LSE Department of Statistics

LSE Department of Statistics
Research Showcase 2026

26-28 May

LSE Department of Statistics

Promoting excellence in research and education in statistics and data science.

Welcome

It is my great pleasure to welcome you to the LSE Department of Statistics Research Showcase 2026.

This annual event is our opportunity to celebrate the breadth and depth of research taking place across our Department, and to bring together colleagues, collaborators, students, partners and friends in a spirit of intellectual curiosity and exchange.

We highlight not only excellence, but also the importance of connection, between fields and disciplines across statistics, data science, AI, and machine learning; and between people, from new PhD students to some of our most experienced Professors. This packed, three-day programme reflects that shared ambition, representing the full range of research conducted in the Department, spanning four areas: data science; statistical learning and time series; probability in finance and insurance; and social statistics. One of the things I am most proud of about the Department is that these are not siloed communities, there is real intellectual exchange across them, and the showcase is designed to reflect that.

The first day brings together researchers working at the forefront of AI in both industry and academia, including speakers from organisations such as AstraZeneca, Barclays, Meta, and Microsoft Research, as well as the Psychometrics Centre at the University of Cambridge and the LSE Department of Statistics, so there will be a genuinely exciting mix of perspectives from people pushing the boundaries of what is possible. The second and third days focus on presentations of research conducted within the Department.

Our Department is immensely proud to contribute to the LSE's vibrant research culture, where methodological innovation meets real-world application. A collaborative ethos runs through so much of the work we do. A central aim of this showcase is to create space for dialogue. Through talks, posters and informal discussions, we hope you will discover new perspectives, forge connections, and explore opportunities for future collaboration. Across all three days, the event is designed to be as interactive and engaging as possible, so we encourage everyone to ask questions, visit the poster sessions, and make the most of the networking opportunities.

Events such as this play a vital role in strengthening our research community - supporting early-career researchers, showcasing emerging talent, developing skills in communicating research and engaging wider audiences, and enabling us all to learn from one another.

I would like to extend my sincere thanks to everyone who has contributed to making this event possible: the organising team, our presenters, and all those who support research within and beyond the Department.

Whether you are joining us as a researcher, student, collaborator or guest, I hope you find the next few days both stimulating and inspiring. Thank you for being part of our research community, and have an enjoyable and rewarding showcase.

With all best wishes,

Professor Milan Vojnović

Head of Department
LSE Department of Statistics

DAY 1 - 26 MAY

Workshop on generative AI: Research advances and applications

AI image by Luke Jones, Unsplash

9.50am - 10am

Opening remarks

Speaker

Milan Vojnović, Head of Department, LSE Department of Statistics.

Speaker biography

Milan Vojnović's research focuses on machine learning and optimisation, involving the development of novel algorithms and the analysis of their theoretical guarantees to support the design of efficient intelligent systems. He has made significant contributions to scalable optimisation methods for machine learning, multi-armed bandits, multi-agent systems, algorithms under uncertainty, game theory, and network system control and optimisation. His work has been applied across a variety of domains, including online platforms, computer networks, and machine learning systems.

He has received several best paper awards at leading conferences, as well as the ACM SIGMETRICS Rising Star Researcher Award and the ERCIM Cor Baayen Award. He is also the author of Contest Theory, published by Cambridge University Press in 2016. Milan has held a visiting scientist position at Meta. Prior to joining LSE, he spent 13 years as a researcher at Microsoft Research, working on a wide range of projects. He also held a two-year appointment as an affiliated lecturer at the Statistical Laboratory, University of Cambridge. He was awarded his PhD in 2003 by the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

LinkedIn

linkedin.com/in/milan-vojnovic-70705b5

10am - 10.45am

Advancing scientific research with AI research agents

Speaker

Tatiana Shavrina, Research Scientist Manager, Meta.

Abstract

Large Language Model (LLM) agents are poised to transform the landscape of scientific research by automating complex, multi-stage workflows.

To accelerate progress in this domain, AI Research Agents are applied to more and more modeling sciences, the first being AI Research itself. The talk will cover recent works in the domain, including frontier scaffolds, LLMS, and benchmarks.

We will look at how the current agentic tasks encompass a broad spectrum of scientific challenges, including language modeling, mathematics, bioinformatics, and time series forecasting.

We will rigorously evaluate agentic capabilities across the entire research lifecycle—spanning ideation, experimental analysis, and iterative refinement—without providing baseline code. Our empirical results reveal that while agents surpass human state-of-the-art (SOTA) performance in four tasks, they fall short in sixteen, and even the best-performing agents do not reach the theoretical task ceilings. These findings highlight that AIRS-Bench remains unsaturated, offering significant headroom for future advancements.

Speaker biography

Tatiana Shavrina is a Research Scientist Manager working on large language models at FAIR, with a background in linguistics and natural language processing. Her research focuses on multilingual modeling, evaluation of LLMs, and Accelerating scientific research with AI agents and LLMs. She has published at major venues including ACL and NeurIPS and actively works on bridging AI agents and the methodology of science.

LinkedIn

linkedin.com/in/tatiana-shavrina

10.45am - 11.15am

Break

11.15am - 12noon

GenAI in finance: An overview of applications and skill sets

Speaker

Dimitrios Emmanoulopoulos, Director, Head of AI/ML and Quantum Technologies, Data Engineering Domain, Chief Technology Office, Barclays.

Abstract

Generative AI is revolutionizing financial services by automating complex tasks, enhancing analytical capabilities, and improving operational efficiency across front, middle, and back-office functions. This presentation explores how GenAI is being deployed to transform customer interactions, streamline workflows, support decision-making processes, and accelerate information processing. The discussion examines the critical skill sets emerging for finance professionals, including prompt engineering, technical literacy, model interpretation, and the integration of AI capabilities with financial domain expertise. Understanding these applications and required competencies enables organisations to effectively harness GenAI's transformative potential.

Speaker biography

Dimitrios is the Head of AI/ML and Qauntum Technologies for Barclays Chief Technology Office. For the last eight years he has been productionizing machine learning solutions for a variety of different business use cases: card fraud detection, recommendation engines, credit card delinquency predictions, customer sentiment analysis, compliance and investment bank markets. Moreover, Dimitrios has been building cost effective GPU hardware solutions for machine learning projects, that deal with petabytes of data, across the entire bank and he is developing the next generation state-of-the-art AI models (using Deep Neural Networks) and generative AI frameworks. Currently he is leading the tech adoption of GenAI models across Barclays (i.e. aligning GenAI tech to particular business use cases). In parallel, he is also actively involved in the Cloud based technologies, that will be implemented in the near future by Barclays, and he is following very closely the recent developments around Quantum Computing and Quantum AI. Finally, he is actively benchmarking 3rd party hardware and software solutions around AI and has been representing Barclays in various discussion panels and forums. Dimitrios completed a Ph.D. on the development of predictive algorithms for black hole astrophysics and quantum gravity detection.

LinkedIn

linkedin.com/in/dimitrios-emmanoulopoulos-384147107

12noon - 12.45pm

Generative AI and creativity

Speaker

David Stillwell, Professor of Computational Social Science at Cambridge Judge Business School, University of Cambridge.

Abstract

This talk examines how large language models perform on creative tasks and how they compare with humans individually and in groups. David Stillwell will discuss a study measuring the creativity of LLMs relative to human participants, as well as ongoing work on human–AI collaboration in creative problem solving. The talk will also compare teams of LLMs with teams of humans working on the same creative tasks, focusing on differences in idea generation, collaboration, and performance. The findings contribute to broader questions about creativity, collective intelligence, and the role of AI systems in knowledge work.

Speaker biography

David Stillwell is Professor of Computational Social Science at Cambridge Judge Business School, University of Cambridge, where he researches the intersection of artificial intelligence, psychology, and big data. His work explores how digital footprints — including social media activity and online behaviour — can be used to better understand human personality, preferences, and decision-making. He is widely known for research demonstrating how machine learning models can predict psychological traits from digital data, with implications for marketing, organisational behaviour, and public policy. His current interests include generative

AI, responsible AI adoption, digital behaviour, and the future relationship between humans and intelligent systems.

Alongside his academic research, David works closely with industry and public-sector organisations on the practical application of AI and data science. He regularly teaches executives and senior leaders about generative AI, helping organisations understand both the opportunities and limitations of emerging technologies. His work emphasises translating complex technical concepts into actionable insights for decision-makers.

LinkedIn

linkedin.com/in/david-stillwell

12.45pm - 2pm

Lunch and poster session

2pm - 2.45pm

Longitudinal lesion segmentation and its application in patient response prediction

Speaker

Yi Wei, Lead, Deep Learning Team, AstraZeneca.

Abstract

Predicting individual patient response to therapy is central to precision medicine, directly informing clinical trial design, patient selection, and treatment decisions. In oncology, estimating survival and treatment benefit remains challenging because outcomes evolve over time and depend on complex, patient-specific disease dynamics. Longitudinal medical data capture this temporal evolution and offer a richer basis for modelling response than single time-point assessments.

In this talk, I will present our approach to generating high-quality longitudinal lesion segmentations from serial CT scans and show how these time-resolved features can improve patient response prediction. I will discuss the end-to-end pipeline—data curation, lesion detection and segmentation, temporal alignment across scans, and feature extraction. Finally, I will illustrate how these longitudinal biomarkers enhance predictive performance for survival and treatment effect estimation, and outline implications for clinical decision-making and trial optimization.

Speaker biography

Yi Wei is the Director of the Deep Learning team in R&D IT at AstraZeneca, where he leads efforts in patient response prediction and molecule toxicity prediction to accelerate data-driven drug discovery. Previously, he served as Chief Technology Officer at a hedge fund, developing machine learning–based time-series models for trading equities and futures. Earlier in his career, Yi worked at Microsoft Research Cambridge, where he created a code retrieval and synthesis algorithm later integrated into the Bing search engine and received the MSR Technology Transfer Award for the project. He earned his PhD in Software Engineering from ETH Zurich in 2012.

LinkedIn

linkedin.com/in/jasonweiyi

2.45pm - 3.30pm

Multicalibration for real-world machine learning

Speaker

Niek Tax, Staff Research Scientist, Tech Lead, Meta.

Abstract

Machine learning models are increasingly used to make decisions that affect people, content, and business outcomes. For these systems, calibration—ensuring that predicted probabilities match real-world outcomes—is essential for trust, fairness, and optimal decision-making. However, global calibration alone is insufficient. Even a well-calibrated model on average can systematically overestimate or underestimate for specific segments (also known as protected groups)—such as users in a particular country, transactions on a certain device, or content of a certain type. These hidden calibration gaps lead to unfair, unreliable, or suboptimal decisions for affected segments.

In this talk we present MCGrad, an algorithm that calibrates an ML model globally as well as in arbitrary subgroups, that is deployed at scale and used in production in many ML models across Meta. MCGrad is available open source (see https://mcgrad.dev). We also present metrics to quantify multicalibration, and discuss various applications where multicalibration is of meaningful importance for the quality of automated business decisions.

Speaker biography

Niek Tax is a Machine Learning researcher, staff research scientist and tech lead at Meta, based on London. His work focuses on scalable, production-ready Machine Learning methods, especially model calibration, uncertainty quantification, active learning, and trustworthy AI systems.

Before Meta, he worked in applied R&D in Machine Learning at Philips Research and as a machine learning scientist at Booking.com. He has authored more than 40 peer-reviewed publications with thousands of citations across machine learning, process mining, and predictive modelling.

At Meta, Niek has led work on multicalibration and model correction, including MCGrad, an open-source framework for improving the reliability of machine learning predictions at production scale.

He combines academic research depth with practical engineering experience, focusing on methods that are theoretically grounded, scalable, and useful in real-world systems.

LinkedIn

linkedin.com/in/niektax

3.30pm - 4pm

Break

4pm - 4.45pm

Agentic repository automation with GitHub agentic workflows

Speaker

Don Syme, Principal Researcher, GitHub Next and Visiting Professor, King’s College London.

LinkedIn

linkedin.com/in/don-syme

Abstract

Agentic repository automation revolutionizes how we build software and process information: adding “Continuous AI” (CAI) to CI/CD. The future of development is individuals and teams equipped with automated agentic workflows (repository-bound agents) scouting ahead, cleaning up behind, proactively improving and relentlessly validating code. AI slop is solved by automated AI code improvement, repository maintenance is is solved by automated AI repository assistants, and difficult, neglected software engineering such as performance engineering and formal verification becomes more tractable.

Speaker biography

Don Syme is a Principal Researcher at GitHub Next and Visting Professor at Kings College London, specializing in AI assisted software development. He’s the designer of the F# language, the co-originator of async/await, and a contributor to the C# design and GitHub Copilot. He was awarded the Royal Engineering Society Silver Medal in 2016 for his contributions to computing.

4.45pm - 5.30pm

Demystify LLM reasoning through U-statistics theory

Speaker

Chengchun Shi, Associate Professor, LSE Department of Statistics.

Abstract

Group relative policy optimization (GRPO), a core methodological component of DeepSeekMath and DeepSeek-R1, has emerged as a cornerstone for scaling reasoning capabilities of large language models. Despite its widespread adoption and the proliferation of follow-up works, the theoretical properties of GRPO remain less studied. This talk provides a unified framework to understand GRPO through the lens of classical U-statistics. We demonstrate that the GRPO policy gradient is inherently a U-statistic, allowing us to characterize its mean squared error (MSE), derive the finite-sample error bound and asymptotic distribution of the suboptimality gap for its learned policy. Our findings reveal that GRPO is asymptotically equivalent to an oracle policy gradient algorithm -- one with access to a value function that quantifies the goodness of its learning policy at each training iteration -- and achieves asymptotically optimal performance within a broad class of policy gradient algorithms. Furthermore, we establish a universal scaling law that offers principled guidance for selecting the optimal group size. Empirical experiments further validate our theoretical findings, demonstrating that the optimal group size is universal, and verify the oracle property of GRPO.

Speaker biography

Chengchun is an Associate Professor in the Department of Statistics at LSE. He works at the interface of RL, LLMs and statistics, with applications to ride-sharing and healthcare. His work brings to light the relevance and significance of statistical learning in AI, and demonstrates the usefulness of RL as a framework for policy evaluation and A/B testing in two-sided marketplaces. Chengchun has published over 70 papers, with majority of them accepted in prestigious statistical journals (JRSSB, JASA, AoS) and top machine learning venues (NeurIPS, ICML, KDD, JMLR). His outstanding contributions have been recognized with esteemed awards such as the Peter Gavin Hall IMS Early Career Prize, IMS Tweedie Award and the Royal Statistical Society Research Prize. He has served as the area chair of NeurIPS and associate editors of prestigious journals JRSSB, JASA and AoAS.

5.30pm - 6.15pm

From typist to orchestrator: Rethinking core skills in programming education in the age of AI agents

Speaker

Marcos Barreto, Associate Professor, LSE Department of Statistics.

Abstract

Programming is undergoing a profound shift—from the manual construction of code to AI-assisted generation driven by large language models, “vibe coding” practices, and the rapid adoption of agent-based frameworks in industry. In this evolving landscape, the role of the programmer is moving from code authoring toward system orchestration, specification, and oversight.

This transformation poses a critical challenge for educators: how to preserve essential foundations in programming—such as abstraction, debugging, and computational thinking—while equipping students with the skills needed to design, guide, and critically evaluate AI-generated solutions.

This talk explores the implications of agentic development for programming education, highlighting both risks (e.g., superficial understanding, over-reliance on automation) and opportunities (e.g., accelerated development, increased focus on higher-level reasoning). Drawing on recent research and use cases from data science and industry, it invites discussion on what baseline competencies should define programming education today—particularly for graduates entering workplaces where AI agents are increasingly embedded in development workflows.

Speaker biography

Marcos Barreto is an Associate Professor in Data Science in the Department of Statistics at LSE. He has experience designing and delivering undergraduate and postgraduate courses in Computing and Data Science, including databases, big data, and AI. His current research investigates the role of AI agents in software development and their implications for teaching and curriculum design in quantitative disciplines.

LinkedIn

linkedin.com/in/marcosennesbarreto

6.15pm - 7.30pm

Drinks reception and poster session

DAY 2 - 27 MAY

Graph image by Yashowardhan Singh, Unsplash

10am - 10.20am
A brief overview of psychometrics as a framework for AI assessment

Speaker
Irini Moustaki, LSE Department of Statistics

10.20am - 10.40am
Pairwise comparisons without stochastic transitivity

Speaker
Sze Ming Lee, LSE Department of Statistics

10.40am - 11am
A generalized additive partial-mastery cognitive diagnostic model

Speaker
Camilo Cárdenas-Hurtado, LSE Department of Statistics

11am - 11.20am
Multi-output gaussian processes with network information

Speaker
Xinhui Liu, LSE Department of Statistics

11.20am - 12noon
Break

12noon - 12.20pm
Consumption-investment problem in rank-based models

Speaker
David Itkin, LSE Department of Statistics

12.20pm - 12.40pm
Stochastic factors can matter: improving robust growth under ergodicity

Speaker
Paul Mangers Bastian, LSE Department of Statistics

12.40pm - 1pm
Testing for endogeneity of irregular sampling schemes

Speaker
Giulia Livieri, LSE Department of Statistics

1pm - 2.20pm
Lunch

2.20pm - 2.40pm
Multi-scale online aggregation for non-stationary tabular prediction

Speaker
Yutong Wang, LSE Department of Statistics

2.40pm - 3pm
Detecting structural breaks in high-dimensional functional time series factor models

Speaker
Caixia Xu, LSE Department of Statistics

3pm - 3.20pm
Deep independent component analysis for time series via invertible neural networks

Speaker
Han Yan, LSE Department of Statistics

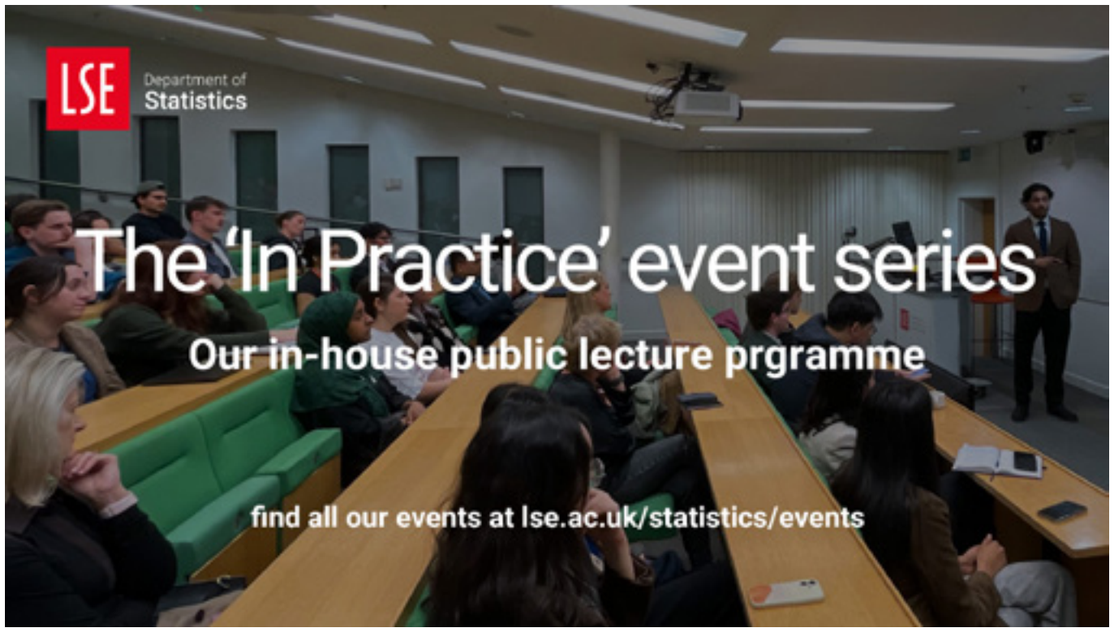
3.20pm - 4pm
Break

4pm - 4.20pm
Batchwise advantage shaping: Shared inter-query signal for LLM policy optimization

Speaker
Kai Ye, LSE Department of Statistics

4.20pm - 4.40pm
Interpretable time series analysis with Gumbel dynamics

Speaker
Yiliu Wang, Allen Institute



DAY 3 - 28 MAY

Chart image by Khampha Phimmachak, Vecteezy.com

10am - 10.20am
Maximum softly penalized likelihood for factor analysis

Speaker
Philipp Sterzinger, LSE Department of Statistics

10.20am - 10.40am
Advancing modelling of spatio-temporal effects in political science

Speaker
Cécile Richetta, University of Geneva

10.40am - 11am
Optimal in-context adaptivity and distributional robustness of transformers

Speaker
Tengyao Wang, LSE Department of Statistics

11am - 11.20am
Frontier estimation with functional inputs

Speaker
Shakeel Gavioli-Akilagun, City University Hong Kong and LSE Department of Statistics

11.20am - 12noon
Break

12noon - 12.20pm
Kernelised integrated R^2

Speaker
Pouya Roudaki, LSE Department of Statistics

12.20pm - 12.40pm
Training provably-robust neural networks

Speaker
Alessandro De Palma, LSE Department of Statistics

12.40pm - 1pm
Beyond vintage rotation: Bias-free sparse representation learning with oracle inference

Speaker
Yunxiao Chen, LSE Department of Statistics

1pm - 2.20pm
Lunch

2.20pm - 2.40pm
Detecting change regions on spheres

Speaker
Di Su, LSE Department of Statistics

2.40pm - 3.25pm

Closing talk:How a foundation in OR shaped a career in business - anecdotes and case studies

Speaker

Akin Adamson, Middle East Managing Director, Ricardo.

Abstract

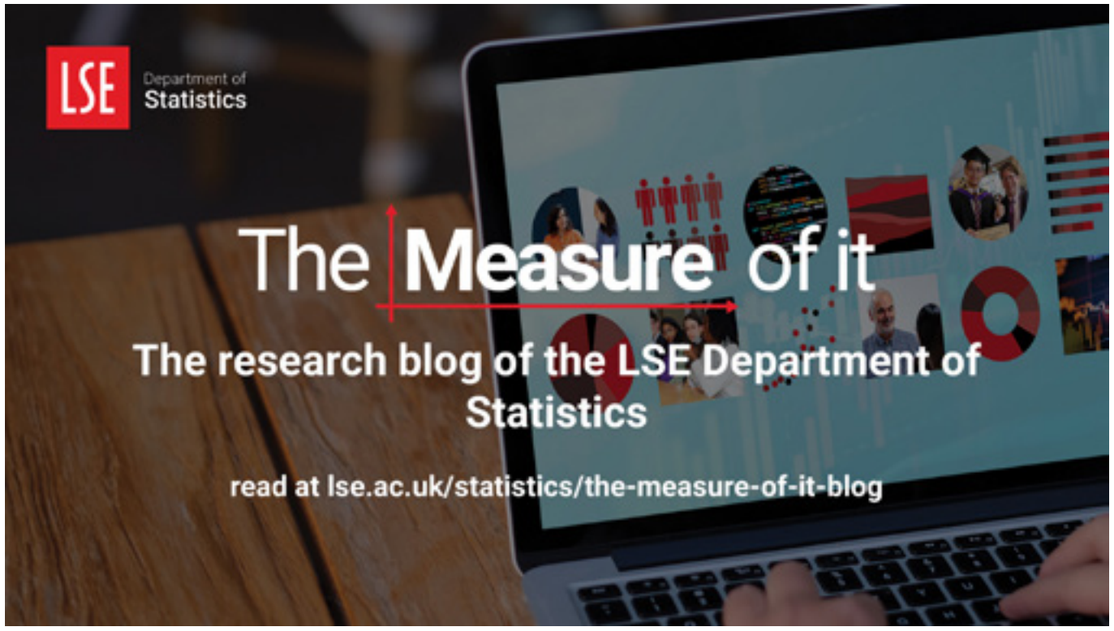
This presentation will illustrate how a grounding in Operational Research, Decision Sciences and Mathematical Modelling has shaped the presenter's career and the wider implications for how the presenter has dealt with uncertainty when making decisions in business.

Speaker biographer

Akin Adamson is the Middle East Managing Director for Ricardo - a global strategic engineering and environmental consultancy with offices across the Middle East. Ricardo advises public and private sector organisations on Air Quality, Energy, Water, Waste, Sustainability and Mobility strategies, technologies, and operations.Previously, he was the UK's Transport Research Laboratory's (TRL's) Regional Director for the Middle East. Before that, he held senior Director level roles at Nokia Siemens Networks, Telstra and A.T. Kearney.

LinkedIn

linkedin.com/in/akin-adamson



Find us online

lse.ac.uk/statistics
linkedin.com/school/lse-department-of-statistics
@LSEStatistics
@lsedepofstatistics.bsky.social

LSE Department of Statistics

Promoting excellence in research and education in statistics and data science.