Rashinda Wijethunga

rashinda.wijethunga@gmail.com · (+1) 437-985-4184 · London, ON (Relocating to Toronto)
LinkedIn: linkedin.com/in/rashinda · GitHub: github.com/RashindaW · Portfolio: rashindaw.github.io

Professional Summary

Machine Learning Ph.D. candidate (Western University, expected Aug 2026) who builds and deploys **production-grade Generative and Agentic AI systems**. Through the Mitacs Accelerate program, embedded in a Canadian SME to deliver two ML systems from open problem to production at Sobeys supermarkets (early refrigerant leak detection). Independently built a guardrailed hybrid-RAG assistant and co-built a 5-agent provisioning system (tool-use, memory, human-in-the-loop), both backed by formal evaluation harnesses and test suites. Author of 6 peer-reviewed publications, including top-tier Q1 journals.

Technical Skills

Generative AI & RAG: RAG (hybrid retrieval, Reciprocal Rank Fusion, re-ranking), prompt engineering, vector embeddings & semantic search, grounding & guardrails, evaluation (RAG metrics, faithfulness), synthetic data generation, fine-tuning; Anthropic Claude, Gemini, Hugging Face Transformers, LlamaIndex, Gradio
Agentic AI: autonomous & multi-agent systems, LLM tool-use / function calling, memory, human-in-the-loop; Google ADK, Model Context Protocol (MCP), agent-to-agent (A2A)
ML & Deep Learning: PyTorch, PyTorch Geometric (PyG), GNNs, scikit-learn, XGBoost, CatBoost, LightGBM
Deployment & MLOps: FastAPI, Docker, Azure Functions, Google Cloud (Cloud Run, Cloud SQL), PostgreSQL + pgvector, CI/CD (GitHub Actions)
Languages & Data: Python, SQL, TypeScript, PySpark, Spark SQL

Education

- **Ph.D., Electrical and Computer Engineering** — *Sep 2021 – Expected Aug 2026*
  Western University — *Canada*
  - Dissertation: *Trustworthy Anomaly Detection using Graph Neural Networks and Counterfactual Explanations*
- **B.Sc. Eng. (Hons) in Electronic and Telecommunication Engineering** — *Dec 2015 – Jan 2020*
  University of Moratuwa — *Sri Lanka*

Selected Projects

- **BuildRight AI – Guardrailed, Agentic RAG Commerce Assistant** | Code | Live Demo — *Jun 2026*
  - Built and deployed a full-stack AI commerce assistant (FastAPI + React) on **Hugging Face Spaces**, driven by an agentic **Claude** tool-use loop over a 10K+ SKU catalog: "paint my living room" triggers **clarifying questions**, then computes materials, shortlists products, and builds a **one-step bill-of-materials cart**.
  - Engineered a **hybrid RAG** pipeline combining lexical, dense (**BGE**/pgvector), and visual (**CLIP**) retrieval via **Reciprocal Rank Fusion** + cross-encoder re-ranking, with a hit@k/MRR/nDCG evaluation harness (hit@5 improved 0.86 → 0.93).
  - Designed deterministic **guardrails** validating every price and citation against retrieved data (so a fabricated price can't render), plus a **Haiku-to-Sonnet model router** that cuts LLM cost by keeping simple queries on the cheaper model.
  - Added **multimodal** input (image-based product search, voice ordering via **Whisper**, and handwritten OCR with confirm-then-audit writes), plus persistent per-user **memory**.
  - Shipped production infrastructure as a single **Docker** container: 4-tier RBAC, JWT + CSRF, Stripe checkout, per-turn cost observability, and **267 automated tests** with CI/CD across SQLite and PostgreSQL + pgvector.
- **LLM-Powered Multi-Agent Chatbot for Enterprise User Provisioning** | Code | Demo — *Dec 2025*
  - Co-developed a 5-agent system (Google ADK, Gemini) as the capstone of the Google/Kaggle 5-Day AI Agents Intensive, automating user onboarding/offboarding via a supervisor delegating to policy, identity, and approval agents, **replacing a manual, breach-prone process with a policy-aligned workflow**.
  - Integrated agents with live **Microsoft Entra ID** operations via a custom **MCP server**, enabling agent-to-agent (A2A) delegation, SQLite session memory, and tiered model usage (Gemini Pro for supervisor, Flash for workers) to control cost.
  - Implemented **human-in-the-loop** approval gates on high-risk actions, policy-aware RAG over PDF rulebooks, role-based access control, and regression testing using pytest + ADK's AgentEvaluator for safe, auditable workflows.

Experience

- **Machine Learning Intern (Phase II)** *Jan 2025 – Dec 2025*
  Neelands Group *Burlington, ON, remote*
  - Built and deployed an end-to-end **anomaly detection system** for **CO_2-based refrigeration** (the customer's latest architecture) at **Sobeys**, monitoring 170+ subsystems per site.
  - Processed multi-year sensor telemetry (∼210M rows) using **PySpark** and **Spark SQL** to build a scalable curation and feature-engineering pipeline.
  - Translated ambiguous business requirements into technical specifications through collaboration with domain engineers and operations stakeholders; communicated model performance and limitations back in business terms.
  - Trained per-site **XGBoost** models with residual-based detection layered into a state-machine architecture; achieved **87.5% detection rate** within weeks of onset (vs. months previously undetected).
  - Deployed via **Azure Functions** with **CI/CD (GitHub Actions)**, **automated retraining**, and drift monitoring.

- **Machine Learning Intern (Phase I)** *May 2022 – Apr 2023*
  Neelands Group *Burlington, ON, remote*
  - Designed a **sensor-independent leak detection framework** for **HFC-based refrigeration** (the legacy architecture), using only standard temperature and pressure telemetry.
  - Engineered physics-informed features with **SHAP**-driven selection; trained a **CatBoost** model with dynamic thresholding that reduced detection latency from **months to 5 days** with **100% precision** (F1 0.92).
  - Deployed via **Azure Functions** with CI/CD; success led to a **second engagement** at the same client.

- **Ph.D. Researcher** *Sep 2021 – Expected Aug 2026*
  Western University *London, ON*
  - Built an unsupervised **temporal graph neural network (GNN)** for **anomaly detection on multivariate time series**, outperforming 9 baselines by up to 7% in AUC. (Code)
  - Engineered a **GNN framework** for supervised fault detection and localisation under noise; added a **1D-CNN compression module** that cut training time **24x** (296 h to 12 h) at <1% accuracy cost. (Code)
  - Built a **robust explainability method** for tree-based ensemble models (XGBoost, LightGBM, Random Forest) that stays stable under retraining; ranked **#1 in 158/248** robustness scenarios. (Code)
  - Developed a novel method to improve **GNN-based anomaly detection** trustworthiness through uncertainty estimation, reducing false positives by **17%**.

- **Systems Engineer** *Mar 2020 – Aug 2021*
  Millennium IT ESP *Colombo, Sri Lanka*
  - Engineered the migration of legacy data center infrastructure to **Cisco's application-centric (ACI) architecture** for **SLT** (Sri Lanka's largest telecom provider).

Certifications & Awards

- 5-Day AI Agents Intensive Course, Google & Kaggle (Certificate)
- Mitacs Accelerate Scholarships: Phase II $30K (2025 Jan–Dec) + Phase I $20K (2022–23)
- IEEE ComSoc Student Competition Honorary Mention (World Top 15, 2019)

Publications

(* indicates equal contribution) Google Scholar

- **R. Wijethunga**, A. Sadhu, J. Samarabandu, "Dual-View Spectral-Temporal Graph Fusion for Incipient Fault Detection," *Information Fusion* (IF 15.5), 2026 (under review).
- **R. Wijethunga**, J. Samarabandu, A. Sadhu, "Robust and Efficient Dual-Graph Neural Networks for Structural Damage Detection," *Engineering Structures* (Q1, IF 6.4), 2025.
- **R. Wijethunga***, V. Galwaduge*, et al., "Evaluating Robustness of Counterfactual Methods for Tree-based Ensembles," *IEEE CJECE*, 2026.
- **R. Wijethunga**, H. Nouraei, et al., "Precision Leak Detection in Supermarket Refrigeration Systems," *Energies* (Q1, IF 3.2), 2024.
- P. Singh, **R. Wijethunga**, et al., "Expert Evaluation System for Pothole Defect Detection," *Expert Systems with Applications* (Q1, IF 7.5), 2025.
- V. Galwaduge*, **R. Wijethunga***, et al., "A Novel Objective Function for Counterfactual Explanations Using Conic Optimization," *IEEE ICAC*, 2025.