Henry W. Leung Ph.D.

Astronomy & Astrophysics Researcher at the University of Toronto

henrysky.github.io henryskyleung@gmail.com henrysky Henry Leung

Bilingual in English & Chinese Python & C Canadian & Hong Konger

SUMMARY

I am a recent graduate in Astronomy and a Data Science Institute doctoral fellow at the University of Toronto, specializing in the application of advanced **deep learning** methods, including Transformers and denoising diffusion, to build **multi-modal foundation models for science**. My research, presented at conferences such as *NeurIPS* and *ICML*, involved analyzing large, cross-domain astronomical datasets, contributing to a deeper understanding of the dynamics of our Galaxy. I am passionate about open science, with the majority of my code and models being well-tested, well-documented, and openly available to the community. I am eager to apply my expertise in machine learning, data analysis, and software development to solve complex problems and drive innovation in industry.

PROFESSIONAL EXPERIENCE

University of Toronto Sept 2019 — Oct 2024
Graduate Researcher & **Data Science Institute Doctoral Fellow**

- Led an independent, data-driven research program, integrating cutting-edge deep learning techniques such as Transformers, Denoising Diffusion models, and Large Language Models (LLMs) to develop multi-modal foundation models for scientific applications.
- Created and maintained robust, well-documented, and thoroughly tested open-source software mainly written in `Python`, `C` and `SQL`, contributing both to personal projects and to the wider scientific community.
- Delivered spotlight talks and posters at major conferences, including *NeurIPS*, *ICML*, and *ESA AI in Astronomy*, and collaborated with community-led initiatives like the Multimodal Universe project.
- Curated a value-added catalogue of deep learning-derived stellar parameters and uncertainties, for the Sloan Digital Sky Survey (SDSS) collaboration.
- Built and optimized machine learning models using frameworks like `PyTorch` and `TensorFlow` trained on large astronomical datasets, leveraging tools such as `Docker` and `Postgresql` as well as deploying them on Canada's research supercomputing clusters `Narval` equipped with Nvidia A100 GPUs.
- Peer-reviewing articles in journals like American Astronomical Society journal.
- Mentored undergraduate students in research projects, providing guidance on data analysis, software development, and scientific writing.

University of Toronto Jan 2018 — Dec 2024
Teaching Assistant

- Developed homework assignments, including multiple choice questions and Python modules. Assisted with grading Python codes, written lab reports, and term projects.
- Led weekly tutorials and practical lab sessions, providing instructional support and clarifying course concepts. Answered student queries via email and discussion boards.
- Organized and hosted review/help sessions, invigilated midterms/exams and observation nights at the campus observatory.

EDUCATION

Ph.D. in Astronomy & Astrophysics, University of Toronto 2020 — 2024
Dissertation: "Exploring the Milky Way with Deep Learning" with Prof. Jo Bovy

M.Sc. in Astronomy & Astrophysics, University of Toronto 2019 — 2020

H.B.Sc. in Physics & Astronomy, University of Toronto 2014 — 2019

PUBLICATION OVERVIEW

I am an author on **14 refereed papers** that have **2590+** citations (h-index=11). Excluding 2 collaboration papers, there are **12 refereed papers** that have **760+** citations. Details of my ORCID (0000-0002-0036-2752) associated publications can be accessed on Astrophysics Data System (ADS). Some publications are also accepted for talks and posters at conferences like NeurIPS (see this Section).

CONFERENCE

My research has been presented at various international conferences and workshops. Here are some of the highlights:

NeurIPS (Conference on Neural Information Processing Systems) Vancouver, Canada
Datasets and Benchmarks Track December 2024
Collaboration poster on "The Multimodal Universe: Enabling Large-Scale Machine Learning with 70TBs of Astronomical Scientific Data"

ICML (International Conference on Machine Learning) Vienna, Austria
Organized by Workshop on Foundation Models in the Wild July 2024
Poster on "Estimating Probability Densities with Transformer and Denoising"

NeurIPS (Conference on Neural Information Processing Systems) New Orleans, US
Organized by Machine Learning and the Physical Sciences Workshop Dec 2023
Talk on "Towards an Astronomical Foundation Model for Stars"

Debating the Potential of Machine Learning in Astronomical Surveys New York, US
Organized by Flatiron Institute & Institut Astrophysique de Paris Nov 2023
Talk on "Towards an Astronomical Foundation Model for Stars with a Transformer-based Model"

Artificial Intelligence for Astronomy Garching, Germany
Organized by European Southern Observatory (ESO) July 2019
Talk on "Mapping the Milky Way Galaxy with Deep Learning"

SOFTWARE OVERVIEW

Most of my research are open-sourced including codes for publications: `https://github.com/henrysky`. This includes a few software packages used by the community that are well tested using continuous integration with GitHub Actions and well documented with docstrings and user guides, for example:

astroNN
Deep Learning for Astronomers with `Keras`

Galaxy10
A CIFAR10-like galaxy image dataset for educational and research purposes

milkyway_plot
A handy visualization tool ge for plotting face-on and all-sky MilkyWay with `Matplotlib` and `Bokeh`

MyGaiaDB
A data management package to setup local serverless multi-terabytes astronomical databases using `SQLite` and run query locally with `Python`