# 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, NumPy, Pandas
**Machine Learning & AI Systems:** Federated learning, RAG, MCP, LLMs, PyTorch, TensorFlow, Scikit-learn, MLflow, FastAPI
**Data & DevOps Infrastructure:** PostgreSQL, FAISS, ETL pipelines, AWS EC2, Docker, GitHub Actions, Linux, CI/CD

## Work Experience

**Graduate Research Assistant** — Sep 2022 – June 2025, Waterloo, ON
*University of Waterloo*

- Designed a scalable federated learning framework integrating Q-learning based adaptive encryption, achieving a 24% reduction in model convergence time across 1,000+ simulated clients while maintaining system stability under heterogeneous environments.
- Enhanced privacy and computational efficiency through adaptive CKKS homomorphic encryption, improving convergence performance by 30% and preserving full 256-bit security guarantees across training nodes.

**Teaching Assistant** — Sep 2022 – June 2023, Waterloo, ON
*University of Waterloo*

- Led tutorials for CS114–CS135 in Python, SQL, and functional programming for 300+ students, creating modules on algorithmic optimization and data preprocessing best practices.

**Machine Learning Researcher** — Jan 2020 – Sep 2022, Toronto, ON
*Toronto Metropolitan University*

- Developed and deployed deep learning architectures (TensorFlow, BiGRU) for stripe noise removal on large-scale image datasets, improving PSNR by 22.99% and SSIM by 3.79% over prior state of the art and validating performance across multiple test sets.
- Designed distributed training and evaluation pipelines on AWS EC2 for 100GB+ data, reducing end-to-end runtime by 35% through dynamic batching, optimized data loaders, and parallelized model tuning.

**Business Insights and Analytics Intern** — June 2022 – Aug 2022, Toronto, ON
*TD Canada Trust*

- Built end-to-end analytics pipelines in Python and Tableau to track and visualize 5k+ unresolved support tickets across departments, improving cross-team accountability and reducing reporting latency by 15% in monthly executive summaries.
- Led internal data learning sessions and deployed lightweight ML models for anomaly detection, cutting manual review time by 25% and enabling proactive identification of recurring ticketing issues.

**Software Engineering Fellowship** — Jan 2022 – June 2022, Toronto, ON
*Royal Bank of Canada*

- 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 (RAG) system integrating SentenceTransformers, FAISS, and Ollama, with an MCP-based FastAPI backend enabling structured tool use and contextual reasoning across private document corpora.

**HERL: Tiered Federated Learning with Adaptive Homomorphic Encryption using Reinforcement Learning** | *TPS 2025*

- Developed a tiered federated learning framework using adaptive homomorphic encryption and Q-learning to address stragglers and client heterogeneity, achieving a 20% increase in model accuracy and 30% reduction in communication overhead.

**Towards Taming the Resource and Data Heterogeneity in Federated Learning** | *USENIX OPML 2019*

- Investigated resource and data heterogeneity in federated learning, using AWS EC2 simulations with 20 clients and achieving 40% reduction in per-epoch training time, optimizing model quality in non-homogeneous environments.