Rajit Sareen

US Citizen | rajitsareen@gmail.com | rajit-sareen.vercel.app | github.com/rajit-sareen | linkedin.com/in/rajit-sareen

Education

University of Washington — Seattle, WA

Bachelor of Science, Computer Science, GPA: 3.75/4.0 — *Expected Graduation:* ***Jun 2028***

- CS Coursework: Data Structures and Algorithms, Distributed Systems, Computer Architecture, Software Design
- Math/ML Coursework: Linear Algebra, Differential Equations, Probability and Statistics, Deep Learning

Experience

FlexTrade Systems — Jun 2026 – Aug 2026

Software Engineering Intern — *New York, NY*

- Designed a multi-agent system with MCP plugins to route issues across **12** categories, cutting response time **5x**
- Exposed backend logs to agents via MCP server plugins for faster issue diagnosis
- Built a RAG ingestion pipeline for **1K+** documents, optimizing semantic search and cutting retrieval time **2x**
- Built an account management tool for the sales and client services division, supporting **225+** clients
- Maintained order-routing logic in a live multi-threaded C++ platform; fixed a null-pointer production outage via GDB

Data Analysis & Intelligence Lab, University of Washington — Jun 2025 – Jan 2026

Undergraduate Research Assistant — *Seattle, WA*

- Contributed to DeepTracer, protein structure prediction research presented at **NeurIPS 2025**
- Trained and evaluated **3+** generative architectures (diffusion, flow-matching, U-Net) in PyTorch on cryo-EM data
- Parallelized training across **10 GPUs** on **160,000** samples, tracking **100+** experiments in Weights & Biases
- Fixed a 3D-to-2D tensor bug silently dropping icosahedral proteins, adding **hundreds** of proteins back to training

Projects

SmartFridge | React Native, TypeScript, FastAPI, PostgreSQL, YOLOv8, AWS EC2

- Built a full-stack mobile app (React Native, TypeScript, FastAPI) for real-time fridge inventory tracking
- Trained a YOLOv8 object detection model to auto-identify fridge items from live camera input
- Designed a PostgreSQL schema with JWT auth and deployed the backend on **AWS EC2**

ContextOpt | Python, Google Cloud (Vertex AI), Hugging Face, RAG, XGBoost — GitHub

- Built an evaluation framework comparing **6** LLM context strategies (RAG, summarization, compression)
- Benchmarked Llama-3.1-8B and Gemini 2.5 on LongBench across quality, latency, cost, and token efficiency
- Building an XGBoost/MLP selector that picks the best context strategy per query, aiming to beat fixed baselines

Open Source Contributions | C++, Python — Stan Math / librosa

- Contributed to Stan Math (C++ autodiff library, **100K+** users) and librosa (**8.6K+** GitHub stars)
- Fixed an infinite-input edge case in Stan Math's object-oriented autodiff using template metaprogramming and Eigen
- Expanded librosa's parameter-validation tests and added spectral display and fade-windowing functions

Distributed Job Runner | C++, TCP, Distributed Systems, Linux — GitHub

- Built a distributed job orchestrator in C++ with a hand-rolled thread pool and TCP coordinator
- Implemented heartbeat-based failure detection and automatic job rescheduling for downed nodes
- Designed a lock-free ring buffer (atomic CAS), verified with **4 producers/4 consumers** and **40K+** operations

Achievements

- Dean's List, University of Washington
- AWS Certified Cloud Practitioner

Skills

Languages: Python, C++, C, Java, TypeScript, SQL
Backend & Cloud: FastAPI, REST APIs, PostgreSQL, AWS (EC2), Docker, GitLab CI/CD, Linux, React Native
AI/ML: RAG, Multi-Agent Systems, MCP, LLM Evaluation, Hugging Face, PyTorch, XGBoost, YOLOv8
Concepts: Distributed Systems, Concurrency, Operating Systems, Algorithms, Object-Oriented Programming
Tools: Git, GDB, Bash, Weights & Biases