

## ÉDUCATION

- Maîtrise en informatique (spécialisation en apprentissage automatique), Institut MILA / Université de Montréal** 2025  
Cours pertinents : Apprentissage de représentations, Modèles graphiques probabilistes, Science des données, Fondements de l'apprentissage automatique
- Baccalauréat en informatique, Université de Montréal** 2023  
**Session d'études à l'étranger (ERASMUS+), Politecnico di Milano** 02/2022 – 07/2022

## EXPÉRIENCE PROFESSIONNELLE

**DÉVELOPPEUR ML** 01/2025 — présent  
Centre de Recherche Informatique de Montréal (CRIM) Montréal, Canada

- **Génération de données tabulaires synthétiques pour jeux de données déséquilibrés**
    - Conception d'un VAE basé sur des Transformers pour la génération conditionnelle de données tabulaires synthétiques à partir de données réelles rares et déséquilibrées, **améliorant l'utilité de 22% et la préservation de la vie privée de 50%**.
    - Passage à l'échelle de 1 à 4 entraînements de modèles par heure via des jobs distribués Hydra sur une infrastructure SLURM.
    - Conception d'un cadre de modélisation extensible permettant des expérimentations modulaires (« plug-and-play ») avec diverses architectures (Transformers, VAE, Diffusion, GNN).
    - Développement d'un cadre d'entraînement et d'évaluation passant de **1 à plus de 60** combinaisons jeu de données / modèle.
  - **Génération de CV synthétiques multilingues représentatifs du Québec**
    - Échantillonnage probabiliste conditionnel créant 30k profils de travailleurs servant de contexte générant des CV via LLM.
    - Optimisation de prompts à l'aide de DSPy et de résultats d'évaluations humaines pour un système d'évaluation LLM-as-Judge.
  - **Système de recommandation**
    - Construction d'un graphe de connaissances combinant relations métier-compétence et formation-compétence, permettant l'inférence par plus court chemin et l'analyse des écarts de compétences pour l'orientation de carrière.
  - **Numeria : première ligne d'analyse des besoins clients pour des collaborations potentielles et le développement des affaires**
    - Participation à l'extraction des besoins d'affaires et scientifiques des clients afin de guider et proposer des solutions sur mesure respectant leurs contraintes budgétaires.
- Outils utilisés : Python, PyTorch, SDV, MLflow, Hydra, Azure OpenAI, GCP, AWS, DSPy, SLURM

**ANALYSTE-PROGRAMMEUR (DIVISION RETSCREEN)** 09/2022 — 08/2024  
CanmetÉNERGIE Varennes, Canada

- RETScreen : logiciel de gestion de l'énergie utilisé pour évaluer des projets d'énergie propre et en assurer le suivi de performance.
  - Conçu des outils automatisant le reporting performance/visualisation des données, améliorant les flux d'analyse entre modules.
  - Résolu les problèmes signalés par l'équipe assurance qualité, assurant des versions stables, fiables pour les utilisateurs.
- Tools used: C#, XAML, Azure DevOps

**ASSISTANT DE RECHERCHE POUR UN PROJET MULTIDISCIPLINAIRE (SUPERVISÉ PAR DRE SHIRIN A. ENGER)** 05/2021 – 12/2021  
Institut Lady Davis / Hôpital général juif Montréal, Canada

- Objectif : Développer un pipeline pour la reconstruction 3D de cathéters et optimiser les traitements de curiethérapie.
  - Implémentation d'un U-Net 2D, traitant les données médicales avec un générateur de données optimisant la mémoire.
  - Aidé l'organisation de la saison inaugurale des McMedHacks workshop et hackathon virtuel.
- Outils utilisés : Python, fichiers DICOM, TensorFlow

## PROJETS

**ALIGNEMENT AUTO-ADAPTATIF DES VLMs AVEC UN FOCUS SUR LA MODALITÉ VISUELLE (rapport complet ici)** 01/2024 – 04/2024

- Amélioré l'alignement des composantes visuelles et textuelles avec des données synthétiques, **hausse de 13% du score CLIP**.
- Amélioration **de plus de 10%** des de la précision de localisation, après intégration des données synthétiques.

Outils utilisés : Python, CLIP, LoRA, Stable Diffusion, PyTorch

**ANALYSE DES TIRS DE LA LNH (blogue complet ici)** 09/2023 – 12/2023

- Développement d'un live game client pour chercher les tirs de la LNH, effectuant l'ingénierie des features.
- Création d'un tableau de bord interactif avec Streamlit pour visualiser en temps réel les prédictions de tirs pendant les matchs.

Outils utilisés : Python, TensorFlow, Comet ML, NHL API, Streamlit, Flask, Docker

**AMÉLIORATION DE CYCLENET POUR LA PRÉDICTION DE SÉRIES TEMPORELLES (rapport complet)( poster)** 10/2024 – 12/2024

- **Amélioration du MAE de 4-12%** en intégrant Time-GNN pour modéliser les dépendances spatio-temporelles des features.

## PUBLICATIONS ET PRIX

- CTTVAE: Latent Space Structuring for Conditional Generation on Imbalanced Datasets (article complet ici) 3 février 2026
- Création du premier dataset public de CVs synthétiques multilingues représentatif du Québec (rédaction en cours)
- Bourse d'excellence du DIRO (Département d'Informatique en Recherche Opérationnelle) Automne 2020
- Gagnant de la journée de recherche virtuelle du Lady Davis Institute pour étudiants 08/2021