Youssef Samaan

438-404-7439 | ysamaan2001@gmail.com | linkedin.com/in/youssefsamaan | github.com/YoussefSamaan | Website

Education

McGill University Montreal, CA

Bachelor of Science in Computer Science, Minor in Statistics — GPA: 3.8/4 *Sep. 2021 – Apr. 2025*

- Relevant Coursework: Reinforcement Learning, Natural Language Processing, Applied Machine Learning, Algorithm Design, Data Structures, Software Design, Stochastic Processes, Probability, Statistics, Linear Algebra.

Research Experience

McGill University Montreal, CA

Researcher in Reinforcement Learning & NeuroAI *Sep. 2024 – Present*

- RL Algorithm Innovation: Investigating sparse reward RL algorithms (PPO, SAC, TD3) in AnimalAI environments to mimic biological learning processes, aiming to enhance algorithm efficiency and applicability in novel domains.

Machine Learning Research Projects

Generalization and Preprocessing for Sarcasm Detection

- LLM Benchmarking: Benchmarked NLP model generalization (LLMs, Classical ML) on multiple sarcastic datasets; LLMs (ChatGPT-o1) showed 17% accuracy gain while classical models trained faster.
- NLP Preprocessing Optimization: Demonstrated minimal text preprocessing (lowercase, punctuation removal) maintains performance, reduces training time (30-99%), highlighting word sequence importance in sarcasm detection.

Effect of Noisy Rewards on RL-Agent Performance

- Novel RL Reward Mechanism Study: Evaluated the impact of controlled reward noise (normal/uniform distributions) on the convergence and performance of Q-learning, Expected SARSA, DQN, and DDQN algorithms across Gymnasium environments (Cart Pole, Acrobot).
- Convergence Acceleration Discovery: Discovered that strategic introduction of reward noise accelerates RL agent convergence and facilitates faster discovery of optimal policies across varying noise levels.

Machine Learning Algorithm Implementations

- Algorithm from-scratch Development: Developed and implemented core ML algorithms from scratch, including Linear Regression, Logistic Regression, KNN, and Multi-Layer Perceptron in Python, demonstrating a deep understanding of fundamental ML principles.
- Advanced Model Building & Analysis: Built a CNN in TensorFlow for CIFAR-10 image classification and fine-tuned BERT (Hugging Face) for sentiment analysis, including attention matrix analysis, showcasing expertise in modern DL architectures and NLP techniques.

Work Experience

Ericsson Montreal, CA

Machine Learning Intern *January - August 2024*

- Scalable Data Pipeline Architecture: Architected and implemented an end-to-end data pipeline using PySpark to process 100+GB daily for multiple clients, significantly improving data accessibility for 100+ engineers and reducing retrieval time by 99%.
- Data Engineering & Agile Development: Designed robust data pipelines for data extraction, parsing, transformation, and storage (Amazon OpenSearch, S3), employing Agile methodologies for iterative development and customer-focused improvements.

DermBiont Boston, US

Software Developer and Data Scientist *May - August 2022*

- Data Processing & Automation: Developed Python programs for processing corrupt data files, automating data classification and report generation, and ensuring data integrity.
- Efficiency & Accuracy Improvements: Automated HTML file renaming improving hyperlink functionality by 35% and rectified 25%+ data discrepancies between Excel files and databases, enhancing overall data accuracy and efficiency.

Technical Skills

Languages: Python, Java, C/C++, JavaScript, TypeScript, HTML/CSS, SQL/NoSQL, MATLAB.
Tools & Frameworks: Numpy, Pandas, Scikit-learn, TensorFlow, PyTorch, HuggingFace, Transformers, Gymnasium, Django, stable-baselines3, Matplotlib, React, Next.js, REST APIs, PySpark, AWS, Git, Docker, Linux.