

Rahul Surya

Edinburgh, UK · +44 7742 977485 · rahulsurya021@outlook.com · [LinkedIn](#) · [GitHub](#) · [Portfolio](#)

SUMMARY

Machine Learning Engineer with hands-on research ML experience at ISRO and advanced parallel computing expertise from an MSc in High-Performance Computing with Data Science at the University of Edinburgh (EPCC). Proficient in PyTorch, Python, C++, CUDA, and cloud infrastructure (Azure Certified). Demonstrated ability to build scalable deep learning pipelines and real-time inference systems. Seeking ML Engineer roles in the UK from August 2026.

SKILLS

- **Languages:** Python, C++, C, Java, Go, SQL
- **ML Frameworks:** PyTorch, TensorFlow, Hugging Face Transformers, Scikit-learn, LLMs, NLP, Computer Vision, Time Series Analysis
- **MLOps & Infra:** FastAPI, REST APIs, Docker, Kubernetes, Prometheus, Grafana, Git, CI/CD, Microsoft Azure, CUDA
- **HPC:** MPI, OpenMP, ARCHER2 Supercomputer, Slurm, Parallel Algorithm Design
- **Data:** PostgreSQL, TimescaleDB, MongoDB, Redis, Apache Spark, Tableau

PROFESSIONAL EXPERIENCE

Machine Learning Engineer Intern · ISRO – National Remote Sensing Centre Oct 2023 – Jan 2024

- Developed end-to-end research ML pipeline for lightning prediction using **ConvLSTM-Seq2Seq** in PyTorch, achieving **92% forecast accuracy** on live meteorological data
- Optimised deep learning pipeline on 500GB+ of WRF simulation data using Variational Autoencoders (VAE), reducing model training time by **35%**
- Built React.js dashboard with REST APIs serving real-time predictions to **15+ meteorologists**
- Applied software engineering best practices: version control (Git), testing, documentation, code reviews

Data Science Intern · Clustor Computing Jun 2023 – Sep 2023

- Built ensemble ML models for stock price prediction and designed Tableau dashboards for financial analytics on client datasets

Software Engineer Intern · Develoscope Software Solutions Jun 2023 – Sep 2023 (*concurrent*)

- Delivered four Java web applications (JSP, Tomcat); improved client satisfaction by 25%

PROJECTS

Dissertation: ML-Based Memory Leak Detection in Containerised Environments · Python, Prometheus, eBPF, TimescaleDB, Kubernetes

Building an early-detection system for memory leaks using ML on time-series metrics collected via Prometheus and eBPF from containerised workloads on an EIDF HPC cluster.

LLM Inference Engine · Python, FastAPI, CUDA, PyTorch, Docker, Prometheus

Built a scalable inference server for transformer architectures with GPU memory management, continuous batching, and Prometheus-based observability. Containerised with Docker and deployed via Kubernetes.

Massively Parallel Cellular Automata Simulation · C, MPI, OpenMP

Implemented 2D-domain decomposed simulation on ARCHER2 supercomputer achieving **85%+ parallel efficiency** across 64 CPU cores with optimised halo exchange.

EDUCATION

University of Edinburgh – *MSc High-Performance Computing with Data Science* Sep 2025 – Aug 2026

Coursework: Advanced Parallel Programming (MPI/OpenMP), GPU Programming (CUDA), Machine Learning, Data Science

G H Rasoni Institute of Engineering – *B.Tech Artificial Intelligence (First Class with Distinction)* Aug 2020 – Jul 2024

CERTIFICATIONS

Microsoft: Azure DevOps Engineer Expert · Azure Developer Associate · Fabric Analytics Engineer · AI Fundamentals · Data Fundamentals **Oracle:** OCI Generative AI Professional **Coursera:** Machine Learning – Stanford Online · Google Cybersecurity Professional