

Kastan Day

kastanday.com | kastanday@gmail.com | (206) 801-0466 | github.com/kastanday

EDUCATION

University of Illinois at Urbana-Champaign May 2023 MS Computer Science in Applied Machine Learning

- Relevant coursework: ML Engineering, Advanced NLP, Meta-Learning in ML and ML for Bioinformatics
- Recruited 10 direct reports for my thesis project and thrived while managing & inspiring technical excellence.

Swarthmore College Sep 2016 – May 2020 BA Computer Science, BA Cognitive Science (double major) Relevant Coursework:

- Machine Learning, Adaptive Robotics, Data Structures and Algorithms, Software Engineering, SQL Database Design
- Calculus, Linear Algebra, Statistics 1 & 2, Applied Statistics

Phillips Academy Andover high school Andover, MA

SKILLS

Natural and demonstrated team player and leader of technical teams. I'm obsessed, mission driven, and eagerly coachable. I ship great code.

Proficient in Machine learning engineering and data engineering, Python & Linux (7 years daily), C/C++ (2 years).

ML Engineering Deeply engaged in PyTorch and Huggingface ecosystems. Aspiring to JAX/Hydra and Julia.

Distributed ML PyTorch FSDP, DeepSpeed, Ray.io, Huggingface Accelerate, KubeFlow, Colossal-AI.

Data Eng. Spark, Hadoop, Hive, Dask, Zarr, Parquet, PyArrow, NetCDF4, HDF5, Xarray, np & pd. OpenCV & PCL. Distributed ETL pipelines.

Infrastructure Kubernetes, Ray.io, HPC/SLURM, AWS, IBM. Strong experience in Docker & fixing Python envs.

WORK EXPERIENCE

National Center for Supercomputing Applications (NCSA)

May 2023 – Present

Research Software Engineer

- Leading NCSA's effort on scalable ML workloads and developing expertise in HPC Infra, LLM ops. Implemented unique distributed ML training and inference for climate modeling. Scaling scientific discovery: petabytes of data, 800 GPUs, 200-gig networking, 0 bottlenecks.

IBM Research

May 2022 – Sep 2022

Visiting Scholar in Foundation Model Training Infrastructure

Yorktown, NY

- Systematically experimented on the effect of transformer quantization techniques on model-parallel training in Colossal-AI and PyTorch.
- Implemented CodeFlare, an open-source developer tool for foundation model training via Ray & PyTorch on SLURM & Kubernetes.
- **Skills:** Systems-level compute optimization on Kubernetes & HPC. Deep understanding of LLMs and PyTorch parallelism.

National Center for Supercomputing Applications (NCSA)

Sep 2021 – May 2023

Research Assistant in Distributed Machine Learning Training – Funded by the NSF

Sarcos Robotics – The world's most advanced [human exoskeleton](#) ([demo](#))

May 2021 – Sep 2021

Machine Learning R&D Intern – Funded by DARPA & the US Air Force

- Independently developed machine learning vision segmentation models running on edge computers (Nvidia Jetson on-robot) and sensor fusion with 3D point clouds to understand 3D scenes with **pixel-per-pixel segmentation**, at 100+ FPS, for highly accurate robotics.
- **Tech:** PyTorch, custom ML engineering & training, transfer learning, research engineering for 50x speedup by re-writing in TensorRT.

Crescent Health (startup) – Personalized sleep coaching <https://crescent.co/>

Jan 2018 – Dec 2019

Co-Founder

Silicon Valley, CA

My co-founder and I built Crescent to solve the global crisis of sleeplessness, as enabled by personalized and preventative care using wearables data.

- **Infrastructure dev:** I orchestrated cloud infra for ML inference, our proprietary ML & time-series forecasting and production-grade ETL.
- **Communication:** my writing and pitch secured Y-Combinator and Trinity Ventures on Sand Hill Road interviews, and a YC-120 invite.
- **Customer obsessed:** I conducted 350 customer interviews, and personally onboarded each of our first 150 customers. This hands-on approach enabled max-speed iteration by understanding customer pain points and building retention features to see next day improvements.

NASA ([LaRC Autonomy Incubator](#))

May 2017 – Oct 2017 and May 2018 – Oct 2018

Software Engineering Intern

- I was wholly responsible for the computer vision software running a prototype robotic arm to assemble satellites while in Earth's orbit.
- I developed highly parallel camera data filtering and intelligent smoothing in C++. I contributed to two open source projects to increase the performance of my algorithm 10x from the standard PCL implementation. I thrived in a highly collaborative NASA Rapid Research group.
- **Tech:** production quality C++, [OpenCV](#) and [PCL](#), [ROS](#), Linux system admin (Bash/Python scripting), [Intel Realsense 3D](#), Doxygen.

SOFTWARE PROJECTS

Alexa 5th Chatbot Grand Challenge ([announcement here](#))

Oct 2022 – Present (Expected June 2023)

- Co-author of grant that was awarded \$250,000 for Alexa's 5th grand challenge, competing for \$1 million grand prize.

AI Hackathon in Molecular Dynamics – Argonne National Labs

Feb 2022

1st place winner for developing a novel AI solution for molecular structure prediction.

HackMIT – [2nd consecutive win!](#) Best use of machine learning and \$1,500 (Microsoft), best use of natural language processing (Quora)

Sep 2018

HackMIT – [1 of 10 overall winners](#) out of 400+ teams and 1,250 students

Sep 2017

- Best hack for the Social Good and \$1,500 (Baidu) and Most Interesting use of Data (Hudson River Trading)