Updated on 2024 May 28

Henry W. Leung

PhD candidate at Department of Astronomy & Astrophysics
University of Toronto – 50 St. George Street, Toronto, Ontario, Canada M5S 3H4

henrysky.github.io henrysky.leung@utoronto.ca henrysky Henry Leung

Bilingual in English & Chinese Python, C Canadian

RESEARCH INTERESTS

My research broadly focused on how to adopt and adapt **deep learning** methodology to analyze big cross-domain cross-survey datasets to help us better understand the formation history and **dynamics of our MilkyWay Galaxy**. I utilize a wide range of machine learning methods in my research from simple supervised models to self-supervised Transformers and diffusion models. I am interested in big questions like how would **foundation models** like "Large Astronomy Models" play a role in data-driven astronomy as well as how **artificial intelligence** would look like in astronomy in the future. Most of my codes and models are well tested, well documented and open sourced to support open science.

EDUCATION

University of Toronto **In Progress**
PhD in Astronomy & Astrophysics *2020-2024*
Thesis advisor: Prof. Jo Bovy

University of Toronto
MSc in Astronomy & Astrophysics *2019-2020*
Thesis advisor: Prof. Jo Bovy & Prof. Abigail Crites

University of Toronto
HBSc in Astronomy & Physics *2014-2019*

MAJOR AWARDS & FELLOWSHIPS

Data Science Institute Doctoral Student Fellowship
University of Toronto *2023-2027*
CAD $75,000

PUBLICATION OVERVIEW

I am an author on **13 refereed papers** that have **2270+** citations (h-index=10). Excluding 2 collaboration papers, there are **11 refereed papers** that have **690+** citations. Details of my ORCID (0000-0002-0036-2752) associated publications can be accessed on Astrophysics Data System (ADS).

TALKS AT CONFERENCE/WORKSHOPS

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

Debating the Potential of Machine Learning in Astronomical Surveys **New York, USA**
Organized by Flatiron Institute & Institut Astrophysique de Paris *Nov 2023*

Talks on "Towards an Astronomical Foundation Model for Stars with a Transformer-based Model"

Asteroseismology in the Era of Surveys from Space and the Ground — **Leuven, Belgium**
Organized by KU Leuven — *July 2022*
Talks on "Stellar Age for Giant Stars with Deep Learning"

Gaia Hike — **Vancouver, Canada**
Organized by The Canadian Institute for Theoretical Astrophysics (CITA) — *June 2022*
Talks on how to access 220 millions multi-terabytes Gaia BP/RP spectra

Stellar Stats Workshop — **Toronto, Canada**
Organized by University of Toronto — *May 2022*
Talks on "Understanding the Milky Way Galaxy with Deep Learning'

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

SOFTWARE

Most of my research are open-sourced including codes for publications: https://github.com/henrysky. This includes a few software packages that are well tested and well documented, for example:

astroNN
Deep Learning for Astronomers with Tensorflow

milkyway_plot
A handy python package to do plotting on a face-on/edge-on/allsky map milkyway with matplotlib and bokeh

Galaxy10
A CIFAR10-like galaxy image dataset

MyGaiaDB
Setup local SQL serverless ESA Gaia / 2MASS / ALLWISE / CATWISE databases and run query locally with python

I have also contributed to several open-source software packages, for example:

galpy
Galactic Dynamics in python
I have implemented an explicit Runge-Kutta method of order 8(5,3) numerical integer DOP-853 in Python and C, as well as improving 2D animation rendering performance and implementing 3D animation using plotly.js

mwdust
Dust maps in the Milky Way
I have implemented necessary Hierarchical Equal Area isoLatitude Pixelation of a sphere (HEALPix) functionality in C for cross-platform compatibility as well as improved out-of-the box user experience

python-fsps
Python bindings to Flexible Stellar Population Synthesis (FSPS) Fortran code
I have fixed various compilation issues such that the code is usable on Windows.

REFEREEING & REVIEW:

Referee for: *American Astronomical Society journal*, *Astrophysics and Space Science journal*

MENTORSHIPS

Yiwei Jiang

Summer Undergraduate Data Science Research with Prof. Jo Bovy *Summer 2024*
Dating stars with contrastive learning

Peter Shi
Summer Undergraduate Data Science Research with Prof. Jo Bovy and Dr. Yingyi Song *Summer 2023*
Mapping the chemical structure of the Milky Way with neural nets and SDSS-V

Rohan Ashar
Undergraduate Mentorship Program *Fall 2022*

OUTREACH

UofT AstroTours Team
Website director and help with monthly public talks by graduate student *2020-2023*

UofT AstroTours
Public talk on "Exploring our Milky Way Galaxy with Deep Learning" *13 January 2023*

PUBLICATIONS

First/Second Author (ordered by date):

[10 cites] 2024, MNRAS.527.1494L / arXiv:2308.10944
Towards an astronomical foundation model for stars with a transformer-based model
Henry W. Leung & Jo Bovy

[14 cites] 2023, MNRAS.522.4577L / arXiv:2302.05479
A variational encoder-decoder approach to precise spectroscopic age estimation for large Galactic surveys
Henry W. Leung, Jo Bovy, J. Ted Mackereth & Andrea Miglio

[28 cites] 2023, MNRAS.519..948L / arXiv:2204.12551
A measurement of the distance to the Galactic centre using the kinematics of bar stars
Henry W. Leung, el al.

[145 cites] 2019, MNRAS.490.4740B / arXiv:1905.11404
Life in the fast lane: a direct view of the dynamics, formation, and evolution of the Milky Way's bar
Jo Bovy, **Henry W. Leung**, el al.

[126 cites] 2019, MNRAS.489.2079L / arXiv:1902.08634
Simultaneous calibration of spectro-photometric distances and the Gaia DR2 parallax zero-point offset with deep learning
Henry W. Leung & Jo Bovy

[156 cites] 2019, MNRAS.483.3255L / arXiv:1808.04428
Deep learning of multi-element abundances from high-resolution spectroscopic data
Henry W. Leung & Jo Bovy

[4 cites] 2017, JAVSO..45...30P / arXiv:1611.03334
Studies of the Long Secondary Periods in Pulsating Red Giants. II. Lower-Luminosity Stars
J. R. Percy & **Henry W. Leung**

Contributing Author (ordered by date):

[3 cites] 2023, MNRAS.526.1997P / arXiv:2306.09319
Decoding the age-chemical structure of the Milky Way disc: an application of copulas and elicitable maps
Aarya A. Patil, Jo Bovy, Sebastian Jaimungal, Neige Frankel, **Henry W. Leung**, el al.

[39 cites] 2022, ApJS..260...32W / arXiv:2108.08860
Chemical Cartography with APOGEE: Mapping Disk Populations with a 2-process Model and Residual Abundances
David H. Weinberg, el al. (includes **Henry W. Leung**)

[582 cites] 2022, ApJS..259...35A / arXiv:2112.02026
The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data
Abdurro'uf, el al. (Collaboration paper; includes **Henry W. Leung**)

[1005 cites] 2020, ApJS..249....3A / arXiv:1912.02905
The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra
Romina Ahumada, el al. (Collaboration paper; includes **Henry W. Leung**)

[11 cites] 2020, MNRAS.494.2268W / arXiv:1910.01646
Searching for solar siblings in APOGEE and Gaia DR2 with N-body simulations
Jeremy J. Webb, Natalie Price-Jones, Jo Bovy, Simon Portegies Zwart, Jason A. S. Hunt, J. Ted Mackereth, **Henry W. Leung**, el al.

[152 cites] 2019, MNRAS.489..176M / arXiv:1901.04502
Dynamical heating across the Milky Way disc using APOGEE and Gaia
J. Ted Mackereth, Jo Bovy, **Henry W. Leung**, el al.

RESEARCH INTERESTS

My research broadly focused on how to adopt and adapt **deep learning** methodology to analyze big cross-domain cross-survey datasets to help us better understand the formation history and **dynamics of our MilkyWay Galaxy**. I utilize a wide range of machine learning methods in my research from simple supervised models to self-supervised Transformers and diffusion models. I am interested in big questions like how would **foundation models** like "Large Astronomy Models" play a role in data-driven astronomy as well as how **artificial intelligence** would look like in astronomy in the future. Most of my codes and models are well tested, well documented and open sourced to support open science.

EDUCATION

University of Toronto **In Progress**
PhD in Astronomy & Astrophysics *2020-2024*
Thesis advisor: Prof. Jo Bovy

University of Toronto
MSc in Astronomy & Astrophysics *2019-2020*
Thesis advisor: Prof. Jo Bovy & Prof. Abigail Crites

University of Toronto
HBSc in Astronomy & Physics *2014-2019*

PUBLICATION OVERVIEW

I am an author on **13 refereed papers** that have **2270+** citations (h-index=10). Excluding 2 collaboration papers, there are **11 refereed papers** that have **690+** citations. Details of my ORCID (0000-0002-0036-2752) associated publications can be accessed on Astrophysics Data System (ADS).

MAJOR AWARDS & FELLOWSHIPS

Data Science Institute Doctoral Student Fellowship
University of Toronto *2023-2027*
CAD $75,000

SOFTWARE

Most of my research are open-sourced including codes for publications: `https://github.com/henrysky`. This includes a few software packages that are well tested and well documented, for example:

astroNN
Deep Learning for Astronomers with Tensorflow

milkyway_plot
A handy python package to do plotting on a face-on/edge-on/allsky map milkyway with matplotlib and bokeh

Galaxy10
A CIFAR10-like galaxy image dataset

MyGaiaDB
Setup local SQL serverless ESA Gaia / 2MASS / ALLWISE / CATWISE databases and run query locally with python

I have also contributed to several open-source software packages, for example:

galpy
Galactic Dynamics in python
I have implemented an explicit Runge-Kutta method of order 8(5,3) numerical integer DOP-853 in Python and C, as well as improving 2D animation rendering performance and implementing 3D animation using plotly.js

mwdust
Dust maps in the Milky Way
I have implemented necessary Hierarchical Equal Area isoLatitude Pixelation of a sphere (HEALPix) functionality in C for cross-platform compatibility as well as improved out-of-the box user experience

python-fsps
Python bindings to Flexible Stellar Population Synthesis (FSPS) Fortran code
I have fixed various compilation issues such that the code is usable on Windows.

PUBLICATIONS

First/Second Author (ordered by date):

[10 cites] 2024, MNRAS.527.1494L / arXiv:2308.10944
Towards an astronomical foundation model for stars with a transformer-based model
Henry W. Leung & Jo Bovy

[14 cites] 2023, MNRAS.522.4577L / arXiv:2302.05479
A variational encoder-decoder approach to precise spectroscopic age estimation for large Galactic surveys
Henry W. Leung, Jo Bovy, J. Ted Mackereth & Andrea Miglio

[28 cites] 2023, MNRAS.519..948L / arXiv:2204.12551
A measurement of the distance to the Galactic centre using the kinematics of bar stars
Henry W. Leung, el al.

[145 cites] 2019, MNRAS.490.4740B / arXiv:1905.11404
Life in the fast lane: a direct view of the dynamics, formation, and evolution of the Milky Way's bar
Jo Bovy, **Henry W. Leung**, el al.

[126 cites] 2019, MNRAS.489.2079L / arXiv:1902.08634
Simultaneous calibration of spectro-photometric distances and the Gaia DR2 parallax zero-point offset with deep learning
Henry W. Leung & Jo Bovy

[156 cites] 2019, MNRAS.483.3255L / arXiv:1808.04428
Deep learning of multi-element abundances from high-resolution spectroscopic data
Henry W. Leung & Jo Bovy

[4 cites] 2017, JAVSO..45...30P / arXiv:1611.03334
Studies of the Long Secondary Periods in Pulsating Red Giants. II. Lower-Luminosity Stars
J. R. Percy & **Henry W. Leung**

Contributing Author (ordered by date):

[3 cites] 2023, MNRAS.526.1997P / arXiv:2306.09319
Decoding the age-chemical structure of the Milky Way disc: an application of copulas and elicitable maps
Aarya A. Patil, Jo Bovy, Sebastian Jaimungal, Neige Frankel, **Henry W. Leung**, el al.

[39 cites] 2022, ApJS..260...32W / arXiv:2108.08860
Chemical Cartography with APOGEE: Mapping Disk Populations with a 2-process Model and Residual Abundances
David H. Weinberg, el al. (includes **Henry W. Leung**)

[582 cites] 2022, ApJS..259...35A / arXiv:2112.02026
The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data
Abdurro'uf, el al. (Collaboration paper; includes **Henry W. Leung**)

[1005 cites] 2020, ApJS..249....3A / arXiv:1912.02905
The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra
Romina Ahumada, el al. (Collaboration paper; includes **Henry W. Leung**)

[11 cites] 2020, MNRAS.494.2268W / arXiv:1910.01646
Searching for solar siblings in APOGEE and Gaia DR2 with N-body simulations
Jeremy J. Webb, Natalie Price-Jones, Jo Bovy, Simon Portegies Zwart, Jason A. S. Hunt, J. Ted Mackereth,

Henry W. Leung, el al.

[152 cites] 2019, MNRAS.489..176M / arXiv:1901.04502
Dynamical heating across the Milky Way disc using APOGEE and Gaia
J. Ted Mackereth, Jo Bovy, **Henry W. Leung**, el al.