Jean Charle Yaacoub

j.yaacoub@mail.utoronto.ca | (343) 989-1792
linkedin.com/in/jean-yaacoub | github.com/jyaacoub

EDUCATION

Master of Science in Applied Computing (MScAC) *Sep. 2022 – Dec. 2023*
University of Toronto
Artificial Intelligence Concentration
Courses: CSC2559 Trustworthy ML, CSC2552 Topics in Computational Social Science, CSC2231 Visual and Mobile Computing Systems, and CSC2545 Advanced Topics in ML – Causal Learning
cGPA: **3.85/4.0**

Bachelor of Computing (Honors) *Sep. 2018 – June 2022*
Queen's University (Kingston, ON)
Artificial Intelligence Specialization
cGPA: **4.17/4.3**

WORK EXPERIENCE

Princess Margaret Cancer Center - UHN *Jan. 2024 – Present*
ML Research Analyst (full-time)
Continued research with GNNs and **large language models** for drug discovery, **mentored rotation students** and engaged in project management for collaborative works.

Princess Margaret Cancer Center - UHN *May 2023 – Dec. 2023*
Co-op Master's Student (intern)
- Worked on project to improve precision oncology using GNNs (see MutDTA project)
- Effectively communicated ideas clearly through multiple presentations in and outside the lab, including final poster presentation for ARIA. Currently working on a paper for publication.

Vancouver Prostate Centre - UBC *Aug. 2020 – Dec. 2021*
Undergraduate Academic Assistant (part and full-time)
- Helped improve the performance of Deep Docking (DD) which was designed to accelerate drug discovery utilizing AI and physical docking programs like AutoDock Vina. Optimized performance of code to run up to **3x faster** and improved model accuracy by **5%**.
- Co-lead in the design and development of a GUI application that made DD **more accessible** to lab members and other researchers.
- Wrote and reviewed papers for **publication**.

PROJECTS

MScAC Thesis – MutDTA | PyTorch Geometric, Ray[Tune], Graph Networks *May 2023 – Dec. 2023*
GNNs with Protein Dynamics for Enhanced Drug Targeting – github.com/jyaacoub/MutDTA
- Researched, designed, and iterated on models under **limited resources** requiring creative solutions to overcome with distributed **multi-node computing** and leveraging pretrained ESM-2 **foundation protein language models**.

CSC2231 – Visual and Mobile Computing Project | TensorFlow-Federated, FLower, Computer Vision (CV) *Winter 2023*
Federated Learning with Vision Transformers – github.com/jyaacoub/FL-ViT
- Researched the performance of novel ViT models under challenging **federated learning** environments for private **distributed** learning with **non-IID** conditions. Found that distilled ViTs were up to **2x faster** in training with less memory consumption.
- The attention mechanisms employed in the ViT architecture were found to be more effective in handling the challenges of non-IID data.

CSC2559 –Trustworthy ML Project | HuggingFace, Natural Language Processing (NLP) *Fall 2022*
Cross-Domain Attacks in NLP – github.com/jyaacoub/ Cross-Domain-Attacks-NLP
- Researched transferability of adversarial examples across problem domains in **NLP** to understand why such examples exist.
- Found that transferability across domains was weak causing drops in performance of only **5-12%** under different domains indicating that adversarial examples arise from "non-robust features" (same as with CV).

OpenAI Hackathon for Climate Change | Natural Language Processing (NLP) *Fall 2022*
Net Zero AI – github.com/jyaacoub/CSR_summarizer *Nov. 11-14*
- Team lead for a tool designed to eliminate purposeful obfuscation behind lengthy Corporate Social Responsibility reports.
- Carefully **recruited, organized, and delegated tasks** to team members for effective work. The team consisted of a UX designer, software engineer, and environmentalist.
- Designed backend and pdf parser utilizing **OpenAI** API for semantic search and summarization **with GPT-3.**
- Resulted in fully fleshed out prototype and demo video as part of the submission in less than **3 days**.

SKILLS

Languages: Python, JavaScript, Java, C, and MATLAB
Machine Learning: PyTorch, PyTorch Geometric, Lightning, Scikit-learn, Matplotlib, Pandas, Numpy, HuggingFace
HPC and Distributed Learning: SLURM, Ray[Tune,Train], FLower