James Park

📞 +1(647)-244-0322 — ✉ jp7dec23@gmail.com — 🔗 LinkedIn — 🐙 Github

Skills

Languages	Python	**Tech Stack**	AWS, Docker, Databricks, Git/Github Actions
ML Libraries	PyTorch, JAx, sklearn, SciPy, NumPy, Pandas, Ray, MLFlow, Holoview, Composer	**Environment**	Linux, Unix

Experience

Xanadu.ai

Senior Machine Learning Scientist   *10/2024 - Present*

- 100x speed up of Xanadu's hardware analysis pipeline by converting the underlying analysis model into it's algebraic form using SymPy and rewriting the L-BFGS optimization step from PyTorch to JAx.

Machine Learning Scientist   *07/2023 - 10/2024*

- Led the development of a novel 1D CNN Transformer Model along with an end-to-end optimization workflow to quantify the performance quality of over thousands of photonic chips at Xanadu. Demonstrated a 25% increase in precision for the loss measurement, resulting in a **recovery of $5MM worth of hardware from previous unmeasurable chips.**
- Deployed Ray clusters on AWS to scale up simulations as an HPC alternative to the Niagara Supercomputer cluster and visualized the results in real-time using Databricks.
- MLOps development using AWS, Docker, Databricks, MLFlow and Github Actions. Followed best practices such code review, integration & unit testing, and ensuring backward compatibility.

Hallow

Data Scientist (Part Time Contract)   *04/2023 - 12/2023*

- Led the development of a SVD recommendation model targeting over 300,000 new users on trial demonstrating a lift in user trial subscription by 10%, **leading to an increase in $500,000 ARR**.
- Researched and proposed the use case to upper management to utilize state of the art audio AI models to improve the processing efficiencies for audio files.

Toronto Dominion Bank

Applied Machine Learning Scientist   *05/2022 - 07/2023*

- Using PyTorch, developed a Multitask classification model used to label missing privacy data across different lines of business at TD. Projected to **save approximately $2MM in annual costs** from manual labeling.
- Developed an encoder-decoder Transformer model that uses product key-codes to predict missing product names in TD's item database, saving over **$500k from manual labelling costs annually**.
- Tech lead for an internal research team to develop an in-house a nano-GPT model for internal document parsing.

Queen's University

Research Scientist   *05/2020 - 05/2022*

- Developed a compression optimization pipeline capable of efficiently compressing images taken in space up to a factor of 500%. This pipeline was deployed on the NASA BLAST-TNG space telescope and estimated to **save over $30MM USD annually** from data down-link costs.
- Created PyDCF, a statistical model used to measure the magnetic field strengths in galaxies and stars. Improved the original DCF method from 50% → 77% accuracy, resulting in 3 scientific publications.
- Developed various image ETL processing scripts using Fourier transforms and Gaussian filters, resulting in a 30 % recovery of unusable noisy images and saving over 30 hours/week in manual work for other scientists.

Education

Queen's University

Masters of Science in Theoretical Physics   *2022*

University of Toronto

Bachelor of Science in Physics   *2020*