Rambod Azimi

Software Engineer and Machine Learning Researcher

(438) 928-2042 rambod.azm@gmail.com Canada
LinkedIn GitHub Google Scholar

Summary

Software engineering graduate from McGill University, with a CGPA of 3.67/4, specializing in backend development, cloud computing, database management, the development of AI-powered software solutions, automation systems, and scalable web applications. Worked at Quantiphi, McGill, NRC, and MILA on several impactful projects ranging from software engineering to machine learning. Proficient in Python, Java, C, C++, Flask, TensorFlow, PyTorch, SQL, Docker, and more. Previously interned at Ericsson as Network Engineer Intern, and Walter Surface Technologies as IT Intern.

Professional Experience

Research Engineer Intern, *(Quantiphi Analytics)* **Toronto, Canada** *02/2025 - 07/2025*
- Developed a modular knowledge distillation pipeline supporting SeqKD, Multitask, GRPO, and ORPO, across diverse fine-tuning strategies.
- Engineered dataset preprocessing workflows to incorporate rationale annotations into popular benchmarks (e.g., ARC, GSM8K).
- Conducted large-scale experiments with teacher-student LLMs, evaluating performance trade-offs across multiple distillation techniques.

Machine Learning Engineer Intern, *(NRC - National Research Council Canada)* **Ottawa, Canada** *10/2024 - 08/2025*
- Optimized neural networks for image-based deep learning, focusing on uncertainty reduction and model calibration.
- Implemented generative adversarial networks (GANs) to simulate uncertainty in photonic device fabrication.
- Leveraged variational autoencoders (VAEs) to improve data augmentation.

Machine Learning Researcher, *(Mila - Quebec AI Institute)* **Montreal, Canada** *04/2024 - 12/2024*
- Built an AI-agent for automatic evaluation of excavator operators and answering questions using the simulator's documentation.
- Implemented various LLM fine-tuning methods, including parameter efficient fine-tuning (PEFT) and knowledge distillation.
- Developed AI tools combining LLMs with simulation data for operator performance analysis and documentation query handling.

Research Assistant, *(McGill University)* **Montreal, Canada** *05/2024 - 10/2024*
- 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.

Information Technology Intern, *(Walter Surface Technologies)* **Montreal, Canada** *05/2023 - 08/2023*
- Developed an AI-powered software to identify essential fields in an invoice file and automate the order management process.
- Collaborated with teams in Cloud architecture, Cloud storage and databases, Security and compliance, and Microsoft Azure.

Network Engineer Intern, *(Ericsson)* **Ottawa, Canada** *05/2022 - 08/2022*
- Collaborated with teams in disaster planning, CIQ and intake validations, and automation tools.
- Automated the data handling process for networks, working with Python and bash scripting, and other automation tools.

IT Assistant, *(McGill University)* **Montreal, Canada** *09/2020 - 03/2024*
- Coordinated with departments to systematize information of students and track their status.
- Conducted data analysis tools to extract insights from complex datasets.

Skills

- **Programming Languages:** Python, Java, C, C++, OCaml, SQL, Assembly, HTML, CSS, JavaScript, Bash, MATLAB
- **AI & Data:** TenserFlow, PyTorch, Pandas, NumPy, Transformers, Image Segmentation, PowerBI
- **Infrastructure:** AWS, Azure, Git, Gitlab, Jira, Bitbucket, Linux, Unix, OS, DevOps
- **Backend:** Java, SpringBoot, Python, Flask, SQL, REST, Django, Ruby, PostgreSQL, MVC, UML/ER Modeling
- **QA Testing:** Junit, Selenium, pytest, Postman, Cucumber, CI/CD

Education

Bachelor of Software Engineering - GPA: 3.67/4 *McGill University* **Montreal, Canada** *2020-2024*

Relevant Courses: Algorithms, Data Structures, Database Systems, Programming Languages, Software Requirements, Signals and Networks, OS, AI, Computer Vision, Parallel Computing, Model-Based Programming, Software Validation

Publications

- KD-LoRA: A Hybrid Approach to Efficient Fine-Tuning with LoRA and Knowledge Distillation - NeurIPS ENLSP 2024
- SEMU-Net: A Segmentation-based Corrector for Fabrication Process Variations of Nanophotonics with Microscopic Images - WACV 2025