

## EDUCATION

### Masters in Computer Science (ML specialization), **MILA Institute**/Université de Montréal

Fall 2025

Relevant coursework: Representation Learning, Probabilistic Graphical Models,  
Data Science, Fundamentals of Machine Learning

### Bachelor of Computer Science, Université de Montréal

Spring 2023

Semester abroad (ERASMUS+), Politecnico di Milano

02/2022-07/2022

## EXPERIENCE

### MACHINE LEARNING DEVELOPER

01/2025 – present

Montreal, Canada

#### • Synthetic Tabular Data Generation for Imbalanced Datasets

- Designed of a Transformer-based VAE for conditional synthetic tabular data generation for scarce and imbalanced real data, improving utility by 22% and privacy preservation by 50%.
- Scaled from 1 to 4 models training per hour using Hydra's distributed jobs on slurm framework infrastructure.
- Designed a modeling framework enabling extendable plug-and-play experimentation with diverse architectures (Transformers, VAEs, Diffusion, GNNs).
- Built training/evaluation framework scaling from 1 to 60+ combinations dataset/model.

#### • Representative Multilingual Synthetic CV Generation

- Used conditional probability sampling to generate 30K Quebec labor force profiles as context for LLM-based CV generation.
- Cost efficient prompt optimization using DSPy and human evaluation results for an LLM-as-Judge evaluation system.

#### • Recommendation System

- Built a knowledge graph combining occupation–skill and training–skill relationships, enabling shortest-path inference and skill-gap analysis for career recommendation.

#### • Numeria: First line for client requirements analysis for potential project collaborations and business development

- Participated in the extraction of business and scientific client requirements used to guide and propose tailored solutions within their budget.

Tools used: Python, PyTorch, SDV, MLflow, Hydra, Azure OpenAI, GCP, AWS, DSPy, SLURM

### PROGRAMMER ANALYST

09/2022 – 08/2024

Varennes, Canada

CanmetENERGY (RETScreen Division)

- RETScreen: energy management software used to evaluate clean energy projects and monitor performance.
- Designed automation tools for performance reporting and data visualization, improving analytical workflows across modules.
- Diagnosed and resolved software issues reported by QA teams, ensuring stable and reliable releases for international users.

Tools used: C#, XAML, Azure DevOps

### RESEARCH ASSISTANT FOR MULTIDISCIPLINARY PROJECT (DR. SHIRIN A. ENGER'S LAB)

05/2021 – 12/2021

Montreal, Canada

Lady Davis Institute/Jewish General Hospital

- Objective: 3D catheter reconstruction for Brachytherapy patients to improve treatment planning and accuracy.
- Implemented a 2D U-Net model, processing medical images and using a custom data generator for efficient memory usage.
- Helped organize the inaugural season of McMedHacks' workshops and virtual hackathon.

Tools used: Python, DICOM files, TensorFlow

## PROJECTS

### SELF-ALIGNING VLMs WITH A FOCUS ON IMAGE MODALITY (click here to view full paper)

01/2024 – 04/2024

- Improved the alignment between visual and textual components with synthetic data, increasing CLIP score by 13%.
- Improved object localization task accuracy by over 10% after incorporating our synthetic dataset.

Tools used: Python, CLIP, LoRA, Stable Diffusion, PyTorch

### NHL SHOT ANALYSIS (click here to view associated blog)

09/2023 – 12/2023

- Built a real-time sports analytics pipeline fetching live data from NHL APIs and performing feature engineering.
- Created an interactive dashboard using Streamlit for real-time shot prediction visualization during live games.

Tools used: Python, TensorFlow, Comet ML, NHL API, Streamlit, Flask, Docker

### ENHANCING CYCLENET FOR TIME SERIES FORECASTING (click here to view full paper)( click here for poster) 10/2024 – 12/2024

- Improved mean MAE by 4-12% by integrating a Time-GNN backbone to model spatio-temporal feature dependencies.

## PUBLICATIONS & AWARDS

- CTTVAE: Latent Space Structuring for Conditional Tabular Data Generation on Imbalanced Datasets ([paper here](#))

Feb 3rd 2026

(coming soon)

- Creation of first multilingual Quebec-representative synthetic CVs

Fall 2020

- DIRO (Département d'Informatique en Recherche Opérationnelle) Scholarship of Excellence

08/2021

- Winner of the Lady Davis Institute Undergraduate Research Day