Proceedings of the Oxford Computer Science Conference 2025

General Chairs: Mathias Jackermeier and Siyi Sun
Conference Committee: Hunar Batra and Jake Masters
With assistance from: Sarah Retz-Jones and Jodie Mattioli

26th June 2025

Contents

Programme

09:30	Registration and Coffee		
10:00	Introduction by Deputy Head of Department (Research)	Dave Parker	
10:30	Oral Session I: Theory	Jake Masters	Logic of Computation with Orbit-Finite Sets
		Jack Liell-Cock	An algebraic theory of concurrency
11:10	Coffee break		
11:30	Lightning Session I: Advances in LLMs	Akshat Naik	Rethinking Safety Evaluations for AI Agents
		Julian Manyika	Mechanism Design for Alignment via Human Feedback
		Thierry Blankenstein	Uncovering and Mitigating Selection Bias in Tool-Augmented Large Language Models
		Amit Levy	Language Models Encode Numbers Using Digit Representations in Base 10
12:00	Oral Session II: Security	Jessica Richards	Proving security of Signal
		Ethan D'Alessandro	Vulnerabilities in SAR and AIS Fusion Algorithms in Ship Detection
12:40	Lunch and Posters	Ari Hernawan	Gossiping Multiparty Session Protocols
		Jerred Chen	Image as an IMU: Estimating Camera Motion from a Single Motion-Blurred Image
		Haoxuan Yin	A Trace Model of an Imperative Multi-Stage Language
		Sydney Reis	Responsible Innovation Frameworks for Tech Workers at the Intersection of Geopolitics
		Lia Yeh	The Focked-up ZX Calculus: Picturing Continuous-Variable Quantum Computation
		Mark Denhoed	Synthesis of Code-Reuse Attacks from P-Code Programs
		Vivek Kothari	Information Theoretic Decomposition of Network Homophily Measures
14:00	OxWoCS Keynote Speaker	Ekaterina Komendantskaya	Proof-Carrying Neuro-Symbolic Code
15:00	Coffee break		
15:20	Lightning Session II: AI Applications	Kyrylo Boiko	Counter-drone Defense Using Multi-Agent Reinforcement Learning
		Sophie Fischer	Causes of Inter-Operator Variability in Semantic Segmentation for Medical Imaging
15:35	Oral Session III: Algorithms	Maria-Alexa Tudose	Optimal Deterministic Online Algorithms for Chasing Small Sets
		Giannis Tyrovolas	Fraud-Proof Revenue Division on Subscription Platforms
16:15	Coffee break		
16:30	Oral Session IV: Information Governance	Alastair McCullough	Information Governance, Domain Specificity, and the Data Mesh Paradigm
		Zhilin Zhang	Trouble in Paradise? Understanding Mastodon Admin's Motivations, Experiences, and Challenges Running Decentralised Social Media
		Keyu Zhang	Bridging Privacy and Practicality in Blockchain
17:30	End of conference. Walk over to Kellogg College		
18:00	Drinks reception at Kellogg College		
19:00	Conference dinner at Kellogg College		

OxWoCS Keynote Speaker

This year's Keynote Speaker was kindly organised in conjuction with Oxford Women in Computer Science.

Proof-Carrying Neuro-Symbolic Code

Ekaterina Komendantskaya

Proof Carrying Code is a long tradition within programming language research, broadly referring to methods that interleave verification with executable code, thus avoiding the inevitable discrepancies that arise when the code and the proofs are handled in different languages. Although the term was coined by Necula almost three decades ago, with time, it grew to encompass any languages that are powerful enough to handle both the coding and the proving. Examples are dependently-typed (Agda, Idris, Coq/Rocq) and refinement-typed (F*, Liquid Haskell) languages. In the last decades, both families of languages have seen substantial successes in the proof-carrying code direction.

An equally impressive, though very different in nature, revolution has happened elsewhere in computer science: machine learning methods grew in quality, diversity and quickly proliferated to different applications. Machine learning attracted attention of programming language researchers, and there have been recent advances in functional algorithms for automated differentiation, probabilistic programming, as well as neuro-symbolic programming and neuro-symbolic theorem proving. The latter two look into, respectively, methods of merging machine learning code and standard (symbolic) code; and automating theorem proving with the help of machine learning deployed at the stage of proof search or lemma conjecturing.

The natural question to ask is: does the concept of proof-carrying code bear any value, or indeed meaning, in the age of the neuro-symbolic paradigm shift? In this invited talk I will answer both questions in the positive. I will introduce the concept of "proof-carrying neuro-symbolic code" and explain its meaning and value, from both the "neural" and the "symbolic" perspectives. I will outline the first successes and challenges that this new area of research faces.

Abstracts of Talks

Oral Session I: Theory

Logic of Computation with Orbit-Finite Sets

Jake Masters

The theory of orbit-finite sets allow us to perform computation over infinite objects that are finite up to symmetries. Defining orbit-finite computation involves strict constraints on the atoms and symmetries used. The logic of orbit-finite sets of certain atoms and symmetries that do not satisfy the constraints can be equivalent to the logic for atoms and symmetries that do, allowing the computation theory to apply in new situations.

An algebraic theory of concurrency

Jack Liell-Cock

Concurrency is important in an ever-growing distributed world. We present an algebraic theory that captures dynamically created, Unix-like threads. We move beyond classical algebraic theories to parameterised algebraic theories, which allow for parameter binding to accommodate dynamic creation. We propose an operational semantics mirroring Unix, and provide a denotational semantics via monads. We show that the monad is presented by the algebraic theory, and it is adequate with respect to the operational semantics.

Oral Session II: Security

Proving security of Signal

Jessica Richards

The Signal protocol is an end to end encryption protocol underpinning several modern messaging applications, including Signal, WhatsApp and Facebook Messenger. The protocol has two interesting security guarantees of forward secrecy and post compromise security, which ensure secrecy of messages sent before and after a breach of state, outside of a narrow window around the breach time. We prove these properties in ProVerif, a symbolic protocol verifier. We also consider three adaptations to the model - header encryption, reuse of keys and composition with the initial key exchange protocol - and view how these affect the proofs.

Vulnerabilities in SAR and AIS Fusion Algorithms in Ship Detection

Ethan D'Alessandro

Ship detection is extremely important to Maritime Domain Awareness, enabling the discovery of illegal activities on the ocean. Cooperative data like AIS is fused with non-cooperative data like SAR in order to give a better picture of the ocean environment while limiting the ability of malicious adversaries to fly under the radar. However, SAR and AIS fusion algorithms still have vulnerabilities adversaries can exploit. My talk will include a background on current SAR and AIS fusion algorithms, common goals that modern adversaries want to achieve, attack primitives that allow adversaries to accomplish subgoals even in the presence of AIS and SAR monitoring, and a case study of how they can combine these attack primitives in order to accomplish a more complex goal. Finally, we will conclude with recommendations to improve these fusion algorithms to make them more resilient to attackers.

Oral Session III: Algorithms

Optimal Deterministic Online Algorithms for Chasing Small Sets

Maria-Alexa Tudose

We study the small set chasing problem (also known as metrical task systems), where an online algorithm must respond to a sequence of incoming requests – each consisting of a set of at most k points in a metric space – by moving to one of the requested points, aiming to minimize the total travel distance. This is a classic online decision-making problem under uncertainty, as future requests are unknown. We evaluate online algorithms using competitive analysis, comparing their performance to that of an optimal offline algorithm with full knowledge of the future.

We present a deterministic online algorithm with a competitive ratio of $O(2\hat{k})$, closing the previous 30-year gap between known lower and upper bounds. We also slightly improve the best known exact lower bound. Notably, our algorithm is directly inspired by the construction used in our lower bound.

Fraud-Proof Revenue Division on Subscription Platforms

Giannis Tyrovolas

We study a model of subscription-based platforms where users pay a fixed fee for unlimited access to content, and creators receive a share of the revenue. Existing approaches to detecting fraud predominantly rely on machine learning methods, engaging in an ongoing arms race with bad actors. We explore revenue division mechanisms that inherently disincentivize manipulation. We formalize three types of manipulation-resistance axioms and examine which existing rules satisfy these. We show that a mechanism widely used by streaming platforms, not only fails to prevent fraud, but also makes detecting manipulation computationally intractable. We also introduce a novel rule, ScaledUserProp, that satisfies all three manipulation-resistance axioms. Finally, experiments with both real-world and synthetic streaming data support ScaledUserProp as a fairer alternative compared to existing rules.

Oral Session IV: Information Governance

Information Governance, Domain Specificity, and the Data Mesh Paradigm

Alastair McCullough

This talk presents a novel critical review of the design of enterprise and domain specificity in information governance operating models–or target business architectures-by looking closely at the well-known data mesh sociotechnical method to understand a localised, domain specific approach. The analysis builds upon a new definition of enterprise information governance, defined as acting through control mechanisms to assure accountability in managing decision rights over information and data assets in organisations. The paper uses a graphic representation of such governance as a framework to consider the nature of strategic and tactical policies and standards that form the basis for data mesh thinking. It includes definitions of data objects and data products, and defines a technical use case, anchored in standard corporate accounting practice and software engineering, to exemplify data products. The paper will support both further scholarly endeavor and practitioners in the field aiming to utilise data mesh concepts in structuring domain-specific regulatory business architectures for governance.

Trouble in Paradise? Understanding Mastodon Admin's Motivations, Experiences, and Challenges Running Decentralised Social Media

Zhilin Zhang

Decentralised social media platforms are increasingly being recognised as viable alternatives to their centralised counterparts. Among these, Mastodon stands out as a popular alternative, offering a citizen-powered option distinct from larger and centralised platforms like Twitter/X. However, the future path of Mastodon remains uncertain, particularly in terms of its challenges and the long-term viability of a more citizen-powered internet. In this paper, following a pre-study survey, we conducted semi-structured interviews with 16 Mastodon instance administrators, including those who host instances to support marginalised and stigmatised communities, to understand their motivations and lived experiences of running decentralised social media. Our research indicates that while decentralised social media offers significant potential in supporting the safety, identity and privacy needs of marginalised and stigmatised communities, they also face considerable challenges in content moderation, community building and governance. We emphasise the importance of considering the community's values and diversity when designing future support mechanisms.

Bridging Privacy and Practicality in Blockchain

Keyu Zhang

The data transparency inherent to blockchain limits its applicability in scenarios involving sensitive information. Although previous research has proposed various solutions, these approaches generally lack support for inter-contract calls—an essential feature for practical, complex applications that cannot be implemented as a single monolithic program. Our research, RaceTEE, offers a novel architecture to addressing this limitation.

Abstracts of Lightning Talks

Lightning Session I: Advances in LLMs

Rethinking Safety Evaluations for AI Agents

Akshat Naik

The presentation will be about a suite of evaluations to measure different misalignment behaviours (e.g. scheming, resisting shutdown, etc.) from agentic LLMs and how they differ from LLMs purely in the chat setting.

Mechanism Design for Alignment via Human Feedback

Julian Manyika

Ensuring the faithfulness of human feedback is crucial for effectively aligning large language models (LLMs) with human supervision, as low-effort or dishonest reporting can significantly undermine the quality of this feedback and, consequently, the alignment process. The challenge of faithfully modeling pairwise feedback can framed it as a mechanism design problem, in which the process of eliciting preferences can be made to incentivize high effort and truthfulness. In our work, we present a new principal-agent model for preference elicitation, and outline new and existing mechanism frameworks, and then derive the constraints that allow them to satisfy desirable incentive compatibility properties for faithful preference alignment. We discuss the implications of these findings and outline future research directions in the design of robust mechanisms for preference elicitation.

Uncovering and Mitigating Selection Bias in Tool-Augmented Large Language Models

Thierry Blankenstein

The presentation will be about a suite of The rapid advancement in augmenting large language models with external APIs has significantly broadened their capabilities, enabling them to effectively interact with various tools to accomplish complex real-world tasks. However, the tool selection phase may inadvertently introduce biases, which my MSc thesis is investigating. Typically, this selection occurs in two stages: a neural retriever first identifies the top-k candidate APIs, and then the LLM selects from these retrieved options. Bias could emerge in either stage due to factors such as API

ordering, description phrasing, or content type, potentially causing unintended disparities in API usage and compensation.

This presentation will describe the progress I have made in researching the biases in prevalent API selection methods used by tool-augmented LLMs, aiming to identify whether the neural retriever or the LLM itself disproportionately favors certain APIs. By systematically analyzing these selection mechanisms, we aim to uncover underlying biases and, once found, explore effective mitigation strategies, such as fine-tuning with parameter-efficient methods. Ultimately, this work seeks to improve the fairness and robustness of API selection, thus supporting a more ethically sound integration of tool-augmented LLMs in real-world applications.

Language Models Encode Numbers Using Digit Representations in Base 10

Amit Levy

Large language models (LLMs) frequently make errors when handling even simple numerical problems, such as comparing two small numbers. A natural hypothesis is that these errors stem from how LLMs represent numbers, and specifically, whether their representations of numbers capture their numeric values. We tackle this question from the observation that LLM errors on numerical tasks are often distributed across the digits of the answer rather than normally around its numeric value. Through a series of probing experiments and causal interventions, we show that LLMs internally represent numbers with individual circular representations per-digit in base 10.This digit-wise representation, as opposed to a value representation, sheds light on the error patterns of models on tasks involving numerical reasoning and could serve as a basis for future studies on analyzing numerical mechanisms in LLMs.

Lightning Session II: AI Applications

Counter-drone Defense Using Multi-Agent Reinforcement Learning

Kyrylo Boiko

UAV attacks have become commonplace on battlefields in recent years. Targeted or scattered, they have proven an effective yet cheap tool to reach cities and objects far away from the frontlines (Hunder, Goodman, and Fenton 2025). As such weapons become cheaper and easier to manufacture, we must produce effective countermeasures to keep infrastructure and the public secure against military and terrorist threats. Using multi-agent reinforcement learning, we aim to train defensive agents to interact with each other as a team to stop UAV opponents from breaching the protected perimeter.

The hostile warzone environment poses two major challenges to the use of UAVs: localization and communication. The typical commercial approach of using GPS localization is untenable in warzones, as electronic warfare (EW) weapons can easily prevent a target UAV from receiving satellite signals (Wang, Liu, and Song 2020). Meanwhile, communication can be disrupted, corrupted, or entirely denied (Kim, Cho, and Sung 2019). As such, any autonomous system must be robust to these challenges.

Causes of Inter-Operator Variability in Semantic Segmentation for Medical Imaging

Sophie Fischer

When comparing an automated method of semantic segmentation to a ground truth, we must also consider how much the ground truth could vary between annotators. In the case of medical imaging the boundaries of an object are often unclear, leading to massive variation between observers. We believe that this variation is not random however, and by considering the causes of inter-operator variability we can not only reliably predict disputed areas, but integrate them into the final evaluation against an automated method.

Abstracts of Posters

Gossiping Multiparty Session Protocols

Ari Hernawan

Describes randomness communication in gossiping while maintaining deadlock freedom.

Image as an IMU: Estimating Camera Motion from a Single Motion-Blurred Image

Jerred Chen

In many robotics and VR/AR applications, fast camera motions cause a high level of motion blur, causing existing camera pose estimation methods to fail. In this work, we propose a novel framework that leverages motion blur as a rich cue for motion estimation rather than treating it as an unwanted artifact. Our approach works by predicting a dense motion flow field and a monocular depth map directly from a single motion-blurred image. We then recover the instantaneous camera velocity by solving a linear least squares problem under the small motion assumption. In essence, our method produces an IMU-like measurement that robustly captures fast and aggressive camera movements. To train our model, we construct a large-scale dataset with realistic synthetic motion blur derived from ScanNet++v2 and further refine our model by training end-to-end on real data using our fully differentiable pipeline. Extensive evaluations on real-world benchmarks demonstrate that our method achieves state-of-the-art angular and translational velocity estimates, outperforming current methods like MASt3R and COLMAP.

A Trace Model of an Imperative Multi-Stage Language

Haoxuan Yin

MetaML is a classic multi-stage programming language. It allows programmers to split a program into different stages and convert a general-purpose program into a specialized one. A crucial concern is whether the staging transformation is faithful, which can only be settled if we have a notion of program equivalence. In this paper, we give the first full abstraction result for imperative MetaML. Our trace model is given by operational game semantics, where the meaning of a program is modelled by its possible interactions with the environment. As evaluations can go under a lambda in MetaML, the model utilizes a novel partially closed instances of use approximation. We then display several

applications, including proving new equational rules and specifying a sufficient condition under which staging is faithful.

Responsible Innovation Frameworks for Tech Workers at the Intersection of Geopolitics

Sydney Reis

As technology development becomes more politicised and less scrutinised through standards and law, especially in the most consequential jurisdictions, responsible innovation approaches should be readapted. This presentation will present the necessity for, and a way forward on, a responsible innovation framework that engages tech workers directly to think about international and geopolitical impacts of their technologies.

The Focked-up ZX Calculus: Picturing Continuous-Variable Quantum Computation

Lia Yeh

While the ZX and ZW calculi have been effective as graphical reasoning tools for finite-dimensional quantum computation, the possibilities for continuous-variable quantum computation (CVQC) in infinite-dimensional Hilbert space are only beginning to be explored. In this work, we formulate a graphical language for CVQC. Each diagram is an undirected graph made of two types of spiders: the Z spider from the ZX calculus defined on the reals, and the newly introduced Fock spider defined on the natural numbers. The Z and X spiders represent functions in position and momentum space respectively, while the Fock spider represents functions in the discrete Fock basis. In addition to the Fourier transform between Z and X, and the Hermite transform between Z and Fock, we present exciting new graphical rules capturing heftier CVQC interactions. We ensure this calculus is complete for all of Gaussian CVQC interpreted in infinite-dimensional Hilbert space, by translating the completeness in affine Lagrangian relations by Booth, Carette, and Comfort. Applying our calculus for quantum error correction, we derive graphical representations of the Gottesman-Kitaev-Preskill (GKP) code encoder, syndrome measurement, and magic state distillation of Hadamard eigenstates. Finally, we elucidate Gaussian boson sampling by providing a fully graphical proof that its circuit samples submatrix hafnians.

Synthesis of Code-Reuse Attacks from P-Code Programs

Mark Denhoed

We present a new method for automatically synthesizing code-reuse attacks—for example, using Return Oriented Programming—based on mechanized formal logic. Our method reasons about machine code via abstraction to the p-code intermediate language of Ghidra, a well-established software reverse-engineering framework. This allows it to be applied to binaries of essentially any architecture, and provides certain technical advantages. We define a formal model of a fragment of p-code in propositional logic, enabling analysis by automated reasoning algorithms. We then synthesize code-reuse attacks by identifying selections of gadgets that can emulate a given p-code reference program. This enables our method to scale well, in both reference program and gadget

library size, and facilitates integration with external tools. Our method matches or exceeds the success rate of state-of-the-art ROP chain synthesis methods while providing improved runtime performance. This work has been accepted to Usenix Security 2025.

Information Theoretic Decomposition Network homophily measures

Vivek Kothari

Existing network homophily measures quantify the tendency of connected nodes to be similar but fail to identify the underlying sources of this similarity. Understanding whether homophily stems from intrinsic node features, neighbor interactions, or their combination is crucial for comprehending network structure and predicting Graph Neural Network (GNN) performance, especially in challenging heterophilous settings. This work proposes an information theoretic framework utilizing Partial Information Decomposition (PID) to dissect network homophily. By treating a node's label as the target variable and node/neighbor attributes as source variables, PID allows us to decompose the mutual information, revealing the unique, redundant, and synergistic contributions of these sources to homophily.

This decomposition provides a fine-grained understanding of what drives homophily in a network. We hypothesize that the specific composition of these information-theoretic components, rather than just the overall homophily level, dictates GNN effectiveness. The research aims to apply this PID-based analysis to characterize diverse networks and correlate the resulting information profiles with various GNNs' performance. The goal is to establish a clearer link between the origins of homophily and GNN behavior, ultimately informing the development of more robust and interpretable GNN architectures.