

Shayan Shafquat

📍 Nottingham, UK 📞 +44-7741926490 ✉ shafquat.shayan@gmail.com in shayan-shafquat 🌐 shayanshafquat

Senior Data Scientist with 2.5+ years of experience, delivering production-grade AI/ML solutions in insurance, transportation & security domains. Skilled in multimodal learning, deep learning (NLP, CV, RL) and deploying LLM applications (RAG, Guardrails). Proficient with PyTorch/TensorFlow, serverless infra, and CI/CD pipelines for scalable LLM deployments.

Work Experience

Allstate | SENIOR DATA SCIENTIST

Northern Ireland, UK (Hybrid) | Jan 2025 – Present

- Trained **deep networks** on imbalanced time-series for total loss prediction, improving capture rate by **1.8%** in prod
- Enhanced LLM-powered Prototype with token-aware context handling, **Map-Reduce** summarization & **evaluation**
- Designing scalable ML infra for **200K daily calls** using Lambda, Bedrock, Jenkins; optimized for **latency & uptime**

Enkrypt AI, Inc | AI RESEARCH CONSULTANT

Boston, US (Remote) | Jun 2023 – Sep 2023

- Developed a **malicious file scanner** for Python scripts and Jupyter notebooks, focusing on securing the ML lifecycle
- Curated adversarial datasets; tested LLM pipelines under **prompt injection** and **backdoor attacks** using Rebuff
- Designed a security layer using **RAGAS** and **GuardRails** to defend LLM pipelines from **toxicity, bias and jailbreaks**

ANI Technologies Private Limited (OLA) | DATA SCIENTIST

Bangalore, IN | Sep 2020 – Jan 2022

- **Pay after ride user identification:** Impact: Increased cashless ride by **5%** with no change in default rate (**2.5%**)
 - Enhanced the **feature store** API to analyze the behaviour of **15M+** users to establish trusted user base of 0.5M
 - Deployed bagging **ensemble model** (AUC: 0.91) to infer user trust levels for expanding cashless ride coverage
 - Modeled **default likelihood** across non-trusted users, whitelisting **8.7M** while adhering to **risk** tolerance limits
- **Improvements in peak pricing module:** Impact: Reduced conversion prediction error by **2.2%**
 - Designed recency-weighted regression **fallback** for real-time **pricing** ensuring continuity during infra failures
 - Led regression-driven **dynamic pricing** automation for bikes using spatiotemporal cab's **demand-supply** signals
 - Optimized modules for **hotspot pricing** and in-trip cab inclusion in **supply signal**, with automated reporting
- **User level pricing:** Impact: **+1.5%** GMV/bookings and improved net completion during COVID-19
 - Automated user-specific coefficients from conversion rates via **cron jobs**, enhancing **demand signal** accuracy
 - Validated fare sensitivity and quantified **churn risk** through a **control-test study**, driving pricing optimizations
- **Traffic Lights Optimization:** Achieved an 18.3% reduction in wait times by optimizing traffic light signals with a Deep Q-learning RL agent in the SUMO simulation, effectively minimizing congestion
- **System:** Enhanced the in-house utility package, streamlining data preprocessing and retrieval from **AWS S3**

ANI Technologies Private Limited (OLA) | RESEARCH ENGINEER- INTERN

Bangalore, IN | May 2019 - Jul 2019

- Engineered driver login behaviour models using **decision trees** to segment cohorts, optimizing incentive allocation
- Conducted **A/B testing** on incentive strategies in Pune and Kochi, minimizing burn while maintaining engagement

Key Technical Skills

Programming and Scripting Languages: Python, C++, SQL, MATLAB, R, Bash, HTML/CSS, JavaScript

Frameworks and Libraries: LangChain, LlamaIndex, CrewAI, PyTorch, Tensorflow, Scikit-learn FastAPI, Seaborn, SciPy, Terraform

Tools and Platforms: Git, PySpark, SageMaker, Lambda, GuardRails, HuggingFace, Docker, Git Actions, FAISS, Jenkins, Ollama, Streamlit

Projects

Intelligent RAG-Based Research Summarization and Q&A Platform

OCT 2024 - PRESENT

- Built modular RAG platform integrating **OpenAI/Llama** with external web & document retrieval for structured **Q&A**
- Developed a full pipeline with **UI & CI workflows** for evaluating retrieval-generation outputs using GitHub Actions

Stochastic models of exploration in patch foraging tasks

MAY 2024 - SEP 2024

Master's Dissertation | Guide: Prof. Mark Humphries | University of Nottingham

- Simulated foraging behaviours with resource depletion models and compared leave time predictions to **MVT**
- Evaluating **stochastic action selection** algorithms like epsilon-greedy, mellowmax using human patch-foraging data

Machine Learning in Science | COURSEWORK

University of Nottingham | Oct 2023 - May 2024

- **Optimizing CNNs for real-time autonomous driving on AutoPicar**
 - Developed **CNN** models using **transfer learning** for speed and angle prediction, addressing class imbalance
 - Enhanced model performance through data augmentation, **architectural modifications** and data collection
 - Deployed the **tensorflow lite** models on toy car with **TPU** and camera, testing on three tracks and 12 scenarios
- **Developing 2D drone navigation with reinforcement learning**
 - Tuned heuristic approaches and created a **discretized action space** for rapid movement and stable landing
 - Trained and evaluated **DQN, Q-learning** models on average steps taken, fuel consumption and average thrust

Conversion of large-scale cortical models - INCF

GOOGLE SUMMER OF CODE 2022

Mentor: Prof. Padraig Gleeson, Ankur Sinha | UCL London

- Converted the channels, morphological and biophysical properties in L5 pyramidal cell from **NEURON** to **NeuroML**
- Implemented **GitHub Actions** to ensure the multi-compartmental cell model's behavior during development

AI Ethics and NLP Research | SELF-MOTIVATED

IIT KGP | Dec 2018 - Feb 2020

- **Fairness-aware data representation:** Reimplemented the **iFair** framework, minimizing **utility and fairness loss**, improving fairness by 3% and 9% with a 7% and 1% accuracy trade-off on Census and German credit datasets
- **Rumour detection in tweets:** Utilized **one-hot encoding** on characters to represent both tweets and comments, trained a **CNN** on named entities, and classified using a **GRU**, achieving a 68% accuracy
- **Hyperpartisan News Detection:** Designed and evaluated a **self-attention model** for bias classification in the **SemEval 2019** competition, outperforming the benchmarks set by **HCAN** and **Empath** models

Education

MSc. Computational Neuroscience, Cognition and AI

UNIVERSITY OF NOTTINGHAM

Nov 2024 | Nottingham, UK

GRADE: DISTINCTION

Integrated MSc. Mathematics and Computing

INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR

Jun 2020 | Kharagpur, IN

CGPA: 7.2 / 10.0

Academic and Technical Training

- **Mathematics:** Linear Algebra, Stochastic Process, Operations Research, Statistics for Applications
- **Machine Learning & AI:** Regression & Time Series Models, Neural Computation, AWS (SageMaker & Bedrock)
- **Software & Cloud Engineering:** Data Structures & Algorithms, Systems Design, AWS (Lambda & IAM), Terraform

Achievements

- Graduated with **Distinction**; ranked **top for MSc thesis** focused on RL and stochastic human foraging models
- Gold winning LSTM model comparing India's top mutual fund houses in the Inter Hall **Data Analytics** 2018
- Achieved **99.5** percentile score among 0.2M students who appeared in the **JEE Advance** examination
- Eligible to receive the **INSPIRE** scholarship by the Ministry of Science and Technology for the undergraduate study
- Part of the Gold winning **Football** team in the Inter Hall Sports Championship for the year 2017 and 2019