

Jeevan Parmar

(778) 389-2213 | j29parma@uwaterloo.ca | [LinkedIn](#) | [GitHub](#)

TECHNICAL SKILLS

Languages: Java, Python, Kotlin, JavaScript, SQL, R, C#

Frameworks: LangChain, LangGraph, Llama.cpp, OpenAI, Cohere, HuggingFace, Express.js, Redux, JUnit

Libraries: Pytorch, pandas, NumPy, Matplotlib, Selenium, Scikit-learn, Node.js, React.js, OpenCV

Developer Tools: Git, Docker, Azure, AWS, GCP, Firebase, Jenkins, MySQL, MongoDB, ChromaDB

EXPERIENCE

Norwegian University of Science and Technology — NorwAI

Trondheim, Norway

Social Robotics Programmer - Researcher

Jan 2025 - April 2025

- Developed a conversational agent using Adaptive, Corrective, and Self RAG patterns, boosting accuracy by 15%
- Integrated Python agent with custom Kotlin Furhat skills for NorwAI's first agent-based robot-connected backend
- Optimized locally deployed models with llama.cpp, achieving sub-8s responses, cutting cloud costs by 15%
- Deployed Dockerized FastAPI on an EC2 instance, leveraging extra compute to cut generation time to sub-4s

Cognite

Austin, Texas

Software Engineer Co-op

May 2024 — Sept 2024

- Developed Cognite's first industrial agent for the Atlas AI program, integrating tools for troubleshooting workflows
- Boosted doc-parser's keyword extraction accuracy to 90% using advanced embedding and cross-encoding techniques
- Integrated Gemini model into Cognite's doc-parser, enabling GCP users access and enhancing overall functionality
- Implemented the Tail Generation Pattern to generate summaries, optimizing long-term memory recursively

XCare

Toronto, Ontario

AI Engineer

Oct 2023 — Oct 2024

- Fine-tuned Dense CNNs and Vision Transformers for X-ray diagnosis, achieving 90% accuracy
- Developed a RAPTOR-AI pipeline, increasing retrieval accuracy to 95%, graded by medical professionals
- Architected a RAG-AI pipeline delivering personalized rehabilitation info with references from medical sources
- Wrote, presented, and published a paper on the tool at the Canadian Undergraduate Conference on AI

Genellipse Inc.

Toronto, Ontario

Software Engineer Co-op

Sept 2023 — Dec 2023

- Optimized MongoDB architecture: enabling vector similarity search, enhancing data efficiency across 13 collections
- Boosted data processing accuracy by 75% with Adobe and RAG, while decreasing runtime to sub-3 minutes
- Implemented MNN and RNN Pytorch models, leading to a R^2 of 0.85 and 0.95, respectively

TD Bank — Technology

Toronto, Ontario

Automation Quality Engineer Co-op

Jan 2023 — April 2023

- Introduced Recursive and Object-Oriented testing methods using Java Selenium, reducing runtime by 58%
- Implemented 30+ end-to-end test cases on TD's DEFI Environment, successfully closing a 2+ year project
- Deployed automation scripts on Jenkins leading to 20+ bugs found within TD's DEFI Environment

PROJECTS

Search Engine | *Python*

Sept 2024 — Dec 2024

Audio Transcriber | *Python, HuggingFace, OpenAI*

Oct 2024 — Oct 2024

Nodal Price Forecast Algorithm | *MongoDB, Python, Scikit-learn, Pytorch*

Jan 2024 — April 2024

Meal Stream | *MySQL, Firebase, JavaScript, Node.js, React.js, Redux, Express.js*

Jan 2024 — April 2024

IMDB Clone | *MySQL, JavaScript, Node.js, React.js, Express.js, REST APIs*

June 2023 — July 2023

NBA Fantasy Projection Project | *MySQL, Python, HTML/CSS, Scikit-learn*

June 2022 — Aug 2022

EDUCATION

University of Waterloo

Waterloo, Ontario

Honours Co-op BAsc Management Engineering, Artificial Intelligence Option

Sept 2021 — Present