Shirley Xiaoqi Liu

Department of Statistics, University of Oxford, UK | shirley.liu@stats.ox.ac.uk | shirleyliuxq.github.io

My research develops **statistical and information-theoretic foundations** that make machine learning more **efficient and reliable**, especially when data is heterogeneous, high-dimensional, or changing over time.

RESEARCH EXPERIENCE & EDUCATION

Postdoctoral Researcher, University of Oxford Oct 2024 – present

Host: Prof. Patrick Rebeschini

- Mathematically understand common heuristics and phenomena in model training, such as early stopping [2], benign overfitting, and feature learning. Use this understanding to design better models and algorithms.
- Sequential decision-making with rigorous guarantees: design algorithms to make decisions when the best choice shifts over time [5]; train model cascades that adaptively route inputs to models for the best cost-accuracy tradeoff [3].
- Quantify uncertainty under minimal assumptions, using e-values, conformal prediction, and prediction-powered inference.

PhD in Information Engineering, University of Cambridge Oct 2019 – Sep 2024

Advisor: Prof. Ramji Venkataramanan

Thesis: Message passing algorithms for statistical estimation and communication (link)

- Localise when a latent signal changes in high-dimensional regression data [8, 9].
- Establish fundamental limits, and design efficient encoding and decoding schemes, for data storage [4] and communication in large user networks [6, 7].
- Recover large structured matrices from a near-minimum number of measurements, far fewer than their ambient dimensions [10].

BA and MEng, University of Cambridge Oct 2015 – Jun 2019

Honours with Distinction in Information and Computer Engineering: ranked top 3 – 7% each year in a cohort of 300+

OTHER WORK EXPERIENCE

Research Intern, Microsoft Research, Cambridge Apr – Jun 2023

- Designed a novel error correction coding scheme for Project Silica (a glass-based data storage system), increasing transmission rate by 5 – 10%
- Clarified the key information-theoretic performance metric for selecting the optimal coding scheme
- Co-authored a resulting *Nature* publication [4]

Software Development Intern, MediaTek Inc, Cambourne Jun – Sep 2017

Structural Design Engineer, WSP | Parsons Brinckerhoff, Cambridge Jun – Aug 2016

SELECTED PREPRINTS & PUBLICATIONS

[1] **S. X. Liu**, T. Adhikari, P. Pandit, "Precise Asymptotics of Non-Convex M-Estimators", 2026. (preprint)

[2] A. Buna, **S. X. Liu**, P. Rebeschini, "Minimax Optimal Early-Stopped Gradient Descent for Gaussian Mixture Classification", *Advances in Neural Information Processing Systems (NeurIPS)*, 2026. (preprint, talk)

[3] **S. X. Liu***, F. Regol*, A. Valkanas*, P. Rebeschini, M. Coates, "Bayes-Consistent Losses for Adaptive Inference in Classification", 2026. (preprint)

[4] J. Allison†, P. Anderson†, ..., **X. Liu**†, ..., et al., "Laser writing in glass for dense, fast and efficient archival data storage", *Nature*, 2026. (paper)

[5] **X. Liu**, D. Baudry, J. Zimmert, P. Rebeschini, A. Akhavan, "Non-stationary Bandit Convex Optimization: A Comprehensive Study", *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. (paper, talk, poster)

[6] **X. Liu**, P. Pascual Cobo, R. Venkataramanan, "Many-User Multiple Access with Random User Activity: Achievability Bounds and Efficient Schemes", *IEEE Transactions on Information Theory*, 2025. (paper, conference version, talk, poster)

[7] **X. Liu**, K. Hsieh, R. Venkataramanan, "Coded Many-User Multiple Access via Approximate Message Passing", *Information Theory, Probability and Statistical Learning: A Festschrift in Honor of Andrew Barron*, 2025. (preprint, conference version, talk)

[8] G. Arpino, **X. Liu**, J. Gontarek, R. Venkataramanan, "Inferring Change Points in High-Dimensional Regression via Approximate Message Passing", *Journal of Machine Learning Research*, 2025. (paper, talk, code)

[9] G. Arpino, **X. Liu**, R. Venkataramanan, "Inferring Change Points in High-Dimensional Linear Regression via Approximate Message Passing", *International Conference on Machine Learning (ICML)*, 2024. (paper, poster)

* Equal contribution.

† Alphabetical order.

[10] **X. Liu**, R. Venkataramanan, "Sketching Sparse Low-Rank Matrices With Near-Optimal Sample- and Time-Complexity Using Message Passing", *IEEE Transactions on Information Theory*, 2023. (paper, conference version)

SELECTED TALKS

[1] "When to Stop Gradient Descent: Minimax Optimal Early Stopping for Classification" at:
- **Upcoming:** Machine Learning Theory Seminar, University of Cambridge, Feb 2027
- **Upcoming:** IMS International Conference on Statistics and Data Science (ICSDS), Croatia, Dec 2026
- **Upcoming:** Computational Statistics and Machine Learning Reading Group, Imperial College London, Dec 2026
- Department of Statistics, University of Oxford, Nov 2026
- Statistics Seminar, School of Mathematics, University of Bristol, Oct 2026 (slides)

[2] "Bayes-Assisted Confidence Sequences for CDFs", Safe Anytime-Valid Inference (SAVI) Conference, University of Twente, Jul 2026.

[3] "Mind the U Curve: Why We Stop Gradient Descent Early", Meet Our Group Seminar Series, Department of Statistics, University of Oxford, May 2026. (slides)

[4] "Learning Under Non-Stationarity: Statistical Inference & Bandit Convex Optimization", Algorithms and Computationally Intensive Inference Seminar, Department of Statistics, University of Warwick, Mar 2026. (slides)

[5] "Tight Confidence Sequences From Universal Portfolio", Hypothesis Testing with E-values Reading Group, Department of Statistics, University of Oxford, Mar 2026. (slides)

[6] "Non-Stationary Bandit Convex Optimization: Algorithms and Links to Coin Betting", Algorithmic Statistics Workshop, Department of Statistics, University of Oxford, Nov 2025. (slides)

[7] "Sequential Anytime-Valid Inference Using E-processes", Hypothesis Testing with E-values Reading Group, Department of Statistics, University of Oxford, Oct 2025. (slides)

[8] "Loss landscapes and optimization in over-parameterized neural networks", Learning Theory and Statistical Optimization Reading Group, Department of Statistics, University of Oxford, Nov 2024. (slides)

[9] "Communication over many-user channels via Approximate Message Passing", Information Theory Seminar, Department of Mathematics, University of Cambridge, May 2024. (slides)

SERVICE

Reviewer: NeurIPS (top reviewer 2024), ICML, ICLR, TMLR, ISIT, ITW
Organiser, Cambridge Information Engineering Divisional Conference, Mar 2022
Postgraduate Representative, Information Engineering Division, 2020 – 2022

TEACHING & MENTORING

Lecturer, Oxford Statistics in AI CDT Module, Nov 2025: "Tutorial on Sequential Anytime-Valid Inference Using E-processes" (slides)

Research project supervision, 2023 – present:
- Oxford MSc projects, 2025 – present: sequential inference for time-to-unsafe monitoring, prediction-powered inference, and in-context learning in language models (proposal)
- Oxford UNIQplus research internship, 2026: sequential inference
- Cambridge MEng projects, 2023 – 2024: high-dimensional statistical estimation for heterogeneous data

Subject supervision, 2019 – 2024:
- Cambridge third-year courses: 3F7 Information Theory & Coding, and 3F4 Data Transmission, 2019 – 2024
- Cambridge AI+ Programme, 2022 – 2023

SCHOLARSHIPS & AWARDS

Cambridge Trust International PhD Scholar (£49,000/year)
2020 British Education Award, Nov 2019
Best Thesis Presentation Award, Jun 2019
Newnham College Scholar, Cambridge, 2016 – 2019