

# Muhammad Faaiz Taufiq

☎ +44.7459.833939 | 🔗 faaizt.github.io | ✉ faaiztaufiq@gmail.com | 📍 Oxford, UK

## Education

### University of Oxford

Oct 2020 – Present

*PhD in Statistical Machine Learning, Google DeepMind Scholar*

Oxford, UK

- **Supervisors:** Rob Cornish, Arnaud Doucet and Yee Whye Teh
- **Research topics:** LLM alignment, distribution-free uncertainty quantification, causal inference, fairness, explainability, reinforcement learning.

### University of Cambridge

Oct 2017 – Jul 2018

*Master of Mathematics (MMath)*

Cambridge, UK

- **Result:** Honours pass with Distinction – Ranked among top 5% of the cohort.
- **Modules:** Machine Learning in Practice, Modern Statistical Methods, Topics in Convex Optimisation, Advanced Probability and Advanced Financial Models.
- **Research Essay:** Generative Adversarial Networks (Score: 93/100).

### University of Cambridge

Oct 2014 – Jul 2017

*BA Mathematics*

Cambridge, UK

- **Result:** Class I (each of the three years)
- **Modules:** Applied Probability, Statistics, Markov Chains, Principles of Statistics, Statistical Modelling, Stochastic Financial Models and Optimisation.

## Experience

### ByteDance

May 2023 – Jan 2024

*Research Intern, Responsible AI Team*

London, UK

- **Dataset Fairness:** Proposed a computationally efficient framework for understanding the fairness-accuracy trade-offs in machine learning tailored to datasets across various modalities, including computer vision and natural language datasets. This work is under review at ICML 2024.
- **Trustworthiness of LLMs:** Co-authored a comprehensive survey on LLM trustworthiness, and conducted measurement studies to evaluate alignment and adherence to social norms in widely-used models.

### Amazon Web Services

May 2022 – Nov 2022

*Research Intern, AWS Causality Team*

Tübingen, Germany

- **Robust Explainability Methods:** Used ideas from causal inference to develop a robust Shapley method for explaining model predictions. This work was accepted to AISTATS 2023.
- **Conformal Root Cause Analysis:** Proposed a novel root cause analysis methodology with strong theoretical guarantees based on Conformal Prediction.

### Morgan Stanley

Sep 2018 – Sep 2020

*Associate, Application Developer*

London, UK

- Designed and developed software for a proprietary risk calculation platform to increase computational efficiency.
- Partnered with sales teams to develop an award-winning data intelligence engine to streamline analysis.
- Completed a comprehensive training program spanning a range of programming languages and technologies including Java, C++, Python, Scala, C#, SQL and Linux.

## Citigroup

Quantitative Placement Analyst, Exotics Trading Desk

Jun 2018 – Aug 2018

London, UK

- Created a Risk Analytics Tool to carry out daily risk metric calculations.
- Automated the calculation, visualisation and distribution of trading metrics to facilitate timely data analysis.
- Collaborated with traders on a project to mark volatilities for long-dated exotic products on single stocks.

## J.P. Morgan

Cyber Security Summer Intern

Jun 2017 – Aug 2017

London, UK

- Led and managed the delivery of a complex analytical tool to understand the Cyber Controls environment.
- Designed, developed and delivered software for the visualisation of Cyber Threats and Controls Data.

## Dyson Ltd

Research Intern

Jun 2016 – Sep 2016

Wiltshire, UK

- **Research Project:** An investigation into the resonant frequencies of metallic structures.
- Developed models to simulate the structural vibration properties for metals.
- Leveraged research findings to devise techniques for structural noise reduction and presented these insights to the company's founder, Sir James Dyson.

## Publications

---

- [1] Achievable Fairness on Your Data with Utility Guarantees. **M.F. Taufiq**, J.F. Ton, Y. Liu. *Under review at NeurIPS 2024.*
- [2] Marginal Density Ratio for Off-Policy Evaluation in Contextual Bandits. **M.F. Taufiq**, A. Doucet, R. Cornish, J.F. Ton. *NeurIPS 2023.*
- [3] Trustworthy LLMs: A Survey and Guideline for Evaluating Large Language Models' Alignment. Y. Liu, Y. Yao, J.F. Ton, X. Zhang, R.G.H. Cheng, Y. Klochkov, **M.F. Taufiq**, H. Li. *NeurIPS 2023 Workshop on Socially Responsible Language Modelling Research (SoLaR).*
- [4] Manifold Restricted Interventional Shapley Values. **M.F. Taufiq**, P. Blöbaum, L. Minorics. *AISTATS 2023.*
- [5] Conformal Off-Policy Prediction in Contextual Bandits. **M.F. Taufiq\***, J.F. Ton\*, R. Cornish, Y.W. Teh, A. Doucet. *NeurIPS 2022* & Spotlight presentation at *ICML 2022 Workshop on Distribution-Free Uncertainty Quantification.*
- [6] Causal Falsification of Digital Twins. R. Cornish\*, **M.F. Taufiq\***, A. Doucet, C. Holmes. *Under review at Biometrika.*

## Skills

---

**Programming Languages** : Python, Java, Scala, C++, C#, SQL, R, MATLAB

**Frameworks** : Pytorch, Scikit-learn, Numpy, Pandas, Pyro, Keras

## Invited Talks

---

**On The Robustness of XAI Methodologies.** The appliedAI Institute for Europe gGmbH, Germany

Sep 2023

## Awards

---

Google DeepMind Scholarship to study PhD at Wolfson College, Oxford	2020
Hugo de Balsham Prize for Exceptional Academic Distinction. <i>Peterhouse, Cambridge</i>	2018
Charles Clemmow/Harry Benton Talbot Prize for Mathematics. <i>Peterhouse, Cambridge</i>	2017
Senior Scholarship in Mathematics and Mathematics College Prize. <i>Peterhouse, Cambridge</i>	2014 – 2017
800th Anniversary Scholarship to study Mathematics at Peterhouse, Cambridge. <i>Cambridge Trust</i>	2014 – 2018
Feroze Khan Noon Scholarship awarded for academic excellence. <i>Noon Foundation</i>	2014 – 2015
Honourable Mention Award in International Mathematical Olympiad (IMO)	2013
Honourable Mention Award in Asian Pacific Mathematics Olympiad (APMO)	2013