JAMES COLLINS

collijk1@gmail.com
+1 (206) 452-9090
github.com/collijk
linkedin.com/in/jamescollins11
orcid.org/0000-0001-7633-7804

EXPERIENCE

Sr. Software Engineer, Rapid Response and Analytics **2023-Present**
Institute for Health Metrics and Evaluation (IHME) *Seattle, Washington*

- Leads the development of quarterly, high-resolution gridded population product in collaboration with Planet and Microsoft's AI for Good lab.
- Built the technical architecture and pilot implementation for a database of census estimates linked to detailed administrative boundaries.
- Designed and implemented statistical models using PyTorch to estimate population counts from buidling density estimates, census data, and geospatial covariates with a per-run throughput of 250TB.
- Built a comprehensive database of downscaled and bias-corrected climate forecasts from ERA5 historical data and CMIP6 projections.
- Developed or aided in the development of multiple climate-related models of human health and climate-driven migration.
- Collaborated on an AI research project to leverage LLMs to determine cause of death from verbal autopsies.
- Collaborated on an AI research project to use neural network ODE models to do infectious disease prediction in the presence of endogenous response (like mask-wearing) and policy action (like school closures).

Sr. Software Engineer, COVID-19 Response **2020-2023**
Institute for Health Metrics and Evaluation (IHME) *Seattle, Washington*

- Led the development of the COVID-19 infection, death, and hospitalization forecasting model methodology and implementation.
- Designed and implemented the ETL pipelines, provenance systems, and parallelization tools that allowed dozens of researchers and engineers collaborate to produce weekly COVID forecasts in the early pandemic period.
- Worked with research and engineering management to develop communication channels and technical documentation.
- Managed and weekly (2020) and monthly (2021, 2022) production cycles, delivering regular findings to senior managment, government officials, and global health leaders.
- Co-authored half a dozen research papers on the institute's COVID-19 work.

Team Lead, Simulation Science Engineering **2017-2020**
Institute for Health Metrics and Evaluation (IHME) *Seattle, Washington*

- Hired, mentored, and lead a engineering team, developing an agile software management practice from scratch.
- Led the design and development of Vivarium, an open-source individual-based modeling framework for public health simulations.
- Designed and led the implementation of a from-scratch overhaul of the team's ETL pipeline, providing a unified data access API for team members with robust validation, reducing developer time spent managing data by more than 80%.
- Built a CI pipeline for our open-source libraries supporting automated testing, deployment, documentation building, and branch synchronization for multi-library dependencies.
- Designed and implemented a distributed system for robustly parallelizing tens of thousands of simulations across an HPC cluster.
- Scoped, designed, and implemented or led the implementation of more than a dozen simulation projects of population health interventions.
- Developed an operational framework for the research and engineering teams that boosted the overall team output from 2-3 research projects annually to 6-10 projects.
- Co-authored 4 papers and gave several presentations to groups external to our research team.
- Initiated and led weekly training sessions in data structures, algorithms, and software enginnering best practices for junior employees at the institute.

Software Engineer, Simulation Science **2016-2017**
Institute for Health Metrics and Evaluation (IHME) *Seattle, Washington*

- Re-engineered a prototype simulation framework into the production-ready Python modeling suite, Vivarium.
- Developed methodology and implementation for population disaggregation to produce demographically accurate simulation populations.
- Expanded an implementation of Common Random Numbers to support counterfactual analysis in open cohort simulations using a custom hashing implementation.
- Co-led a Python bootcamp for an incoming class of 25 IHME Post-Bachelor Fellows.
- Initiated and co-led a programming book club, fostering a culture of continuous learning and skill development.

EDUCATION

BSc in Physics, Applied Mathematics **2014-2016**
University of Washington *Seattle, Washington*

ASc General **2013-2014**
North Seattle College *Seattle, Washington*

BA in Theater, English (incomplete) **2003-2008**
Lamar University *Beaumont, Texas*

SKILLS

Programming

- Python (expert)
- Fortran, C, C++, Java, C, Prolog, others (novice or dated)
- NumPy, Pandas, SciPy, Numba, Cython, Dask, Xarray, PyTorch, TensorFlow, WandB

Scientific Computing

- HPC and large scale distributed systems (SGE/UGE, Slurm)
- Containerization (Docker and Singularity)
- Data Storage (SQL, Parquet, HDF5, NetCDF)

Math/Machine learning/AI

- Dynamical systems modeling (ODEs, PDEs, Agent-based modeling)
- Statistical modeling (classic supervised and unsupervised models)
- Deep learning (MLPs, CNNs, Recurrent networks, autoencoders, transformers) Some experience with mechanistic interpretability