: jshen.net
: shenjeff[@]princeton.edu

Jeff Shen

: github.com/al-jshen
: linkedin.com/in/jeff-shen

Research Interests

I research deep neural networks that learn efficiently and generalize robustly through self-supervision and principled, domain-informed architectural biases. I am particularly interested in multimodal foundation models and generative approaches that combine disparate data and transfer across domains.

Education

M.A., Ph.D. — Princeton University 2022 - 2027
Specialization: Machine learning for the physical sciences.
Advisor: Shirley Ho.

H.B.Sc. — University of Toronto 2018 - 2022
Majors: Astrophysics and Statistics.

Experience

AI/ML Research Intern — Netflix May 2026 - Aug 2026
Content Representation Models team.

Researcher — Polymathic AI Oct 2023 - Present
Self-supervised multimodal foundation models for the physical sciences.

Graduate Researcher — Princeton University Sep 2022 - Present

Awards & Honours

2025 - 2026	**GQS PhD Fellowship in Physics** ($100,000 USD) Citadel GQS; awarded only to one PhD student in the country
2024 - 2027	**NSERC Postgraduate Scholarship - Doctoral (PGS-D)** ($113,667 CAD) Natural Sciences and Engineering Research Council of Canada
2022 Jan	**Astrostatistics Student Paper Competition Finalist** ($1,100 USD) Astrostatistics Interest Group of the American Statistical Association
2021 - 2022	**Dean's List Scholar** University of Toronto
2021 May - 2021 Aug	**Undergraduate Summer Research Fellowship** ($8,396 CAD) Canadian Institute for Theoretical Astrophysics, University of Toronto
2020 May - 2020 Aug	**Summer Undergraduate Research Program Award** ($9,500 CAD) Dunlap Institute for Astronomy and Astrophysics, University of Toronto

Publications [arXiv] [Google Scholar] ∗ indicates equal contribution.

First Author

8. **Jeff Shen**, Joshua S. Speagle, Shirley Ho (2026). Homogeneous Stellar Parameters from Heterogeneous Spectra with Deep Learning. Under review at ApJ. arXiv: 2604.25786.

7. **Jeff Shen***, Lindsay M. Smith* (2025). ALICE: An Interpretable Neural Architecture for Generalization in Substitution Ciphers. Preprint. arXiv: 2509.07282. Project page: https://jshen.net/alice.

6. **Jeff Shen**, Francois Lanusse, Liam Parker, et al. (2025). Universal Spectral Tokenization via Self-Supervised Panchromatic Representation Learning. Accepted at NeurIPS Workshop on Machine Learning and the Physical Sciences on 23 Sep 2025. arXiv: 2510.17959.

5. Liam Parker*, Francois Lanusse*, **Jeff Shen***, et al. (2025). AION-1: Omnimodal Foundation Model for Astronomical Sciences. Accepted at NeurIPS 2025 Main Conference on 18 Sep 2025 and as an oral presentation at NeurIPS 2025 AI4Science Workshop on 23 Sep 2025. arXiv: 2510.17960.

4. **Jeff Shen**, Peter Melchior (2023). Multiscale Feature Attribution for Outliers. Accepted at NeurIPS Workshop on Machine Learning and the Physical Sciences on 27 Oct 2023. arXiv: 2310.20012.

3. **Jeff Shen**, Joshua S. Speagle, J. Ted Mackereth, Yuan-Sen Ting, Jo Bovy (2023). Disentangling Stellar Age Estimates from Galactic Chemodynamical Evolution. ApJ 960 84. arXiv: 2305.15634.

2. **Jeff Shen**, Gwendolyn M. Eadie, Norman Murray, Dennis Zaritsky, Joshua S. Speagle, Yuan-Sen Ting, Charlie Conroy, Phillip A. Cargile, Benjamin D. Johnson, Rohan P. Naidu, Jiwon Jesse Han (2022). The Mass of the Milky Way from the H3 Survey. ApJ 925 1. arXiv: 2111.09327.

1. **Jeff Shen**, Allison W. S. Man, Johannes Zabl, Zhi-Yu Zhang, Mikkel Stockmann, Gabriel Brammer, Katherine E. Whitaker, and Johan Richard (2021). Molecular gas in a gravitationally lensed galaxy group at $z = 2.9$. ApJ 917 79. arXiv: 2105.11572.

Major Contributions

2. Rudy Morel, Francesco P. Ramunno, **Jeff Shen**, et al. (2025). Predicting partially observable dynamical systems via diffusion models with a multiscale inference scheme. Accepted at NeurIPS 2025 Main Conference on 18 Sep 2025. arXiv: 2511.19390.

1. The Multimodal Universe Collaboration, including **Jeff Shen** (2024). The Multimodal Universe: Enabling Large-Scale Machine Learning with 100TB of Astronomical Scientific Data. Accepted at NeurIPS 2024 Track on Datasets and Benchmarks on 26 Sep 2024. arXiv: 2412.02527.

Coauthor

7. Siavash Golkar et al., including **Jeff Shen** (2026). MIMIC: A Generative Multimodal Foundation Model for Biomolecules. Preprint. arXiv: 2604.24506.

6. Michael McCabe et al., including **Jeff Shen** (2025). Walrus: A Cross-Domain Foundation Model for Continuum Dynamics. Preprint. arXiv: 2511.15684.

5. Ruben Ohana et al., including **Jeff Shen** (2024). The Well: a Large-Scale Collection of Diverse Physics Simulations for Machine Learning. Accepted at NeurIPS 2024 Track on Datasets and Benchmarks on 26 Sep 2024. arXiv: 2412.00568.

4. Neige Frankel, David W. Hogg, Scott Tremaine, Adrian Price-Whelan, **Jeff Shen** (2024). Iron Snails: non-equilibrium dynamics and spiral abundance patterns. ApJ 987 81. arXiv: 2407.07149.

3. Yan Liang, Peter Melchior, ChangHoon Hanh, **Jeff Shen**, Andy Goulding, Charlotte Ward (2023). Outlier Detection in the DESI Bright Galaxy Survey. ApJL 956 L6. arXiv: 2307.07664.

2. Seery Chen, Deborah M. Lokhorst, **Jeff Shen**, Imad Pasha, Evgeni I. Malakhov, Roberto Abraham, Pieter van Dokkum (2022). The Dragonfly Spectral Line Mapper: design and first light. Proc. SPIE 12182, Ground-based and Airborne Telescopes IX, 121824E. arXiv: 2209.07489.

1. Deborah M. Lokhorst, Seery Chen, Imad Pasha, **Jeff Shen**, Evgeni I. Malakhov, Roberto G. Abraham, Pieter van Dokkum. The pathfinder Dragonfly Spectral Line Mapper: pushing the limits for ultra-low surface brightness spectroscopy. Proc. SPIE 12182, Ground-based and Airborne Telescopes IX, 121821T. arXiv: 2209.07487

Talks

2025 Oct **Schwab Group Meeting, CUNY/Princeton**
Invited talk: "ALICE: An Interpretable Neural Architecture for Generalization in Substitution Ciphers."

2025 Sep **Princeton University Machine Learning Journal Club, Princeton**
Invited talk: "ALICE: An Interpretable Neural Architecture for Generalization in Substitution Ciphers."

2025 Aug **Multimessenger Astronomy in the Era of Foundational AI, Nashville**
Invited talk: "Multimodal Foundation Models for Astrophysics."

2022 Aug **JSM 2022, Washington, D.C.**
Invited talk: "The Mass of the Milky Way from the H3 Survey."

2021 Aug **SDSS 2021 Collaboration Meeting, JHU/Online**
Lightning talk: "Predicting the ages of stars with machine learning."

2021 May **Stellar Stats Workshop, UofT/Online**
Lightning talk: "Estimating the mass distribution of the Milky Way with Bayesian multilevel models."

2021 May **Canadian Astronomical Society (CASCA) AGM, NRC Herzberg/Online**
Lightning talk: "Molecular gas in a gravitationally lensed galaxy group at z = 2.9."

2021 Apr **Vancouver Area Cosmology Meeting, UBC/SFU/TRIUMF/Online**
"Molecular gas in gravitationally lensed galaxies at z = 2.9."

2020 Oct **REQUIEM Galaxies Telecon, Online**
"Molecular gas in gravitationally lensed galaxies at z = 2.9."

Posters

2025 Dec **NeurIPS 2025, San Diego**
"AION-1: Omnimodal Foundation Model for Astronomical Sciences."

2025 Dec **Machine Learning and the Physical Sciences Workshop, NeurIPS 2025, San Diego**
"Universal Spectral Tokenization via Self-Supervised Panchromatic Representation Learning."

2025 Oct **Meta PhD Forum 2025, Menlo Park**
Invited poster: "AION-1: Omnimodal Foundation Model for Astronomical Sciences."

2024 Dec **NeurIPS 2024, Vancouver**
"The Multimodal Universe: Enabling Large-Scale Machine Learning with 100 TBs of Astronomical Scientific Data."

2023 Dec **Machine Learning and the Physical Sciences Workshop, NeurIPS 2023, New Orleans**
"Multiscale feature attribution for outlier detection models."

2023 Jan **241st Meeting of the American Astronomical Society, Seattle**
"Stellar parameters for 220m stars from *Gaia* DR3 BP/RP spectra."

2021 May **Canadian Astronomical Society (CASCA) AGM, NRC Herzberg/Online**
"Molecular gas in a gravitationally lensed galaxy group at z = 2.9."

2020 Aug **Summer Undergraduate Research Program Poster Session, UofT/Online**
"Molecular gas in a gravitationally lensed galaxy group at z = 2.9." Awarded best poster.

Technical Skills

Pytorch, JAX, Rust, Stan, SQL, R, C++, LaTeX, regex, bash, git, Docker, Javascript, React

Teaching

2023	**Teaching Assistant**, Princeton University AST 255: Life in the Universe (astrobiology course for physics, chemistry, and biology students)
2015 - 2018	**Python Workshop Lead**, Sir Winston Churchill Secondary, Vancouver, B.C. Coordinated and delivered workshops to prepare students for national programming contest.

Professional service

Reviewer for NeurIPS 2025, 2026, ICLR 2026, RAS Techniques and Instruments.