

EDUCATION AND CERTIFICATIONS

M.Sc. in Artificial Intelligence, Université de Montréal + MILA + CRCHUM 2023 - Present
*Thesis Masters with a focus on **Concept-based Representation Learning for LLMs***
Co-Supervisors: Prof. Bang Liu and Prof. Quoc Dinh Nguyen
Founder and President of the UdeM AI club
GPA: 4.1/4.3

Bachelor in Computer Science and Mathematics, Université de Montréal 2020 - 2023
Specialization in Data Science

SKILLS

Research Interest	Interpretability, Alignment, AI Safety, Explainability, AI4Science
Tools and Languages	Python, PyTorch, HuggingFace, R, TensorFlow, SQL, SAS
Communication	French, English, Mandarin (fluent)

SCHOLARSHIPS AND AWARDS

Collective Engagement Prize , Université de Montréal	2023
Excellence Scholarship , Ministère de l'éducation Supérieure	2022
DIRO Scholarship , Université de Montréal	2021

PUBLICATIONS AND RESEARCH WORKS

Concept-based Activation Steering for Conditional Molecular Generation with LLMs In Review @ ICLR 2026

- Research that aims to improve generation of larger molecules through the use of **foundation models**
- Leverage **Concept-based learning** to improve alignment of properties for **conditional** and **controllable generation**
- Leverage insights from the success of small molecules optimization for larger molecules and for improved properties prediction
- Accepted at NeurIPS AI4Mat Workshop

Concept-based Calibration of LLMs in Healthcare

- Research that aims to better understand **model's uncertainty** through their internal activations
- Leverage insights from **concept-based interpretability** to understand LLM's sensitivity and feature representation space

SHADES: Towards a Multilingual Assessment of Stereotypes in Large Language Models Published @ NAACL 2025

- Assisted in developing a dataset for evaluating stereotypes present in LLMs
- M. Mitchell, ...**J. Qin**...Aurélie Névél, Zeerak Talat et al. "SHADES: Towards a Multilingual Assessment of Stereotypes in Large Language Models", NAACL 2025

Enhancing Healthcare LLM Trust with Atypical Presentations Recalibration Published @ EMNLP 2024

- Developed a recalibration method that leverages **atypical presentations** for Medical QA datasets
- **J. Qin**, B. Liu and Q. Nguyen "Enhancing Healthcare LLM Trust with Atypical Presentations Recalibration", EMNLP 2024 Findings arXiv:2409.03225

Scaling Affordances in Deep Reinforcement Learning

Presented @ ICML 2024

- Leverage the concept of affordance for better action selection in large discrete spaces
- Showed outperformance in system recommender tasks
- L. Cherif, E. Meriaux*, **J. Qin***, V. Patel, M. Klissarov, D. Precup, and K. Khetarpal "Leveraging Affordances for Reinforcement Learning in Large Discrete Action Spaces," In Women in Machine Learning (WiML) Symposium at ICML 2024

Conferences Reviewer

- Reviewer for ICML 2025 NewInML Workshop
- Reviewer for NeurIPS Socially Responsible Language Modelling Research 2024 Workshop

TEACHING EXPERIENCE

Graduate Teaching Assistant (Representation Learning and Machine Learning) 2024-2025
Université de Montréal - Mila MTL, QC

- Assisted teaching for multiple graduate level courses such as Representation Learning, Machine Learning and Advanced Project in Machine Learning
- Helped design theoretical and practical assignment questions and grading
- Supported graduate students with office hours and tutorials