

Education

M.sc. Computer Science - Machine Learning

l'Université de Montréal & Mila

- Focused Studies: Deep Learning, Machine Learning, Data Science, Data Structures & Algorithms Analysis
Current GPA: 3.9/4.3

B.Eng. Mechatronics

l'Université du Québec à Trois-Rivières

- Focused Studies: Finite Element Analysis (FEA), Automation & Programming, Circuit Design, Microcontrollers, Advanced PLC Techniques
GPA: 3.8/4.3

Work Experience

Data Science Specialist Intern

May 2023 – March 2024

CIUSSS du Centre-Sud-de-l'Île-de-Montréal

Montreal, Canada

- Developed and deployed a computer vision-based diagnostic aid platform to assist radiologists, utilizing PyTorch, MMDetection, and YOLOv5&v8 toolboxes.
- Implemented the Barlow-twins CXR self-supervised learning strategy to address the challenge of insufficient labeled data, achieving a 3% improvement in mAP on the test set and publishing a related paper.
- Designed and developed a novel AUC-ROC metric to more accurately assess detector performance, which has been adopted by the research team.
- Model training acceleration through distributed training techniques on Linux servers.
- Leverage Azure DevOps for version control, continuous integration, and work management to enhance team collaboration efficiency.

Automation Engineering Intern

May 2021 – September 2021

HYVA Machinery Co., Ltd.

Yangzhou, China

- Maintained and optimized an automated production line, focusing on optimizing its performance. Included programming tasks for materials handling transfer robots.
- Participated in the design and construction of automatic welding production lines, handled programming and debugging responsibilities for new industrial robots

Project Experience

RAG Local Assistant Chatbot

Spring 2024

Developing a Retrieval-Augmented Generation (RAG) Local Assistant Chatbot using LLaMA2-7B, GEMMA-2B, LangChain, and Hugging Face Transformers for secure, offline functionality on personal computers. Leverage Steamlit to create a user-friendly interface for the chatbot, package it with Docker for easy deployment and distribution.

Ongoing projects. Continuous analysis to optimize response accuracy and efficiency using a combination of NLP benchmarks and custom evaluation metrics.

LLM-based Text Summarization Generation

Fall 2023

Explore large language model-based, automated text summarization techniques using CNN/Daily Mail datasets to fine-tune GPT-2, T5, and GPT-3.5 models.

Rigorous model performance evaluation and analysis through ROUGE and BERT scores and human feedback. Created a technical report explaining the importance of various evaluation methods and provided insights to enhance the summary model.

NHL Shot Analysis & Predictive Model

Fall 2022

Developed a goal-scoring prediction system from scratch, encompassing data collection, preprocessing, feature engineering, and model training and deployment.

Collected data using APIs, built robust pipelines to retrieval and preprocess shot datasets.
Analyzed shot and goal distributions through data visualization, deriving novel insights and feature ideas.
Built a demo app using Flask and Streamlit to showcase findings and containerized it with Docker for seamless deployment.

Movie Review Sentiment Analysis Text

Fall 2022

Trained robust textual sentiment classification models using over one million textual data.
Collaborated within the team. Applied Natural Language Processing (NLP) techniques combined with machine learning and deep learning algorithms to systematically preprocess and embedding data.
Performed careful hyperparameter tuning to optimize model performance and achieve 91% accuracy.

Recyclable garbage Recognition Project

Fall 2020

Developed models to help sorting centers categorize various collected materials.
Implemented data augmentation techniques to augment the size of limited datasets. Trained a Convolutional Neural Network (ResNet-50 architecture) to classifier multiple types of recyclable waste. Got more than 95% accuracy on the test set.

Publication

Sheng, H., Ma, L., Samson, J.-F., & Liu, D. (2024). BarlowTwins-CXR: Enhancing Chest X-Ray abnormality localization in heterogeneous data with cross-domain self-supervised learning. arXiv preprint arXiv:2402.06499.

Technical skills

Programming Language	Python (comfort), Java, C
ML & DL Tools	Pytorch, Keras, Scikit-learn, MMdetection, transformers, Langchain, Weight & Bias, etc.
Data science Tools	Pandas, NumPy, Matplotlib, Seaborn, SQL, NoSQL
Other	Linux, Git, Docker, Matlab, Simulink, SolidWorks, AutoCAD, etc.

Personal skills

Langues	English, French, Mandarin.
Soft skills	Quick learning, team player, time management