Micah **Bowles**

Oxford, UK

micah.bowles@physics.ox.ac.uk | mb010.github.io | mb010 | Micah Bowles

Overview

- Applied AI researcher handling real world noisy multi-sensor and multimodal data. Fluent English and German.
- Experienced with computer vision, efficient pre-training, generative modelling, multimodal modelling, NLP, interpretability, custom evaluations, and vector embeddings.
- Well versed in evaluating, designing, and implementing deep learning models with PyTorch (and some tensorflow and JAX), lightning, HuggingFace, transformers, and HF Datasets.
- Strong research engineer experienced with project management, git, experiment logging, high performance/cloud computing (SLURM, AWS), distributed training (FSDP, DDPM), and data pipelines (Python, BASH, SCALA).

Work Experience

Schmidt AI in Science Fellow (Independant Research Fellow)

Oxford, UK

University of Oxford — *May 2024 - Present*

- Awarded £140k in competitive research funding to research large scale models for astronomy.
- Construction of largest astronomical multimodal dataset (70TB) for large scale scientific pre-training.
- Investigating vision transformer masked image modelling systematic failures. Proposed and evaluated a tractable solution to reduce the error rate by over 50% in this regime for fine-tuned models (on downstream tasks) whilst maintaining scalability.
- Large scale pre-training for representation learning across various sensor data and modalities.
- Building and leading a team of international researchers to investigate conditional generative models for science.
- Construction and evaluation of native resolution vision transformers for efficient pre-training and pixel scale aware models.
- Score based modells (diffussion models) for inverse problems in astronomy.

Research Engineering Consultant

UK

Polymathic AI — *August 2024 - Present*

- Consulting on research driven approaches to massively multimodal foundation model for science.
- Supporting in training and evaluation of the large scale foundation model (billions of parameters) for science.
- Discuss and implement research driven ideas for a production ready astrophysics models.

Quantum Innovation (Placement)

UK

DiRAC — *(3m) Winter 2023*

- Design and development of an easily extensible quantum machine learning framework for training quantum-classical hybrid neural networks.

Machine Learning Scientist (Intern)

London, UK

Expedia Group — *(3m) Summer 2023*

- Personalising recommendation systems for Expedia Group search results (10 million searches per day).
- Custom data pipeline for historical personalised recommendation data from the internal data lake through with AWS clusters.
- Trained models from the personalised recommendation data.
- Models resulted in a $1.36 M/yr measured uptick in revenue.

Education

PhD (AI for Science)
Advancing Artificial Intelligence in Astronomy

Manchester, UK

The University of Manchester, Funded by The Alan Turing Institute — *Oct 2020 - April 2024*

- PhD in AI for science. Research implementing and enabling the first NLP methods and applications in astronomy, and advancing the state of the art in computer vision methods for astronomy resulting in publications at NeurIPS, ICML, MNRAS, JOSS, and RASTI.
- Developed a novel NLP method to derive semantic taxonomy (pre-LLM era). Proposed approach to adapt the method to the LLM era.
- Enabled the first NLP based analyses in astronomy, by enabling plain English to be used in place of technical terminology. This resulted in me organising a large scale data collection effort without the need for domain specific training and enabling language based semantic search for scientists.
- Identified novel failure mode in vision transformer (ViT) based masked image modelling hindering scientific applications and proposed a novel solution enabling the application of these state of the art pre-training methods.
- Developed multiple HPC-centric data pipelines, including enabling a multi-TB scale science data imaging pipeline to run dynamically on any SLURM based cluster.
- Worked within a team to advance self-supervised learning approaches achieving state of the art performance for noisy astronomy data, achieving an error reduction of 19%.
- Constructed equivariant convolutional self-attention based model, reducing overfitting, improving prediction consistency across augmentations, and improves the interpretability of the respective attention maps (saliency maps).

MScR, Distinction

Manchester, UK

The University of Manchester — *2019 - 2020*

- First in my year.
- A year of AI in Science research and classes.
- Constructed and evaluated self-attention mechanisms in convolutional neural networks for scientific computer vision, employing the attention mechanism as an interpretable saliency map and improving upon the performance of the previous state of the art approaches in the field. Leading to a publication in MNRAS.

BSc Physics

Cologne, Germany

The University of Cologne — *2012 - 2018*

- Writing scientific software to simulate self-gravitating galaxy mergers, leading to a publication in A&A.

Professional Activities

Softwares

- Zoobot: Open source framework designed to have sufficient abstraction for scientists to have easy access to large scale pre-trained domain specific neural networks, including finetuning for both tensorflow and pytorch.
- Multimodal Universe: Data cleaning and hosting for 70TB of scientific multimodal data from a large variety of instruments.
- Cata2Data: Framework to dynamically and quickly turn any sets of astronomical images and catalogues into machine learning ready data loaders drawing from raw astronomical data at speed.
- AstroAugmentations: Custom imaging augmentations for instrument and science specific deep learning based data augmentations.
- ProcessMeerKAT: Multi-node and multi-threaded TB scale data pipeline for interferometric imaging for any SLURM based system.

Public Speaking

- Seven invited colloquia and seminars, including at the Flatiron Institute (NY), and the Universities of Birmingham (UK) and Manchester (UK) for which expenses were included. As well as online talks at the University of Montrèal (CA) and the Open University (UK). *2021 - Present*
- Initialised special graduate seminars on sessions on teaching on software development practices, version tracking, data management, machine learning, and neural networks.
- For a full list of talks, including conferences and paid events, see my website.

Miscellaneous

- Supervised an MSc student's research project (statistical science), two summer research interns, and co-supervisor of one undergraduate thesis. *2021 - Present*
- Reviewer for MNRAS, NeurIPS ML4Phys Workshop, ICML ML4Astro Workshop, and the Astrophysical Journal (ApJ).

Selected Publications

- Cover All Your Bases: Representations of high frequency features from masked auto encoding, in-prep, lead author.
- Scaling Laws for Galaxy Images, Submitted, `https://arxiv.org/abs/2404.02973`, co-author.
- The Multimodal Universe: Enabling Large-Scale Machine Learning with 70TBs of Astronomical Scientific Data, submitted, coordinator.
- For a full list of my publications see my google scholar profile: `https://scholar.google.com/citations?user=Q7ziv7YAAAAJ&hl=en`