07471 812947
London
lucas.curtin@gmail.com

Lucas Curtin

Data Scientist

GitHub: lucas-curtin
Website: lucas-curtin

Objective

I am an MPhil student in Data Intensive Science at the University of Cambridge and a Data Scientist at Kantar, specialising in predictive modelling, Bayesian inference and high-performance computing. I seek to apply my expertise in scenario simulation, anomaly detection and statistical analysis to risk modelling and mitigation, partnering with cross-functional teams to deliver actionable insights and robust risk management strategies.

Skills

Programming & Tools: Python, SQL, C++, Git, Docker, Bash, AWS, TensorFlow, PyTorch
Quantitative Techniques: Time-series modelling, Monte Carlo simulation, stochastic calculus, regression and multivariate analysis, optimisation and numerical methods, principal component analysis, Bayesian inference
Data & Visualisation: Databricks, Snowflake, Spark, Tableau

Education

University of Cambridge **Oct 2024 – Present**
MPhil in Data Intensive Science (predicted First Class Honours)

- **Advanced Statistical Methods**: Bayesian and frequentist inference, likelihood estimation, Gaussian mixture modelling and high-dimensional joint probability projection
- **Machine Learning Theory & Application**: design and deployment of complex machine learning architectures such as fine-tuning LLMs and image analysis pipelines.
- **High-Performance Computing**: parallelisation with MPI and OpenMP, GPU-accelerated workflows and development of performance-critical kernels in C++
- **Software Engineering**: Linux and Bash automation, Python package development and release of a Cython-based automatic differentiation library
- **Capital Markets Fundamentals**: completion of Cambridge Summer School in Finance, covering fixed income, derivatives pricing and market risk management

University College London **Oct 2019 – Jul 2023**
Physics MSci (Second Class Honours, Upper Division)

- **Advanced Statistical Mechanics**: stochastic methods, Wiener processes, Brownian motion, and partition function theory
- **Advanced Quantum Theory**: perturbation theory, scattering matrices, relativistic wave equations
- **Practical Machine Learning for Physicists**: supervised and unsupervised techniques applied to experimental datasets, feature engineering and model validation
- **Particle Physics Experiments**: calculated the approximate lifetime of muons using a FPGA and plastic scintillator
- **Achievements**: Awarded first prize for a cohort-wide statistical analysis of maritime scrubber emissions

Professional Experience

Kantar, London **Jul 2023 – Present**
Data Scientist – Forecasting & Analytics

- **End-to-End Panel Health Pipeline & Risk Detection**: designed and maintained a Databricks-Snowflake pipeline ingesting 50M+ records weekly, automating data transformation, displaying dashboard metrics to monitor panel health, and developed risk metrics to minimise meter recycling costs.
- **Lead Developer & Client Delivery**: oversaw multiple international data services, delivered tailored analytics and predictive insights to optimise content strategies, and mitigated operational risks concerning demographic panel coverage
- **Software Engineering & Architecture**: refactored and modularised legacy Python and SQL codebases, building reusable libraries for use across multiple services
- **Data Science & Visualisation**: engineered features and trained XGBoost and LightGBM ensembles for panel scoring, and created interactive Tableau and Power BI dashboards to communicate KPIs and risk metrics to stakeholders
- **High-Performance Data Processing**: optimised Spark and Pandas workflows with parallelism and vectorisation, integrating intensive routines into SLURM jobs for timely analytics

Amazon, East London **Jul 2022 – Oct 2022**
Operations & Logistics Associate

- **ML Pipeline & Risk Prediction**: developed a scraping workflow on internal dashboards in order to preprocess logistics data and predict package misrouting incidents
- **Data Integration & Compliance Checks**: integrated geospatial compliance checks for delivery routes, detecting anomalous patterns and enforcing standards to mitigate fraud risk
- **Visualisation & Stakeholder Reporting**: built daily dashboards tracking case escalations, enabling data-driven interventions by branch managers
- **Cross-Functional Analysis & Recognition**: authored and presented a detailed analytics whitepaper on pipeline methodology and impact, earning an Operational Excellence award

Personal Projects & Hackathons

Event Timetable Extraction for Labour Party Conference **2025**
Contracted Web-Scraping Project

- Scraped dynamic conference data using Python, BeautifulSoup and Selenium.
- Parsed session titles, speakers, times and venues into structured JSON/CSV for downstream analytics.

IBM Quantum Challenge **2024**
Quantum Computing Personal Project

- Completed IBM's Quantum Challenge by designing and optimising quantum circuits using Qiskit on IBM Quantum Experience.
- Implemented VQE for molecular energy estimation and Grover's search on real quantum hardware, applying error-mitigation techniques.
- Used Qiskit Terra and Aer for simulation, then benchmarked results on IBM Quantum backends.

Custom Parachain for Insurance Workflows (Hackathon Winner) **2023**

- Conceptualised and prototyped a Polkadot parachain using zero-knowledge proof contracts to combine NHS records, employment history and smart-car telematics for privacy-preserving premium calculation
- Authored test scenarios and demo scripts illustrating secure ZKP-based information flows and GDPR-compliant data exchanges
- Designed an on-chain identity verification and claims pipeline to reduce fraud, streamline underwriting and dynamically adjust premiums

Dynamic Art NFT Platform on Tezos (Hackathon Runner Up) **2022**

- Prototype NFT ecosystem where artworks adapt their metadata in response to simulated audience engagement metrics
- Implemented on-chain triggers to showcase evolving art pieces, emphasising how digital tokens can drive physical gallery and event attendance
- Demonstrated the concept of merging blockchain interactivity with cultural experiences in a post-pandemic context

Research Experience

T3Net: Joint PDF & SMEFT Inference **Jul 2025**
MPhil Thesis

- Created neural-network parameterisation to extract non-singlet PDF and SMEFT Wilson coefficient
- Applied replica-based Monte Carlo training with Lagrange penalties to effectively propagate uncertainties
- Tested and verified ongoing global QCD analysis

LLM-Based Time-Series Forecasting **Apr 2025**
LoRA Fine-Tuning of Qwen2.5-Instruct

- Adapted large language model for forecasting of Lotka-Volterra dynamics
- Implemented percentile shift-scale preprocessing alongside custom text token encoding
- Employed LoRA for parameter-efficient attention fine-tuning

2D Heat Diffusion Simulation **Dec 2024**
HPC Optimisation & Parallelisation

- Rewrote finite-difference solver in C++ with SIMD intrinsics; 3$\times$ speedup over scalar code
- Applied MPI domain decomposition and OpenMP threading
- Validated performance on 64-node cluster which sustained 80% parallel efficiency

Quantum Track Reconstruction **Apr 2023**
UCL MSci Thesis

- Reformulated LHC track reconstruction as graph classification
- Designed hybrid quantum-classical GNNs integrating parameterised quantum circuits
- Introduced 'entangling dropout' to investigate classical analogies in quantum machine learning