

Zeshan Fayyaz

• 647 551 XXXX • zeshanf59@hotmail.com • zeshanfayyaz.com • zeshan-fayyaz • ZeshanFayyaz

EDUCATION

University of Waterloo | Master of Mathematics, Computer Science Sep 2022 - June 2025, Waterloo, ON
Toronto Metropolitan University | Bachelor of Engineering, Computer Engineering Sep 2016 - April 2022, Toronto, ON

SKILLS

Programming Languages & Tools: Python, SQL, C, Java, Bash, Git, LaTeX, Regex, REST APIs, FastAPI, NumPy, Pandas
Machine Learning & AI: Federated learning, RAG, MCP, LLMs, GRPO, Reinforcement Learning, PyTorch, TensorFlow, MLflow
Data & DevOps Infrastructure: PostgreSQL, FAISS, ETL pipelines, AWS EC2, Docker, GitHub Actions, Linux, Distributed Systems

WORK EXPERIENCE

| | |
|---|------------------------------------|
| Graduate Research Assistant <i>University of Waterloo</i> | Sep 2022 – June 2025, Waterloo, ON |
| • Designed a scalable distributed machine learning framework that used reinforcement learning to coordinate over 1,000 heterogeneous clients, reducing model convergence time by 24% and improving overall system stability under heterogeneous conditions. | |
| • Improved privacy and computational efficiency by developing an adaptive encryption mechanism that optimized CKKS encryption levels during training, increasing convergence performance by 30% while preserving full security guarantees across all clients. | |
| • Led tutorials for CS114, CS115, and CS135 focused on Python, SQL, and functional programming for over 300 students. | |
| ML Research Engineer <i>Toronto Metropolitan University</i> | Jan 2020 – Jan 2023, Toronto, ON |
| • Built a production grade image restoration model by implementing deep learning architectures in TensorFlow and BiGRU, improving PSNR by 22.9% and SSIM by 3.79% on large scale datasets and validating model robustness across multiple test suites. | |
| • Built a reliable large scale training pipeline on AWS EC2 to handle over 100GB of image data, engineering dynamic batching, optimized data loaders, and parallel model tuning workflows that eliminated bottlenecks and reduced end to end runtime by 35%. | |
| Data Analytics Intern <i>TD Canada Trust</i> | June 2022 – Aug 2022, Toronto, ON |
| • Built reliable end to end analytics pipelines in Python and Tableau that processed and visualized over 5k unresolved support tickets across multiple departments, improving cross team accountability and reducing reporting latency by 15% in executive summaries. | |
| • Led company wide data learning sessions and deployed lightweight machine learning models for anomaly detection in JIRA, reducing manual review effort by 25% | |
| Software Engineer Intern <i>Royal Bank of Canada</i> | Jan 2022 – June 2022, Toronto, ON |
| • Built Git-integrated features including credential caching, commit diff visualization, and unsaved change warnings that improved developer workflow reliability and reduced user-reported errors by 35%. | |
| • Integrated CI/CD automation using Docker and GitHub Actions to streamline deployment workflows and cut setup time by 40%. | |

RELEVANT PROJECTS AND PUBLICATIONS

Author of 5 peer-reviewed papers with 650+ total citations on distributed machine learning, recommendation systems, and privacy-aware AI.

| | |
|--|--|
| Knowledge Assistant: RAG + MCP Pipeline for Context-Grounded LLM Systems | |
| • Developed a retrieval augmented generation system using SentenceTransformers, FAISS, and Ollama, with an MCP powered FastAPI backend that supported structured tool use, efficient retrieval pipelines, and contextual reasoning over domain specific knowledge. | |
| GRPO Driven Optimization for Distributed LLM Workflows | |
| • Building a GRPO driven optimization pipeline that applies reinforcement learning to reduce straggler effects in distributed LLM workloads, with support for reward shaping, logging, evaluation, and PPO baseline comparisons. | |
| Secure Computation Benchmarking for Encrypted ML Workloads | |
| • Built a scalable benchmarking platform with automated pipelines, parameter sweeps, data processing modules, performance logging, and report generation to evaluate system behavior and efficiency for machine learning workloads using secure computation tools. | |
| HERL: Tiered Federated Learning with Adaptive Homomorphic Encryption using Reinforcement Learning TPS 2025 | |
| • Developed a tiered federated learning framework that used adaptive encryption, reinforcement learning, and clustering techniques to manage stragglers and heterogeneous clients, increasing model accuracy by 20% and reducing communication overhead by 30%. | |