Joel Smith

SUMMARY

Data Science & Artificial Intelligence MSc graduate with a strong focus on building reliable, well-structured machine learning systems. Experienced in translating mathematical models into clean, reproducible code, and in designing end-to-end pipelines spanning data, training, evaluation, and deployment. Demonstrated depth across generative modelling and applied ML systems, with hands-on experience using PyTorch and modern ML tooling.

Education

University of Liverpool Sept 2024 - Sept 2025

MSc in Data Science & Artificial Intelligence, Distinction

- Thesis: *Manipulation Lab* — a configuration-first benchmarking framework for robotic manipulation in Isaac Sim/Isaac Lab (80%).
- Addressed benchmark fragmentation by designing a single execution harness where benchmarks are specified as shareable Hydra configuration bundles, reducing setup overhead across tasks, sensors, and evaluation protocols.

University of Chester Sept 2020 - July 2023

BSc in Economics, First Class

- Awarded the University Prize for Economics.

Selected Projects

Alchemy Lab GitHub Repo

- Built an installable diffusion-model research framework in PyTorch (pyproject/uv) with CLI entrypoints for training/evaluation/sampling, enabling reproducible runs outside of notebooks.
- Implemented configuration-driven experimentation, checkpoint/resume and EMA, W&B/TensorBoard logging, and reproducibility primitives (seed control, config snapshots) with automated evaluation hooks (sampling/metric logging).

Manipulation Lab GitHub Repo

- Built an installable benchmarking and data-collection toolkit (pyproject) for robotic imitation learning in NVIDIA Isaac Sim/Isaac Lab, parameterising task variants and sensor configurations via Hydra.
- Implemented dataset recording and playback for expert demonstrations (trajectory/state/action logging) and standardised evaluation (success rate, completion time), enabling fair comparisons across policies and observation setups.
- Reduced experiment implementation overhead by separating environment/task definitions from training/evaluation code, allowing new tasks or observation stacks to be added with minimal code changes.

Fin-LLM GitHub Repo

- Designed and implemented a full-stack LLM-based financial Q&A system to reduce information asymmetry in retail finance, integrating retrieval-augmented generation with real-time market APIs and vector search (ChromaDB/pgvector) to ground responses in up-to-date, verifiable data.
- Engineered a production-oriented inference pipeline with a FastAPI backend and React frontend, containerised via Docker and deployed on a local Kubernetes (kind) cluster to support modular services and reproducible serving.

Technical Writing

Generative Modelling Blog

- Authored a blog series of technical studies on generative models (e.g., mixture models, VAEs, diffusion models) developing detailed mathematical derivations and explaining core probabilistic concepts from first principles.
- Produced supporting PyTorch implementations and visualisations to illustrate the theory and compare learned generative distributions against real data.

Experience

Commercial Finance Analyst June 2024 – Sept 2024
Matalan Liverpool, UK

- Owned weekly sales performance reporting to the British Retail Consortium (BRC), using SQL to query, reconcile, and validate 10M+ row datasets across multiple tables before producing channel breakdowns (e.g., online vs in-store revenue) under fixed deadlines.
- Produced monthly sales and promotion performance reporting for senior stakeholders/executives, combining repeatable SQL extracts with ad-hoc analysis to support pricing and promotional decisions.
- Maintained lightweight dashboards (e.g., Tableau/Excel) for ongoing sales and promotion tracking.

Team Supervisor Sept 2020 – January 2024
Costa Coffee Liverpool, UK

- Supervised day-to-day operations in a high-throughput store, developing leadership and stakeholder communication while completing a full-time degree.

Skills

Machine Learning: PyTorch, scikit-learn, Hydra
Data & Experimentation: PostgreSQL, Google BigQuery, Microsoft SQL Server, Weights & Biases
LLM & Retrieval: OpenAI API, ChromaDB, pgvector
Systems & Deployment: FastAPI, Docker, Kubernetes (kind), Git, Linux, AWS (EC2)
Languages: Python, SQL