Henry W. Leung Ph.D.

Astronomy & Astrophysics Researcher at the University of Toronto

🌐 henrysky.github.io   ✉ henryskyleung@gmail.com   henrysky   in Henry Leung

🔤 Bilingual in English & Chinese   </> Python & C   🏴 Canadian & Hong Konger

SUMMARY

I am a recent PhD graduate in Astronomy and a Data Science Institute doctoral fellow at the University of Toronto, applying deep learning methods behind GenAI to build **multi-modal foundation models for science**. My research presented at **NeurIPS** and **ICML**, involved analyzing multi-terabyte, cross-domain datasets of billions of stellar objects. Majority of my code and models being well-tested, well-documented, and open sourced. 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**

- Explore cutting-edge deep learning methods such as Transformers, denoising diffusion probablistic models, and Large Language Models (LLMs) to develop multi-modal foundation models for astronomy. Built and optimized machine learning models using frameworks like `PyTorch` and `TensorFlow` trained on large multi-terabytes datasets with billions of stellar objects, leveraging tools such as `Docker` and `Postgresql` as well as deploying them on national research supercomputing clusters equipped with Nvidia A100 GPUs.
- Delivered multiple spotlight talks and posters at major conferences like *NeurIPS*, *ICML*, and collaborated with community-led initiatives like the Multimodal Universe project. Created and maintained well-documented and thoroughly tested open-source softwares mainly written in `Python`, `C` and `SQL`, contributing both to personal projects and to the wider scientific community.
- Curated a value-added catalogue of deep learning derived stellar parameters and uncertainties, for the Sloan Digital Sky Survey (SDSS) with more than 10% improvement on stellar parameters accurancy to low signal-to-noise data compared to astrophysical driven pipeline.
- Serve as peer reviewer to peer-review articles in journals like American Astronomical Society journal, Astronomy & Astrophysics jornals. Mentored undergraduate students by providing guidance on data analysis, software development and scientific writing.

University of Toronto   Jan 2018 − Dec 2024
Teaching Assistant

- Developed Python assignment modules, grading codes and lab reports. Organized weekly tutorials and observation nights at the campus observatory. Answered students emails and discussion boards.

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 the first/second author on **9 refereed papers** that have **560+** citations. In total, I am an author on **14 refereed papers** that have **2650+** citations (h-index=11). Details of my ORCID (0000-0002-0036-2752) associated publications can be accessed on Astrophysics Data System (ADS).

My research has been presented at international conferences and workshops. Here are some of the highlights (first-author unless noted as part of a collaboration):

NeurIPS (2024)
Collaboration poster on "The Multimodal Universe: Enabling Large-Scale Machine Learning with 100TBs of Astronomical Scientific Data"

ICML (2024)
Poster on "Estimating Probability Densities with Transformer and Denoising Diffusion"

NeurIPS (2023)
Talk on "Towards an Astronomical Foundation Model for Stars"

Debating the Potential of Machine Learning in Astronomical Surveys (2023)
Talk on "Towards an Astronomical Foundation Model for Stars with a Transformer-based Model"

Artificial Intelligence for Astronomy (2019)
Talk on "Mapping the Milky Way Galaxy with Deep Learning"

SOFTWARE OVERVIEW

I am comfortable programming in `Python` and `C` and familiar with tools around high performance computing and `SQL` databases. I am currently learning `Rust` and `C++` by taking initiatives to implement wishlist features in other open source projects written in those languages.

Most of my research are open-sourced including codes for publications are hosted on my Github. 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`