John D. Soltis
https://johnsoltis.github.io/ | https://www.linkedin.com/in/john-soltis/

EDUCATION:

Johns Hopkins University — Baltimore, MD
Ph.D in Astronomy and Astrophysics — 2019 - 2025
M.A. in Physics — 2019 - 2023
University of Michigan — Ann Arbor, MI
B.S. in Physics and Mathematical Physics — 2014 - 2018

TECHNICAL SKILLS:

Software: Python, Pytorch, Tensorflow, Scikit-Learn, Numpy, SciPy, bash, LaTeX, Excel
Methods: Bayesian Analysis, Deep Learning, ML Interpretability, Uncertainty Quantification

WORK EXPERIENCE:

LASKER DATA SCIENCE FELLOW
Space Telescope Science Institute — *Baltimore, MD* — *2025 - present*

- Researching machine learning applications to astronomy, with a focus on foundation models.

RESEARCH ASSISTANT
Johns Hopkins University — *Baltimore, MD* — *2019 - 2025*

- Designed, trained, and tested a normalizing flows model with convolutional layers on ∼12,000 simulated 4×32×32 pixel observations to predict an intrinsic astrophysical property. Model predictions were ∼2 times more accurate compared to non-ML method. Model-estimated uncertainties were accurate to within ∼7%. Interpreted model behavior with physically-motivated image decomposition method. Designed the model using Pytorch Python package.

- Designed, trained, and tested a convolutional neural network on ∼3000 3×256×256 pixel observations. Used simulated short duration, realistic observations to infer longer duration, noise-free idealized observations. Constructed a novel loss metric, combining mean absolute error with physically motivated summary statistics, to prioritize accuracy in those statistics. Model predicted observations were more up to four times more accurate in astrophysics-relevant statistics than non-ML method. Designed the model using Tensorflow Python package.

- Cleaned and analyzed newly released satellite data to estimate the distance to a cluster of stars. Identified an initial set of ∼100,000 cluster members and reduced it to a set of ∼67,000 high-data-quality stars. Used the distance to obtain highly-cited independent estimate of the cosmologically important Hubble constant, with comparable precision to existing estimates. Project completed within ∼2 weeks of initial data release. Analysis done in Python.

- Trained ML models on Johns Hopkins University's high performance computing cluster, including using CPU and GPU cores.

PRE-DOCTORAL RESEARCH FELLOW
Flatiron Institute — *New York, NY* — *2023 - 2024*

- Used Flatiron Institute's high performance computing cluster to analyze data from a cosmological simulation with 4096^3 particles. Determined bounds of simulation accuracy, as function of time and space, to assess suitability for future machine learning simulation-based inference.