

Himmat Rai

himmatrai20@gmail.com | LinkedIn | GitHub | Website

Education

BSc Computer Science (Hons) – 1:1 Predicted
De Montfort University, Leicester, UK

Oct 2023 – Sep 2027

Experience

Software Engineer (Placement) – NEXT plc.
Enderby, UK

Aug 2025 – Aug 2026

- Developed and deployed production features for a supply-chain logistics platform that processes and tracks millions of items daily, using C#, ASP.NET Core, SQL Server, Azure, and REST APIs.
- Designed SQL Server schemas and implemented complex T-SQL stored procedures to support transactional workflows, data access, and backend business logic for production logistics services.
- Automated the migration of SQL Server databases from Windows Server 2012 to 2022, transferring logins and role mappings and rebinding users through secure dynamic SQL to resolve orphaned accounts.
- Modernised legacy services to support SOAP 1.2 while preserving SOAP 1.1 compatibility for third-party label suppliers, maintaining interoperability during a staged integration upgrade.
- Migrated legacy WCF integrations from HTTP to HTTPS, improving transport security for internal and third-party services and documenting the rollout for future maintenance.
- Took features from requirements and design through implementation, testing, deployment, and documentation in collaboration with business analysts, testers, product owners and engineering team.

IT Officer Intern (AI Engineering) – Dr. Shafi & Partners (NHS)
Leicester, UK

Jun 2025 – Jul 2025

- Researched and developed an AI triage prototype for GP front-desk patient intake, using the Infermedica Engine API to conduct structured symptom interviews and provide recommended levels of care to a clinician dashboard.
- Delivered a real-time conversational voice bot in Python (Pipcat, FastAPI, SQLite) on Cerebrum, integrating an STT, LLM, and TTS pipeline to conduct symptom assessment interviews over voice via Daily.

Projects

Containerised Parallel Compute Cluster

- Designed a parallel prime factorisation pipeline for 10,000 large integers using Python, Celery, and Redis; partitioned work across four Dockerised workers and reduced runtime by 26% from 101.75s to 74.99s.

Self-Hosted LLM Inference Service

- Built and deployed a Dockerised self-hosted LLM inference service on a Linux homelab server, exposing HTTP APIs through Tailscale and serving quantised 4B-parameter models under a 4 GB VRAM constraint.
- Automated build and remote deployment with GitHub Actions over SSH.

Hackathons and Leadership

Let's Hack for Good 25/26 – University of Leicester
Leicester, UK

Dec 2025 / Apr 2026

- Helped organise the hackathon as a mentor representing NEXT, providing students with technical guidance on project scoping, idea validation, pitch preparation, and implementation.

Technical Skills

Programming: Python, C#, SQL, JavaScript

Concepts: Algorithms, Data structures, object-oriented programming

Systems and infrastructure: Linux, Docker, Docker Compose, SSH, Bash, Tailscale

Backend: ASP.NET Core, FastAPI, REST APIs, Multi-Tenancy

Database: Microsoft SQL Server, PostgreSQL, SQLite, MongoDB

CI/CD & Automation: GitHub Actions, Azure DevOps Pipelines, Bash, PowerShell