

James Ngai

✉ jamesdngai@gmail.com | 🌐 james-ngai | 📺 ngai-james

Education

Carnegie Mellon University

Pittsburgh, PA

B.S. Computer Science, Concentration in Machine Learning, GPA: 3.81/4.00

Aug 2022 - May 2026

- **Courses:** 11-785 Deep Learning, 10-701 Intro Machine Learning, 15-213 Computer Systems, 15-210 Parallel Algorithms, 15-259 Probability

Work Experience

National Security Agency

Fort Meade, MD

Machine Learning Engineering Intern

Jun 2024 - Aug 2024

- Develop a **Quantum Machine Learning Model**, trained on **high-performance computing (HPC)**, for deployment at **Google Quantum AI**.
- Author and present a **research report** on **Neural Ordinary Differential Equation Recurrent Neural Networks** to **senior NSA officials**.
- Operate NSA's **Slurm job scheduler** to train models on **\$5 million clusters** comprising over **100 state-of-the-art Nvidia GH200 GPUs**.
- **Technical Skills:** Python, JAX, Slurm, PyTorch, Recurrent Neural Networks (RNNs), Linux, High-Performance Computing (HPC).

Carnegie Mellon University

Pittsburgh, PA

Machine Learning Researcher

May 2023 - Aug 2023

- Developed **Python** code to simulate Agent-Based **Reinforcement Learning** model to find optimum trading prices in **real-world** forest graphs.
- **Improved model convergence** from **53%** to **99%**, surpassing prior PhD students' work and resolving a **significant bottleneck** that had persisted in the research lab for **3 years** prior to my arrival.
- Accelerated the research feedback loop by developing a **systematic testing framework** utilizing **Docker containers**, **Linux scripts**, and **multithreading** on **Google Cloud**, enabling 24/7 model testing.
- **Technical Skills:** Python, NumPy, Matplotlib, Pandas, Ubuntu Linux, Linux tools, Shell Scripting, Git, Google Cloud (GCP).

Exprii

Pittsburgh, PA

Teaching Assistant

Feb 2022 - Sep 2022

- Instructed and mentored a team of **20+ TAs**, resulting in consistently **high performance evaluations**.
- Identified class-specific misconceptions, improving teaching clarity through **in-depth analysis** of student test results using **R**.
- Taught **advanced combinatorics** and **probability** courses for **100+ students** in several different classes for **AMC** and **MATHCOUNTS** contests.
- **Technical Skills:** R, Probability, Combinatorics, Algebra.

Allstate

Seattle, WA

Software Engineering Intern

Jun 2018 - Jun 2022

- Realized **significant cost savings** of over **\$20,000 annually**, by leading the end-to-end development of a comprehensive **data pipeline** to automate Allstate's advertising processes in **Python** and **C++**.
- Achieved a **33%** improvement in the quality of SalesGenie data through the implementation of **cross-validation** techniques using **REST APIs** and **PostgreSQL**, ensuring **robustly accurate advertising data**.
- **Technical Skills:** Python, C++, Matplotlib, Pandas, Scikit-learn, Ubuntu Linux, Linux tools, SQL, PostgreSQL, Git, CI/CD.

University Projects

AI Safety Research: Robust Control on Transformers

Pittsburgh, PA

Center for AI Safety

Feb 2024 - Present

- Engineered an **evaluation pipeline** tailored for **language generation** tasks, used by **20+ researchers**.
- Tested the **robustness** of **Representation Engineering** against current methods, included in evaluation within **published paper**, by comparison with **Direct Preference Optimization (DPO)** and **Supervised Fine-Tuning (SFT)**.
- **Technical Skills:** Python, PyTorch, HuggingFace Transformers, Natural Language Processing (NLP), Linux, Docker.

eBay Machine Learning Competition (Token Classification)

Pittsburgh, PA

eBay

Aug 2023 - Nov 2023

- Ranked **12th** out of 887 teams, **surpassing** primarily **PhD student-led teams**.
- Spearheaded data preprocessing for **10 million rows** of **labeled data**, leveraging **Pandas** and **Hugging Face** to create **training** and **evaluation** pipelines in **Python**.
- Optimized **NLP model** training scripts using **LoRA**, **Quantization**, and **DeepSpeed** to facilitate efficient **multi-GPU** training, resulting in a **55% reduction in memory consumption** while preserving **competitive performance** on **AWS EC2** clusters.
- **Technical Skills:** Python, Pytorch, HuggingFace Transformers, Natural Language Processing (NLP), AWS, Google Cloud (GCP), DeepSpeed.

Skills

Programming Python (Pandas, PyTorch, NumPy, Scikit-learn), R(ggplot2), JAX, C/C++, Java, SQL, HTML/CSS, JavaScript.

Miscellaneous Linux, Docker, GCP, AWS, Shell, Bash, Tableau, Git.