# Shayan Shafquat

+44-7741926490 shafquat.shayan@gmail.com linkedin.com/in/shayan-shafquat github.com/shayanshafquat

## Education

**MSc. Computational Neuroscience, Cognition and AI** — Sep 2024 | Nottingham, UK
University of Nottingham — Tentative Grade: Distinction

**Integrated MSc. Mathematics and Computing** — Jun 2020 | Kharagpur, IN
Indian Institute of Technology, Kharagpur — CGPA: 7.2 / 10.0

## Work Experience

**Enkrypt AI, Inc** | AI Research Consultant (Part-Time, Remote) — Boston, US | Jun 2023 – Sep 2023

- Developed a **malicious file scanner** for python scripts and jupyter notebooks, focusing on securing the ML lifecycle
- Surveyed LLM security, emphasizing **prompt injection, backdoor attacks** and strategies for adversarial robustness
- Integrated **RAGAS**, **NeMo GuardRails** and **rebuff** into a package for future seamless LLM security and monetization

**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
  - Trained an **ensemble model** (balanced bagging classifier) on trusted users, achieving an **AUC score** of 0.91
  - Predicted default probabilities on **non-trusted** base, whitelisting **8.7M** users while maintaining the default rate
- **Improvements in peak pricing module**: Impact: Reduced conversion prediction error by **2.2%**
  - Developed a recency-weighted heuristic **fallback** model for peak pricing during outages, minimizing **loss**
  - Built and **deployed** a bike peak pricing module for bikes using spatiotemporal cab data and **regression** model
  - Maintained 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 the hypothesis that abrupt fare changes cause regular **user churn** through a **control-test** study
- **Collaborative key involvements:**
  - **Traffic Lights Optimization:** Trained a Deep Q-learning RL agent in the SUMO environment using a reward function focusing on minimizing wait times and congestion, achieving an 18.3% reduction in wait times
  - **System:** Enhanced the in-house **utility package** for data preprocessing, streamlining retrieval from **AWS S3**
  - **Ola Foods (Grocery):** Developed a baseline model using item-based kNN, predicting next basket items
  - **Technical Interviews:** Led coding, statistics, and ML interviews for the initial round of the tech hiring process

**ANI Technologies Private Limited (OLA)** | Research Engineer- Intern — Bangalore, IN | May 2019 - Jul 2019

- **Feature-engineered** partner's ride data, using **decision trees** to predict login hours and form homogenous cohorts
- **Optimized** incentives for each cohort independently, minimizing **burn** while considering constraints on login hour
- Conducted **A/B testing** in Pune and Kochi to validate and ensure the effectiveness of the incentive optimization

## Technical Skills

**Experience with:** Python, C++, Git, AWS, Docker, Kubernetes, SQL, Linux, Bash, R, Hive, Spark, LaTeX, NEURON, OpenSourceBrain
**Python Packges:** Tensorflow, LangChain, LlamaIndex, Numpy, Pandas, Scipy, Sklearn, Scikit-learn, Seaborn, PyTorch, PyNeuroML, NLTK
**Modelling:** DQN, Dyna-Q, FineTuning, Transformers, RNN/LSTM/GRU, VAE, GANs, CNN, Gradient Boosting, Linear Regression, SVM

## Projects

**Stochastic models of exploration in patch foraging tasks** PRESENT

Guide: Prof. Mark Humphries | School of Psychology | 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 - Present

- 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
- 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

**Conversion of large-scale cortical models - INCF** GOOGLE SUMMER OF CODE 2022

- Converted the channels, morphological and biophysical properties in L5 pyramidal cell from **NEURON** to **NeuroML**
- Simulated the cell's expected behaviour and shared it on **Open Source Brain** for visualization and experimentation
- Implemented **CI/CD testing** to ensure the integrity of the multi-compartmental cell model throughout development

**iFair - AI and Ethics** FEBRUARY 2020

- Learned a **generalized data representation** preserving **fairness-aware similarity** between individual records
- Developed the combined objective function involving **utility** and **fairness loss** and minimized that using L-BFGS
- Applied the method on two classification tasks of **Census** and **German credit** dataset with **gender**, **age** as protected group resulting in a gain of **3%**, **9%** in **consistency** (fairness) and a drop of **7%** and **1%** in accuracy respectively

**Portfolio Optimization involving System of Linear Interval Equations** JANUARY 2020

Guide: Prof. Geetanjali Panda | Department of Mathematics | IIT Kharagpur

- Designed a problem of **portfolio optimization** involving equations of return, risk and utility with interval parameters
- Investigated and programmed the concepts of **regularity** in interval matrices as a necessary assumption condition
- Used **least squares** to obtain cost function for the problem and minimised that using iterative **gradient descent**

**NLP research projects** | SELF-MOTIVATED IIT Kharagpur | Dec 2018 - May 2019

- **Rumour detection in tweets**: Guide: Prof. Pawan Goyal | Department of Computer Science
  - Transformed tweets and comments using **one-hot encoding** on each character to feed as input for the **CNN**
  - Trained a CNN with label as named entities using **entity tagger tool**, finally obtaining the pre-final learned layer
  - Classified tweets by sequentially parsing the pre-final layer of the tweet and comments onto the **GRU** model
- **Hyperpartisan News Detection** | Competition: SemEval 2019
  - Implemented **Hierarchical Convolutional Attention Networks** for classifying articles on 5 classes of biasness
  - Implemented **Empath** model which analyzes articles on **lexical categories** using them for feature engineering

## Relevant Coursework

- **Offline:**
  - Regression and Time series model
  - Data Structure and Algorithm
  - Neural Computation
  - Practical Biomedical Modelling
  - Object Oriented Systems Design
  - Stochastic Process
- **Online:**
  - Linear Algebra by Prof. Gilbert Strang
  - Statistics for Applications by Prof. Philippe Rigollet

## Achievements

- Eligible to receive the **INSPIRE** scholarship by the Ministry of Science and Technology for the undergraduate study
- 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