

Sumukh Adiraju

London, UK | +44 7768 893901 | sadiraju06@gmail.com | [linkedin.com/in/sumukh-adiraju-07235b31b](https://www.linkedin.com/in/sumukh-adiraju-07235b31b) | github.com/sadir06 | sumukhadiraju.com

EDUCATION

Imperial College London, MEng Electronic and Information Engineering 2024 – 2028

- Relevant coursework: Instruction Set Architecture, Compilers, Software Systems, Probability & Statistics

The Institute of Education, Irish Leaving Certificate

Sept 2021 – Jun 2024

- Achieved 100% in all four of Mathematics, Applied Mathematics, Physics, and Business; 97.2% in Economics; 88% in English

LEADERSHIP & AWARDS

- Selected as 1 of 8 from over 2,000 applicants for the inaugural AI Engine Room cohort (Chapter One); raised £5,000 to develop and prototype an AI startup idea from concept to working MVP
- Finalist, Top 6 of roughly 50 teams, Imperial IdeaForge; developed and pitched a startup concept to judges at final demo day
- Selected participant, Y Combinator Startup School (2025)
- Meta Engineer of the Future Award; CREST Gold Award
- Technical Director, Google Developer Student Club, Imperial College London

EXPERIENCE

Canopy Labs, Machine Learning Engineer

Mar 2026 – Jun 2026

- Designed distributed video preprocessing pipelines across hundreds of GCP CPU/GPU instances using Terraform, multiprocessing, staged execution, batching, and PostgreSQL-backed orchestration, increasing preprocessing throughput by approximately 10×
- Optimised inference for a 1.4B-parameter video diffusion transformer and VAE decoder on H100 GPUs, improving throughput to 2–4× realtime using torch.compile, FlashAttention, BF16 inference, batching strategies, and KV-cache optimisation
- Built an end-to-end data pipeline converting 1000+ hours of raw video and audio into training-ready latent and token representations using an in-house VAE, pose tokenizer, and Qwen ASR + forced aligner
- Designed and scaled GPU-accelerated speaker diarisation pipelines using NVIDIA NeMo (SoftFormer) and Pyannote to process over 1 million hours of audio, filtering high-quality training data at over 2000× realtime on H100 GPUs
- Developed multi-stage video quality evaluation pipelines combining YOLO, ViTPose, Qwen VLMs, OCR, and speaker diarisation to curate high-quality multimodal training datasets with approximately 10% acceptance rates

Vosyn Inc, Software Engineering Intern

Aug 2025 – Oct 2025

- Built automation and AI-powered web systems, including Selenium-based ad-posting pipelines and a custom AI-powered chatbot

PROJECTS

FPGA-Accelerated Crypto Trading System | [GitHub](#)

Feb 2026 – Mar 2026

- Designed a real-time crypto trading system using dual PYNQ-Z1 FPGA nodes streaming Binance market data, with cloud-synchronised state coordination via Flask and AWS
- Implemented FPGA-accelerated linear regression training using incremental outer-product accumulation, achieving approximately 68× faster weight updates versus a software baseline
- Built a full microstructure feature pipeline (17 features at 200ms cadence) with real-time WebSocket ingestion, NumPy vectorisation, and z-score normalisation
- Achieved 57.6% directional accuracy and 112% paper-trading return under leveraged simulation; optimised the FPGA datapath using a 512-bit AXI bus, 16-bit fixed-point arithmetic, and a DSP-heavy parallel design (77% DSP utilisation)

Reinforcement Learning Smart Grid Controller | [GitHub](#)

May 2026 – Jun 2026

- Designed and trained a PPO-based reinforcement learning agent in PyTorch to optimise energy arbitrage and deferred-load scheduling within a closed-loop smart grid driven by live electricity market data
- Developed a custom Gymnasium simulation modelling a physical microgrid with battery storage, photovoltaic generation, stochastic electricity prices, and hardware constraints
- Engineered a 21-dimensional state representation and continuous control policy, iterating through four simulation prototypes to improve training stability and reduce the simulation-to-hardware gap
- Deployed the trained policy on a physical laboratory smart grid, integrating Azure market data, Raspberry Pi Pico hardware control, and a Flask dashboard for live telemetry

RISC-V RV32I Processor, SystemVerilog

Oct 2025 – Dec 2025

- Architected a full RV32I processor, evolving from a single-cycle design to a 5-stage pipelined CPU (IF/ID/EX/MEM/WB) with hazard detection and data forwarding
- Built a 2-way set-associative write-back data cache with LRU replacement and a multi-cycle FSM for miss handling, writeback, and refills
- Debugged pipeline and cache correctness issues using waveform analysis, introducing shadow address latching to stabilise cache refills under stalls

C Compiler, Imperial EIE Coursework

Feb 2026 – Mar 2026

- Implemented a full C compiler in C++ (lexer, parser, AST, semantic analysis, code generation), supporting core C constructs with a modular front-end/back-end architecture

s3 Shell, POSIX-style UNIX shell in C

Nov 2025 – Dec 2025

- Implemented a UNIX shell supporting process creation, command execution, job control, I/O redirection, pipelines, batched commands, and built-in shell functionality
- Designed a recursive parser for nested subshells and complex pipelines, and built redirection infrastructure using fork(), execvp(), pipe(), and dup2(), with custom quote handling and error recovery

TECHNICAL SKILLS

Languages: Python, C++, C, SystemVerilog, SQL, Bash, Lua

Machine Learning: PyTorch, TensorFlow, Transformers, NumPy, Reinforcement Learning, PPO

Tools & Technologies: FPGA, RISC-V, Computer Architecture, WebSockets, Multithreading, Flask, FastAPI, REST APIs, MongoDB, Pandas, OpenCV, MediaPipe, Linux, UNIX, Git, Docker, Terraform, AWS, GCP, PostgreSQL, Neovim, Tmux