Alex J. Chan

alexjchan.com | +44 (0) 777 112 8977 | me@alexjchan.com
LinkedIn: alexjchan | GitHub: XanderJC

I am a PhD student interested in developing **safe** and **robust** machine learning systems that behave in the way we expect, and thus able to work with humans effectively. My work has often explored **inverse reinforcement learning** and **imitation learning** in order to learn from humans and how we work, as well as **synthetic data** for more robust evaluation. I'm currently thinking about how we can effectively leverage **large language models** for optimal **Human-AI collaboration** given the huge amount of information encoded in them.

Education

University of Cambridge — Cambridge, UK

PhD Machine Learning — *Oct. 2020 – Present*

- Penultimate year PhD student under the supervision of Professor Mihaela van der Schaar, working on machine learning with applications in medicine. My main area of focus is probabilistic methods for understanding decision making through interpretable imitation learning and inverse reinforcement learning.

MPhil Machine Learning and Machine Intelligence — *Oct. 2019 – Sep. 2020*

- Awarded with Commendation and an average of 79%.
- 92% for my thesis "Interpretable Policy Learning" - developing interpretable imitation learning algorithms for decision making in high stakes environments.
- Taught component focusing on probabilistic machine learning, with modules on natural language processing, reinforcement learning, and computational neuroscience.

University College London — London, UK

BSc Statistics — *Oct. 2016 – June 2019*

- 1^{st} Class Honours - 81% average. Achieved 84% on my final year project on probabilistic deep learning, focusing on flexible Bayesian approximations in deep neural networks.
- Modules include significant mathematical and statistical content covering probability, linear models, mathematical analysis, and advanced linear algebra.

Hampton School — London, UK

A-Levels/GCSEs — *2016 – 2015*

- A*A*A*A* at A-Level in Maths, Further Maths, Chemistry, and Physics. Further AA at AS-Level in Biology and General Studies. 9A*1A at GCSE. Gold, Silver and Bronze Duke of Edinburgh Awards.

Experience

Stanford Existential Risk Initiative — Berkeley, CA

Research Intern — *Jan. – Mar. 2023*

- Working with Owain Evans (Future of Humanity Institute - University of Oxford) examining large language models (LLMs), in particular their ability to be truthful and articulate about their output.
- Designed and implemented experiments in Python for various LLMs using OpenAI and HuggingFace APIs.

University of Cambridge — Cambridge, UK

Research Assistant — *Dec 2019. – Aug. 2020*

- Research assistant in the van der Schaar lab, working on uncertainty calibration in Bayesian neural networks.
- Developed a novel method for improving calibration in transfer learning - resulted in publication at ICML

University College London — London, UK

Research Assistant — *Jun. – Aug. 2018*

- Worked with Dr Sam Livingstone on a research project during the summer on Markov chain Monte Carlo, specifically comparing theoretical bounds with traditional practical convergence diagnostics.
- Implemented a variety of MCMC samplers and diagnostic methods in Python, and ran experiments to compare their output with recently proposed theoretical bounds.

Awards/Prizes

Stanford Existential Risk Initiative Machine Learning Alignment Theory Scholarship

- Scholarship funding for research project on how large language models can lie when articulating decision rules.

Microsoft Research PhD Scholarship

- Received the award for full funding of my PhD co-supervised with Microsoft Research (Dr Aditya Nori, and Dr Danielle Belgrave) "A Smart Care System for Healthcare using Contextual Reinforcement Learning".

G-Research PhD Prize in Maths and Data Science - 2nd Place

- Runner up in the G-Research competition for best draft PhD dissertation.

Optiver Hex Cambridge Hackathon Challenge - 1st Place

- Developed an algorithmic trading strategy for market making a dual listed product on an Optiver proprietary exchange, beating out about 40 teams.

EPSRC Vacation Bursary

- Awarded funding grant by the Engineering and Physical Sciences Research Council to conduct a research project during the summer on Markov chain Monte Carlo methods.

Skills

Machine Languages: Python, R, MATLAB, PostgreSQL, HTML
Human Languages: English, Conversational French
Libraries/Tools: PyTorch, JAX, TensorFlow, pandas, NumPy, Matplotlib, Git

Supervision

University of Cambridge Cambridge, UK
MPhil Machine Learning and Machine Intelligence Theses *Mar. – Aug. 2021*

- **Tennison Liu:** *Fair Policy Learning.* (Work published at AFCP 2022)
- **Alizée Pace:** *Adaptive Decision Tree Policies* (Resulted in a Spotlight at ICLR 2022).

University of Oxford Oxford, UK
MSc Statistical Science Thesis *Mar. – Aug. 2021*

- **Yuling Chen:** *Clustered Bayesian Inverse Reinforcement Learning Via Variational Inference.*

Peer-Reviewed Publications

Synthetic Model Combination: An Instance-wise Approach to Unsupervised Ensemble Learning
A. J. Chan and M. van der Schaar. *Advances in Neural Information Processing Systems (NeurIPS)* 2022.

Practical Approaches for Fair Learning with Multitype and Multivariate Sensitive Attributes
T. Liu, **A. J. Chan**, B. van Breugel, and M. van der Schaar. *Algorithmic Fairness through the Lens of Causality and Privacy (AFCP) workshop at NeurIPS* 2022.

Inverse Online Learning: Understanding Non-Stationary and Reactionary Policies
A. J. Chan, A. Curth, and M. van der Schaar. *International Conference on Learning Representations (ICLR)* 2022.

POETREE: Interpretable Policy Learning with Adaptive Decision Trees
A. Pace, **A. J. Chan**, and M. van der Schaar. *International Conference on Learning Representations (ICLR)* 2022.

The Medkit-learn(ing) Environment: Medical Decision Modelling through Simulation
A. J. Chan, I. Bica, A. Hüyük, D. Jarrett, and M. van der Schaar. *Proceedings of the Neural Information Processing Systems (NeurIPS) track on Datasets and Benchmarks* 2021.

Scalable Bayesian Inverse Reinforcement Learning
A. J. Chan and M. van der Schaar. *International Conference on Learning Representations (ICLR)* 2021.

Generative Time Series Modelling with Fourier Flows
A. M. Alaa, **A. J. Chan**, and M. van der Schaar. *International Conference on Learning Representations (ICLR)* 2021.

Unlabelled Data Improves Bayesian Uncertainty Calibration under Covariate Shift
A. J. Chan, A. M. Alaa, Z. Qian, and M. van der Schaar. *International Conference on Machine Learning (ICML)* 2020.

In Preperation / Preprints

Assessing and Enforcing Agent Fairness in Sequential Decision Making
A. J. Chan and M. van der Schaar.

Optimising Human-AI Collaboration by Finding Convincing Explanations
A. J. Chan, A. Hüyük, and M. van der Schaar.

Service

AAAI Workshop on Representation Learning for Responsible Human-Centric AI
Invited Area Chair *2023*

NeurIPS SyntheticData4ML Workshop
Program Committee / Area Chair *2022*

NeurIPS Workshop on Causality for Real-world Impact
Invited Reviewer *2022*

ICML/ICLR/NeurIPS
Invited Reviewer ICML21, NeurIPS21, ICLR21-22 *2021 – Present*

University of Cambridge Cambridge, UK
Club Captain, Wolfson College Boat Club *Aug. 2021 – Aug. 2022*

- As Captain of the boat club, I was in charge of the overall running of the club, organising the training of the members as well as broader events and the alumni network.

University College London London, UK
Vice President/Treasurer - Pure Krav Maga Society *Oct. 2018 – June 2019*

- I oversaw the organisation and finances behind sessions while helping to run classes as a trainee instructor.

Welfare Officer, Effective Altruism Society *Oct. 2018 – June 2019*

- I was responsible for engaging with the wider community to develop more of an understanding of the aims of Effective Altruism as well as looking out for the welfare of our members and helping develop the society further.

Electric Eels Swimming Club Windsor, UK
Volunteer Swimming Coach *2011 – 2015*

- I spent four years volunteering with the club, which aims to provide special coaching for children with Down syndrome, coaching both groups and 1-on-1 at a range of swimming ability
- I became ASA Lv1 certified in Teaching Aquatics, allowing me to develop my technical and communication skills to be a more effective coach.