Shrey Shingala

437-980-5133 | sshingal@uwaterloo.ca | linkedin.com/in/shrey-shingala1 | github.com/ShreyShingala | shreyshingala.me

Skills

Languages: Python, JavaScript, TypeScript, C++, C, C#, SQL, Swift, Bash
Technologies: Node.js, React, FastAPI, Next.js, Docker, PostgreSQL, MongoDB, Pandas, Pytest, ROS 2, YOLOv8

Experience

Software Developer | Python, FastAPI, Pandas, Typescript — Jan. 2026 – Present
PivoHub — *(Remote) Montréal, QC*

- Cut customer onboarding time from **multiple days to <1 hour** by engineering an AI-powered catalog import pipeline, processing catalogs of **1,000+ products**
- Eliminated manual data entry by designing an **LLM-driven** mapping engine to automatically detect headers and generate transformation plans from messy supplier spreadsheets with **90%+** mapping accuracy
- Improved complex bilingual product classification across **29 categories** by developing a hybrid pipeline combining **keyword matching** and **embedding cosine similarity**
- Implemented **20+ regex-based** extraction transforms (weight, volume, pack count, other patterns), eliminating per-row LLM inference costs while maintaining **95%+** extraction accuracy
- Developed the pipeline as a **FastAPI service** for seamless integration into PivoHub's import workflow

Software Engineer | Python, React, TypeScript — Sep. 2025 – Present
Waterloo Rocketry — *Waterloo, ON*

- Eliminated **3+ second** UI freezes in rocket telemetry dashboard by rebuilding visualization using **D3.js** SVGs, enabling smooth rendering of **6 high-frequency streams** and ensuring launch readiness
- Implemented high-frequency web telemetry by creating a bidirectional **WebSocket-ZeroMQ bridge** in Python **Flask-SocketIO**, handling **100+ msg/s** to expose low-level rocket bus data to modern web clients
- **Prevented downstream application** crashes from malformed sensor data by enforcing **Pydantic** runtime validation across **60+** unique sensor fields, replacing untyped dictionary outputs and boosting system reliability
- Improved maintainability by **overhauling 4 legacy data parsers** from procedural code to a unified **object oriented** design with **polymorphism**
- Facilitated safe refactoring and achieved **98% coverage** across critical low-level modules such as bitwise decoding and CRC validation by establishing test infrastructure from scratch using **Pytest**

Projects

Pokémon Card Scanner | *Python, FastAPI, YOLOv8, Docker* — GitHub

- Attained **95%+** identification accuracy across **20,000+** Pokémon cards by developing a holistic computer vision pipeline combining **YOLOv8** object detection, **CLIP+FAISS** embedding search, and **OCR**-based validation
- Reduced latency for multi-card images by **50%** by designing a **parallel processing system** to distribute work, enabling faster scans for large collections
- Deployed a production full-stack application with **dockerized FastAPI** backend on Hugging Face Spaces, **Next.js** frontend, Google OAuth authentication, and **PostgreSQL (Supabase)** database

Big Brother | *Python, FastAPI, MongoDB, Neo4j* — GitHub

- Won **Runner-Up** at UofT Hacks 2026 (Backboard.io Adaptive AI Challenge) by building an **autonomous browser agent** that orchestrates **5** LLMs to execute complex web tasks from natural language prompts
- Reduced inference costs by **40%** and average response latency from **3.2s to 1.8s** (**44%** improvement) by implementing **intelligent model routing** and semantic caching with **embeddings**
- Improved automation accuracy by creating an **adaptive memory system** using **Neo4j** graph database with **RAG-based** retrieval, enabling the learning of user preferences through successful actions

Education

University of Waterloo — Waterloo, ON
*Bachelor of Software Engineering (**GPA: 4.0/4.0**)* — Sep. 2025 – Apr. 2030