

Thomas Bale

thomasbale.com | tokbale@outlook.com | linkedin.com/in/thomas-bale

Experience

Demonstrator and Graduate Teacher, University of Bristol Aug 2025 –

- Delivering Lectures and Workshop Exercises. Mentoring a small team of students through their year long project. Helping students in workshops. (Software Engineering Project)
- Helping students in workshops. (Computer Systems A, Programming Languages and Computation)

UK HPC Student Team, UKSCC May 2025 – Jun 2025

- Represented the UK at ISC SCC. Optimised OpenMX and LLMs (llama) on a 208 core, 8xH100 cluster.
- Experience with load balancing, networking, OpenMP, MPI and CUDA programming.
- Worked with SLURM, LAPACK, BLAS, ScaLAPACK, FFTW, ELPA, OpenBLAS, Intel MKL and custom profiling.
- Explored LoRA, QORA, DoRA, transformer engine and FP8, flash attention 3 for llama 8B.

Machine Learning Research Assistant, University of Bristol Feb 2025 – Aug 2025

- Developed a scalable ML workflow for generating photorealistic, emotional faces for psychological research.

Machine Learning Software Engineer Intern, DigitalU3 Sep 2024 – Mar 2025

- Engineered a machine learning-based system with a strong focus on efficiency and scalability.
- Collaborated in an AGILE team, contributing in sprints, code reviews, and iterative delivery.
- Developed a scalable web application with integrated backend and structured database.

Education

University of Bristol, BSc in Computer Science Sep 2023 – Current

- 85.3% avg in Final Year; First Class
- Treasurer, Planning Control Lead in FSAI; Founder, President UoB Quantum Computing Society

Colchester Royal Grammar School, A-Levels Sep 2021 – Jul 2023

- A*A*AA - Computer Science (ranked 1st in cohort; 100% NEA), Maths, Further maths, Physics

Technical Projects

Quantum Cross-Chain Arbitrage - \$5000 Hackathon Win

- Built a hybrid quantum-classical arbitrage bot for executing cross chain flash loans.
- By using QAOA with custom graph construction, liquidity/slippage modelling, and flash loans with Vyper.

YukiGPT

- Built a decoder-only GPT-style Transformer from scratch in Python (following Attention Is All You Need).
- Implemented the self-attention mechanism using scaled dot-product attention within a decoder-only Transformer to model long-range token dependencies without recurrence or convolution.

Technologies and Skills

Languages: Python, Go, Java, C, C#, Haskell, HTML/CSS, JavaScript,

Frameworks/Tools: SQL, API/REST, Git/Github, React Native, PyTorch, TensorFlow, Qiskit, AWS, Cursor

Skills: AGILE & Test-Driven Development, Quantum Programming, Machine Learning (ML), Artificial Intelligence (AI), High performance computing (HPC), CAD, Generative AI, Computer Vision, LLM, A/B Testing, NLP