\star$ and z on Fig. 2's diagonal and a subunity slope for sSFR—which is what we recover. By definition, sSFR $\equiv$ SFR$/M_\star$, so HSC sSFR carries information about stellar mass as well as star formation. If both modalities encode mass in a shared, modality-invariant way, then the $M_\star$ component of HSC sSFR should align with JWST's mass representation. Because JWST carries little SFR information but traces $M_\star$ cleanly, we therefore expect HSC sSFR to correlate more strongly with JWST $M_\star$ than with JWST sSFR—and indeed we find ρ(HSC sSFR, JWST $M_\star$) $= 0.915$ vs ρ(HSC sSFR, JWST sSFR) $= 0.649$. Taken holistically, our crossmodal correlations are consistent with our models having internalised modality invariant astrophysical content rather than instrument-specific image features, and foreshadows our further crossmodal analysis in §3.2.

Relative orientations of probe directions are consistent across models and with known physics. We assess the relative orientations of the redshift, stellar mass, and sSFR directions in embedding space by measuring their pairwise cosine similarities. Fig. 3 shows the cosine similarity matrices for HSC (left) and JWST (right), averaged across all models. Within the star-forming population, galaxies follow the well-established main sequence, $\log \text{SFR} = \alpha \log M_\star + \beta$, with a sub-linear slope $\alpha \lesssim 1$ (Speagle et al., 2014). As sSFR $\equiv$ SFR$/M_\star$ it follows that $\log \text{sSFR} \propto (\alpha - 1) \log M_\star$, and so we would expect the sSFR-$M_\star$ correlation to be mildly negative. This trend is reinforced by the rising quiescent (non-star forming) galaxy fraction with stellar mass, which shifts more massive galaxies to lower sSFR, and both effects are weak relative to the several orders of magnitude that sSFR spans at fixed mass (Wetzel et al., 2012). All of which is consistent with the small negative cosine recovered across all our tested models in Fig. 3. The correlation between $M_\star$ and redshift arises primarily from a Malmquist-style selection effect (Malmquist, 1922): since higher-redshift galaxies appear fainter, only intrinsically luminous—and therefore more massive—galaxies exceed the detection threshold of telescopes, biasing high-redshift samples toward larger stellar masses. Finally, higher redshift galaxies are observed to have higher star-forming rates (Madau & Dickinson, 2014), which explains the sSFR and redshift correlation. In short, the correlations between these three parameters result from intrinsic galaxy physics and selection effects, yet they are nonetheless captured by all models despite all-but-one of our models having been pre-trained on natural imagery or text.

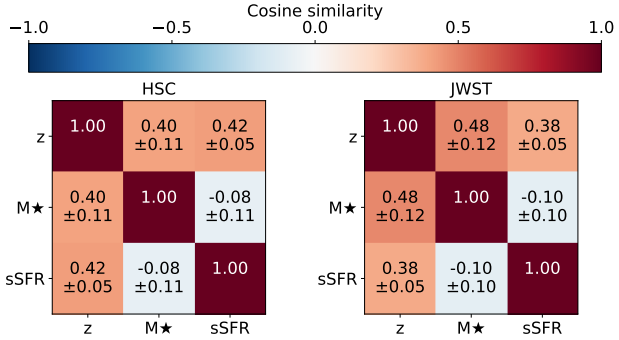


Figure 3. Cosine similarity matrices between the redshift (z), stellar mass ($M_\star$), and sSFR probe weight vectors for HSC (left) and JWST (right), averaged over all models. The relationships between the three probe weight vectors are consistent across all models, suggesting that the embedding spaces encode physics via similar internal organisation.

Taken together, the results in this subsection establish three

facts about how foundation models encode galaxy physics. First, capacity translates into physical knowledge: larger models recover more of the underlying astrophysics from images alone, despite only one model in our basket being substantially pre-trained on galaxy imagery. Second, this capacity is a property of the model rather than the modality: probe performance is tightly correlated between HSC and JWST across all three physical parameters, indicating that a model's ability to linearly encode galaxy physics transfers across observed wavelength ranges. Third, the geometric organisation of physics within embedding space is shared across architectures: the relative orientations of the redshift, stellar mass, and sSFR probe directions agree across models to within small scatter, and the agreement reflects genuine astrophysical and selection-effect correlations rather than arbitrary embedding choices.

3.2. Global and local embedding alignment experiments

Having established that models encode physics to varying degrees, we now ask whether that physical content is correlated to geometric convergence by adopting the approaches described in Huh et al. (2024); Gröger et al. (2026). Specifically, we test global and local embedding similarity scaling within architecture families, between astronomical modalities, and across architecture families. The results in this section are built upon two experimental frameworks: **(1)** cross-model embedding comparison, where embedding pairs are generated by passing the same astronomical mode data into a pair of different models; and **(2)** cross-modality embedding comparison, where embedding pairs are generated by passing different astronomical modality data into the same model. In all cases we measure the embedding similarity via the calibrated MKNN and CKA metrics (§2.3; (Gröger et al., 2026)).

DINOv3 reveals a distinct scaling regime. Our simplest test asks whether MKNN and CKA scores increase in the direction the PRH predicts. We construct ordered pairs in two ways. The method takes consecutively sized pairs of pre-trained models, ordered by model size within a family (Tab. 5); the PRH predicts that embedding similarity of pairs should increase with scale, since better-performing models sit closer to the Platonic Ideal. The second method pairs a single model's embeddings of two distinct astronomical modalities (e.g. JWST and HSC imagery, Tab. 6); here we expect better-performing models to produce more aligned cross-modal embeddings. We treat each ordered pair as an independent Bernoulli trial: success if the similarity score increases along the predicted direction (larger model pair in the intra-architectural case, the larger model in the crossmodal case), failure otherwise. Under the null that scale carries no information about alignment, the success rate is 50%, and we test the basket rate against this null via a binomial test. Across our full basket, including DINOv3, the MKNN trend is in the predicted direction (60/96 = 62%, $p = 0.02$), confirming the PRH's basic prediction. However, this effect is concentrated outside DINOv3, and excluding it sharpens the signal substantially (51/69 = 74%, $p = 0.006$) while DINOv3 alone trends in the opposite direction (9/27 = 33%). A Fisher exact test confirms that DINOv3 sits in a statistically distinct scaling regime from the rest of the basket ($p = 4e{-}4$ for MKNN, $p = 0.03$ for CKA). CKA shows no significant trend in either case.

We view the DINOv3 anomaly as a probe of the PRH's underlying mechanism. The original hypothesis attributes convergence to three drivers: task generality, model capacity, and simplicity bias (Huh et al., 2024). None of these straightforwardly predict that distilled model families should converge under scale. If anything, distillation pins all students near a fixed teacher representation, decoupling capacity from the independent optimization pressure that the PRH posits as the driver of convergence. In other words, varying student capacity around a frozen 7B teacher is not the same kind of scaling the PRH was formulated around. We therefore propose a refinement: representational convergence under scale is a property of independently trained models, and may not extend to families produced by capacity-varied distillation from a shared teacher. Given this and the results from the Fisher exact test, we exclude DINOv3 from the remaining tests in this section[5].

Within-family local embedding alignment correlates with physics performance. We further test whether embedding similarities of models within the same architecture family tracks how well those models recover galaxy physics. Concretely, for each model family we pair adjacent size rungs (e.g. ConvNeXtv2 Nano vs. Tiny, Tiny vs. Base, Base vs. Large), compute their MKNN and CKA scores on each of our three imaging surveys (JWST, Legacy, HSC; §2.1), and compare these pairwise alignment scores against the physical-probe R^2 of the larger of the two models in the pair (§3.1). The PRH predicts that better-performing pairs should also be more internally aligned, since both models in a high-performing pair should sit closer to the Platonic Ideal. Fig. 4 shows this is borne out for MKNN: across all three surveys we find a strong, significant Spearman correlation ($\rho = 0.679$–0.724, $p < 0.005$). CKA shows no significant trend in any survey ($p \geq 0.3$) Per-property breakdowns (mass, sSFR, redshift) and the full table of pairwise scores are reported in App. C.

Local embedding alignment across modalities correlates with physics performance. We now test whether

[5]We note that PaliGemma's crossmodal MKNN also trends negatively with scale, plausibly due to its vision tower being held at fixed-capacity across model sizes. However, the same Fisher exact test we applied to DINOv3 does not reach significance here ($p = 0.74$ for MKNN), so we retain PaliGemma in the remaining analyses.

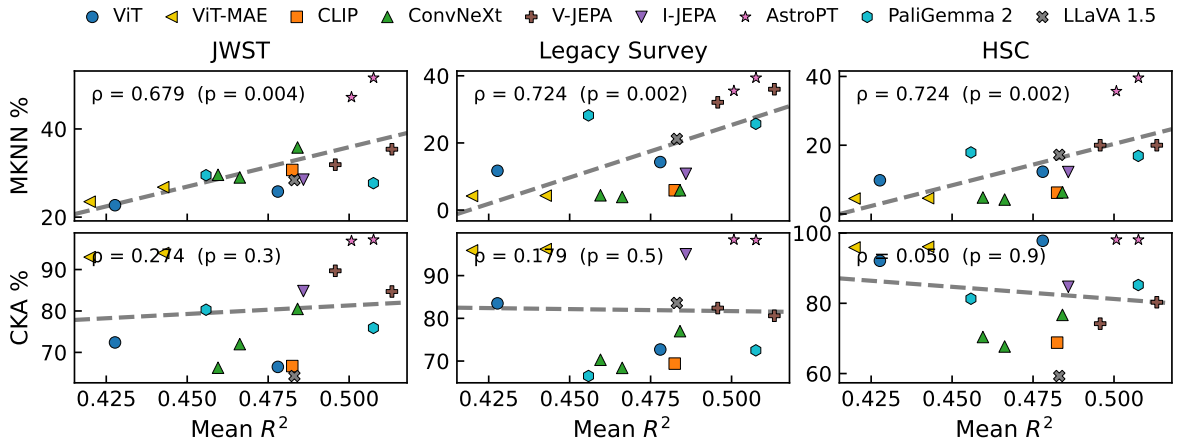


Figure 4. Intra-architectural scaling results. We plot our MKNN and CKA similarity values within each architecture family against the mean R^2 probe performance across the physical tasks described in §3.1. In each pane we state the ρ and p values as measured via a Spearman's rank test.

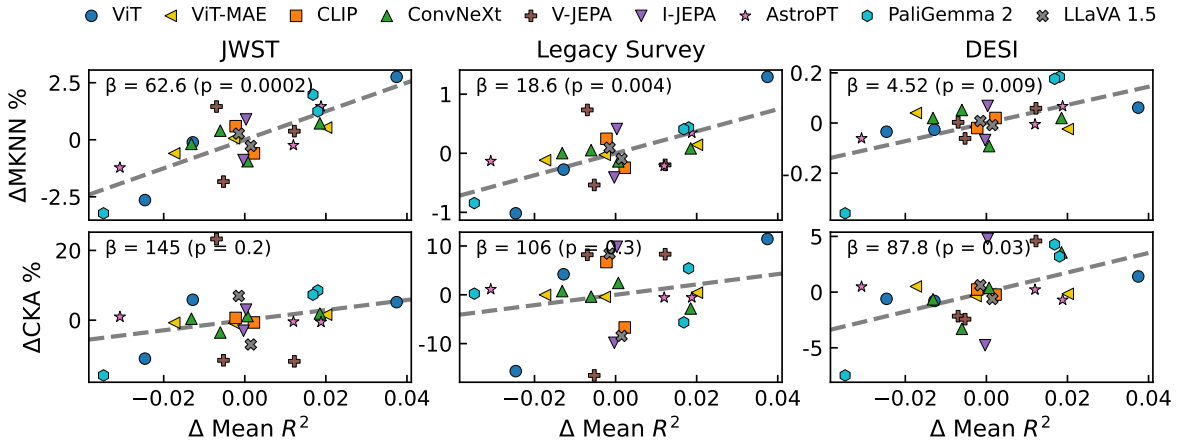


Figure 5. Crossmodal MKNN and CKA metrics vs R^2 (§3.1). R^2 and MKNN/CKA are demeaned per architecture family to show the trend without architecture family confounders. We state β and p from the ANCOVA tests $R^2 \sim \text{MKNN} + C(\text{family})$, and $R^2 \sim \text{CKA} + C(\text{family})$.

cross-modal embeddings of the same galaxies converge as model performance grows. For each model we compute the MKNN and CKA alignment between the embeddings of HSC imagery and the embeddings of a second modality (JWST imagery, Legacy Survey imagery, or DESI spectra as processed by Specformer), and regress these alignment scores against probe R^2 (§3.1). In this test, if two instruments observing the same galaxies result in embeddings that align, the alignment cannot be inherited from raw pixel similarity and must instead reflect shared latent content recovered by the model. Fig. 5 summarizes our results. Controlling for familial effects via an ANCOVA, we find a significant correlation ($p < 0.01$) in all tested MKNN cases (JWST: $\beta = 62.6, p = 2e-4$; Legacy: $\beta = 18.6, p = 0.004$; DESI: $\beta = 4.52, p = 0.009$), and no significant trend in our tested CKA cases. Notably, we find a significant relation for DESI spectra vs HSC imagery: spectra and images share essentially no low-level statistics—a 1D sequence of flux measurement vs a 2D RGB cutout, processed by a spectra model that has never seen an image, and image models that have never seen spectra—so alignment between them must be driven by shared physical content. The MKNN signal here (better-performing models produce more aligned crossmodal embeddings) is direct evidence that convergence is mediated by underlying astrophysics rather than by input correlations or shared inductive biases inherited from natural-image pre-training. Per-property breakdowns are reported in App. D.

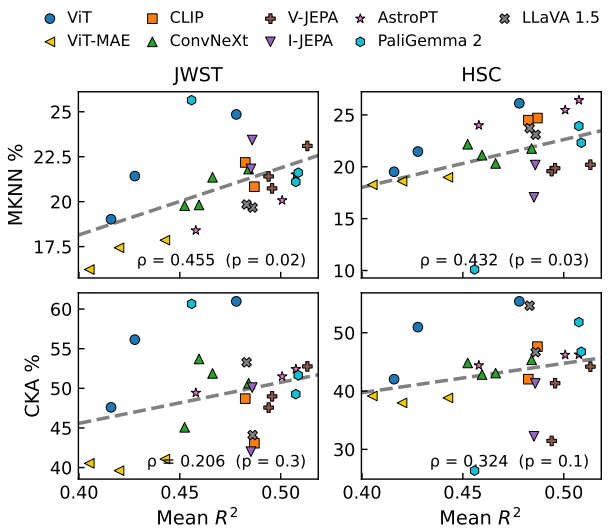


Figure 6. Cross-architecture metric distance (MKNN, CKA), plotted against average R^2 over the three physical properties for HSC and JWST images (§3.1). We find that probe performance is significantly correlated with MKNN, but not with CKA.

Cross-architectural alignment correlates with physics performance. Finally, we ask whether our basket of models is converging toward a single 'ideal' embedding space across architectures. For each model in our basket we compute the MKNN and CKA scores against every other model and take the mean across comparisons, yielding a single per-model alignment score that quantifies how close a model sits to the rest of the basket in embedding space. Fig. 6 plots the mean cross-architectural MKNN and CKA against the average linear probe R^2 as described in §3.1. We restrict this analysis to the JWST-HSC crossmatched MMU dataset for computational tractability, since pairwise comparison across the full basket scales quadratically with model count. We find a significant positive Spearman correlation between MKNN alignment and physics performance on both probe-spaces ($\rho = 0.455, p = 0.02$ for JWST, $\rho = 0.432, p = 0.03$ for HSC), indicating that models with better physics performance sit closer to the basket's average embedding geometry. This trend holds across all tested physical properties (App. E). As in our earlier tests, we do not see a significant trend for CKA vs physics performance ($p > 0.1$ in both probe-spaces).

Across all tests in this subsection—scaling within model families, alignment between astronomical modalities, and convergence across architectures—we find that local embedding similarity (MKNN) tracks physics performance, while global similarity (CKA) does not. The intra-architectural test shows that within a family, better-performing size-pairs are more locally aligned. The crossmodal test extends this to pairs of instruments observing the same galaxies, and crucially holds for DESI spectra vs. HSC imagery—a comparison in which low-level input statistics share essentially nothing, isolating shared physical content as the driver of alignment. The cross-architectural test shows that models that encode more physics sit closer to the basket's mean embedding geometry, suggesting convergence toward a shared representation rather than family-specific solutions. The CKA null result across all four tests indicates that convergence is a property of local neighbourhood structure rather than global embedding geometry, consistent with the Aristotelian rather than the strict Platonic reading of the convergence hypothesis (Gröger et al., 2026).

4. Think local (embedding convergence), not global

Below we summarize and distil our results across all our experiments.

We find strong support for the Aristotelian Representation Hypothesis. Across our tests in §3.2 we find consistently that local neighborhood structure (MKNN) tracks model performance, while global embedding similarity (CKA) does not. We must remind ourselves that, aside from AstroPT and Specformer, our tested models are not significantly pre-trained on astronomical data; that these models identify any correspondence between fundamentally different astronomical observations is remarkable, and suggests that sufficiently scaled up neural networks learn universal structural patterns transcending their training domains. Our local-not-global embedding convergence pattern supports the Aristotelian reading of representational convergence (Gröger et al., 2026): foundation models trained on different objectives, modalities, and data appear to agree on which galaxies are like other galaxies, without agreeing on a shared global coordinate system in which to place them. We can also see that our natural image-trained models achieve embedding alignment that increases with model performance with Specformer's DESI spectral embeddings, therefore showing correspondence between fundamentally different modalities and data types they have never encountered that do not share low-level statistics.

AstroPT does not dominate. AstroPTv2 is the only model in our basket pre-trained substantially on galaxy imagery, and yet it sits comfortably within the trends defined by general-purpose models rather than above them, in physics probe performance and in representational similarity. We read this as evidence that sufficiently scaled general-purpose models already recover representations close to those delivered by domain-specific pre-training—precisely as Sutton's Bitter Lesson would predict (Sutton, 2019). This follows directly from the representational convergence documented throughout our remaining results. Domain-specific architectures may still be useful where the input modality is genuinely outside the natural-image distribution (Specformer

on 1D spectra is a clear example), but for galaxy imaging, scale and data diversity appear to substitute for domain specificity.

DINOv3 is a probe of the convergence mechanism. DINOv3 scales in the opposite direction to the rest of our model basket. The PRH attributes convergence to task generality, capacity, and simplicity bias, none of which obviously predict that capacity-varied students distilled from a frozen 7B teacher should converge under scale as distillation pins students near a fixed reference rather than letting them discover representations under independent optimization pressure. We therefore propose that representational convergence under scale is a property of independently trained model families, and may not extend to capacity-varied distillation. Training self-distilled model variants from scratch at multiple scales would test this directly and is a natural target for follow-up work.

Limitations. While our results provide compelling evidence for the ARH, we note several limitations that suggest avenues for future exploration. Our cross-modal samples are uneven in size (1.67k galaxies for JWST–HSC, 102k for Legacy–HSC, 18.6k for DESI–HSC), and the JWST–HSC overlap in particular may be underpowered to capture the full diversity of galaxy populations. We probe only three physical properties; convergence on morphology, kinematics, or environment may behave differently. Our cross-architectural test is restricted to JWST for tractability. And finally, our analysis is correlational: we cannot distinguish whether better physics drives alignment or whether some third factor drives both. We leave these questions for future studies.

Implications for astronomical foundation modelling. We observe general improvement in representational alignment at larger model scales, suggesting that each architecture is converging towards a shared representation. Taken to its conclusion, this convergence implies that future efforts in astronomical foundation modelling should focus less on astronomy-specific architectures and more on scale and data diversity. It also follows that the astronomy community should embrace pre-trained foundation models rather than training from scratch: if all architectures converge toward the same representations, then starting from models pre-trained on natural images or text—with their billions of parameters and massive computational investment already spent—offers both superior performance and dramatic reductions in environmental impact. The broader open source machine learning community has already invested the GPU-centuries needed to learn general-purpose representations, we need now only gently guide these models toward astronomical use-cases.

Acknowledgements

MJS would like to thank Regina Sarmiento, Rafael Martínez-Galarza, and the AstroAI group for illuminating discussion and comments that improved this paper.

We would like to thank Stella Biderman and EleutherAI for their compute and infrastructure support.

This research made use of the University of Hertfordshire's High Performance Computing facility (https://uhhpc.herts.ac.uk/).

This research used the DeltaAI advanced computing and data resource through allocation number PHY250286, which is supported by the National Science Foundation (award OAC 2320345) and the State of Illinois. DeltaAI is a joint effort of the University of Illinois Urbana-Champaign and its National Center for Supercomputing Applications. Access to DeltaAI was granted through the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by U.S. National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

This research used data obtained with the Dark Energy Spectroscopic Instrument (DESI). DESI construction and operations is managed by the Lawrence Berkeley National Laboratory. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High-Energy Physics, under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: www.desi.lbl.gov/collaborating-institutions. The DESI collaboration is honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation, the U.S. Department of Energy, or any of the listed funding agencies.

This work is based in part on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Tele-

scopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST.

The DESI Legacy Imaging Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS), the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. Pipeline processing and analyses of the data were supported by NOIRLab and the Lawrence Berkeley National Laboratory (LBNL). Legacy Surveys also uses data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), a project of the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. Legacy Surveys was supported by: the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility; the U.S. National Science Foundation, Division of Astronomical Sciences; the National Astronomical Observatories of China, the Chinese Academy of Sciences and the Chinese National Natural Science Foundation. LBNL is managed by the Regents of the University of California under contract to the U.S. Department of Energy. The complete acknowledgments can be found at https://www.legacysurvey.org/acknowledgment/.

Impact Statement

This paper studies whether astronomical foundation models converge toward shared representations of the underlying physics of galaxies. Its primary aim is to advance the field of machine learning and its application to astronomy. A practical implication of our findings is that general-purpose pre-trained models can be reused for astronomical tasks rather than trained from scratch, which can substantially reduce the computational and environmental cost of building astro-foundation models. We do not foresee negative societal consequences specific to this work beyond those generally associated with advancing machine learning, none of which we feel must be specifically highlighted here.

Contributors

As a way to provide transparency in a large, multi-author scientific effort, we adopt the CRediT taxonomy[6] to report individual contributions. In this paper we define the contribution categories as:

Conceptualization [C]: Paper- or section-level intellectual framing: defining scope, research questions, experimental design, and narrative arc.

Methodology [M]: Designing the probe protocol, alignment metrics (MKNN, CKA), statistical tests, cross-matching, and preprocessing pipelines.

Software [S]: Implementing embedding extraction across model families, assembling cross-matched datasets, alignment metric code, probe training, statistical analyses, and figure production.

Project administration [P]: Soliciting and organizing inputs, managing timelines and revisions, integrating contributions, and ensuring internal consistency across contributors.

Writing – Original Draft [W]: Substantive preparation of original text, including drafting new material and/or synthesizing multiple contributions into a coherent section or subsection.

Writing – Review & Editing [R]: Substantive review and revision of the manuscript text, including critical feedback, edits for clarity and correctness, and incorporation of reviewer comments.

Where shown, bracketed scope tags (e.g., [§3.2]) indicate the section(s) associated with a listed role; [all] denotes contributions spanning the full manuscript. Bolded scopes identify primary contributions, and unbolded scopes identify secondary contributions.

[6] CRediT – Contributor Roles Taxonomy: https://credit.niso.org/.

Name	Contribution
Trinidad Borrell	C[§2.3, §3.2, §4], S[§2.3, §3.2], W[§3.2], R[§2, §3, §4]
Steven Dillmann	C[§2.3, §3.2], M[§2.3, §3.1], S[§2.3, §3.1, §3.2], R[§2, §3]
Kshitij Duraphe	C[§2.3, §3.2], M[§2.3, §3.1], S[**§2.3, §3.1, §3.2**, §2.1], R[§2, §3]
Furkan Eris	S[§3.2, §2.2], R[§2]
Ashod Khederlarian	C[§2.1, §3.1], M[§2.1, §3.1], S[§2.1, §3.1], W[§2.1, §3.1], R[§2, §3]
Aman Kumar	C[§2.3, §3.2], S[§2.2, §2.3, §3.2], W[§3.2], R[§2]
Giovanni Marraffini	C[§2.3 ,§3.2], S[§2.3, §3.2], W[§3.2], R[§2, §3, §4]
Michael J. Smith	C[**all**], M[**all**], S[**all**], P[**all**], W[**all**], R[**all**]
Shashwat Sourav	C[§2.1, §3], M[§2, §2.1, §2.3, §3], S[§2.3], P[all], W[§C], R[§2, §3]
Rocco Di Tella	C[§3.2, §4], S[**§3.2, §C**], R[§2, §3, §4]
John F. Wu	C[§2.3, §3.2], M[§2.1,§2.3], P[**all**], W[§2, §3]

References

Adorf, H. M. and Johnston, M. D. Artificial neural nets in astronomy. In *Arbeitspapier der Gesellschaft für Mathematik and Datenverarbeitung*, volume 329 of *Arbeitspapier der Gesellschaft für Mathematik and Datenverarbeitung*, 1988.

Angel, J., Wizinowich, P., Lloyd-Hart, M., and Sandler, D. Adaptive optics for array telescopes using neural-network techniques. *Nature*, 348(6298):221–224, 1990. ISSN 0028-0836. doi: 10.1038/348221a0.

Arnouts, S., Cristiani, S., Moscardini, L., Matarrese, S., Lucchin, F., Fontana, A., and Giallongo, E. Measuring and modelling the redshift evolution of clustering: the hubble deep field north. *Monthly Notices of the Royal Astronomical Society*, 310:540–556, 1999. doi: 10.1046/j.1365-8711.1999.02978.x.

Assran, M., Duval, Q., Misra, I., Bojanowski, P., Vincent, P., Rabbat, M., LeCun, Y., and Ballas, N. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2301.08243.

Assran, M., Bardes, A., Fan, D., Garrido, Q., Howes, R., Mojtaba, Komeili, Muckley, M., Rizvi, A., Roberts, C., Sinha, K., Zholus, A., Arnaud, S., Gejji, A., Martin, A., Hogan, F. R., Dugas, D., Bojanowski, P., Khalidov, V., Labatut, P., Massa, F., Szafraniec, M., Krishnakumar, K., Li, Y., Ma, X., Chandar, S., Meier, F., LeCun, Y., Rabbat, M., and Ballas, N. V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning. *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2506.09985.

Beyer, L., Steiner, A., Pinto, A. S., Kolesnikov, A., Wang, X., Salz, D., Neumann, M., Alabdulmohsin, I., Tschannen, M., Bugliarello, E., Unterthiner, T., Keysers, D., Koppula, S., Liu, F., Grycner, A., Gritsenko, A., Houlsby, N., Kumar, M., Rong, K., Eisenschlos, J., Kabra, R., Bauer, M., Bošnjak, M., Chen, X., Minderer, M., Voigtlaender, P., Bica, I., Balazevic, I., Puigcerver, J., Papalampidi, P., Henaff, O., Xiong, X., Soricut, R., Harmsen, J., and Zhai, X. PaliGemma: A versatile 3B VLM for transfer. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2407.07726.

Caplar, N., Beebe, W., Branton, D., Campos, S., Connolly, A., DeLucchi, M., Jones, D., Juric, M., Kubica, J., Malanchev, K., Mandelbaum, R., and McGuire, S. Using LSDB to enable large-scale catalog distribution, cross-matching, and analytics. *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2501.02103.

Casey, C. M., Kartaltepe, J. S., Drakos, N. E., Franco, M., Harish, S., Paquereau, L., Ilbert, O., Rose, C., Cox, I. G., Nightingale, J. W., Robertson, B. E., Silverman, J. D., Koekemoer, A. M., Massey, R., McCracken, H. J., Rhodes, J., Akins, H. B., Allen, N., Amvrosiadis, A., Arango-Toro, R. C., Bagley, M. B., Bongiorno, A., Capak, P. L., Champagne, J. B., Chartab, N., Chávez Ortiz, Ó. A., Chworowsky, K., Cooke, K. C., Cooper, O. R., Darvish, B., Ding, X., Faisst, A. L., Finkelstein, S. L., Fujimoto, S., Gentile, F., Gillman, S., Gould, K. M. L., Gozaliasl, G., Hayward, C. C., He, Q., Hemmati, S., Hirschmann, M., Jahnke, K., Jin, S., Khostovan, A. A., Kokorev, V., Lambrides, E., Laigle, C., Larson, R. L., Leung, G. C. K., Liu, D., Liaudat, T., Long, A. S., Magdis, G., Mahler, G., Mainieri, V., Manning, S. M., Maraston, C., Martin, C. L., McCleary, J. E., McKinney, J., McPartland, C. J. R., Mobasher, B., Pattnaik, R., Renzini, A., Rich, R. M., Sanders, D. B., Sattari, Z., Scognamiglio, D., Scoville, N., Sheth, K., Shuntov, M., Sparre, M., Suzuki, T. L., Talia, M., Toft, S., Trakhtenbrot, B., Urry, C. M., Valentino, F., Vanderhoof, B. N., Vardoulaki, E., Weaver, J. R., Whitaker, K. E., Wilkins, S. M., Yang, L., and Zavala, J. A. COSMOS-Web: An Overview of the JWST Cosmic Origins Survey. *ApJ*, 954(1):31, September 2023. doi: 10.3847/1538-4357/acc2bc.

Charnock, T., Lavaux, G., and Wandelt, B. D. Automatic physical inference with information maximizing neural networks. *Physical Review D*, 97(8):083004, 2018. ISSN 2470-0029. doi: 10.1103/PhysRevD.97.083004.

Chechik, G., Sharma, V., Shalit, U., and Bengio, S. Large scale online learning of image similarity through ranking.

Journal of Machine Learning Research, 11(36):1109–1135, 2010. URL http://jmlr.org/papers/v11/chechik10a.html.

Conroy, C. Modeling the Panchromatic Spectral Energy Distributions of Galaxies. *Annual Review of Astronomy and Astrophysics*, (Volume 51, 2013):393–455, 2013. doi: 10.1146/annurev-astro-082812-141017.

de Haan, T., Ting, Y.-S., Ghosal, T., Nguyen, T. D., Accomazzi, A., Wells, A., Ramachandra, N., Pan, R., and Sun, Z. AstroMLab 3: Achieving GPT-4o Level Performance in Astronomy with a Specialized 8B-Parameter Large Language Model. *ArXiv e-prints*, 2024. doi: 10.1038/s41598-025-97131-y.

DESI Collaboration, Adame, A. G., Aguilar, J., Ahlen, S., Alam, S., Aldering, G., Alexander, D. M., Alfarsy, R., Prieto, C. A., Alvarez, M., Alves, O., Anand, A., Andrade-Oliveira, F., Armengaud, E., Asorey, J., Avila, S., Aviles, A., Bailey, S., Balaguera-Antolínez, A., Ballester, O., Baltay, C., Bault, A., Bautista, J., Behera, J., Beltran, S. F., BenZvi, S., Silva, L. B. e., Bermejo-Climent, J. R., Berti, A., Besuner, R., Beutler, F., Bianchi, D., Blake, C., Blum, R., Bolton, A. S., Brieden, S., Brodzeller, A., Brooks, D., Brown, Z., Buckley-Geer, E., Burtin, E., Cabayol-Garcia, L., Cai, Z., Canning, R., Cardiel-Sas, L., Rosell, A. C., Castander, F. J., Cervantes-Cota, J. L., Chabanier, S., Chaussidon, E., Chaves-Montero, J., Chen, S., Chen, X., Chuang, C., Claybaugh, T., Cole, S., Cooper, A. P., Cuceu, A., Davis, T. M., Dawson, K., de Belsunce, R., de la Cruz, R., de la Macorra, A., Costa, J. D., de Mattia, A., Demina, R., Demirbozan, U., DeRose, J., Dey, A., Dey, B., Dhungana, G., Ding, J., Ding, Z., Doel, P., Doshi, R., Douglass, K., Edge, A., Eftekharzadeh, S., Eisenstein, D. J., Elliott, A., Ereza, J., Escoffier, S., Fagrelius, P., Fan, X., Fanning, K., Fawcett, V. A., Ferraro, S., Flaugher, B., Font-Ribera, A., Forero-Romero, J. E., Forero-Sánchez, D., Frenk, C. S., Gänsicke, B. T., García, L. Á., García-Bellido, J., Garcia-Quintero, C., Garrison, L. H., Gil-Marín, H., Golden-Marx, J., Gontcho, S. G. A., Gonzalez-Morales, A. X., Gonzalez-Perez, V., Gordon, C., Graur, O., Green, D., Gruen, D., Guy, J., Hadzhiyska, B., Hahn, C., Han, J. J., Hanif, M. M. S., Herrera-Alcantar, H. K., Honscheid, K., Hou, J., Howlett, C., Huterer, D., Iršič, V., Ishak, M., Jacques, A., Jana, A., Jiang, L., Jimenez, J., Jing, Y. P., Joudaki, S., Joyce, R., Jullo, E., Juneau, S., Karaçaylı, N. G., Karim, T., Kehoe, R., Kent, S., Khederlarian, A., Kim, S., Kirkby, D., Kisner, T., Kitaura, F., Kizhuprakkat, N., Kneib, J., Koposov, S. E., Kovács, A., Kremin, A., Krolewski, A., L'Huillier, B., Lahav, O., Lambert, A., Lamman, C., Lan, T.-W., Landriau, M., Lang, D., Lange, J. U., Lasker, J., Leauthaud, A., Le Guillou, L., Levi, M. E., Li, T. S., Linder, E., Lyons, A., Magneville, C., Manera, M., Manser, C. J., Margala, D., Martini, P., McDonald, P., Medina, G. E., Medina-Varela, L., Meisner, A., Mena-Fernández, J., Meneses-Rizo, J., Mezcua, M., Miquel, R., Montero-Camacho, P., Moon, J., Moore, S., Moustakas, J., Mueller, E., Mundet, J., Muñoz-Gutiérrez, A., Myers, A. D., Nadathur, S., Napolitano, L., Neveux, R., Newman, J. A., Nie, J., Nikutta, R., Niz, G., Norberg, P., Noriega, H. E., Paillas, E., Palanque-Delabrouille, N., Palmese, A., Pan, Z., Parkinson, D., Penmetsa, S., Percival, W. J., Pérez-Fernández, A., Pérez-Ràfols, I., Pieri, M., Poppett, C., Porredon, A., Pothier, S., Prada, F., Pucha, R., Raichoor, A., Ramírez-Pérez, C., Ramirez-Solano, S., Rashkovetskyi, M., Ravoux, C., Rocher, A., Rockosi, C., Ross, A. J., Rossi, G., Ruggeri, R., Ruhlmann-Kleider, V., Sabiu, C. G., Said, K., Saintonge, A., Samushia, L., Sanchez, E., Saulder, C., Schaan, E., Schlafly, E. F., Schlegel, D., Scholte, D., Schubnell, M., Seo, H., Shafieloo, A., Sharples, R., Sheu, W., Silber, J., Sinigaglia, F., Siudek, M., Slepian, Z., Smith, A., Soumagnac, M. T., Sprayberry, D., Stephey, L., Suárez-Pérez, J., Sun, Z., Tan, T., Tarlé, G., Tojeiro, R., Ureña-López, L. A., Vaisakh, R., Valcin, D., Valdes, F., Valluri, M., Vargas-Magaña, M., Variu, A., Verde, L., Walther, M., Wang, B., Wang, M. S., Weaver, B. A., Weaverdyck, N., Wechsler, R. H., White, M., Xie, Y., Yang, J., Yèche, C., Yu, J., Yuan, S., Zhang, H., Zhang, Z., Zhao, C., Zheng, Z., Zhou, R., Zhou, Z., Zou, H., Zou, S., and Zu, Y. The Early Data Release of the Dark Energy Spectroscopic Instrument. *Astronomical Journal*, 168(2):58, 2024. ISSN 1538-3881. doi: 10.3847/1538-3881/ad3217.

Dey, A., Schlegel, D. J., Lang, D., Blum, R., Burleigh, K., Fan, X., Findlay, J. R., Finkbeiner, D., Herrera, D., Juneau, S., et al. Overview of the DESI Legacy Imaging Surveys. *Astronomical Journal*, 157(5):168, 2019. ISSN 1538-3881. doi: 10.3847/1538-3881/ab089d.

Dieleman, S., Willett, K. W., and Dambre, J. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, 450(2):1441–1459, 2015. doi: 10.1093/mnras/stv632.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv e-prints*, 2020. doi: 10.48550/arXiv.2010.11929.

Euclid Collaboration, Siudek, M., Huertas-Company, M., Smith, M., Martinez-Solaeche, G., Lanusse, F., Ho, S., Angeloudi, E., Cunha, P. A. C., Sánchez, H. D., Dunn, M., Fu, Y., Iglesias-Navarro, P., Junais, J., Knapen, J. H., Laloux, B., Mezcua, M., Roster, W., Stevens, G., Vega-Ferrero, J., Aghanim, N., Altieri, B., Amara, A., Andreon, S., Auricchio, N., Aussel, H., Baccigalupi, C., Baldi, M.,

Bardelli, S., Battaglia, P., Biviano, A., Bonchi, A., Branchini, E., Brescia, M., Brinchmann, J., Camera, S., Cañas-Herrera, G., Capobianco, V., Carbone, C., Carretero, J., Casas, S., Castander, F. J., Castellano, M., Castignani, G., Cavuoti, S., Chambers, K. C., Cimatti, A., Colodro-Conde, C., Congedo, G., Conselice, C. J., Conversi, L., Copin, Y., Courbin, F., Courtois, H. M., Cropper, M., Da Silva, A., Degaudenzi, H., De Lucia, G., Di Giorgio, A. M., Dinis, J., Dolding, C., Dole, H., Dubath, F., Duncan, C. A. J., Dupac, X., Dusini, S., Escoffier, S., Farina, M., Farinelli, R., Faustini, F., Ferriol, S., Finelli, F., Fotopoulou, S., Frailis, M., Franceschi, E., Galeotta, S., George, K., Gillis, B., Giocoli, C., Gracia-Carpio, J., Granett, B. R., Grazian, A., Grupp, F., Gwyn, S., Haugan, S. V. H., Holmes, W., Hook, I. M., Hormuth, F., Hornstrup, A., Jahnke, K., Jhabvala, M., Keihänen, E., Kermiche, S., Kiessling, A., Kubik, B., Kümmel, M., Kunz, M., Kurki-Suonio, H., Boulc'h, Q. L., Brun, A. M. C. L., Mignant, D. L., Ligori, S., Lilje, P. B., Lindholm, V., Lloro, I., Mainetti, G., Maino, D., Maiorano, E., Mansutti, O., Marcin, S., Marggraf, O., Martinelli, M., Martinet, N., Marulli, F., Massey, R., Maurogordato, S., McCracken, H. J., Medinaceli, E., Mei, S., Melchior, M., Mellier, Y., Meneghetti, M., Merlin, E., Meylan, G., Mora, A., Moresco, M., Moscardini, L., Nakajima, R., Neissner, C., Niemi, S.-M., Nightingale, J. W., Padilla, C., Paltani, S., Pasian, F., Pedersen, K., Percival, W. J., Pettorino, V., Pires, S., Polenta, G., Poncet, M., Popa, L. A., Pozzetti, L., Raison, F., Renzi, A., Rhodes, J., Riccio, G., Romelli, E., Roncarelli, M., Saglia, R., Sakr, Z., Sánchez, A. G., Sapone, D., Sartoris, B., Schewtschenko, J. A., Schneider, P., Schrabback, T., Scodeggio, M., Secroun, A., Seidel, G., Seiffert, M., Serrano, S., Simon, P., Sirignano, C., Sirri, G., Stanco, L., Steinwagner, J., Talladа-Crespí, P., Taylor, A. N., Tereno, I., Toft, S., Toledo-Moreo, R., Torradeflot, F., Tutusaus, I., Valenziano, L., Valiviita, J., Vassallo, T., Kleijn, G. V., Veropalumbo, A., Wang, Y., Weller, J., Zacchei, A., Zamorani, G., Zerbi, F. M., Zinchenko, I. A., Zucca, E., Allevato, V., Ballardini, M., Bolzonella, M., Bozzo, E., Burigana, C., Cabanac, R., Cappi, A., Di Ferdinando, D., Vigo, J. A. E., Gabarra, L., Martín-Fleitas, J., Matthew, S., Mauri, N., Metcalf, R. B., Pezzotta, A., Pöntinen, M., Porciani, C., Risso, I., Scottez, V., Sereno, M., Tenti, M., Viel, M., Wiesmann, M., Akrami, Y., Andika, I. T., Anselmi, S., Archidiacono, M., Atrio-Barandela, F., Benoist, C., Benson, K., Bertacca, D., Bethermin, M., Bisigello, L., Blanchard, A., Blot, L., Brown, M. L., Bruton, S., Calabro, A., Quevedo, B. C., Caro, F., Carvalho, C. S., Castro, T., Charles, Y., Cogato, F., Cooray, A. R., Cucciati, O., Davini, S., De Paolis, F., Desprez, G., Díaz-Sánchez, A., Diaz, J. J., Di Domizio, S., Diego, J. M., Duc, P.-A., Enia, A., Fang, Y., Ferrari, A. G., Ferreira, P. G., Finoguenov, A., Fontana, A., Franco, A., Ganga, K., García-Bellido, J., Gasparetto, T., Gautard, V., Gaztanaga, E., Giacomini, F., Gianotti, F., Gozaliasl, G., Guidi, M., Gutierrez, C. M., Hall, A., Hartley, W. G., Hemmati, S., Hernández-Monteagudo, C., Hildebrandt, H., Hjorth, J., Kajava, J. J. E., Kang, Y., Kansal, V., Karagiannis, D., Kiiveri, K., Kirkpatrick, C. C., Kruk, S., Graet, J. L., Legrand, L., Lembo, M., Lepori, F., Leroy, G., Lesci, G. F., Lesgourgues, J., Leuzzi, L., Liaudat, T. I., Loureiro, A., Macias-Perez, J., Maggio, G., Magliocchetti, M., Magnier, E. A., Mannucci, F., Maoli, R., Martins, C. J. A. P., Maurin, L., Miluzio, M., Monaco, P., Moretti, C., Morgante, G., Murray, C., Naidoo, K., Navarro-Alsina, A., Nesseris, S., Passalacqua, F., Paterson, K., Patrizii, L., Pisani, A., Potter, D., Quai, S., Radovich, M., Sacquegna, S., Sahlén, M., Sanders, D. B., Sarpa, E., Schneider, A., Sciotti, D., Scognamiglio, D., Sellentin, E., Smith, L. C., Tanidis, K., Testera, G., Teyssier, R., Tosi, S., Troja, A., Tucci, M., Valieri, C., Venhola, A., Vergani, D., Verza, G., Vielzeuf, P., Walton, N. A., and Sorce, J. G. Euclid Quick Data Release (Q1) Exploring galaxy properties with a multimodal foundation model. *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2503.15312.

Gardner, J. P., Mather, J. C., Abbott, R., Abell, J. S., Abernathy, M., Abney, F. E., Abraham, J. G., Abraham, R., Abul-Huda, Y. M., Acton, S., Adams, C. K., Adams, E., Adler, D. S., Adriaensen, M., Aguilar, J. A., Ahmed, M., Ahmed, N. S., Ahmed, T., Albat, R., Albert, L., Alberts, S., Aldridge, D., Allen, M. M., Allen, S. S., Altenburg, M., Altunc, S., Alvarez, J. L., Álvarez-Márquez, J., Alves de Oliveira, C., Ambrose, L. L., Anandakrishnan, S. M., Andersen, G. C., Anderson, H. J., Anderson, J., Anderson, K., Anderson, S. M., Aprea, J., Archer, B. J., Arenberg, J. W., Argyriou, I., Arribas, S., Artigau, É., Arvai, A. R., Atcheson, P., Atkinson, C. B., Averbukh, J., Aymergen, C., Bacinski, J. J., Baggett, W. E., Bagnasco, G., Baker, L. L., Balzano, V. A., Banks, K. A., Baran, D. A., Barker, E. A., Barrett, L. K., Barringer, B. O., Barto, A., Bast, W., Baudoz, P., Baum, S., Beatty, T. G., Beaulieu, M., Bechtold, K., Beck, T., Beddard, M. M., Beichman, C., Bellagama, L., Bely, P., Berger, T. W., Bergeron, L. E., Bernier, A.-D., Bertch, M. D., Beskow, C., Betz, L. E., Biagetti, C. P., Birkmann, S., Bjorklund, K. F., Blackwood, J. D., Blazek, R. P., Blossfeld, S., Bluth, M., Boccaletti, A., Boegner, Jr., M. E., Bohlin, R. C., Boia, J. J., Böker, T., Bonaventura, N., Bond, N. A., Bosley, K. A., Boucarut, R. A., Bouchet, P., Bouwman, J., Bower, G., Bowers, A. S., Bowers, C. W., Boyce, L. A., Boyer, C. T., Boyer, M. L., Boyer, M., Boyer, R., Bradley, L. D., Brady, G. R., Brandl, B. R., Brannen, J. L., Breda, D., Bremmer, H. G., Brennan, D., Bresnahan, P. A., Bright, S. N., Broiles, B. J., Bromenschenkel, A., Brooks, B. H., Brooks, K. J., Brown, B., Brown, B., Brown, T. M., Bruce, B. W., Bryson, J. G.,

Bujanda, E. D., Bullock, B. M., Bunker, A. J., Bureo, R., Burt, I. J., Bush, J. A., Bushouse, H. A., Bussman, M. C., Cabaud, O., Cale, S., Calhoon, C. D., Calvani, H., Canipe, A. M., Caputo, F. M., Cara, M., Carey, L., Case, M. E., Cesari, T., Cetorelli, L. D., Chance, D. R., Chandler, L., Chaney, D., Chapman, G. N., Charlot, S., Chayer, P., Cheezum, J. I., Chen, B., Chen, C. H., Cherinka, B., Chichester, S. C., Chilton, Z. S., Chittiraibalan, D., Clampin, M., Clark, C. R., Clark, K. W., Clark, S. M., Claybrooks, E. E., Cleveland, K. A., Cohen, A. L., Cohen, L. M., Colón, K. D., Coleman, B. L., Colina, L., Comber, B. J., Comeau, T. M., Comer, T., Conde Reis, A., Connolly, D. C., Conroy, K. E., Contos, A. R., Contreras, J., Cook, N. J., Cooper, J. L., Cooper, R. A., Correia, M. F., Correnti, M., Cossou, C., Costanza, B. F., Coulais, A., Cox, C. R., Coyle, R. T., Cracraft, M. M., Crew, K. A., Curtis, G. J., Cusveller, B., Da Costa Maciel, C., Dailey, C. T., Daugeron, F., Davidson, G. S., Davies, J. E., Davis, K. A., Davis, M. S., Day, R., de Chambure, D., de Jong, P., De Marchi, G., Dean, B. H., Decker, J. E., Delisa, A. S., Dell, L. C., and Dellagatta, G. The James Webb Space Telescope Mission. *Publications of the Astronomical Society of the Pacific*, 135(1048):068001, 2023. doi: 10.1088/1538-3873/acd1b5.

Gröger, F., Wen, S., and Brbić, M. Revisiting the Platonic Representation Hypothesis: An Aristotelian View. *ArXiv e-prints*, 2026. doi: 10.48550/arXiv.2602.14486.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. Masked Autoencoders Are Scalable Vision Learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18–24. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01553.

Heneka, C., Nieser, F., Ore, A., Plehn, T., and Schiller, D. Large Language Models – the Future of Fundamental Physics? *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2506.14757.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. d. L., Hendricks, L. A., Welbl, J., Clark, A., et al. Training Compute-Optimal Large Language Models. *ArXiv e-prints*, 2022. doi: 10.48550/arXiv.2203.15556.

Huh, M., Cheung, B., Wang, T., and Isola, P. Position: The Platonic Representation Hypothesis. In *International Conference on Machine Learning*, pp. 20617–20642. PMLR, 2024. URL https://proceedings.mlr.press/v235/huh24a.html.

Ilbert, O., Arnouts, S., McCracken, H. J., Bolzonella, M., Bertin, E., Le Fèvre, O., Mellier, Y., Zamorani, G., et al. Accurate photometric redshifts for the cfht legacy survey calibrated using the vimos vlt deep survey. *Astronomy & Astrophysics*, 457:841–856, 2006. doi: 10.1051/0004-6361:20065138.

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. Scaling Laws for Neural Language Models. *arXiv*, 2020. doi: 10.48550/arXiv.2001.08361.

Khederlarian, A., Andrews, B. H., Newman, J. A., Zhang, T., and Dey, B. Optimizing Deep Learning Photometric Redshifts for the Roman Space Telescope with HST/CANDELS. *arXiv e-prints*, art. arXiv:2602.10207, February 2026. doi: 10.48550/arXiv.2602.10207.

Koblischke, N. and Bovy, J. SpectraFM: Tuning into Stellar Foundation Models. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2411.04750.

Kornblith, S., Norouzi, M., Lee, H., and Hinton, G. Similarity of neural network representations revisited. In *International conference on machine learning*, pp. 3519–3529. PMlR, 2019.

Lacoste, A., Luccioni, A., Schmidt, V., and Dandres, T. Quantifying the Carbon Emissions of Machine Learning. *arXiv*, 2019. doi: 10.48550/arXiv.1910.09700.

Leung, H. W. and Bovy, J. Towards an astronomical foundation model for stars with a transformer-based model. *Monthly Notices of the Royal Astronomical Society*, 527(1):1494–1520, 2023. ISSN 0035-8711. doi: 10.1093/mnras/stad3015.

Liu, H., Li, C., Li, Y., and Lee, Y. J. Improved Baselines with Visual Instruction Tuning. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2310.03744.

Lupton, R., Blanton, M. R., Fekete, G., Hogg, D. W., O'Mullane, W., Szalay, A., and Wherry, N. Preparing Red-Green-Blue Images from CCD Data. *Publications of the Astronomical Society of the Pacific*, 116(816):133, 2004. ISSN 1538-3873. doi: 10.1086/382245.

Madau, P. and Dickinson, M. Cosmic Star-Formation History. *ARA&A*, 52:415–486, August 2014. doi: 10.1146/annurev-astro-081811-125615.

Malmquist, K. G. On some relations in stellar statistics. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52, 1922. URL https://ui.adsabs.harvard.edu/abs/1922MeLuF.100....1M/exportcitation.

Mishra-Sharma, S., Song, Y., and Thaler, J. PAPERCLIP: Associating Astronomical Observations and Natural Language with Multi-Modal Models. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2403.08851.

Miyazaki, S., Komiyama, Y., Kawanomoto, S., Doi, Y., Furusawa, H., Hamana, T., Hayashi, Y., Ikeda, H., Kamata, Y., Karoji, H., Koike, M., Kurakami, T., Miyama, S., Morokuma, T., Nakata, F., Namikawa, K., Nakaya, H., Nariai, K., Obuchi, Y., Oishi, Y., Okada, N., Okura, Y., Tait, P., Takata, T., Tanaka, Y., Tanaka, M., Terai, T., Tomono, D., Uraguchi, F., Usuda, T., Utsumi, Y., Yamada, Y., Yamanoi, H., Aihara, H., Fujimori, H., Mineo, S., Miyatake, H., Oguri, M., Uchida, T., Tanaka, M. M., Yasuda, N., Takada, M., Murayama, H., Nishizawa, A. J., Sugiyama, N., Chiba, M., Futamase, T., Wang, S.-Y., Chen, H.-Y., Ho, P. T. P., Liaw, E. J. Y., Chiu, C.-F., Ho, C.-L., Lai, T.-C., Lee, Y.-C., Jeng, D.-Z., Iwamura, S., Armstrong, R., Bickerton, S., Bosch, J., Gunn, J. E., Lupton, R. H., Loomis, C., Price, P., Smith, S., Strauss, M. A., Turner, E. L., Suzuki, H., Miyazaki, Y., Muramatsu, M., Yamamoto, K., Endo, M., Ezaki, Y., Ito, N., Kawaguchi, N., Sofuku, S., Taniike, T., Akutsu, K., Dojo, N., Kasumi, K., Matsuda, T., Imoto, K., Miwa, Y., Suzuki, M., Takeshi, K., and Yokota, H. Hyper Suprime-Cam: System design and verification of image quality. *Publications of the Astronomical Society of Japan*, 70(SP1):S1, 2018. ISSN 0004-6264. doi: 10.1093/pasj/psx063.

Moriwaki, K., Jun, R. L., Osato, K., and Yoshida, N. CosmoGLINT: Cosmological Generative Model for Line Intensity Mapping with Transformer. *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2506.16843.

Nguyen, T. D., Ting, Y.-S., Ciucă, I., O'Neill, C., Sun, Z.-C., Jabłońska, M., Kruk, S., Perkowski, E., Miller, J., Li, J., et al. AstroLLaMA: Towards Specialized Foundation Models in Astronomy. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2309.06126.

Odewahn, S. C., Stockwell, E. B., Pennington, R. L., Humphreys, R. M., and Zumach, W. A. Automated Star/Galaxy Discrimination With Neural Networks. *The Astronomical Journal*, 103:318, 1992. doi: 10.1086/116063.

Ore, A., Heneka, C., and Plehn, T. SKATR: A Self-Supervised Summary Transformer for SKA. *ArXiv e-prints*, 2024. doi: 10.21468/SciPostPhys.18.5.155.

Pan, J.-S., Ting, Y.-S., Huang, Y., Yu, J., and Liu, J.-F. The Scaling Law in Stellar Light Curves. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2405.17156.

Parker, L., Lanusse, F., Golkar, S., Sarra, L., Cranmer, M., Bietti, A., Eickenberg, M., Krawezik, G., McCabe, M., Morel, R., Ohana, R., Pettee, M., Régaldo-Saint Blancard, B., Cho, K., Ho, S., and The Polymathic AI Collaboration. AstroCLIP: a cross-modal foundation model for galaxies. *Monthly Notices of the Royal Astronomical Society*, 531 (4):4990–5011, 2024. ISSN 0035-8711. doi: 10.1093/mnras/stae1450.

Perkowski, E., Pan, R., Nguyen, T. D., Ting, Y.-S., Kruk, S., Zhang, T., O'Neill, C., Jablonska, M., Sun, Z., Smith, M. J., et al. AstroLLaMA-Chat: Scaling AstroLLaMA with Conversational and Diverse Datasets. *Research Notes of the AAS*, 8(1):7, 2024. ISSN 2515-5172. doi: 10.3847/2515-5172/ad1abe.

Plato. *The Republic*. c. 375 BCE.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. Learning Transferable Visual Models From Natural Language Supervision. *ArXiv e-prints*, 2021. doi: 10.48550/arXiv.2103.00020.

Salvato, M., Ilbert, O., and Hoyle, B. The many flavours of photometric redshifts. *Nature Astronomy*, 3:212–222, 2019. ISSN 2397-3366. doi: 10.1038/s41550-018-0478-0.

Sarmiento, R., Huertas-Company, M., Knapen, J. H., Sánchez, S. F., Sánchez, H. D., Drory, N., and Falcón-Barroso, J. Capturing the Physics of MaNGA Galaxies with Self-supervised Machine Learning. *The Astrophysical Journal*, 921(2):177, 2021. ISSN 0004-637X. doi: 10.3847/1538-4357/ac1dac.

Shuntov, M., Akins, H. B., Paquereau, L., Casey, C. M., Ilbert, O., Arango-Toro, R. C., McCracken, H. J., Franco, M., Harish, S., Kartaltepe, J. S., Koekemoer, A. M., Yang, L., Huertas-Company, M., Berman, E. M., McCleary, J. E., Toft, S., Gavazzi, R., Achenbach, M. J., Bertin, E., Brinch, M., Champagne, J., Chartab, N., Drakos, N. E., Egami, E., Endsley, R., Faisst, A. L., Fan, X., Flayhart, C., Hartley, W. G., Hatamnia, H., Gozaliasl, G., Gentile, F., Jermann, I., Jin, S., Kakiichi, K., Khostovan, A. A., Kümmel, M., Laigle, C., Laishram, R., Lambrides, E., Liu, D., Lyu, J., Magdis, G., Mobasher, B., Moutard, T., Renzini, A., Rich, R. M., Sanders, D. B., Sattari, Z., Robertson, B. E., Schefer, M., Scognamiglio, D., Scoville, N., Silverman, J. D., Taamoli, S., Trakhtenbrot, B., Valentino, F., Wang, F., Weaver, J. R., and Yang, J. COSMOS2025: The COSMOS-Web galaxy catalog of photometry, morphology, redshifts, and physical parameters from JWST, HST, and ground-based imaging. *A&A*, 704:A339, December 2025. doi: 10.1051/0004-6361/202555799.

Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., and Bojanowski, P. DINOv3. *arXiv e-prints*, art. arXiv:2508.10104, August 2025. doi: 10.48550/arXiv.2508.10104.

Slijepcevic, I. V., Scaife, A. M. M., Walmsley, M., Bowles, M., Wong, O. I., Shabala, S. S., and White, S. V. Radio galaxy zoo: towards building the first multipurpose foundation model for radio astronomy with self-supervised learning. *RAS Techniques and Instruments*, 3(1):19–32, 2024. ISSN 2752-8200. doi: 10.1093/rasti/rzad055.

Smith, M. J. and Geach, J. E. Astronomia ex machina: a history, primer and outlook on neural networks in astronomy. *R. Soc. Open Sci.*, 10(5):221454, 2023. ISSN 2054-5703. doi: 10.1098/rsos.221454.

Smith, M. J., Geach, J. E., Jackson, R. A., Arora, N., Stone, C., and Courteau, S. Realistic galaxy image simulation via score-based generative models. *Monthly Notices of the Royal Astronomical Society*, 511(2):1808–1818, 2022. doi: 10.1093/mnras/stac130.

Smith, M. J., Roberts, R. J., Angeloudi, E., and Huertas-Company, M. AstroPT: Scaling Large Observation Models for Astronomy. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2405.14930.

Speagle, J. S., Steinhardt, C. L., Capak, P. L., and Silverman, J. D. A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6. *ApJS*, 214(2):15, October 2014. doi: 10.1088/0067-0049/214/2/15.

Sutton, R. The Bitter Lesson. https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf, 2019. URL http://incompleteideas.net/IncIdeas/BitterLesson.html.

The Multimodal Universe Collaboration. The Multimodal Universe: Enabling Large-Scale Machine Learning with 100 TB of Astronomical Scientific Data. *Advances in Neural Information Processing Systems*, 37:57841–57913, 2024. URL https://arxiv.org/abs/2412.02527.

Wetzel, A. R., Tinker, J. L., and Conroy, C. Galaxy evolution in groups and clusters: star formation rates, red sequence fractions and the persistent bimodality. *Monthly Notices of the Royal Astronomical Society*, 424(1):232–243, 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.21188.x.

Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., and Xie, S. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2301.00808.

Wu, J. F. and Peek, J. E. G. Predicting galaxy spectra from images with hybrid convolutional neural networks. *ArXiv e-prints*, 2020. doi: 10.48550/arXiv.2009.12318.

Zaman, S., Smith, M. J., Khetarpal, P., Chakrabarty, R., Ginolfi, M., Huertas-Company, M., Jabłońska, M., Kruk, S., Lain, M. L., Méndez, S. J. R., and Tanoglidis, D. AstroLLaVA: towards the unification of astronomical data and natural language. *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2504.08583.

Zhao, X., Huang, Y., Xue, G., Kong, X., Liu, J., Tang, X., Beers, T. C., Ting, Y.-S., and Luo, A.-L. SpecCLIP: Aligning and Translating Spectroscopic Measurements for Stars. *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2507.01939.

Zuo, X., Tao, Y., Huang, Y., Kang, Z., Chen, H., Cui, C., Pan, J., Kong, X., Tang, X., Han, H., Mu, H., Xu, Y., Fan, D., Xue, G., Luo, A., and Liu, J. FALCO: a Foundation model of Astronomical Light Curves for time dOmain astronomy. *ArXiv e-prints*, 2025. doi: 10.48550/arXiv.2504.20290.

A. Code, data, and compute

A.1. Dataset and code hosting

Here we list the datasets used, as well as the models used for this study. For each dataset and model we provide a link to the publicly available data or weights. Cross-matching between surveys is performed using MMU v1 with a 1" matching radius.

We release all of our code on Github at `github.com/UniverseTBD/platonic-universe`.

Table 2. Foundation models and astronomical datasets used in this study.

Category	Name	Size	Hugging Face source
Models	AstroPTv2	15M (Small)	`Smith42/astroPT_v2.0`
		95M (Base)	`Smith42/astroPT_v2.0`
		850M (Large)	`Smith42/astroPT_v2.0`
	CLIP	86M (Base)	`openai/clip-vit-base-patch16`
		304M (Large)	`openai/clip-vit-large-patch14`
	ConvNeXtv2	15M (Nano)	`facebook/convnextv2-nano-22k-224`
		28M (Tiny)	`facebook/convnextv2-tiny-22k-224`
		89M (Base)	`facebook/convnextv2-base-22k-224`
		198M (Large)	`facebook/convnextv2-large-22k-224`
	DINOv3	21M (ViT-S/16)	`facebook/dinov3-vits16-pretrain-lvd1689m`
		29M (ViT-S+/16)	`facebook/dinov3-vits16plus-pretrain-lvd1689m`
		86M (ViT-B/16)	`facebook/dinov3-vitb16-pretrain-lvd1689m`
		300M (ViT-L/16)	`facebook/dinov3-vitl16-pretrain-lvd1689m`
		840M (ViT-H+/16)	`facebook/dinov3-vith16plus-pretrain-lvd1689m`
		6.7B (ViT-7B/16)	`facebook/dinov3-vit7b16-pretrain-lvd1689m`
	IJEPA	632M (Huge)	`facebook/ijepa_vith14_22k`
		1.0B (Giant)	`facebook/ijepa_vitg16_22k`
	LLaVA-1.5	7B	`llava-hf/llava-1.5-7b-hf`
		13B	`llava-hf/llava-1.5-13b-hf`
	PaliGemma 2	3B	`google/paligemma2-3b-mix-224`
		10B	`google/paligemma2-10b-mix-224`
		28B	`google/paligemma2-28b-mix-224`
	Specformer	43M (Base)	`polymathic-ai/specformer`
	ViT	86M (Base)	`google/vit-base-patch16-224-in21k`
		304M (Large)	`google/vit-large-patch16-224-in21k`
		632M (Huge)	`google/vit-huge-patch14-224-in21k`
	ViT-MAE	86M (Base)	`facebook/vit-mae-base`
		304M (Large)	`facebook/vit-mae-large`
		632M (Huge)	`facebook/vit-mae-huge`
	V-JEPA2	300M (ViT-L)	`facebook/vjepa2-vitl-fpc64-256`
		600M (ViT-H)	`facebook/vjepa2-vith-fpc64-256`
		1.0B (ViT-G)	`facebook/vjepa2-vitg-fpc64-256`
Crossmatches	JWST vs HSC	1.67k	`Smith42/jwst_hsc_crossmatched`
	Legacy vs HSC	102k	`Smith42/legacysurvey_hsc_crossmatched`
	DESI vs HSC	18.6k	`Smith42/desi_hsc_crossmatched`
Embeddings			`UniverseTBD/pu-embeddings`

All of our code, embeddings, and crossmatched data are publicly available: the code is hosted on GitHub at `https://github.com/UniverseTBD/platonic-universe`, and the embeddings and datasets are hosted on Hugging Face under the `UniverseTBD` and `Smith42` organisations (see Tab. 2 for individual links).

A.2. Compute resources and experimental cost

Extracting embeddings across the JWST, HSC, Legacy Survey, and DESI datasets cost the most GPU-hours, particularly for the larger vision-language models (PaliGemma, LLaVa). The experiments required approximately 3000 GPU-hours across various hardware tiers. Various downstream tasks (MKNN/CKA computation, linear probing, and statistical analyses) were computationally cheaper compared to extraction, requiring about one CPU-hour per probe once embeddings were cached.

Table 3. Hardware usage summary by chip type.

Hardware	Approximate Usage
NVIDIA A40 / A80	300 GPU-hours
NVIDIA A100 / A800	600 GPU-hours
NVIDIA GH200 / H100	2000 GPU-hours
NVIDIA RTX 5070 and RTX 3090	50 GPU-hours
Standard CPU cores	≈ 1 CPU-hour per probe
Total	**3,000 GPU-hours**

Using the Machine Learning Impact calculator (Lacoste et al., 2019),[7] we estimate the total carbon footprint of these experiments at approximately $0.8\,\mathrm{t\,CO_2eq}$. We obtain this by multiplying a representative thermal design power (TDP) for each hardware tier ($350\,\mathrm{W}$ for the A40/A80 tier, $400\,\mathrm{W}$ for A100/A800, $700\,\mathrm{W}$ for GH200/H100, and $300\,\mathrm{W}$ for the consumer RTX cards) by the corresponding GPU-hours, giving $\approx 1{,}760\,\mathrm{kWh}$ of GPU energy, and applying the calculator's default OECD-average grid intensity of $0.475\,\mathrm{kg\,CO_2eq\,kWh^{-1}}$.

[7] https://mlco2.github.io/impact/

B. Extended physics results

Here we show the full per-property results for the physics linear-probe experiments. Tab. 4 lists mean and standard deviation R^2 values across 10-fold cross-validation for each model on HSC and JWST imagery, broken down by redshift, stellar mass, and sSFR. Fig. 7 shows probe R^2 against parameter count separately for each of the three physical properties. Figs. 8 and 9 show the per-model cosine similarity matrices between the redshift, stellar mass, and sSFR probe weight vectors for HSC and JWST imagery respectively.

Table 4. Downstream regression performance (R^2) when predicting redshift (z), stellar mass ($\log M_\star$), and specific star formation rate (sSFR) from frozen embeddings of each model, evaluated on 45 000 HSC and JWST galaxies. Values are mean ± standard deviation over 10 k-folds.

	HSC (R^2)			JWST (R^2)		
Model	z	$\log M_\star$	sSFR	z	$\log M_\star$	sSFR
AstroPTv2 15M	$0.387_{\pm 0.013}$	$0.525_{\pm 0.009}$	$0.462_{\pm 0.002}$	$0.349_{\pm 0.010}$	$0.674_{\pm 0.010}$	$0.293_{\pm 0.008}$
AstroPTv2 95M	$0.443_{\pm 0.017}$	$0.556_{\pm 0.012}$	$0.502_{\pm 0.006}$	$0.472_{\pm 0.015}$	$0.771_{\pm 0.008}$	$0.336_{\pm 0.005}$
AstroPTv2 850M	$0.453_{\pm 0.007}$	$0.559_{\pm 0.013}$	$0.510_{\pm 0.007}$	$0.531_{\pm 0.012}$	$0.802_{\pm 0.006}$	$0.360_{\pm 0.005}$
ViT Base	$0.359_{\pm 0.006}$	$0.455_{\pm 0.013}$	$0.434_{\pm 0.006}$	$0.372_{\pm 0.012}$	$0.597_{\pm 0.004}$	$0.306_{\pm 0.012}$
ViT Large	$0.369_{\pm 0.007}$	$0.474_{\pm 0.007}$	$0.440_{\pm 0.005}$	$0.408_{\pm 0.010}$	$0.638_{\pm 0.006}$	$0.340_{\pm 0.009}$
ViT Huge	$0.434_{\pm 0.011}$	$0.520_{\pm 0.008}$	$0.480_{\pm 0.008}$	$0.477_{\pm 0.008}$	$0.722_{\pm 0.004}$	$0.358_{\pm 0.009}$
ViT-MAE Base	$0.343_{\pm 0.010}$	$0.450_{\pm 0.009}$	$0.422_{\pm 0.008}$	$0.290_{\pm 0.011}$	$0.581_{\pm 0.009}$	$0.277_{\pm 0.015}$
ViT-MAE Large	$0.357_{\pm 0.012}$	$0.461_{\pm 0.009}$	$0.441_{\pm 0.007}$	$0.324_{\pm 0.012}$	$0.631_{\pm 0.010}$	$0.295_{\pm 0.011}$
ViT-MAE Huge	$0.372_{\pm 0.009}$	$0.486_{\pm 0.008}$	$0.470_{\pm 0.005}$	$0.357_{\pm 0.011}$	$0.668_{\pm 0.010}$	$0.310_{\pm 0.010}$
DINOv3 vits16	$0.359_{\pm 0.005}$	$0.434_{\pm 0.008}$	$0.418_{\pm 0.008}$	$0.402_{\pm 0.009}$	$0.607_{\pm 0.006}$	$0.334_{\pm 0.008}$
DINOv3 vits16plus	$0.357_{\pm 0.008}$	$0.440_{\pm 0.008}$	$0.426_{\pm 0.008}$	$0.380_{\pm 0.013}$	$0.601_{\pm 0.008}$	$0.336_{\pm 0.011}$
DINOv3 vitb16	$0.383_{\pm 0.005}$	$0.465_{\pm 0.008}$	$0.451_{\pm 0.011}$	$0.441_{\pm 0.008}$	$0.646_{\pm 0.007}$	$0.368_{\pm 0.008}$
DINOv3 vitl16	$0.376_{\pm 0.004}$	$0.454_{\pm 0.006}$	$0.453_{\pm 0.008}$	$0.443_{\pm 0.008}$	$0.626_{\pm 0.006}$	$0.367_{\pm 0.012}$
DINOv3 vith16plus	$0.366_{\pm 0.010}$	$0.450_{\pm 0.006}$	$0.439_{\pm 0.008}$	$0.464_{\pm 0.012}$	$0.630_{\pm 0.007}$	$0.379_{\pm 0.011}$
DINOv3 vit7b16	$0.421_{\pm 0.007}$	$0.495_{\pm 0.008}$	$0.489_{\pm 0.013}$	$0.528_{\pm 0.004}$	$0.704_{\pm 0.003}$	$0.421_{\pm 0.011}$
ConvNeXtv2 Nano	$0.392_{\pm 0.005}$	$0.490_{\pm 0.009}$	$0.475_{\pm 0.007}$	$0.457_{\pm 0.010}$	$0.690_{\pm 0.007}$	$0.359_{\pm 0.011}$
ConvNeXtv2 Tiny	$0.412_{\pm 0.009}$	$0.496_{\pm 0.009}$	$0.470_{\pm 0.006}$	$0.455_{\pm 0.016}$	$0.675_{\pm 0.007}$	$0.375_{\pm 0.008}$
ConvNeXtv2 Base	$0.409_{\pm 0.004}$	$0.501_{\pm 0.008}$	$0.488_{\pm 0.011}$	$0.460_{\pm 0.009}$	$0.676_{\pm 0.003}$	$0.386_{\pm 0.006}$
ConvNeXtv2 Large	$0.425_{\pm 0.008}$	$0.522_{\pm 0.010}$	$0.506_{\pm 0.008}$	$0.498_{\pm 0.012}$	$0.713_{\pm 0.004}$	$0.400_{\pm 0.010}$
I-JEPA Huge	$0.429_{\pm 0.008}$	$0.528_{\pm 0.007}$	$0.499_{\pm 0.006}$	$0.524_{\pm 0.011}$	$0.780_{\pm 0.006}$	$0.403_{\pm 0.009}$
I-JEPA Giant	$0.424_{\pm 0.004}$	$0.528_{\pm 0.007}$	$0.506_{\pm 0.012}$	$0.555_{\pm 0.013}$	$0.790_{\pm 0.006}$	$0.416_{\pm 0.008}$
V-JEPA2 Large	$0.421_{\pm 0.008}$	$0.544_{\pm 0.010}$	$0.517_{\pm 0.006}$	$0.525_{\pm 0.010}$	$0.780_{\pm 0.006}$	$0.436_{\pm 0.007}$
V-JEPA2 Huge	$0.420_{\pm 0.006}$	$0.549_{\pm 0.012}$	$0.518_{\pm 0.006}$	$0.521_{\pm 0.011}$	$0.781_{\pm 0.005}$	$0.439_{\pm 0.012}$
V-JEPA2 Giant	$0.450_{\pm 0.006}$	$0.560_{\pm 0.011}$	$0.530_{\pm 0.007}$	$0.565_{\pm 0.011}$	$0.810_{\pm 0.003}$	$0.457_{\pm 0.009}$
CLIP Base	$0.433_{\pm 0.009}$	$0.517_{\pm 0.008}$	$0.511_{\pm 0.007}$	$0.473_{\pm 0.009}$	$0.706_{\pm 0.003}$	$0.358_{\pm 0.011}$
CLIP Large	$0.435_{\pm 0.008}$	$0.515_{\pm 0.010}$	$0.496_{\pm 0.007}$	$0.533_{\pm 0.009}$	$0.736_{\pm 0.004}$	$0.385_{\pm 0.010}$
PaliGemma 2 3B	$0.458_{\pm 0.007}$	$0.539_{\pm 0.008}$	$0.529_{\pm 0.011}$	$0.528_{\pm 0.011}$	$0.764_{\pm 0.005}$	$0.378_{\pm 0.007}$
PaliGemma 2 10B	$0.464_{\pm 0.009}$	$0.537_{\pm 0.011}$	$0.522_{\pm 0.011}$	$0.531_{\pm 0.015}$	$0.773_{\pm 0.004}$	$0.359_{\pm 0.008}$
PaliGemma 2 28B	$0.398_{\pm 0.009}$	$0.492_{\pm 0.009}$	$0.478_{\pm 0.010}$	$0.465_{\pm 0.014}$	$0.722_{\pm 0.006}$	$0.310_{\pm 0.014}$
LLaVA 1.5 7B	$0.431_{\pm 0.009}$	$0.522_{\pm 0.013}$	$0.504_{\pm 0.006}$	$0.528_{\pm 0.014}$	$0.752_{\pm 0.005}$	$0.390_{\pm 0.010}$
LLaVA 1.5 13B	$0.429_{\pm 0.008}$	$0.518_{\pm 0.009}$	$0.502_{\pm 0.004}$	$0.531_{\pm 0.011}$	$0.756_{\pm 0.003}$	$0.394_{\pm 0.012}$

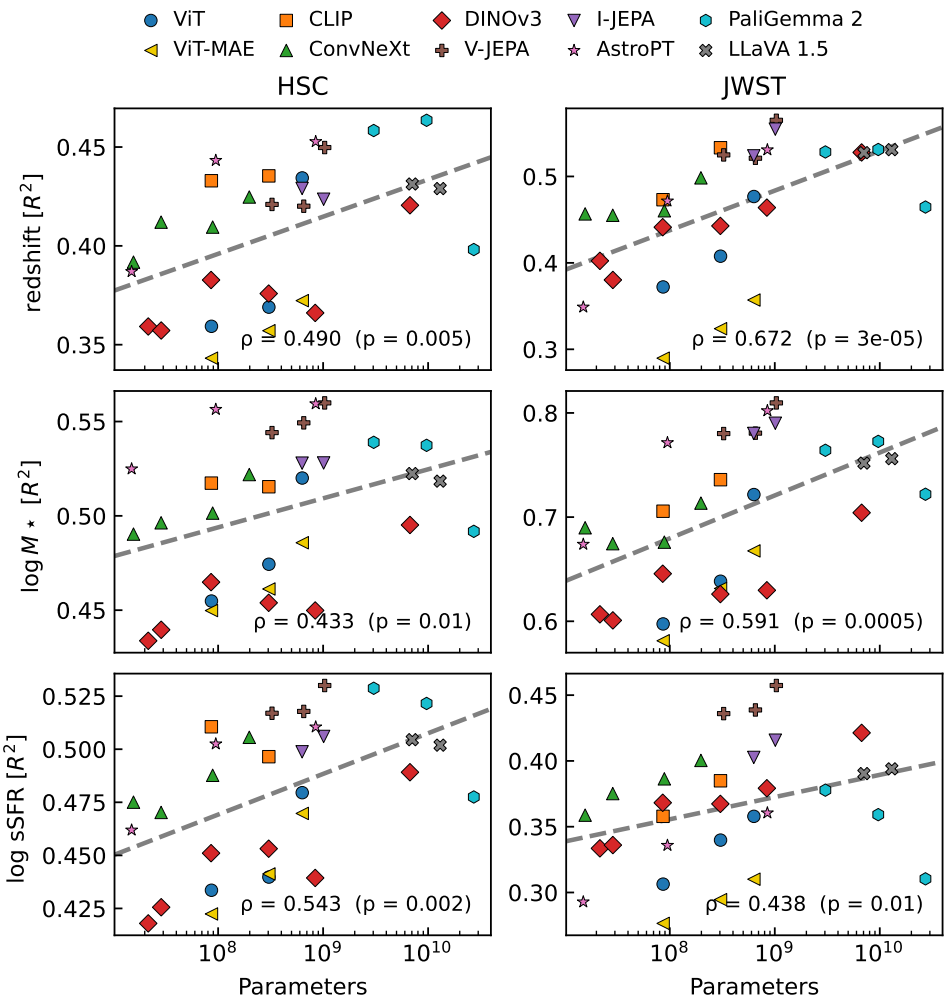


Figure 7. Parameter count vs probe performance for our tested physical properties. Spearman's ρ and p values are stated for each panel. In all tested cases we find a significant correlation between parameter count and physics probe performance.

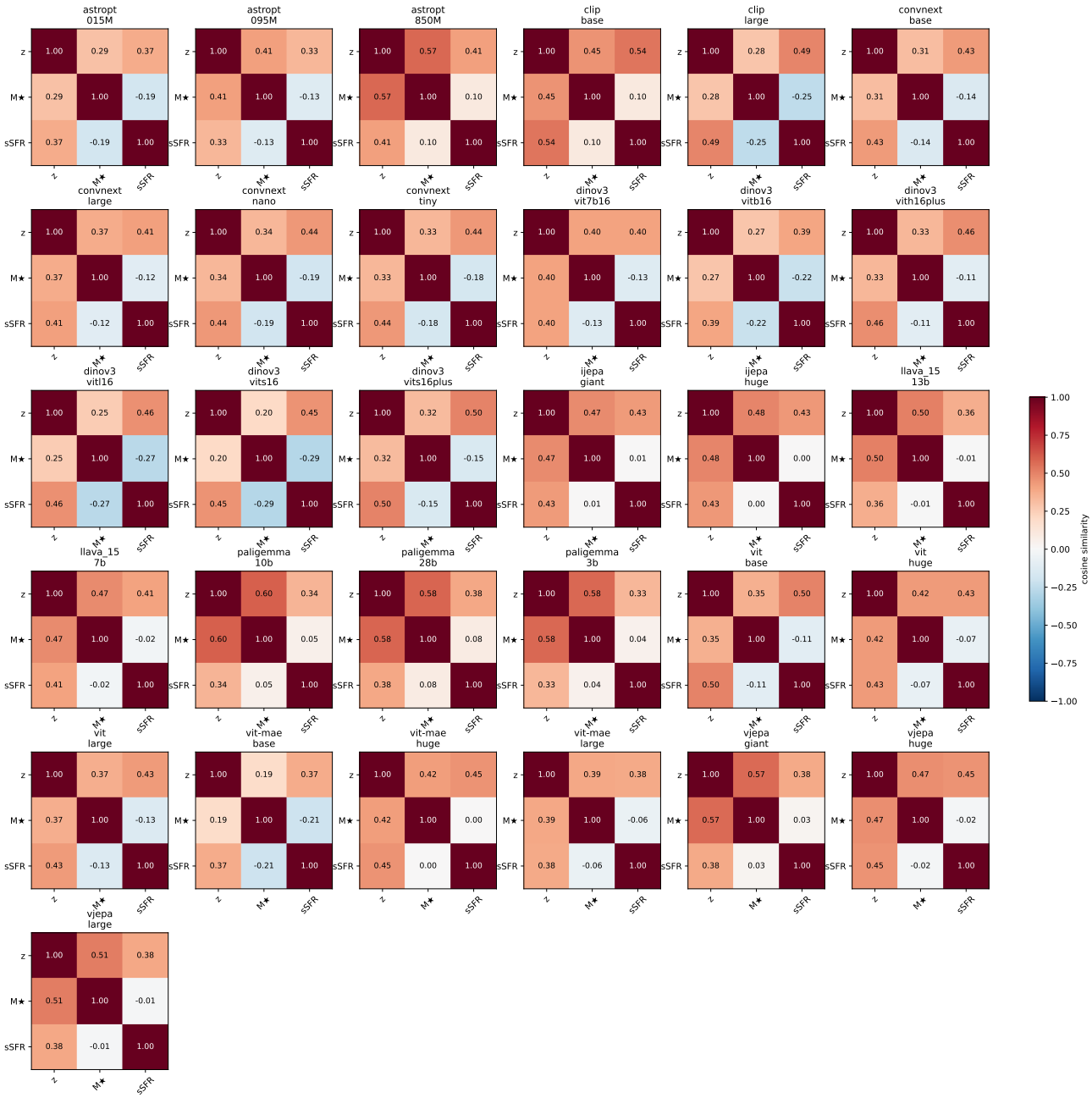


Figure 8. Cosine similarity matrices between the redshift (z), stellar mas ($M_\star$), and sSFR probe weight vectors using HSC images.

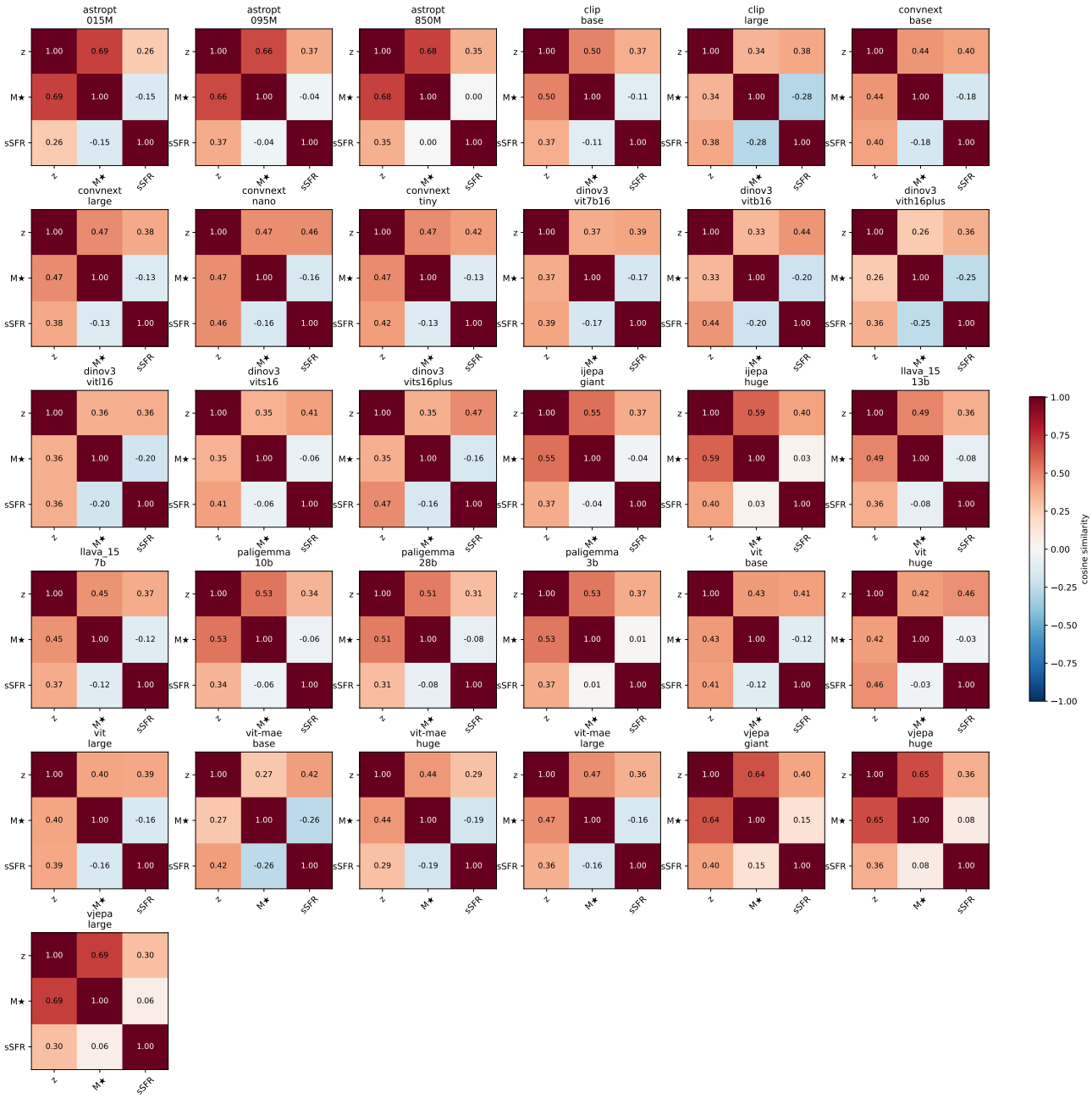


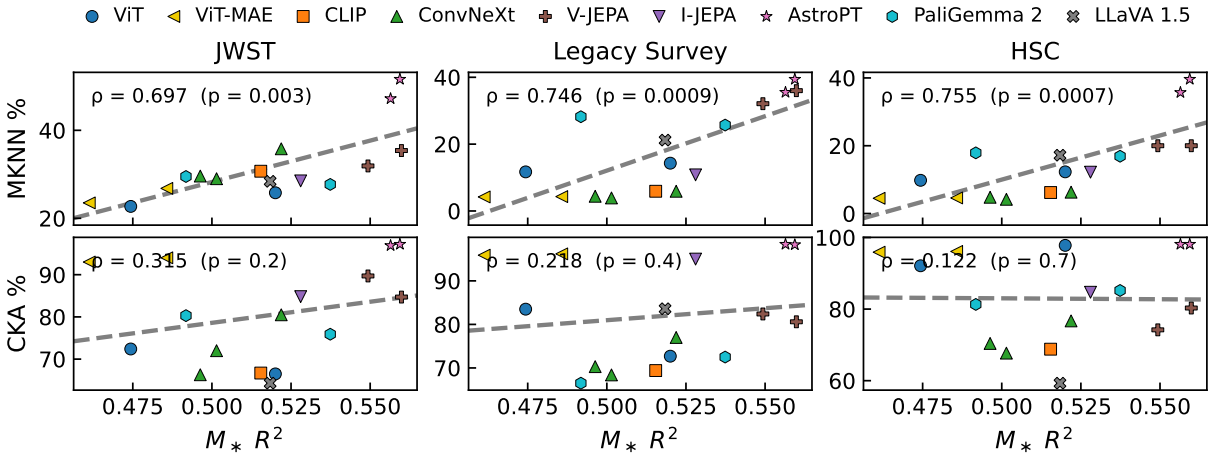
Figure 9. Cosine similarity matrices between the redshift (z), stellar mas ($M_\star$), and sSFR probe weight vectors using JWST images.

C. Extended intra-architectural results

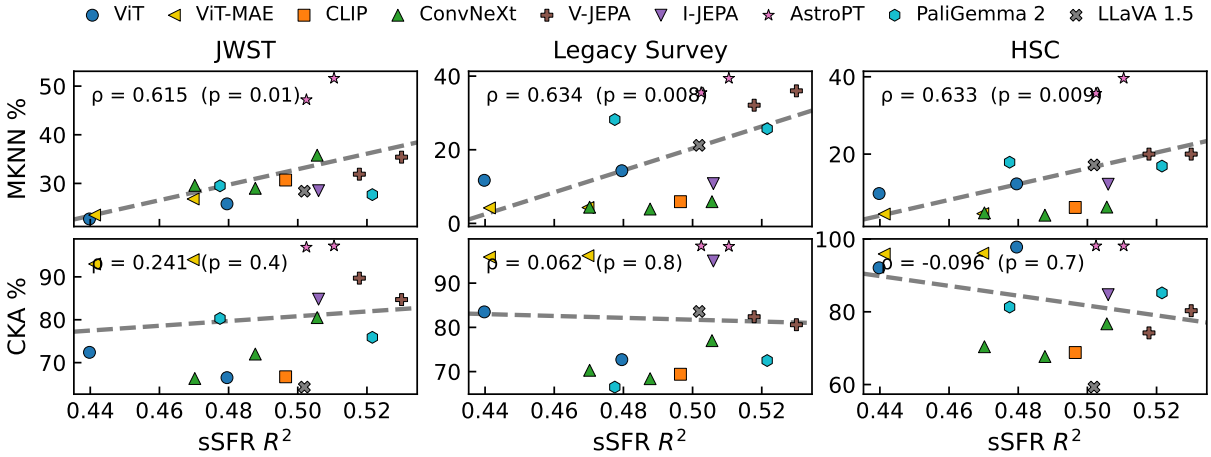
We list the full results across pairs of models within the same family in Tab. 5. Each entry is the MKNN or CKA score as a percentage. Fig. 10 plots intra-architectural MKNN and CKA against linear-probe R^2 for each of redshift, stellar mass, and sSFR individually.

Table 5. Intra-architectural embedding alignment within a model family, measured by MKNN and CKA scores. The PRH predicts that both MKNN and CKA scores will increase as we compare the embeddings of larger model pairs within a model family, since larger models will generate embeddings closer to the Platonic ideal representation.

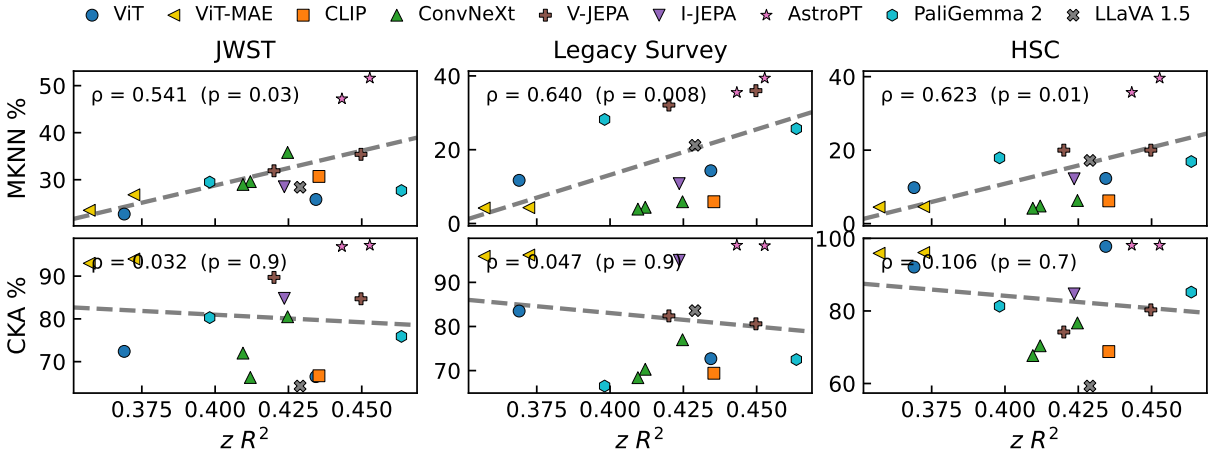
	MKNN (%)			CKA (%)		
Model Pairs	JWST	Legacy	HSC	JWST	Legacy	HSC
AstroPTv2 Small vs Base	47.2	35.5	35.7	96.9	98.4	98.1
AstroPTv2 Base vs Large	51.6	39.4	39.6	97.2	98.3	98.1
CLIP Base vs Large	30.7	5.9	6.2	66.7	69.4	68.8
ConvNeXtv2 Nano vs Tiny	29.6	4.4	4.8	66.3	70.3	70.4
ConvNeXtv2 Tiny vs Base	29.0	3.9	4.2	72.0	68.4	67.7
ConvNeXtv2 Base vs Large	35.8	5.9	6.3	80.5	77.0	76.7
DINOv3 vits16 vs vits16plus	43.6	10.8	16.5	88.5	89.7	86.7
DINOv3 vits16plus vs vitb16	38.2	9.6	14.3	86.5	88.5	86.3
DINOv3 vitb16 vs vitl16	32.7	7.1	9.4	73.6	73.1	68.0
DINOv3 vitl16 vs vith16plus	34.0	6.3	8.8	78.5	68.7	66.6
DINOv3 vith16plus vs vit7b17	40.5	7.8	11.7	81.8	62.1	64.2
IJEPA Huge vs Giant	28.5	10.8	12.2	84.8	95.0	84.7
LLaVA-1.5 7B vs 13B	28.4	21.2	17.2	64.3	83.6	59.3
ViT-MAE Base vs Large	23.5	4.2	4.5	93.0	95.9	95.9
ViT-MAE Large vs Huge	26.8	4.3	4.6	94.0	96.2	96.1
ViT Base vs Large	22.7	11.7	9.8	72.4	83.5	92.1
ViT Large vs Huge	25.8	14.3	12.3	66.5	72.7	97.8
PaliGemma 2 3B vs 10B	27.7	25.7	16.9	75.9	72.5	85.2
PaliGemma 2 10B vs 28B	29.5	28.2	17.9	80.3	66.5	81.3
V-JEPA2 ViT-L vs H	31.9	32.1	20.0	89.7	82.4	74.2
V-JEPA2 ViT-H vs G	35.4	36.0	20.0	84.7	80.6	80.3



(a) Mass ($M_\star$).



(b) Specific Star Formation Rate (sSFR).



(c) Redshift (z).

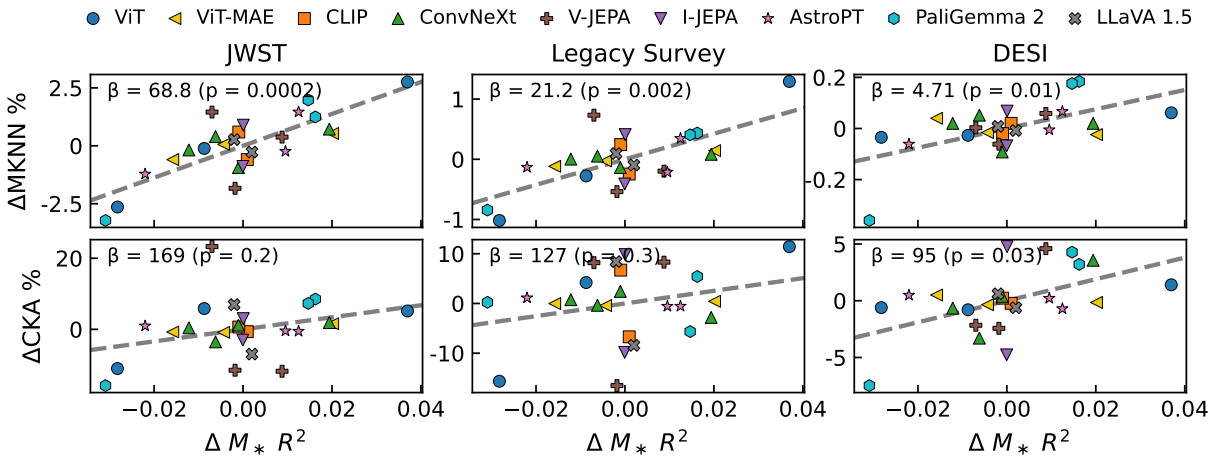
Figure 10. Intra-architectural embedding alignment (MKNN and CKA) vs probe performance for our tested physical properties. Spearman's ρ and p values are stated for each panel. In all tested cases we find a significant correlation between MKNN and physics probe performance. We do not find a significant correlation between CKA and physics probe performance in any tested case.

D. Extended crossmodal results

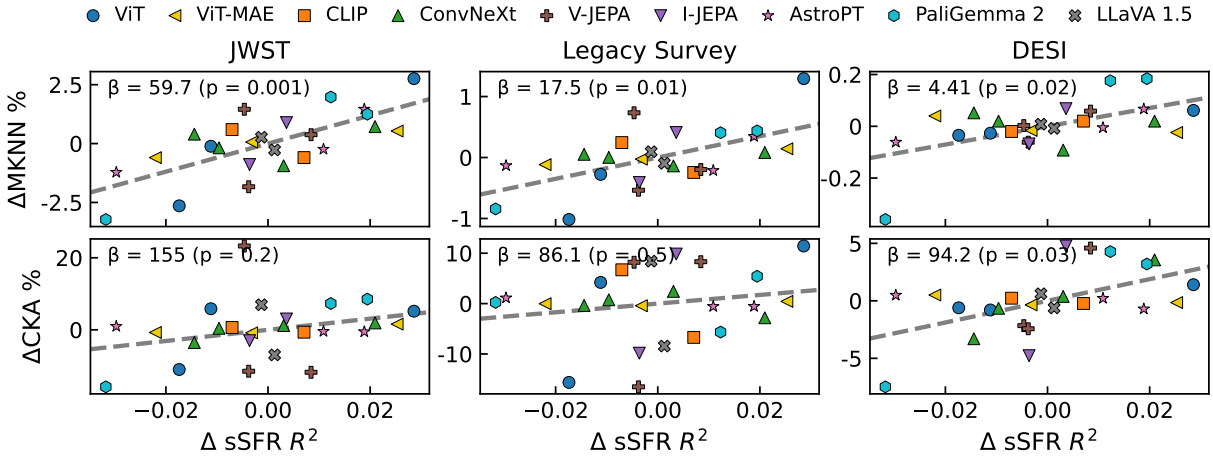
We list the full results across pairs of data modalities for each model variant in Tab. 6. Each entry is the MKNN or CKA score as a percentage. Fig. 11 plots cross modal MKNN and CKA against linear-probe R^2 for each of redshift, stellar mass, and sSFR individually.

Table 6. Crossmodal embedding alignment between a model's embeddings of different astronomical modalities and its embeddings of HSC imaging, measured by MKNN and CKA scores. The DESI spectra column uses Specformer embeddings for the DESI side compared against each listed vision model's HSC image embeddings. The PRH predicts that both MKNN and CKA scores will increase as models within a family grow in size, since larger models should produce embeddings closer to the Platonic ideal representation.

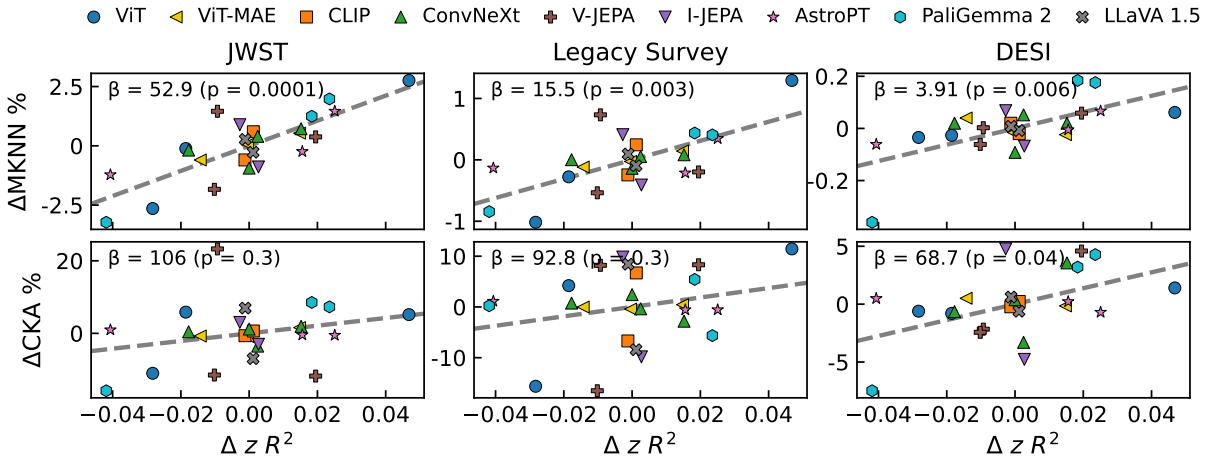
	MKNN (%)			CKA (%)		
Model	JWST	Legacy	DESI	JWST	Legacy	DESI
AstroPTv2 Small	11.62	3.36	1.25	43.4	86.6	45.66
AstroPTv2 Base	12.60	3.28	1.32	42.0	84.9	45.39
AstroPTv2 Large	14.30	3.84	1.41	41.9	84.9	44.47
CLIP Base	12.88	1.79	1.29	30.59	59.32	33.39
CLIP Large	14.07	2.28	1.24	31.89	72.73	33.85
ConvNeXtv2 Nano	11.33	1.13	1.01	32.52	54.06	32.29
ConvNeXtv2 Tiny	11.91	1.18	1.05	28.57	52.93	29.64
ConvNeXtv2 Base	10.57	0.99	0.87	33.28	55.73	33.32
ConvNeXtv2 Large	12.23	1.21	1.01	34.01	50.49	36.50
DINOv3 vits16	14.55	2.42	0.98	52.18	70.23	34.80
DINOv3 vits16plus	13.09	2.09	0.97	48.53	68.06	35.39
DINOv3 vitb16	14.03	2.48	0.92	49.44	70.29	33.25
DINOv3 vitl16	11.80	1.70	0.79	45.31	66.07	30.14
DINOv3 vith16plus	10.35	1.16	0.67	31.58	43.12	22.21
DINOv3 vit7b16	12.62	1.89	0.82	40.14	41.65	30.13
IJEPA Huge	9.85	1.27	0.56	15.22	31.64	15.86
IJEPA Giant	11.63	2.09	0.73	21.25	51.31	25.40
LLaVA-1.5 7B	11.16	1.67	1.04	31.10	56.45	36.88
LLaVA-1.5 13B	11.70	1.86	1.06	45.03	73.30	38.11
PaliGemma 2 3B	11.50	1.84	1.18	43.20	58.58	32.89
PaliGemma 2 10B	12.23	1.81	1.17	41.98	47.51	33.97
PaliGemma 2 28B	7.03	0.56	0.50	31.2	53.39	22.21
ViT-MAE Base	10.09	1.89	1.01	59.34	94.89	29.25
ViT-MAE Large	10.75	1.98	0.94	59.23	94.49	28.39
ViT-MAE Huge	11.22	2.15	0.93	61.66	95.35	28.59
ViT Base	10.48	1.20	1.02	30.13	47.35	35.17
ViT Large	13.01	1.94	1.03	47.00	67.19	34.98
ViT Huge	15.88	3.51	1.14	46.32	74.42	37.18
V-JEPA2 ViT-L	14.27	3.04	0.82	59.63	77.48	23.98
V-JEPA2 ViT-H	10.98	1.77	0.74	24.86	52.76	23.71
V-JEPA2 ViT-G	13.19	2.11	0.89	24.57	77.59	30.73



(a) Mass ($M_\star$).



(b) Specific Star Formation Rate (sSFR).

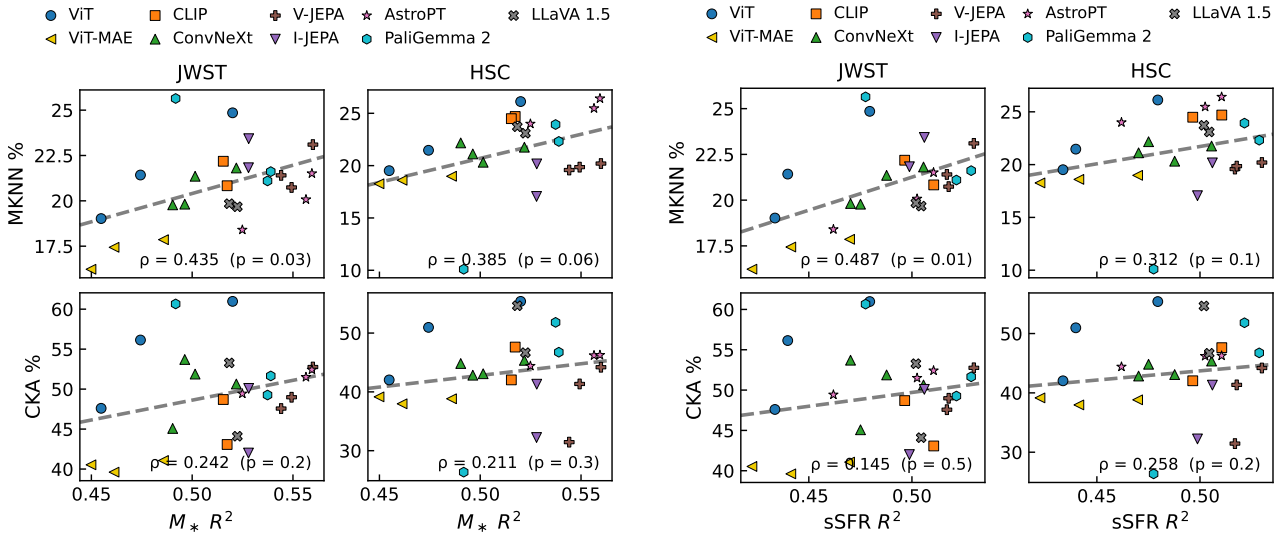


(c) Redshift (z).

Figure 11. Crossmodal embedding alignment between the stated modality and HSC plotted against probe performance for our tested physical properties. ANCOVA β and p values are stated for each panel. In all tested cases we find a significant correlation between MKNN and physics probe performance. We only find a significant correlation between CKA and physics probe performance for DESI Spectra vs HSC.

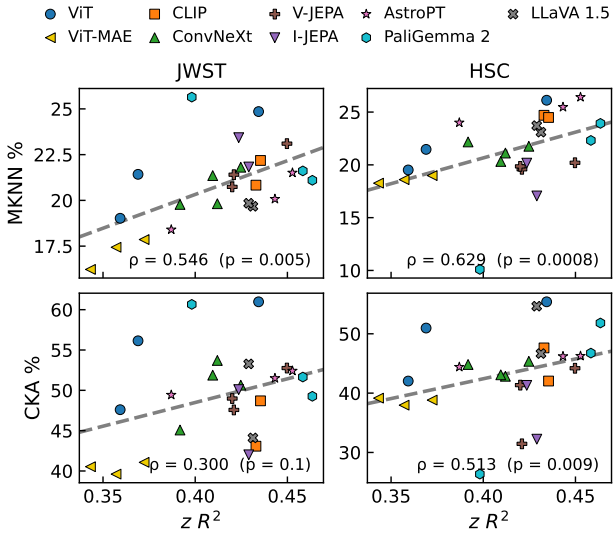
E. Extended cross-architectural results

This appendix expands the cross-architectural analysis. Fig. 12 plots each model's mean MKNN and CKA against the rest of the basket against its linear-probe R^2 for redshift, stellar mass, and sSFR individually, on both JWST and HSC imagery.



(a) Mass ($M_\star$).

(b) Specific Star Formation Rate (sSFR).



(c) Redshift (z).

Figure 12. Cross-architectural embedding alignment between the stated modality and HSC plotted against probe performance for our tested physical properties. Spearman's ρ and p values are stated for each panel. In all tested JWST cases we find a significant correlation between MKNN and physics probe performance. For HSC we find significant or borderline correlations between MKNN and physics probe performance. We only find a significant correlation between CKA and physics probe performance for HSC redshift.