Israel Mason-Williams

+447517672129 israel.mason.williams@gmail.com linkedin.com/in/israelfmw/ https://github.com/IFMW01

Education

Imperial and Kings UKRI Safe and Trusted AI CDT **October 2024 – June 2028**
Doctor of Philosophy (D.Phil) in Computer Science *London,UK*

- **Thesis Title**:"Towards Robust Reasoning in Language Models via Mechanistic Interpretability" under the supervision of Helen Yannakoudakis

The University of Cambridge **October 2023 – June 2024**
Master of Philosophy (M.Phil) in Advanced Computer Science: ***Distinction 79.4%*** *Cambridge,UK*

- **Simms Award for Academic Achievement:** Presented by Lucy Cavendish College
- **Thesis Title**:"Machine Unlearning In Audio Analysis For Data Privacy"
- **Modules Studied**: Machine Learning and The Physical World, Principles of Machine Learning Systems, Practical Research Human-Centred AI, Advanced Topics in Machine Learning (Explainable AI, Multi-modal AI and Narratives of AI) and Theory of Deep Learning.

Queen Mary University of London **September 2020 – June 2023**
Bachelor of Science (BSc) in Computer Science: ***First Class Degree 78.6%*** *London,UK*

- **Notable Positions**: Chair of the Student Staff Liaison Committee and Student Representative for CS.
- **Thesis Title**:"Developing Novel Machine Learning and Data Processing Approaches for Computational Enzymology", a collaboration with the European Bioinformatics Institute (Submitting for publication).
- **Third Year Modules Studied**: Computability, Complexity and Algorithms, Data Mining, Big Data Processing, Bayesian Decision and Risk Analysis, Neural Networks and Deep Learning and Digital Media and Social Networks.

Publications

Knowledge Distillation: The Functional Perspective submitted to NeurIPS 2024 Workshop SciForDL: **Paper/Poster Presentation** (1)
Explicit Regularisation, Sharpness and Calibration submitted to NeurIPS 2024 Workshop SciForDL: **Paper/Poster Presentation** (2)
Neural Network Compression: The Functional Perspective submitted to ICLR 2024 Workshop PML4LRS: **Paper/Poster Presentation** (3)
Neural Networks Are Low Rank Learning submitted to Erlangen AI Hub Conference: **Poster Presentation**
Decomposed Learning an Avenue for Mitigating Grokking submitted to Erlangen AI Hub Conference: **Poster Presentation**

Awards & Extra Curricular

Extra Curricular: Cambridge AI Safety Fellow at the Cambridge AI Safety Hub, LSE Ethics of AI, NVIDIA Fundamentals of Deep Learning and Cambridge Centre for AI in Medicine Summer School Participant.
Awards: NeurIPS Simons Foundation Scholarship 2024, Finalist in The Undergraduate of The Year Through Adversity 2021, Tata Consultancy Services Digital Explorers Scholarship Recipient 2020 and 2021 and Sports Scholarship Recipient at Bryanston School 2019.

Experience

Talos Network - EU AI Policy Fellowship **December 2024 – March 2025**
AI Researcher *Remote*

- Outcomes: Furthering international collaboration on emerging AI governance via an 8-week EU AI Policy Fundamentals Programme and 7-day policymaking summit in Brussels. Research paper submitted to the ICML Workshop on Technical AI Governance.

Queen Mary University of London - Visiting Researcher **September 2024 – September 2025**
AI Fundamentals Researcher *Remote*

- Outcomes: Producing publications for NeurIPS, ICML and ICLR focusing on fundamental approaches to AI research.

University of Exeter - EPSRC Quantitative Modelling in Healthcare **June 2023 – September 2023**
AI Research Scientist *Remote*

- Project: Developing cutting-edge machine learning models for forecasting days of progression-free survival for IPF patients of The Royal Hospital Bath and the North Bristol Trust Hospital.
- Methodology: Integrated mechanistic modelling of lung function with advanced machine learning techniques.
- Outcomes: Successfully created state-of-the-art deep learning models capable of forecasting progression-free survival days for patients to previously unseen accuracy.
- Publication: A forthcoming research paper will be published in 2024. **Best Presentation Award at the Biophysical Modelling in Respiratory Medicine - Network Conference 2024**.

European Bioinformatics Institute - The Thornton Group **September 2022 – Ongoing**

Part-Time AI Research Scientist Intern *Welcome Genome Campus, Cambridge*

- Project: Developing Novel Machine Learning and Data Processing Approaches for Computational Enzymology.
- Methodology: Combine Alphafold-predicted protein structures and existing structural analysis methods to create novel machine learning methods for enzyme functional prediction.
- Outcomes: Created a data pipeline and series of machine learning models to predict the catalytic properties of enzymes.
- Publication: A forthcoming research paper demonstrating methods and results will be published in 2024.

Rolls Royce - Product Cyber Security **June 2021 – September 2021**

Cyber Security Research Internship *Filton, Bristol*

- Researched the National Institute of Science and Technology (NIST) developed Open Security Controls Assessment Language (OSCAL) to provide a preliminary understanding of the out-of-the-box capabilities of OSCAL.
- Produced a white paper on the importance of implementing DevSecOps at Rolls Royce with deep dives into specific technologies to support the adoption of the DevSecOps framework.
- Summarized security regulatory frameworks from NIST, IEEE and ISO to initiate cyber resilience-oriented discussions with product and system developers across the Rolls-Royce portfolio.

References

[1] I. Mason-Williams, G. Mason-Williams, and M. Sandler, "Knowledge distillation: The functional perspective," in *NeurIPS 2024 Workshop on Scientific Methods for Understanding Deep Learning*, 2024. [Online]. Available: https://openreview.net/forum?id=Cgo73ZnAQc

[2] I. Mason-Williams, F. Ekholm, and F. Huszar, "Explicit regularisation, sharpness and calibration," in *NeurIPS 2024 Workshop on Scientific Methods for Understanding Deep Learning*, 2024. [Online]. Available: https://openreview.net/forum?id=ZQTiGcykl6

[3] I. Mason-Williams, "NEURAL NETWORK COMPRESSION: THE FUNCTIONAL PERSPECTIVE," in *5th Workshop on practical ML for limited/low resource settings*, 2024. [Online]. Available: https://openreview.net/forum?id=Q7GXKjmCSB