

Kaili Cao

Research software engineering and data science

Email: kailicao@umich.edu; Website: kailicao.github.io; GitHub: github.com/kailicao; [LinkedIn](#)

Summary

PhD physicist who designs algorithms and builds research software used by teams: a production image-processing package for a NASA mission team, an algorithm 50 to 60 times faster than the one it replaces (in simulations), and a daily service that scores new papers with a large language model.

Skills

Programming: Python (main language; object-oriented, high-performance code with NumPy and Numba), C/C++, Fortran, Shell and Perl, Mathematica; Git and GitHub, continuous integration, documentation and tests

Data and modeling: Large simulated data sets, statistical modeling and parameter inference (Fisher forecasts, validated against Markov chain Monte Carlo), numerical methods, image processing, validation against simulated truth; two large scientific collaborations, 24 papers (10 first-author)

Machine learning and AI: scikit-learn regression in my papers (e.g., a support vector regression emulator; a Theil–Sen calibration); Erdős Institute Data Science Boot Camp (2025); LLMs in a production pipeline (MapHub)

Experience

2026 – present: Schmidt AI in Science Postdoctoral Fellow, Michigan Institute for Data and AI in Society (MIDAS), University of Michigan, Ann Arbor, MI

- Lead [MapHub](#), a team project of the fellowship: designed and built a daily automated service that scores each day's new arXiv preprints for relevance to each participant with a large language model, against a published rubric, and publishes open daily tables and a web viewer (built with AI assistance, Claude Code).

2021 – 2026: PhD Student and Graduate Research Assistant, Physics, The Ohio State University, Columbus, OH

- Led the refactoring of IMCOM, an image-combination code for NASA's Roman Space Telescope, into [PyIMCOM](#), an object-oriented Python package with tests, continuous integration and documentation, cutting computing cost by about an order of magnitude; teammates used it in production for the image coadds of the OpenUniverse2024 simulations.
- Designed and wrote [Effortless](#), a linear image-reconstruction algorithm that is 50 to 60 times faster than IMCOM in simulations of Roman's survey and has a smaller memory footprint and a programming interface general enough for other telescopes.
- Built statistical forecasts of what Roman's weak lensing survey can measure, combining three probes, and showed that outside knowledge of the matter power spectrum improves the results by factors of 1.2 to 3.5.
- Compared stellar models with thousands of stars from a large catalog; helped modernize the Fortran code YREC; supported PyIMCOM users with documentation, onboarding and a tutorial at an internal hack week (2025).

2022 – 2023: Instructor, Online Python Tutorial

- Created and taught a free online Python course: 46 students in four cohorts, mostly first-time programmers from the humanities and social sciences, 28 of them active; gave the lectures, designed the exercises, mentored personal projects, and revised the study guide from students' feedback.

Education

2026: PhD in Physics (MS 2023), The Ohio State University, Columbus, OH

2020: BA in French Language and Literature, Nanjing University, China; 31 credits in Computer Science and Technology (2020 – 2021)