Siu Lun Chau

alanchau@hotmail.co.uk | +44 7415 137484 | chau999 | London, UK

A final year PhD student aspired to conduct research to deepen our understanding of AI methods, advance the state-of-the-art, and bring real-world impact.

Education

Apr 2023 – Sep 2018 | **DPhil in Statistical Machine Learning** | St.Peter's College, University of Oxford
- Thesis: *Towards Trustworthy Machine Learning using Kernel methods and Gaussian Processes.*
- Supervised by Prof. Dino Sejdionvic, Prof. Mihai Cucuringu, and Prof. Xiaowen Dong.

Jul 2018 – Sep 2017 | **MMATH in Mathematics and Statistics** | Lady Margaret Hall, University of Oxford
- Graduated with First Class, ranked 2nd in the year.
- Distinction in Master's Thesis: *Modelling Diseases Trajectories with Infinite Mixture of Gaussian Processes*, supervised by Prof. Mihaela Van Der Shaar.

Jul 2017 – Sep 2014 | **BA in Mathematics and Statistics** | Lady Margaret Hall, University of Oxford
- Graduated with First Class, ranked 1st in the year.
- Distinction in Undergraduate Essay on Boosting methods, supervised by Prof. Francois Caron.

Skills

Software: Python, R, MATLAB, Git, LaTeX | ML Package: PyTorch, Gluon, Scikit-learn | ML Experience: Forecasting, Explainable AI, Causal Inference, Preference Learning, Graph ML, Uncertainty Quantification, Bayesian Optimisation

Work Experience

Present – Jun 2022 | **Applied Scientist Intern** | Amazon, UK
Project: Coherent Multi-granularity Demand Forecasting for the Transportation Service Outbound Network
- Developed **Deep Coherent Probabilistic Forecasts** with respect to the Amazon EU transportation network for logistic optimisation. Solutions are delivered into production-ready AWS infrastructure.

Jun 2022 – Oct 2021 | **Research Intern** | Max Planck Institute of Intelligent System, Germany
Project: Interface between Machine Learning and Economics
- Researched into relaxing restrictive assumptions in **Instrumental Variable Regression** and examined non-parametric hypothesis testing framework for **Regression Discontinuity Design**. Supervised by Dr. Krikamol Muandet.

Oct 2021 – Aug 2017 | **Machine Learning Content Developer (Part Time)** | Cambridge Spark, UK
- Designed and delivered ML projects and courses to up-skill students and companies. Topics covered include: **ML fundamentals, graphs, model explainability using LIME and SHAP, time series forecasting, and Gaussian processes.**

Apr 2019 – Apr 2017 | **Cofounder** & **Managing Director** | Oxford Strategy Group Digital, UK
- Co-founded and managed Oxford first student-led consultancy group with over 50 technical consultants.
- Pitched data science projects to multinational clients and maintained a diverse portfolio ranging from FTSE100 companies to startups, such as P&G, Unilever, Shelter, Biobeats, and National Australia Bank.

Consulting Projects

Apr 2022 | Job Title Alignment using NLP models | Ravio (HR tech startup), UK
- Utilised pre-trained languague models such as **RoBERTa** and **GPT3** to create word embeddings to compare and align job titles across companies to standardise compensations.

Jan 2021 | Explainable Forecasting Spreadsheet Plug-in | gini (Fin tech startup), HK
- Developed a **Gaussian Processes** based explainable time series model using **SAGE** for giniPredict, a forecasting spreadsheet plug-in built for non technical decision-makers.

Jan 2020 | Statistical analysis on crop yield data | Greenvale (Agricultural tech startup), UK
- Conducted statistical analysis on crop yield data to examine seasonal effect on tuber growth across varieties.
- Developed a short-term forecasting model using **Gaussian Processes** for canopy development based on groundcover observations.

Apr 2019 | Markdown price Optimisation | Bonmarché (Fashion retail), UK
- Developed a demand forecasting model to predict pre-markdown sales and solved for the optimal discount and markdown price to reach the user-defined target sell-through.

Invited Talks

Sep 2022	**Spectral Ranking with Covariates** | ECML PKDD 2022
Sep 2022	**Explainability For Kernel Methods** | ELISE Theory Workshop on ML Fundamentals
Jun 2022	**Deconditional Gaussian Processes** | S-DCE Alan Turning Institute seminar
Apr 2022	**Explaining Kernel methods with RKHS-SHAP** | UCL Gatsby Unit
Feb 2022	**Deconditional downscaling with Gaussian Processes** | UCL Statistical Machine Learning Group
Feb 2022	**Shapley values for Model Explanations** | Imperial & Oxford StatML seminar
Jun 2021	**Uncertainty Quantification for Causal Data Fusion** | Warwick ML Group

Awards

Sep 2018	**ESPRC and MRC studentship for DPhil in Statistics and Machine Learning**
Sep 2017	**Department Prize for FHS Mathematics and Statistics Part B (Top of the year)**

Publications

May 2022 | **Explaining Preference with Shapley Values** | NeurIPS 2022 | code | pdf

Siu Lun Chau*, Robert Hu*, Jaime Ferrando Huertas, Dino Sejdinovic

- We proposed a skew-symmetric utility function to build a Shapley value-based explanation framework for preference models.

May 2022 | **Giga-scale Kernel Matrix-Vector Multiplication on GPU** | NeurIPS 2022 | code | pdf

Robert Hu, **Siu Lun Chau**, Dino Sejdinovic, Joan Alexis Glaunès

- Building on top of the *Fast Multipole Method*, we proposed *Faster-Fast and Free Memory Method* (F^3M) to run Kernel Matrix Vector-Multiplication on tall ($n \sim 10^9$) and skinny ($D \leq 7$) data efficiently using a single GPU.

May 2022 | **RKHS-SHAP: Shapley Values for Kernel Methods** | NeurIPS 2022 | code | pdf

Siu Lun Chau, Robert Hu, Javier Gonzalez, Dino Sejdinovic

- We proposed RKHS-SHAP to explain and interpret RKHS functions arose in Kernel methods using the Shapley value paradigm.
- Based on RKHS-SHAP, we proposed a *Shapley Regulariser* that can be used as an attribution prior under the empirical risk minimisation framework to control feature's contribution during the learning procedure.

Mar 2022 | **Spectral Ranking With Covariates** | ECML PKDD 2022 | code | pdf

Siu Lun Chau, Mihai Cucuringu, Dino Sejdinovic

- We proposed three spectral ranking algorithms based on *seriation, low-rank assumption* and *canonical correlation*, to the problem of ranking n players given their incomplete and noisy pairwise comparisons, in light of their player covariate information.

Mar 2022 | **Learning Inconsistent Preference with Gaussian Processes** | AISTATS 2022 | pdf

Siu Lun Chau, Javier Gonzalez, Dino Sejdinovic

- We challenge the usual modelling assumption in preference models that imposes rankability of data items via latent utility function values.
- We proposed *Generalised Preferential Gaussian Process* to model preferences that depart from rankability, a common and strong assumption that is often violated in practice.

Dec 2021 | **BayesIMP: Uncertainty Quantification for Causal Data Fusion** | NeurIPS 2021 | pdf

Siu Lun Chau*, Jean Francois Ton*, Yee Whye Teh, Javier Gonzalez, Dino Sejdinovic

- We proposed *Bayesian Conditional Mean Embedding* to estimate the average treatment effect under a data fusion setting while quantifying model uncertainty.

Dec 2021 | **Deconditional Downscaling with Gaussian Processees** | NeurIPS 2021 | code | pdf

Siu Lun Chau*, Shahine Bouabid*, Dino Sejdinovic

- We devised a Bayesian solution for statistical downscaling which handles unmatched multi-resolution data through the proposed *Deconditional Gaussian Processes*.

Aug 2020 | **Kernel-based Graph Learning from Smooth Signals: A Functional viewpoint** | IEEE 2020 | pdf

Xingyue Pu, **Siu Lun Chau**, Xiaowen Dong, Dino Sejdinovic

- We proposed a kernel ridge regression based graph learning framework to recover topological structure from observed graph signals.