Shawn Xu
431-788-8226 | x57xu@uwaterloo.ca | shawnxu0407.github.io |

Work Experience

Data Science & Data Analyst — Health intelligence Jan 2025 – Apr 2025
Region of Peel

- Create data processing pipeline, perform model selection and validation on public health premises prediction
- Delivered actionable business insights based on model results, enabling data-driven decision making for public health strategies.

Machine Learning Intern Sept 2024 – Dec 2024
Environment and Climate Change Canada

- Assist researchers in a scientific machine learning project, implement complex algorithms in Pytorch
- Implement Transformer Encoder-Decoder Model with Neural ODE Processor for Weather Forecasting train and test model using pytorch-Lightning
- Tested and debugged model training using memorization tests and unit tests to ensure robustness

Machine Learning Intern May 2023 – Aug 2023
Institut Pasteur

- Developed and implemented K-means clustering algorithms to segment high- and low-fluorescence regions for detection of ridge-like spatial patterns in time-series image data.
- Designed and applied piecewise linear regression models to approximate ridge-like patterns and quantitatively measure velocity changes over time.

Education

University of Waterloo Waterloo, Ontario
Master of Mathematics, Applied Mathematics *Sept. 2023 –*

University of Manitoba Winnipeg, Manitoba
Bachelor of Science, Applied Math & Statistics *Sept. 2019 – May 2023*

Coursework and Achievements:
Deep Learning, Statistical Learning, Reinforcement Learning, Probability, Numerical Analysis, Natural Language Processing
Philosophia Mathematica Prize in Applied Mathematics (2022); Award for Highest Standing in Science Faculty (2023)

Technical Skills

Languages & Tools: Python, Pytorch, PySpark, Generative AI, R, Tableau, Java, Docker, ML model deployment, Matlab, MySQL, Data Pipeline, AWS Sagemaker; S3; lambda; ECR

Project Experience

Full Stack Image-to-Text model for Handwriting Text Recognition [GitHub Link] Jan 2025 - April 2025

- Data argumentation and data annotation using Label-Studio, design the data loading pipeline
- Implement cnn-Transformer multimodal model, fine tune with LORA using argumented dataset on Weight & Bias intregrated with Torch-Lightning using AWS sagemaker for cloud training
- Test local deployment using docker with building web UI using gradio and cloud deployment using AWS lambda

Clash Royale Game AI agent with Reinforcement Learning [Blog Link] April 2025 - Aug 2025

- Apply visual fusion (YOLOv8, cnocr, resnet) for offline dataset feature extraction with data pipeline automation
- Train a decision transformer model for policy making through AWS sagemaker
- Deploy visual fusion and policy network on the android phone to play real-time game

Sensitive Weights Quantization for Large Language Model Jan 2024 – April 2024

- Fine-Tune, validating, Hyper-parameter optimization, average gradient and customized loss function on Roberta
- Use sensitive weighted k-means with approximated Hessian to compress the weights of Roberta
- Use Sparse and Dense separation on weights to enhance the weighted k-means
- Obtain 90% reduction in storage and 1% reduction in prediction accuracy through validation