Nicholas Kern

nkern.astro@gmail.com | nkern.github.io | github.com/nkern | linkedin.com/in/nicholas-kern

About

- Seasoned researcher & data scientist with 9+ years experience building end-to-end machine learning systems, and 3+ years **leading a team of engineers**. PhD in data-intensive astrophysics.
- Extensive experience solving abstract & complex problems with **noisy, unlabeled, real-world data**, including image, time-series, and waveform datasets.
- Have worked through the **full ML lifecycle in a production environment** (packaging, integration, testing).
- *Key achievement*: built and deployed a scalable ML pipeline from scratch for TBs of messy time-series data.
- Published over 50 articles spanning applied ML, physics, remote sensing, signal processing, & statistics.

Skills

Code	Python (pytorch, numpy, scipy, scikit-learn, pandas), Bash, SQL | *Familiar:* C++
ML	Deep learning, optimization (CPU/GPU), compression, computer vision, generative
Technical	CI/CD, parallel & high-performance computing, cloud computing, Git, Docker, OpenCV
Scientific	Signal processing, numerical linear algebra, sensor imaging & calibration, probability & statistics, hypothesis testing, data curation, Monte-Carlo simulation, first-principles modeling, reproducibility, clear written & verbal communication

Work Experience

Massachusetts Institute of Technology — *Cambridge, MA*

Research Fellow — *September 2020 -- present*

- Built a cloud-based ML pipeline from scratch for converting raw, "real-world" data (noisy, messy, contaminated, and sparse) into cleaned, inference-ready feature sets. Led a team that used this pipeline to analyze TBs of radio data, setting new SOTA constraints on long-sought cosmological radio signals.
- Served as technical-lead for algorithms & model validation. Designed custom ML frameworks for signal estimation, improving sensitivity by 10x. Developed tools for benchmarking model performance and failure-modes. Worked cross-functionally, and presented to our executive board to gather feedback.
- Built the first end-to-end, auto-diff pipeline for radio cosmology (Kern 2023, NeurIPS ML for Sciences Workshop). Improved runtime by 50x over standard approaches, and enabled uncertainty quantification for the first time. Experience with high-dimensional optimization and multi-GPU training.
- Managed Python code repositories with production standards [1, 2, 3]. Managed unit testing, continuous integration, model and data monitoring, code documentation, GitHub actions, packaging, and pull requests.
- Developed novel wide-field imaging techniques for radio telescope arrays, and improved sensor calibration quality over existing methods by 10x.
- Mentored 4 PhD students and guided their work, leading to journal publications and conference posters.
- Published in top journals (3 first-author; NeurIPS, ApJ, MNRAS), and invited to talk at conferences (9).

Side Projects

- Completed MIT Data-Centric AI Bootcamp (2024) [notebooks]. Built application for xyz [link]
- Built a custom CNN package in PyTorch for 3D radio image segmentation with U-Nets (2022) [code, blog].

University of California Berkeley — *Berkeley, CA*

Graduate Researcher — *September 2015 – August 2020*

- Built a parallelized, end-to-end data pipeline in Python and shell for extracting weak signals in noisy and high-dimensional time-series data. Awarded UC Berkeley Astrophysics' top 2020 PhD dissertation award.
- Created a real-time data monitoring application for diagnosing sensor malfunctions and generating automated data visualizations. Improved turn-around time from "problem-identified" to "problem-solved".

- Developed signal processing algorithms for modeling outlier events in noisy waveform data. Recovered data that were previously contaminated by sporadic outliers.
- Built a toolkit for using ML to accelerate numerical simulations, improving the runtime of our end-to-end inference pipeline by 100x [code, paper].
- Wrote a key "Parameter Inference" section of a funding proposal that garnered $10M in funds from the NSF.
- Created and taught a summer class on Python programming & data science in astrophysics for undergraduates [link]. Garnered multiple teaching awards.

Education

University of California Berkeley — *Berkeley, CA*
***PhD** in Astrophysics, "Novel Algorithms for Analyzing Next-Gen Radio Telescope Data"* — *August 2020*
***MS** in Astrophysics* — *August 2017*

University of Michigan — *Ann Arbor, MI*
***BS** in Physics, Astrophysics* — *May 2015*

Selected Awards

NASA Hubble Prize Research Fellowship ($300k grant), 2023
- One of the most widely recognized and coveted research fellowships across all of astrophysics.

MIT Pappalardo Prize Research Fellowship ($300k grant), 2020
- One of the leading private research fellowships in physics across the US.

UC Berkeley Uhl Award for Best PhD Dissertation in Astrophysics, 2020
- Given annually to the best PhD dissertation across all of astronomy and astrophysics.

UC Berkeley Teaching Awards (2) for Graduate Teaching, 2017
- Awarded for astrophysics teaching innovation (interactive, software-driven labs).