Arnab Mukherjee

arnabmukherjee791@gmail.com | linkedin.com/in/arnab | github.com/arnab

Education

The University of Edinburgh — Edinburgh, UK
Bachelor of Engineering Computer Science — *Sep. 2022 – May. 2026*

National University of Singapore — Singapore
Year-Long Academic Exchange — *Aug. 2024 – May. 2025*

Experience

Google Summer of Code 2024 with CERN — May 2024 – Aug. 2024
LHAPDF: Les Houches Accord PDF — *Remote*

- Creator and Developer of the LHAPDF dashboard- a web-based simulator for visualizing Parton Density Functions (PDFs) supporting 1D, 2D, and 3D plotting with real-time updates and uncertainty bands for physics research at the LHC.
- Integrated CERN backend with the LHAPDF library for the physics data processing, enabling on-the-fly parsing of large datasets (100MB+ per file) and supporting user-defined comparisons of PDFs with runtime validation and error handling.
- Selected for and successfully completed the GSoC program mentored by Professors Andy Buckley (Professor of Particle Physics at the University of Glasgow) and Chris Gutschow (UCL Centre for Advanced Research Computing).

Research Intern — May 2024 – Aug. 2024
ICSA: Institute of Computer Systems Architecture — *The Informatics Forum, Edinburgh*

- Interned at the NetSys Lab under Professor Mahesh Marina (Edinburgh and Johns Hopkins, Ex-Director of ICSA).
- Primary contributor to Morphling, an emulator for distributed training of deep neural networks (DNNs) on edge devices. Implemented low-level function interception (e.g., MKL, OpenBLAS, CUDA) and developed performance models for realistic computation and communication emulation.
- Developed a unified memory management subsystem with static analysis for model checkpointing and shared memory, enabling the emulation of 100+ edge devices on consumer-grade hardware.

INF1A Tutor and Marker — Sep 2023 – Dec 2023
University of Edinburgh — *Edinburgh, UK*

- Taught Haskell and Computational Logic to 40 students as a tutor for the course Introduction to Computation, providing support for weekly tutorials and evaluating all submitted work.
- Achieved a perfect score (100%) on this course during my first year.

Data Engineer Intern — Jun 2023 – Aug 2023
Indian School of Business — *Hyderabad, India*

- Developed a pipeline for collecting, processing, and analyzing 500,000 public policy records, optimizing the data ingestion process to reduce latency by 2.8x using PyTorch and advanced preprocessing techniques.
- Designed and implemented a relational database in PostgreSQL, handling 3 million historical policy entries with optimized indexing and query strategies, reducing data retrieval time by 45%.
- Fine-tuned ETL workflows, integrating Python, SQL, and Pandas, and improved data throughput, enabling faster updates and real-time analytics on public policy trends.

Projects

LLaMAAdapt: Adaptive Fine-Tuning Framework for Low-Resource Applications — September 2024

- Built a fine-tuning framework for LLaMA models focused on memory efficiency, implementing techniques like LoRA (Low-Rank Adaptation) and prefix-tuning to enable training on GPUs with as little as 8GB of memory.
- Designed a modular PyTorch Lightning pipeline with configurable components for dataset preprocessing, model tuning, and checkpoint management, streamlining fine-tuning workflows for varied applications.
- Integrated a real-time monitoring dashboard with Streamlit, providing detailed information into training metrics, attention head activations, and token embedding evolution, significantly reducing debugging and optimization time.