

Aton Kamanda

<https://atonkamanda.github.io/> | atonkamanda@hotmail.com | <https://github.com/atonkamanda> | 438 543-4133

EDUCATION

University of Montreal - Mila

Master of artificial intelligence, 3.95/4.3 GPA

Montreal

Sept. 2021 – Aug. 2023

EXPERIENCE

Machine learning intern

SkalUP

June 2021 – September 2021

Belgium, Namur

- Developed a full-stack web application for code search using Flask and Pytorch.
- The application was designed as a search engine using CodeBERT to retrieve code snippets based on natural language queries, rank them by cosine similarity, and direct the user to the specific GitHub repo related to the snippet. This was the state-of-the-art method at the time before Codex/GPT-4.

Research student in deep learning

GEODES - Software engineering lab

September 2021 – Aug. 2023

Canada, Montreal

- I have been awarded a NSERC grant to research and develop new methods in deep learning.
- Poster and oral presentation of my master thesis on dual process theory at MAIN 2022.

Teacher assistant for a graduate robot learning class

Mila - Montreal institute for learning algorithms

January 2023 – May 2023

Canada, Montreal

- Course focused on state-of-the-art research and composed mainly of PhD students, I have been in charge of creating entirely new assignments with recent research papers, writing automated tests on Gradescope, grading students, and helping students in their research contributions for the final project. More info on the course website.

Machine learning specialist

Caisses Desjardins

April 2023 – August 2023

Canada, Montreal

- Teaching teams with non-technical background high-level reinforcement learning, generative modeling, self-supervised learning, and graph neural networks with an emphasis on finance application.

Machine learning research engineer

VMware

August 2023 – Present

Canada, Montreal

- IVADO partnership with VMware to leverage recent research findings to enhance the large language model (LLM) utilized internally, aimed at bolstering the efficiency of the software engineers for production.
- Some noteworthy improvements include a reduction in inference time, achieving up to 11x increase in inference throughput with performance drops of only 2%, prompt engineering (e.g chain of thought/chain of code), and the implementation of retrieval-augmented generation using CodeLlama, Langchain, and ChromaDB.
- Some of these enhancements were implemented soon after the papers got released and had little to no documentation/code supporting them.

PROJECTS

Dreamer reimplementation | Pytorch

January 2022 - May 2022

- * Reimplementation of the paper Dream to Control: Learning Behaviors by Latent Imagination in Pytorch.
- * We managed to achieve the same result as the base tensorflow implementation and our main codebase has been reused for the paper Stochastic-Marginal-Actor-Critic accepted at ICLR 2023.

VICreg constrained optimization | Pytorch

November 2023 - February 2024

- * Collaborating with Jose Gallego-Posada and Lucas Maes on a paper on using VICreg with constrained optimization for ICML 2024 .
- * VICreg is a compute-intensive architecture that requires proper parallelization to learn efficiently, Meta gave us access to 32 V100 GPUs to run the experiments, and I manage the running of those experiments with Slurm.

TECHNICAL STRENGTHS

Languages : Python, Julia, C/C++, R, SQL

Developer Tools : Docker, Kubernetes

Data: Spark, Hadoop, Pandas, AWS, Google cloud platform, Azure

Machine learning: Pytorch, Jax, TensorFlow, MLFlow, LangChain, ChromaDB, NumPy, Gym, Mujoco