James Ngai

✉ jamesdngai@gmail.com | ⌨ james-ngai | in 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 to 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*
- Solve **unsolved problem** of numerical instability in **Quantum Machine Learning backpropagation** for deployment at **Google Quantum AI**.
- Author and present a **research report** on **Neural Ordinary Differential Equations** to **NSA Director of AI**.
- Engineer **job scheduler** in **C** for **50+ researchers** to train models on **$5 million distributed clusters** comprising **80 Nvidia GH200 GPUs**.
- **Technical Skills**: Python, C, JAX, Slurm, PyTorch, Transformers, LLMs, RNNs, CNNs, Neural ODEs, Apache Spark, SciPy, Diffrax, Linux.

Carnegie Mellon University *Pittsburgh, PA*
Machine Learning Researcher *May 2023 - Aug 2023*
- **Improved model convergence rate** for **reinforcement learning model** from **53%** to **99%**, surpassing the work of prior PhD students and resolving a **critical research bottleneck** that had persisted for **3 years**.
- Developed **Python OOP** code to simulate Agent-Based model to find optimum trading prices in **real-world** forest graphs.
- Accelerated the research feedback loop **30%** 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, Linux tools, Git, Google Cloud (GCP).

Expii *Pittsburgh, PA*
Teaching Assistant *Feb 2022 - Sep 2022*
- Co-taught **combinatorics** and **probability** with **IMO coach Po-Shen Loh** to **100+ students** for AMC and MATHCOUNTS competitions.
- Enhanced student understanding with **15% test score increase** by pinpointing misconceptions with **data analysis** of test results with **R**.
- **Technical Skills:** R, Probability, Combinatorics, Teaching.

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**.
- Reduced **runtime 70%** via **multithreading**, **minimizing network latencies** and increasing computational efficiency.
- **Technical Skills:** Python, C++, Matplotlib, Pandas, Scikit-learn, Linux tools, SQL, PostgreSQL, Git, CI/CD.

University Projects

11-785 Deep Learning Project: Image Restoration with Reinforcement Learning *Pittsburgh, PA*
Carnegie Mellon University *Oct 2023 - Dec 2023*
- Achieved **top 10 paper** in class by peer and faculty review, by improving upon a **novel method** of using **AlphaZero Reinforcement Learning** to restore distorted images.
- Developed a **Monte Carlo Tree Search** algorithm with integrated **Value Network** and **Policy Network** using **CNNs** in **PyTorch**.
- **Technical Skills:** Python, PyTorch, Computer Vision, Convolutional Neural Network (CNN), Matplotlib, Linux.

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**.
- Refactored **Microsoft Research's DeepSpeed source code** in **Python** and **C++** to integrate **LoRA** support for **multi-GPU** training, enabling **20+ billion parameter** models to be trained on the **constrained computational resources** of **AWS EC2 clusters**.
- Spearheaded **data preprocessing** for **10 million rows** of **labeled data**, leveraging **Pandas** and **Hugging Face** to create **training** and **evaluation** pipelines in **Python**.
- **Technical Skills:** Python, C++, Pytorch, HuggingFace Transformers, Natural Language Processing (NLP), AWS, Google Cloud (GCP), DeepSpeed.

Skills

Programming Python, JAX, C/C++, Julia, Java, SQL, R, HTML/CSS, JavaScript.
Miscellaneous Pandas, PyTorch, NumPy, Scikit-learn, Linux, Docker, Slurm, GCP, AWS, Shell, Bash, Tableau, Git.