AstroPT: Scaling Large Observation Models for Astronomy

Michael J. Smith [1 2 3] **Ryan J. Roberts** [1 4] **Eirini Angeloudi** [2 5] **Marc Huertas-Company** [2 5 6 7]

Abstract

This work presents AstroPT, an autoregressive pretrained transformer developed with astronomical use-cases in mind. The AstroPT models presented here have been pretrained on 8.6 million 512×512 pixel *grz*-band galaxy postage stamp observations from the DESI Legacy Survey DR8. We train a selection of foundation models of increasing size from 1 million to 2.1 billion parameters, and find that AstroPT follows a similar saturating log-log scaling law to textual models. We also find that the models' performances on downstream tasks as measured by linear probing improves with model size up to the model parameter saturation point. We believe that collaborative community development paves the best route towards realising an open source 'Large Observation Model'—a model trained on data taken from the observational sciences at the scale seen in natural language processing. To this end, we release the source code, weights, and dataset for AstroPT under the MIT license, and invite potential collaborators to join us in collectively building and researching these models.

1. On 'Large Observation Models'

We find ourselves in a new era of connectionism. This era is dominated by large language models trained on web-scale data, with a rapid parameter growth fertilised by the discovery of predictable 'neural scaling laws'—power laws that can be used to estimate a model's performance given its parameter count and training set size (Cortes et al., 1993; Kaplan et al., 2020). In the language domain, the law that currently best describes model scaling was introduced in Hoffmann et al. (2022). This 'Chinchilla' scaling law established that for every additional neural parameter added to the network, around twenty additional textual data tokens must also be added to train a model in a training compute optimal regime. It is also often wise to 'overtrain' a model—or train at a token to parameter ratio larger than 20:1—which leads to a more performant model at a lower parameter count and therefore reduces compute cost at inference time. Overtraining is both environmentally and economically beneficial if a model's total lifetime compute costs are inference heavy (Touvron et al., 2023a;b; Zhang et al., 2024). The discovery of the Chinchilla scaling law, and the need to minimise compute cost at inference time means that we have nearly exhausted high-quality publicly available reserves of textual data for training large neural models (Sevilla et al., 2022; Xue et al., 2023). Some potential solutions to this 'token crisis' have already been investigated in the literature: for example, one can repeat training dataset epochs multiple times without significant performance degradation (Xue et al., 2023), or one can also turn to the use of synthetic data to generate tokens at scale (Silver et al., 2018; Gunasekar et al., 2023). We believe that multimodality can provide a further solution.

Large autoregressive models can process and digest tokens of different modalities (Reed et al., 2022). We can therefore build a model that is capable of learning from modalities that have an abundance of tokens, an abundance that is particularly notable in the observational sciences (Smith & Geach, 2023; Smith et al., 2023a). To give two representative examples, in astronomy the Vera Rubin Observatory (VRO) will observe over twelve billion 16×16 pixel patch 'tokens' per night when conducting LSST (Ivezić et al., 2019), and in earth observation ESA's Sentinel-2 mission produces over six trillion land-observation tokens on a five day global revisit cadence. There are of course many more missions in the observational sciences besides VRO's LSST and ESA's Copernicus, and a compilation of all useful observational data would certainly dwarf any natural textual dataset in volume. Such a dataset could be combined with aligned textual or other observational data and used to train a foundation model—and so, as large neural models can successfully connect concepts across modalities (e.g. Aghajanyan et al., 2023), unlocking these data would go some way towards solving the token crisis.

We will leave true cross-modal training to future work and concentrate here on using astronomical data as the first step towards the goal of training a general 'Large Observation

[1]Aspia Space [2]Instituto de Astrofísica de Canarias (IAC) [3]UniverseTBD [4]Astrophysics Research Institute, Liverpool John Moores University [5]Departamento Astrofísica, Universidad de la Laguna [6]Observatoire de Paris, LERMA, PSL University [7]Université Paris-Cité. Correspondence to: Michael J. Smith <mike@mjjsmith.com>.

Model' (LOM). Before we dive into describing our model, we will briefly summarise the field as it stands to provide further motivation for our specific approach. For the purposes of this literature review let us consider a 'foundation model' as a model that comprises of two training steps. The first step being a computationally expensive unsupervised 'pretraining' routine, and the second step being a relatively computationally cheap supervised finetuning or conditional prompting routine. If we view the field through this lens, two main avenues[1] come into focus—that is, contrastive learning and generative modelling. Let us discuss these techniques in the subsections below[2].

1.1. Contrastive learning

The first contrastive self-supervised neural network used in astronomy, and perhaps the first work that could be considered an 'astronomical foundation model' in the modern sense is 'SkyNet' (Graff et al., 2014). SkyNet's pretraining routine is driven by contrastive divergence learning, where a model is trained by minimising the difference between the positive pairs of training set examples and reconstructions (Hinton, 2002; Hinton et al., 2006). Once SkyNet is trained, Graff et al. (2014) found that it could be finetuned for downstream astronomical tasks.

We can derive positive and negative pairs directly from training data. Hayat et al. (2021) do precisely this and pretrain a SimCLR model on *ugriz*-band galaxy observations (Chen et al., 2020). Along similar lines, Sarmiento et al. (2021) pretrain a SimCLR on MaNGA galaxy data cubes (Bundy et al., 2014). Slijepcevic et al. (2022) show that the BYOL model can produce meaningful embeddings when trained on only positive radio galaxy pairs (Grill et al., 2020), and Akhmetzhanova et al. (2024) show that the VICReg contrastive framework can learn useful embeddings from cosmological simulations (Bardes et al., 2021).

There have also been cross-modal contrastive approaches to representation learning in astronomy. Lanusse et al. (2023) describe using a CLIP-like method (Radford et al., 2021) to align representations of galaxy images and their spectra. Mishra-Sharma et al. (2024) take a similar approach with their PaperCLIP model. They align textual information and astronomical imagery by training a CLIP criterion on imagery-text pairs derived from telescope observation proposals.

From these studies, we can gather that contrastive learning works across multiple domains even when positive pairs are sourced from different instruments and modalities. The one hindrance to scaling contrastive learning approaches across and within modalities is the need to generate these positive pairs. To do so we either need to impose our domain knowledge onto the data pair generation routine, or otherwise labouriously crossmatch intermodal pairs.

1.2. Generative modelling

We can also use generative modelling to create scientifically useful compressed representations of astronomical objects: Schawinski et al. (2018) pretrained a variational autoencoder (VAE; Kingma & Welling, 2013; Lample et al., 2017) on the task of galaxy image recreation. Their model was able to learn unentangled parameters that described physical properties of the galaxies, and furthermore was able to simulate a realistic galaxy given a set of learnt properties. A similar approach was taken in Spindler et al. (2020), who trained a VAE on the pretraining task of recreating galaxy images. They found that their 'AstroVaDEr' model was capable of learning useful embeddings and of restoring galaxy imagery to a high degree of accuracy. Smith et al. (2022) trained a diffusion model on the pretraining task of replicating galaxy images (Ho et al., 2020). Although the resulting diffusion model was not intended to serve as a foundation for downstream tasks, Karchev et al. (2022) were able to take the pretrained model and use it to perform the out of distribution task of reversing gravitational lensing events, without any further training. Non-textual autoregressive generative approaches have found use across astronomical domains. For instance, Zanisi et al. (2021) show that an autoregressive causally-masked PixelCNN++ model (Oord et al., 2016a;b; Salimans et al., 2017) is capable of quantifying the morphological differences between real and simulated galaxy observations. Similarly, Muthukrishna et al. (2021) show that an autoregressive temporal convolutional neural network is capable of learning embeddings that can then be used to detect outliers in their dataset—in this case stellar transients. While all of the above approaches light viable paths towards building our foundation model, we believe that decoding transformer models pretrained in a causal autoregressive manner are currently the most promising approach to realising a sizable LOM that has open code and weights, and that has been pretrained on openly available data. This is largely due to non-technical sociological factors, and we will discuss our reasoning in detail in §2.1.

Causally-masked language foundation models have also been explored in astronomy. Nguyen et al. (2023) presented AstroLLaMA—a LLaMA-2-7B model (Touvron et al., 2023b) that was finetuned on high-quality astronomical research text. Perkowski et al. (2024) extended AstroLLaMA into a conversational chatbot by further fine-

[1] There are of course other interesting approaches that do not fit into this neat dichotomy, for example Charnock et al. (2018); Jeffrey et al. (2021); Walmsley et al. (2022; 2024).

[2] There are many more applications of contrastive learning and generative modelling in astronomy—far more than we can include in this paper. We therefore direct the interested reader to Huertas-Company et al. (2023) for further reading on contrastive learning, and to §§6–8 of Smith & Geach (2023) for generative modelling.

tuning on a mix of synthetic and human-generated astronomy question-answer pairs. Non-textual transformer-based LOMs have been explored in the literature too. As just one example, Leung & Bovy (2023) describe an encoder-decoder transformer-based model for stellar information extraction (Vaswani et al., 2017). While a robust and innovative approach, Leung & Bovy (2023) leave some open questions which we hope to complement with this work: that is, can we scale neural networks on astronomical observation data just as we have done in the textual domain, and do we need the computational and architectural overhead of pretraining a full encoder-decoder transformer architecture to teach our models scientifically useful information?

This manuscript is structured as follows. In the next section we will describe why we believe that a causal transformer is currently the most promising architectural candidate for building our astronomical foundation model (§2.1), and the dataset we used to train our candidate model (§2.2). In §3 we present our results and discussion. Finally, we bring this paper to an end with some closing remarks in §4.

2. Methods

Here we describe the hyperparameters and training routine of AstroPT, and the dataset we used to train the model.

2.1. AstroPT

Decoding transformer architectures (Vaswani et al., 2017; Radford et al., 2018) have a number of benefits that make them well suited as architectural candidates for training LOMs. Firstly, decoding transformers are efficient at pretraining time due to their causal self-attention mask which ensures that every item in an input sequence creates a signal to be backpropagated. Secondly, the decoding transformer's ubiquity within the open source community has lead to an active 'bazaar' of enthusiasts and researchers providing innovations and contributions to the general architecture (Phang et al., 2022; Touvron et al., 2023a; Liu et al., 2023). As the resource required to develop these large models is substantial, it would be beneficial for the field if the astronomical and other observational sciences follow and contribute back to the upstream community's work as much as possible when developing application-driven machine learning models (Raymond, 1999; Smith & Geach, 2023; Rolnick et al., 2024). Also, we can follow Sutton's (2019) 'Bitter Lesson' to a logical endpoint and arrive at the conclusion that the specific neural architecture used to learn a pretraining task does not matter so long as it scales efficiently and is appropriate for the task at hand (see for example Peng et al., 2023; 2024; Bachmann et al., 2023; Smith et al., 2023b; Huh et al., 2024). Taking a generative approach to embedding extraction removes the need for us to impose external knowledge on the network, something that is required to create physically-consistent positive and negative pairs in a contrastive learning regime. And an autoregressive training process allows us to incorporate multiple modalities via simple token chaining, which is not possible with other contrastive or generative approaches. For these reasons it is prudent to build foundation models off of the task-appropriate general neural architecture that has seen the most development time and that enjoys an active community, and to train the general architecture within an autoregressive generative regime. These criteria are fulfilled by the decoding causally masked transformer model.

With these arguments in mind we propose AstroPT—the Astronomical Pretrained Transformer. AstroPT is a GPT-2 like model that has been repurposed for regression tasks (Radford et al., 2019; Smith et al., 2023a). To this end we replace the textual tokeniser with a multilayer perceptron (MLP) tokeniser that feeds directly into AstroPT, and merge position embeddings with the tokens in the standard additive way. In this way, AstroPT is capable of learning any general autoregressive task. In this work we train on a dataset of galaxy imagery, which we describe in the next subsection. To tokenise our galaxy dataset we follow Dosovitskiy et al. (2020) and define a token as a 16×16 pixel patch, and we follow He et al. (2022) and El-Nouby et al. (2024) by independently standardising each 16×16 pixel galaxy image patch to have a mean of zero and a standard deviation of one before passing them into AstroPT's MLP tokeniser. We define the learning objective for AstroPT as predicting the next token in a sequence under the Huber loss criterion (Huber, 1964). We feed the tokens into AstroPT in a 'spiral' order, as shown in Fig. 1.

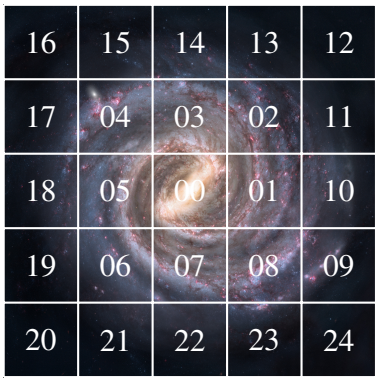


Figure 1. In this work we train AstroPT on the surrogate task of predicting the next token in a 'spiralised' sequence of galaxy image patches. The above image shows the token feed order. As the galaxies are in the centre of each postage stamp, this set up allows us to seamlessly pretrain and run inference on differently sized galaxy postage stamps.

We train a selection of models from 1 million to 2.1 billion parameters, following closely the hyperparameters chosen in Biderman et al. (2023) for comparison's sake. We document our hyperparameters in Tab. 1.

2.2. Dataset and data processing

Our dataset comprises of 8.6 million Dark Energy Spectroscopic Instrument Legacy Survey (DESI-LS) data release 8 galaxies (Dey et al., 2019; Walmsley et al., 2023). DESI-LS is an extensive sky survey comprising three complementary public projects (DECaLS, BASS, and MzLS) aimed at imaging approximately 14 000 deg^2 of the extragalactic sky from both the northern and southern hemispheres, using telescopes at the Kitt Peak National Observatory and the Cerro Tololo Inter-American Observatory. DESI-LS also includes imaging from the Dark Energy Survey (DES), which, although not part of the core DESI-LS, uses the same instrumentation as DECaLS and contributes an additional 5 000 deg^2 of *g,r,z* imaging. The galaxy catalogue used in this work comprises of all extended sources within the DESI-LS source database that have an *r*-band magnitude greater than 19, and a surface brightness less than 18 mag arcsec^{-2} (to avoid stellar contamination). We use this catalogue to build a dataset of source-centred *grz*-band postage stamps of shape 512×512 in JPEG format, at a resolution of 0.262 arcsec pixel^{-1}. Each galaxy is crossmatched with a set of emergent physical properties from the NSA (Aguado et al., 2019), OSSY Type 1 AGN catalogue (Oh et al., 2015), Arecibo Legacy Fast ALFA survey (Haynes et al., 2018), the MPA-JHU SDSS-DR7 derived properties catalogue (Abazajian et al., 2009), the DESI photometric redshift catalogue (Zhou et al., 2021), and Galaxy Zoo morphological classification labels (Walmsley et al., 2023). We can use these crossmatched properties to validate our models' acquired physical knowledge. We downloaded the data directly from DESI-LS using their API, and have reuploaded the resulting galaxy postage stamps and crossmatched property metadata to HuggingFace. We split our galaxy dataset into three sets, a training set that contains 98% (or 8 480 000) of our images, and a test and validation set that each contain 1% (or 86 500) of our images. We pretrain AstroPT on our training set, and finetune our downstream tasks on the validation set, with our downstream task results inferred from the test set.

3. Results and discussion

We present our validation loss plots in Fig. 2. We can see a clear saturating log-log neural scaling law, agreeing with prior work (Cortes et al., 1993; Kaplan et al., 2020; Henighan et al., 2020). We find that the saturation point occurs near our 89 million parameter model, and therefore use our smallest and largest non-saturated models (AstroPT-1M and AstroPT-89M) to generate our embedding plots in Fig. 3. We use all our non-saturated models (that is, AstroPT-{1M, 5M, 12M, 12M, 89M}) to produce our downstream task scaling tests (Fig. 4).

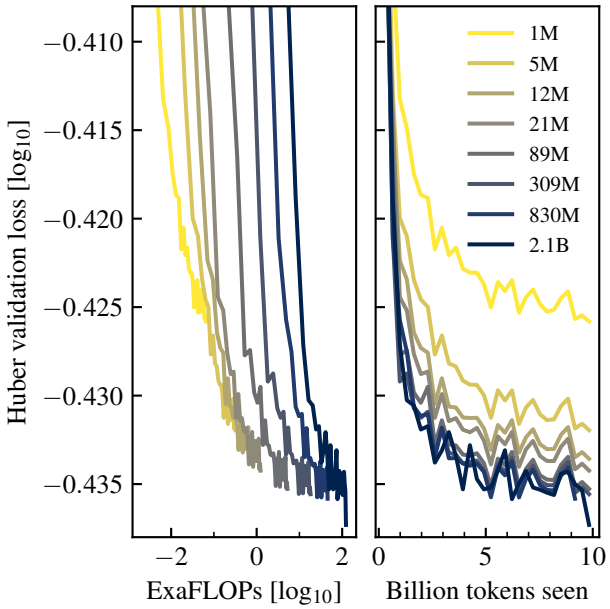


Figure 2. Validation set losses over our full training runs. The left plot shows the validation loss per training floating point operation (FLOP), and the right plot shows the validation loss per 16×16 image patch token seen. Each run is labelled with the total neural parameter count as crossmatched in Tab. 1.

Figure 2 shows a saturating loss after some log-log scaling, which we expect is due to the information density in our chosen dataset (Henighan et al., 2020). In future work we will attempt to confirm our intuition by comparing this result with a LOM trained on other—more information dense—observational modalities such as galaxy spectra, less processed galaxy photometry, stellar time series, or mixed modalities.

To produce an embedding from our AstroPT models we take the averaged outputs of the models' penultimate layer over the central 64 tokens of a galaxy image. In Fig. 3 we show a 'Uniform Manifold Approximation and Projection' (UMAP; McInnes et al., 2018) projection of AstroPT-1M's and AstroPT-89M's embeddings. In these projections each hex bin is coloured by a selected emergent galactic property. We can see a clear structure in each subplot, which suggests that the model has learnt semantically meaningful properties of our galaxy dataset. However, it is difficult to conclude from these projections whether the larger model's embeddings are more informative than those of the smaller model's. We therefore perform linear probing on the embedding spaces of our non-saturated models to quantify

Table 1. Chosen hyperparameters for our AstroPT models. The discrepancy in the number of model parameters between the comparable language models and AstroPT models is due to differences in number of parameters between the models' respective tokenisers. The Pythia family of models are described in Biderman et al. (2023), the GPT-Neo-125M model is described in Black et al. (2021), and the OPT-125M model is described in Zhang et al. (2022).

N PARAMS.	LAYERS	MODEL DIM	HEADS	LEARNING RATE	COMPARABLE LANGUAGE MODELS
1M	4	128	8	10×10^{-4}	—
5M	6	256	8	10×10^{-4}	—
12M	6	384	8	10×10^{-4}	—
21M	6	512	8	10×10^{-4}	PYTHIA-70M
89M	12	768	12	6×10^{-4}	PYTHIA-160M, GPT-NEO-125M, OPT-125M
309M	24	1024	16	3×10^{-4}	PYTHIA-160M
830M	16	2048	8	3×10^{-4}	PYTHIA-1.0B
2.1B	26	2560	32	1.6×10^{-4}	—

how physically informative our model embeddings are, and describe the process in the next paragraph.

To investigate AstroPT's performance per pretraining FLOP on downstream tasks we extract embeddings in the same way as described in the previous paragraph, and use them to train a linear probe mapping from the embedding to a selected set of emergent physical properties. We train our linear probes on the validation set (86 500 galaxies held out during pretraining), and infer on the test set (86 500 further held out galaxies). We perform this probing for all of our models before the loss saturation point that is shown in Fig. 2: the 1M, 5M, 12M, 21M, and 89M parameter models. Once we have the linear probe accuracy of our downstream tasks for each of our tested models, we run a Spearman's ρ statistical test to measure the correlation between total pretraining compute spent and downstream task performance. Our findings are presented in Fig. 4. We find that there is a clear positive correlation between downstream task performance and pretraining FLOP spent.

The results of our linear probe performance per pretraining FLOP are promising, and we expect this result to carry over to other observational modalities across domains, neural architectures, and pretraining routines. Interestingly, we also see 'emergent' abilities (or abilities that suddenly manifest at a certain model scale) here, similarly to what has been shown in language modelling (Wei et al., 2022): the stellar mass estimation (M_*) and tight spiral morphological classification probes see sudden performance improvements at 12M parameters.

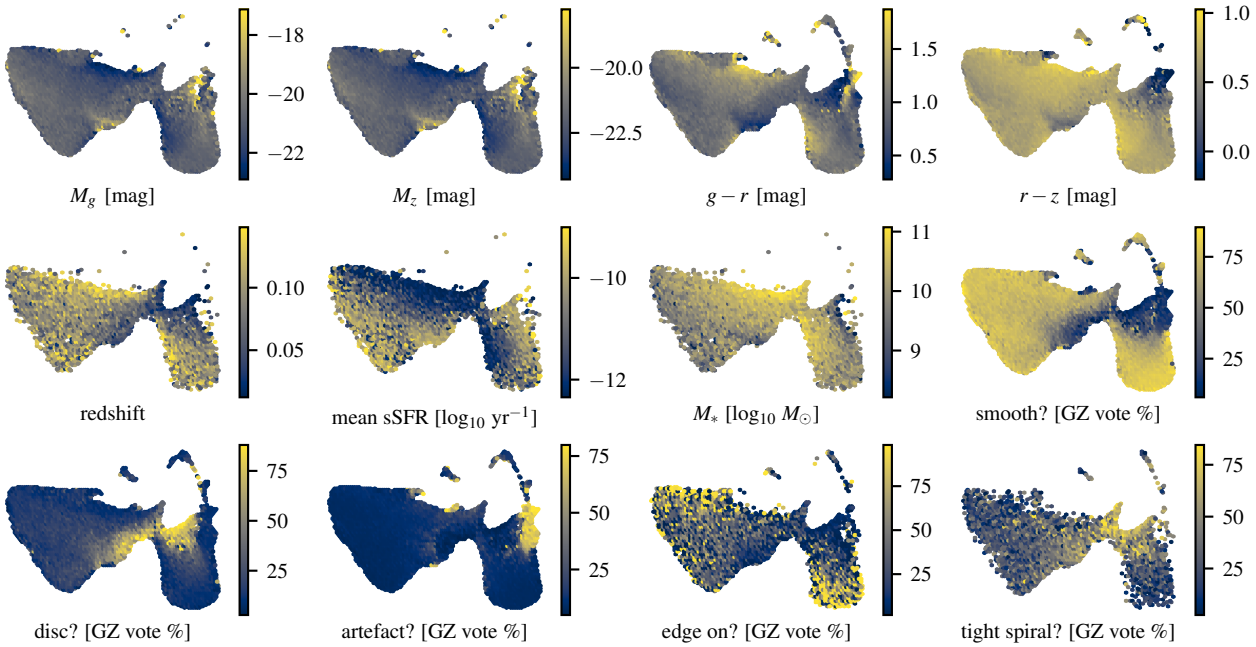
Notably, aside from the properties shown in Fig. 3 and Fig. 4, we can to an extent predict the galaxy location and object ID with a linear probe. This is a result that initially may seem surprising. However, given that the DESI-LS telescopes used for surveying different sections of the sky reportedly have distinct effective response curves (Zhou et al., 2021), it is likely that the model captures the instrumental differences between them. We can also see this effect in our UMAP plots in Fig. 3, where each 'island' corresponds to objects observed with different DESI-LS telescopes. This effect could be mitigated downstream if desired via some domain adaptation finetuning, for example via adversarial domain adaptation (Ganin et al., 2016; Ćiprijanović et al., 2020).

We chose to use linear probing for simplicity—and to robustly show that model scale drives downstream task improvement—but of course we can also use more complicated finetuning methods that would not assume a linear relationship between our embedding space and the downstream task labels. As this study's purpose is to show that self-supervised causal transformer models can scale in downstream performance in the astronomical domain, we leave investigation into optimal finetuning methods and therefore comparison to the state-of-the-art to future work.

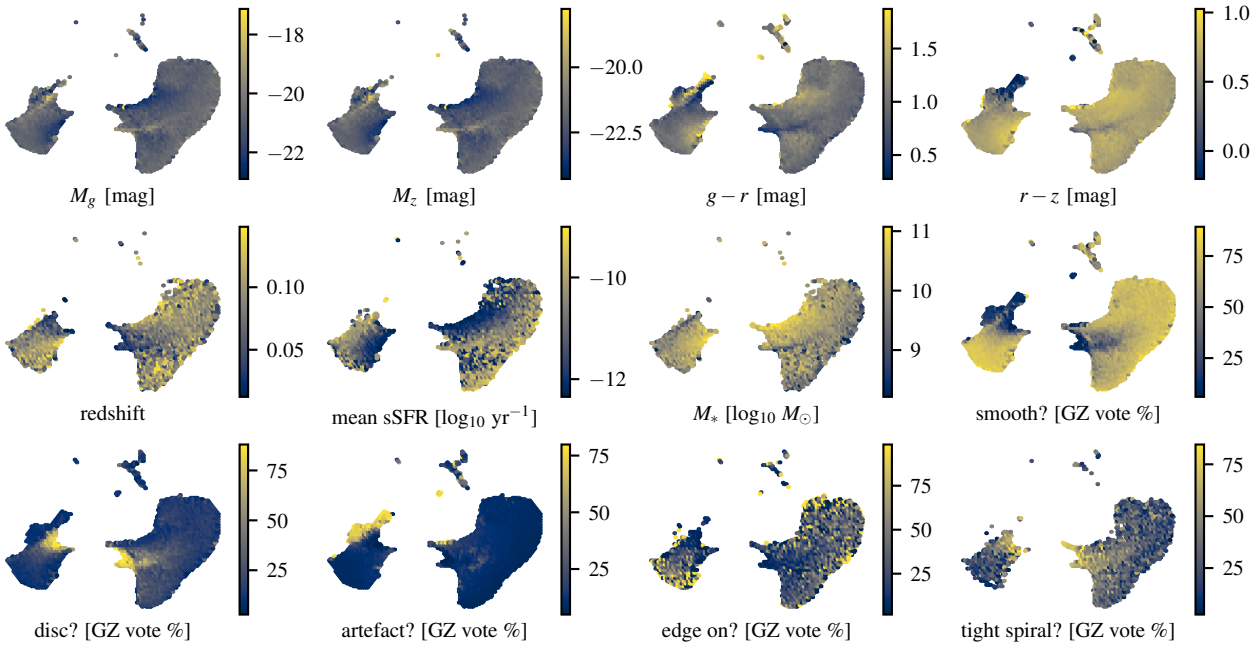
4. Conclusions

To summarise this paper's technical contributions: we demonstrated that simple generative autoregressive models can learn scientifically useful information when pretrained on the surrogate task of predicting the next 16×16 pixel patch in a sequence of galaxy image patches. Furthermore, we showed that our AstroPT model scales predictably in downstream task performance as it is pretrained on more compute, a process that has been shown to be true within natural imagery (e.g. Henighan et al., 2020) and natural language (e.g. Kaplan et al., 2020). This is a promising result that suggests that data taken from the observational sciences would complement data from other domains when used to pretrain a single multimodal LOM (Aghajanyan et al., 2023), and so points towards the use of observational data as one solution to the 'token crisis' (Sevilla et al., 2022; Xue et al., 2023).

As the AstroPT LOM framework is—by design—very flexible, we expect that similar autoregressive models can be used across many observational modalities. And we delib-



(a) Here we show our UMAP projected two dimensional embedding plots for AstroPT-1M.



(b) Here we show our UMAP projected two dimensional embedding plots for AstroPT-89M.

Figure 3. Results from our AstroPT-1M embedding UMAP projections (upper), and AstroPT-89M embedding UMAP projections (lower). We colour the hex bins in both plots with a selected set of emergent physical properties of the galaxies. We find significant structure, signifying that the model has learned physically meaningful representations of the dataset. In the above plots 'M_g' and 'M_z' are the absolute magnitudes in the g and z bands, 'mean sSFR' is the mean specific star formation rate, and 'M_*' is the stellar mass. 'smooth?', 'disc?', 'artefact?', 'edge on?' and 'tight spiral?' are Galaxy Zoo survey responses for these morphological features. Our metadata sources are described further in §2.2.

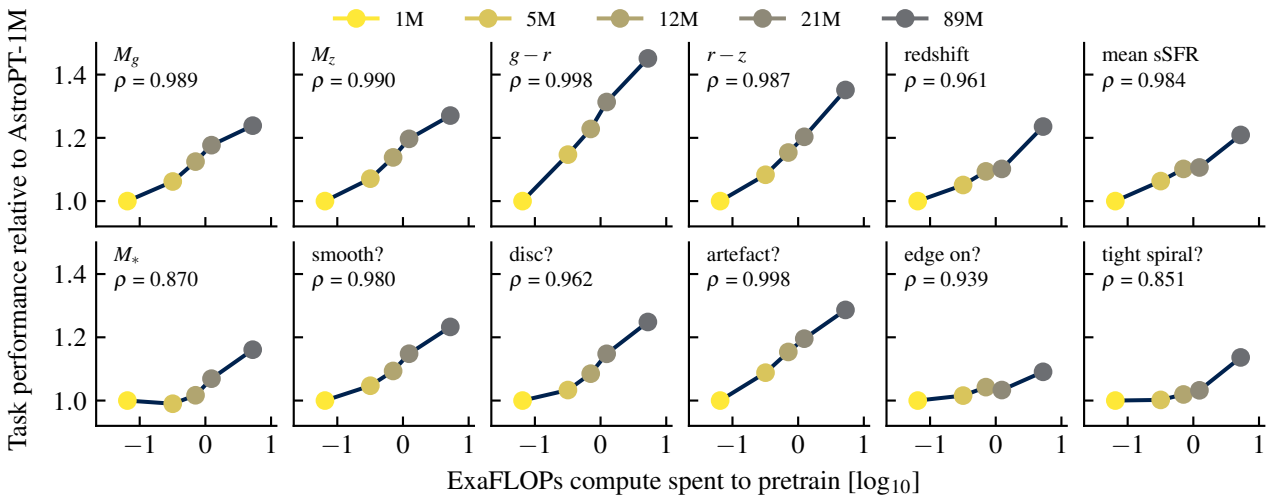


Figure 4. Here we show our relative linear probe performances per pretraining FLOP spent on a selection of scientifically-meaningful downstream tasks. The markers are coloured according to the models' parameter counts. We run a Spearman's ρ fit and find in all cases a strong positive correlation between downstream task performance and model size, meaning that a larger model has more informative embeddings. In this plot 'M_g' and 'M_z' are the absolute magnitudes in the g and z bands, 'mean sSFR' is the mean specific star formation rate, and 'M_*' is the stellar mass. 'smooth?', 'disc?', 'artefact?', 'edge on?' and 'tight spiral?' are Galaxy Zoo survey responses for these morphological features. Our metadata sources are described further in §2.2.

erately designed the AstroPT architecture and generative training regime to stay as close to the current leading community models as possible. We took these decisions in the belief that collaborative community development paves the fastest route towards realising an open source web-scale large observation model. We therefore hope that this work seeds further research in this area, and we release our full dataset, model weights for all trained models checkpointed across the entire training run, and our training code under the MIT license to encourage this. We also welcome direct collaboration—drawing inspiration from EleutherAI's call to do 'science in the open' (Phang et al., 2022) we developed AstroPT in public from inception as open-to-all project, and we will continue to build the AstroPT family of LOMs in public on the UniverseTBD Discord server[3]. We warmly invite potential collaborators to join us.

Code, model weights, and data availability

Our code is available on Github here:
`github.com/Smith42/astroPT`
Our model weights are available on HuggingFace here:
`hf.co/Smith42/astroPT`
The data used to train this model are available here:
`hf.co/datasets/Smith42/galaxies`

[3] `https://discord.gg/CjMBBJKnFH`

Carbon emissions

The training of deep learning models requires considerable energy, contributing to carbon emissions. To counteract further emission from redundant retraining, we follow the recommendations of Strubell et al. (2019) and make available our fully trained models and code. To estimate the carbon equivalent emitted during this work we used the excellent CodeCarbon (`codecarbon.io`), which estimated a total of 120 kg CO_2 eq. across our final pretraining runs.

Acknowledgements

We would like to thank James Geach, Jia-Shu Pan, Kevin Schawinski, Regina Sarmiento, Mike Walmsley, and Zahra Sharbaf for illuminating discussions and comments. MJS would like to thank the Instituto de Astrofísica de Canarias for hosting him while he worked on this project. This study made use of the Liverpool John Moores University's Propsero HPC facility (`prospero-docs.readthedocs.io`).

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Prieto, C. A., An, D., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., et al. The seventh data release of the Sloan Digital Sky Survey. *As-*

trophysical Journal Supplement Series, 182(2):543, 2009. ISSN 0067-0049. doi: 10.1088/0067-0049/182/2/543.

Aghajanyan, A., Yu, L., Conneau, A., Hsu, W.-N., Hambardzumyan, K., Zhang, S., Roller, S., Goyal, N., Levy, O., and Zettlemoyer, L. Scaling Laws for Generative Mixed-Modal Language Models. *ArXiv*, 2023. doi: 10.48550/arXiv.2301.03728.

Aguado, D. S., Ahumada, R., Almeida, A., Anderson, S. F., Andrews, B. H., Anguiano, B., Ortíz, E. A., Aragón-Salamanca, A., Argudo-Fernández, M., Aubert, M., et al. The Fifteenth Data Release of the Sloan Digital Sky Surveys: First Release of MaNGA-derived Quantities, Data Visualization Tools, and Stellar Library. *Astrophysical Journal Supplement Series*, 240(2):23, 2019. ISSN 0067-0049. doi: 10.3847/1538-4365/aaf651.

Akhmetzhanova, A., Mishra-Sharma, S., and Dvorkin, C. Data compression and inference in cosmology with self-supervised machine learning. *Monthly Notices of the Royal Astronomical Society*, 527(3):7459–7481, 2024. ISSN 0035-8711. doi: 10.1093/mnras/stad3646.

Bachmann, G., Anagnostidis, S., and Hofmann, T. Scaling MLPs: A Tale of Inductive Bias. *Advances in Neural Information Processing Systems*, 36:60821–60840, 2023. URL https://papers.nips.cc/paper_files/paper/2023/hash/bf2a5ce85aea9ff40d9bf8b2c2561cae-Abstract-Conference.html.

Bardes, A., Ponce, J., and LeCun, Y. VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. *ArXiv e-prints*, 2021. doi: 10.48550/arXiv.2105.04906.

Biderman, S., Schoelkopf, H., Anthony, Q. G., Bradley, H., O'Brien, K., Hallahan, E., Khan, M. A., Purohit, S., Prashanth, U. S., Raff, E., et al. Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling. In *International Conference on Machine Learning*, pp. 2397–2430. PMLR, 2023. URL https://proceedings.mlr.press/v202/biderman23a.html.

Black, S., Leo, G., Wang, P., Leahy, C., and Biderman, S. GPT-Neo: Large Scale Autoregressive Language Modeling with Mesh-Tensorflow, 2021. URL https://doi.org/10.5281/zenodo.5297715.

Bundy, K., Bershady, M. A., Law, D. R., Yan, R., Drory, N., MacDonald, N., Wake, D. A., Cherinka, B., Sánchez-Gallego, J. R., Weijmans, A.-M., et al. Overview of the SDSS-IV MaNGA survey: mapping nearby galaxies at apache point observatory. *Astrophysical Journal*, 798(1):7, 2014. ISSN 0004-637X. doi: 10.1088/0004-637X/798/1/7.

Ćiprijanović, A., Kafkes, D., Jenkins, S., Downey, K., Perdue, G. N., Madireddy, S., Johnston, T., and Nord, B. Domain adaptation techniques for improved cross-domain study of galaxy mergers. *ArXiv e-prints*, 2020. doi: 10.48550/arXiv.2011.03591.

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.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. A simple framework for contrastive learning of visual representations. In *ICML'20: Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 1597–1607. JMLR.org, 2020. doi: 10.5555/3524938.3525087.

Cortes, C., Jackel, L. D., Solla, S., Vapnik, V., and Denker, J. Learning Curves: Asymptotic Values and Rate of Convergence. *Advances in Neural Information Processing Systems*, 6, 1993. URL https://proceedings.neurips.cc/paper/1993/hash/1aa48fc4880bb0c9b8a3bf979d3b917e-Abstract.html.

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.

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

El-Nouby, A., Klein, M., Zhai, S., Bautista, M. A., Toshev, A., Shankar, V., Susskind, J. M., and Joulin, A. Scalable Pre-training of Large Autoregressive Image Models. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2401.08541.

Ganin, Ya., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., and Lempitsky, V. Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016. ISSN 1533-7928. URL https://jmlr.org/papers/v17/15-239.html.

Graff, P., Feroz, F., Hobson, M. P., and Lasenby, A. SkyNet: an efficient and robust neural network training tool for machine learning in astronomy. *Monthly Notices of the Royal Astronomical Society*, 441(2):1741–1759, 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu642.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Pires, B. A., Guo,

Z. D., Azar, M. G., et al. Bootstrap your own latent a new approach to self-supervised learning. In *NIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, pp. 21271–21284. Curran Associates Inc., Red Hook, NY, USA, 2020. ISBN 978-1-71382954-6. doi: 10.5555/3495724.3497510.

Gunasekar, S., Zhang, Y., Aneja, J., Mendes, C. C. T., Del Giorno, A., Gopi, S., Javaheripi, M., Kauffmann, P., de Rosa, G., Saarikivi, O., et al. Textbooks Are All You Need. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2306.11644.

Hayat, M. A., Stein, G., Harrington, P., Lukić, Z., and Mustafa, M. Self-supervised representation learning for astronomical images. *The Astrophysical Journal Letters*, 911(2):L33, 2021. doi: 10.3847/2041-8213/abf2c7. URL https://doi.org/10.3847/2041-8213/abf2c7.

Haynes, M. P., Giovanelli, R., Kent, B. R., Adams, E. A. K., Balonek, T. J., Craig, D. W., Fertig, D., Finn, R., Giovanardi, C., Hallenbeck, G., et al. The Arecibo Legacy Fast ALFA Survey: The ALFALFA Extragalactic H i Source Catalog. *Astrophysical Journal*, 861(1):49, 2018. ISSN 0004-637X. doi: 10.3847/1538-4357/aac956.

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.

Henighan, T., Kaplan, J., Katz, M., Chen, M., Hesse, C., Jackson, J., Jun, H., Brown, T. B., Dhariwal, P., Gray, S., et al. Scaling Laws for Autoregressive Generative Modeling. *ArXiv e-prints*, 2020. doi: 10.48550/arXiv.2010.14701.

Hinton, G. E. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800, 2002. ISSN 0899-7667. doi: 10.1162/089976602760128018.

Hinton, G. E., Osindero, S., and Teh, Y. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006. ISSN 0899-7667. doi: 10.1162/neco.2006.18.7.1527. URL https://doi.org/10.1162/neco.2006.18.7.1527.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

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.

Huber, P. J. Robust Estimation of a Location Parameter. *Ann. Math. Stat.*, 35(1):73–101, 1964. ISSN 0003-4851. doi: 10.1214/aoms/1177703732.

Huertas-Company, M., Sarmiento, R., and Knapen, J. H. A brief review of contrastive learning applied to astrophysics. *RAS Techniques and Instruments*, 2(1):441–452, 2023. ISSN 2752-8200. doi: 10.1093/rasti/rzad028.

Huh, M., Cheung, B., Wang, T., and Isola, P. The Platonic Representation Hypothesis. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2405.07987.

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873:111, 2019. doi: 10.3847/1538-4357/ab042c.

Jeffrey, N., Alsing, J., and Lanusse, F. Likelihood-free inference with neural compression of DES SV weak lensing map statistics. *Monthly Notices of the Royal Astronomical Society*, 501(1):954–969, 2021. ISSN 0035-8711. doi: 10.1093/mnras/staa3594.

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.

Karchev, K., Montel, N. A., Coogan, A., and Weniger, C. Strong-Lensing Source Reconstruction with Denoising Diffusion Restoration Models. *arXiv*, 2022. doi: 10.48550/arXiv.2211.04365.

Kingma, D. P. and Welling, M. Auto-Encoding Variational Bayes. *ArXiv e-prints*, 2013. doi: 10.48550/arXiv.1312.6114.

Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., and Ranzato, M. Fader networks: manipulating images by sliding attributes. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 5969–5978. Curran Associates Inc., Red Hook, NY, USA, 2017. ISBN 978-1-51086096-4. doi: 10.5555/3295222.3295346.

Lanusse, F., Parker, L., Golkar, S., Cranmer, M., Bietti, A., Eickenberg, M., Krawezik, G., McCabe, M., Ohana, R., Pettee, M., et al. AstroCLIP: Cross-Modal Pre-Training for Astronomical Foundation Models. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2310.03024.

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, Z., Qiao, A., Neiswanger, W., Wang, H., Tan, B., Tao, T., Li, J., Wang, Y., Sun, S., Pangarkar, O., et al. LLM360: Towards Fully Transparent Open-Source LLMs. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2312.06550.

McInnes, L., Healy, J., and Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*, 2018. doi: 10.48550/arXiv.1802.03426.

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.

Muthukrishna, D., Mandel, K. S., Lochner, M., Webb, S., and Narayan, G. Real-time Detection of Anomalies in Multivariate Time Series of Astronomical Data. *ArXiv e-prints*, 2021. doi: 10.48550/arXiv.2112.08415.

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.

Oh, K., Yi, S. K., Schawinski, K., Koss, M., Trakhtenbrot, B., and Soto, K. A new catalog of type 1 AGNs and its implications on the AGN unified model. *Astrophysical Journal Supplement Series*, 219(1):1, 2015. ISSN 0067-0049. doi: 10.1088/0067-0049/219/1/1.

Oord, A. v. d., Kalchbrenner, N., and Kavukcuoglu, K. Pixel Recurrent Neural Networks. *ArXiv e-prints*, 2016a. doi: 10.48550/arXiv.1601.06759.

Oord, A. v. d., Kalchbrenner, N., Vinyals, O., Espeholt, L., Graves, A., and Kavukcuoglu, K. Conditional image generation with PixelCNN decoders. In *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 4797–4805. Curran Associates Inc., Red Hook, NY, USA, 2016b. ISBN 978-1-51083881-9. doi: 10.5555/3157382.3157633.

Peng, B., Alcaide, E., Anthony, Q., Albalak, A., Arcadinho, S., Biderman, S., Cao, H., Cheng, X., Chung, M., Grella, M., et al. RWKV: Reinventing RNNs for the Transformer Era. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2305.13048.

Peng, B., Goldstein, D., Anthony, Q., Albalak, A., Alcaide, E., Biderman, S., Cheah, E., Du, X., Ferdinan, T., Hou, H., et al. Eagle and Finch: RWKV with Matrix-Valued States and Dynamic Recurrence. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2404.05892.

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.

Phang, J., Bradley, H., Gao, L., Castricato, L., and Biderman, S. EleutherAI: Going Beyond "Open Science" to "Science in the Open". *ArXiv e-prints*, 2022. doi: 10.48550/arXiv.2210.06413.

Radford, A., Narasimhan, K., Salimans, T., and Sutskever, I. Improving language understanding by generative pre-training. *OpenAI Whitepaper*, 2018. URL https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

Radford, A., Wu, J., Child, R., Luan, D., A., D., and Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Whitepaper*, 2019. URL https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

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.

Raymond, E. S. *The Cathedral and the Bazaar*. O'Reilly & Associates, Inc., USA, 1st edition, 1999. ISBN 1565927249.

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-maron, G., Giménez, M., Sulsky, Y., Kay, J., Springenberg, J. T., Eccles, T., Bruce, J., Razavi, A., Edwards, A., Heess, N., Chen, Y., Hadsell, R., Vinyals, O., Bordbar, M., and de Freitas, N. A Generalist Agent. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=1ikK0kHjvj.

Rolnick, D., Aspuru-Guzik, A., Beery, S., Dilkina, B., Donti, P. L., Ghassemi, M., Kerner, H., Monteleoni, C., Rolf, E., Tambe, M., et al. Application-Driven Innovation in Machine Learning. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2403.17381.

Salimans, T., Karpathy, A., Chen, X., and Kingma, D. P. PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications. *ArXiv e-prints*, 2017. doi: 10.48550/arXiv.1701.05517.

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.

Schawinski, K., Turp, M. D., and Zhang, C. Exploring galaxy evolution with generative models. *Astronomy & Astrophysics*, 616:L16, 2018. ISSN 0004-6361. doi: 10.1051/0004-6361/201833800.

Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., and Villalobos, P. Compute Trends Across Three Eras of Machine Learning. *arXiv*, 2022. doi: 10.48550/arXiv.2202.05924.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419):1140–1144, 2018. ISSN 0036-8075. doi: 10.1126/science.aar6404.

Slijepcevic, I. V., Scaife, A. M. M., Walmsley, M., and Bowles, M. Learning useful representations for radio astronomy "in the wild" with contrastive learning. *arXiv*, 2022. doi: 10.48550/arXiv.2207.08666.

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., Fleming, L., and Geach, J. E. EarthPT: a foundation model for Earth Observation. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023a. URL https://www.climatechange.ai/papers/neurips2023/2.

Smith, S. L., Brock, A., Berrada, L., and De, S. ConvNets Match Vision Transformers at Scale. *arXiv e-prints*, pp. arXiv:2310.16764, 2023b. doi: 10.48550/arXiv.2310.16764.

Spindler, A., Geach, J. E., and Smith, M. J. AstroVaDEr: Astronomical Variational Deep Embedder for Unsupervised Morphological Classification of Galaxies and Synthetic Image Generation. *Monthly Notices of the Royal Astronomical Society*, 502:985, 2020. ISSN 0035-8711. doi: 10.1093/mnras/staa3670. URL https://doi.org/10.1093/mnras/staa3670. staa3670.

Strubell, E., Ganesh, A., and Mccallum, A. Energy and Policy Considerations for Deep Learning in NLP. *ACL Anthology*, pp. 3645–3650, 2019. doi: 10.18653/v1/P19-1355.

Sutton, R. The Bitter Lesson, 2019. URL http://incompleteideas.net/IncIdeas/BitterLesson.html.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. LLaMA: Open and Efficient Foundation Language Models. *ArXiv e-prints*, 2023a. doi: 10.48550/arXiv.2302.13971.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al. Llama 2: Open Foundation and Fine-Tuned Chat Models. *ArXiv e-prints*, 2023b. doi: 10.48550/arXiv.2307.09288.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention Is All You Need. *arXiv*, 2017. doi: 10.48550/arXiv.1706.03762.

Walmsley, M., Scaife, A. M. M., Lintott, C., Lochner, M., Etsebeth, V., Géron, T., Dickinson, H., Fortson, L., Kruk, S., Masters, K. L., et al. Practical galaxy morphology tools from deep supervised representation learning. *Monthly Notices of the Royal Astronomical Society*, 513(2):1581–1599, 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac525.

Walmsley, M., Géron, T., Kruk, S., Scaife, A. M. M., Lintott, C., Masters, K. L., Dawson, J. M., Dickinson, H., Fortson, L., Garland, I. L., et al. Galaxy Zoo DESI: Detailed morphology measurements for 8.7M galaxies in the DESI Legacy Imaging Surveys. *Monthly Notices of the Royal Astronomical Society*, 526(3):4768–4786, 2023. ISSN 0035-8711. doi: 10.1093/mnras/stad2919.

Walmsley, M., Bowles, M., Scaife, A. M. M., Makechemu, J. S., Gordon, A. J., Ferguson, A. M. N., Mann, R. G., Pearson, J., Popp, J. J., Bovy, J., et al. Scaling Laws for Galaxy Images. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2404.02973.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., et al. Emergent Abilities of Large Language Models. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=yzkSU5zdwD.

Xue, F., Fu, Y., Zhou, W., Zheng, Z., and You, Y. To Repeat or Not To Repeat: Insights from Scaling LLM under Token-Crisis. *ArXiv e-prints*, 2023. doi: 10.48550/arXiv.2305.13230.

Zanisi, L., Huertas-Company, M., Lanusse, F., Bottrell, C., Pillepich, A., Nelson, D., Rodriguez-Gomez, V., Shankar, F., Hernquist, L., Dekel, A., et al. A deep learning approach to test the small-scale galaxy morphology and its relationship with star formation activity in hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 501(3):4359–4382, 2021. ISSN 0035-8711. doi: 10.1093/mnras/staa3864.

Zhang, P., Zeng, G., Wang, T., and Lu, W. TinyLlama: An Open-Source Small Language Model. *ArXiv e-prints*, 2024. doi: 10.48550/arXiv.2401.02385.

Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., Dewan, C., Diab, M., Li, X., Lin, X. V., et al. OPT: Open Pre-trained Transformer Language Models. *ArXiv e-prints*, 2022. doi: 10.48550/arXiv.2205.01068.

Zhou, R., Newman, J. A., Mao, Y.-Y., Meisner, A., Moustakas, J., Myers, A. D., Prakash, A., Zentner, A. R., Brooks, D., Duan, Y., et al. The clustering of DESI-like luminous red galaxies using photometric redshifts. *Monthly Notices of the Royal Astronomical Society*, 501(3):3309–3331, 2021. ISSN 0035-8711. doi: 10.1093/mnras/staa3764.