Rambod Azimi

+1 438-928-2042 | rambod.azimi@mail.mcgill.ca | rambodazimi.com | github.com/rambodazimi | linkedin.com/in/rambod-azimi

Summary

A graduate in Software Engineering from **McGill University**, holding a CGPA of **3.67/4**, with a dedication to research in AI/ML. Currently serving as a **Collaborating Researcher at Mila** and as a **Research Assistant at McGill**. **Two publications** (first author) currently under review: one focusing on efficient fine-tuning of LLMs in **ENLSP 2024**, and another exploring the application of U-Net and its variants to enhance CNN accuracy in **WACV 2024**. Previously interned at Walter Surface Technologies as an IT Intern and Ericsson as a Network Engineer Intern. Teaching a range of courses at McGill University, including Linear Algebra, Programming, and Calculus.

Research Interests:

- **Natural Language Processing (NLP)**
- **Computer Vision (CV)**
- **Deep Learning**

Education

McGill University *Montreal, Canada*

Bachelor of Software Engineering *Fall 2020 - Winter 2024*

- **GPA: 3.67/4**
- **Core Courses:** Artificial Intelligence, Computer Vision, Capstone Design Project, Numerical Methods, Linear Algebra, Operating Systems, Databases, Data Structures, Algorithm Design, Parallel Computing, Software Validation, Signals and Networks, Software Requirements, Computer Organization, Design Principles and Methods, Digital Logic, Software Systems, Model Based Programming
- **Capstone Project:** Machine Learning to Improve Manufacturing of Integrated Photonics Circuits. Supervisor: Professor Odile Liboiron-Ladouceur and Dr. Dusan Gostimirovic. Final Poster

Publications

Efficient Natural Language and Speech Processing (ENLSP) *Under Review*

KD-LoRA: A Hybrid Approach to Efficient Fine-Tuning with LoRA and Knowledge Distillation *Sept 2024*

- First Author
- KD-LoRA integrates LoRA with knowledge distillation to reduce computational costs while maintaining performance.
- KD-LoRA retains 98% of LoRA's performance on GLUE, while being 40% more compact and reducing GPU memory and inference time by 30%.

IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) *Under Review*

SEMU-Net: A Segmentation-based Corrector for Fabrication Process Variations of Nanophotonics with Microscopic Images *Sept 2024*

- First Author
- SEMU-Net segments SEM images to train deep neural networks for predicting and correcting fabrication-induced variations.
- The segmentation U-Net achieves an average IoU of 99.30%, while the corrector attention U-Net reaches 98.67%.

Experience

Mila - Quebec AI Institute *Montreal, Canada*

Collaborating Researcher *April 2024 - Present*

- Supervisor: Professor Samira Ebrahimi Kahou
- Built an AI-agent for automatic evaluation of excavator operators and integrating them LLM to answer questions using the simulator's documentation.
- Researched various LLM fine-tuning approaches, including parameter efficient fine-tuning (PEFT), knowledge distillation, Adapters, and LoRA.

McGill University *Montreal, Canada*

Research Assistant *May 2024 - Present*

- Supervisor: Professor Odile Liboiron-ladouceur
- Conducted research on manufacturing of integrated photonic ICs using ML approaches.
- Developed several CNN models such as U-Net, Attention U-Net, SegNet, EfficientNetB7, DeepLab V3+, and PSP Net.
- Improved the accuracy of the model by using an ensemble of multiple architectures.

McGill University *Montreal, Canada*

Research Assistant *May 2024 - Present*

- Supervisor: Professor Muthucumaru Maheswaran
- Conducted research on real-time Object Tracking using SOTA models (e.g., YOLO).
- Integrated Google Maps APIs for pixel-to-world coordinate conversion.
- Explored NLP techniques to enhance scene understanding and collision detection in LLMs with visualization cues.

Walter Surface Technologies *Montreal, Canada*

Information Technology Intern *May 2023 - Aug 2023*

- Designed and developed an AI-powered software solution to intelligently identify essential fields in an invoice file and automate the generation of a CSV file.
- Collaborated with teams in Cloud architecture and design, Cloud storage and databases, Security and compliance, Microsoft Azure tools, Monitoring and troubleshooting

Ericsson *Ottawa, Canada*

Network Engineer Intern *May 2022 - Aug 2022*

- Collaborated with teams in disaster planning, CIQ and intake validations, and automation tools.
- Automated and optimised the data handling process for networks, working with Python and did shell scripting, and employed other automation tools.
- Reviewed technical specifications from clients and vendors to confirm adequacy, accuracy and functionality.

McGill University | Faculty of Medicine *Montreal, Canada*

IT Assistant *Sept 2020 - April 2024*

- Diligently coordinated with internal departments to systematize standard information of students, track their status, and manage the contract and timely payment of each student.
- Worked in close collaboration with the professors in the faculty of Medicine.
- Proficiently conducted data analysis using Microsoft Excel, employing advanced functions and techniques to derive valuable insights from complex datasets.

Teaching Experience

Academic Tutor *Montreal, Canada*

McGill University *Sept 2021 - Current*

- Helped students prepare for tests and exams by providing guidance, motivation and encouragement.
- Used personalized learning plans to guide students to deeper understanding of course material and learning styles.
- Both online and in person sessions.
- Courses: Linear Algebra, Computer Organization, Java Programming, Python Programming, Calculus 1 and 2,

Tomlinson Engagement Award for Mentoring (TEAMS) *Montreal, Canada*

McGill University *Sept 2023 - Current*

- Provided comprehensive responses to student inquiries on the discussion board.
- Conducted both in-person and online one-on-one sessions with students throughout the semester.
- Course: Model-Based Programming (ECSE 223)
- Course Instructor: Professor Gunter Mussbacher

Computer Skills

Python Libraries	Pytorch, Tensorflow, Transformers, PEFT, Scikit-learn, OpenCV, Pandas, Numpy, Matplotlib, Optuna, MLFlow
Programming Languages	Python, Java, C/C++, HTML/CSS, OCaml, Bash, Swift, VHDL/Verilog, MATLAB, Octave, LaTeX, JavaScript
IDEs/Tools	VS Code, PyCharm, Eclipse, XCode, Android Studio, IntelliJ, CUDA, VIM, Adobe Dreamweaver
Frameworks	Spring (Java), Vue.js, JUnit (Java), JavaFX, Wordpress, REST API
Operating Systems	MacOS, Windows, Linux (ubuntu), RasperryPi, iOS, Android