The Alan Turing Institute

2019 Doctoral Students Yearbook

Foreword

One of the Institute's goals is to build skills for the future. This yearbook demonstrates this achievement and is a celebration of your individual success as well as your collective contribution to the success of the Institute.

As the final cohort of the Institute's Doctoral Programme, you have been instrumental in shaping the Turing's student experience and research landscape. We are extremely grateful for your input to the community and ongoing dialogue with our team over the years.

You showed great resilience in managing the significant challenges of the pandemic so soon after commencing your PhDs. Despite the disruption to face-to-face collaboration (and an array of other things!) for nearly 2 years, you found alternative ways to collaborate and progress your research. At the time of writing, our 2019 doctoral cohort produced over 60 papers and developed work with industry and public bodies through internships at organisations such as Meta, Amazon, Deep Mind, NHSX and GCHQ.

Since graduation, some of your peers are now working as post-docs in academia or researchers at leading companies across the world, using your skills to solve real-world problems.

We are proud of you all and thank you for enriching our Turing community.

The Academic Services team

Georgia Koumara, Sam Selvarajah and Tom Chandler

Notes

"Prateek Gupta (2017 Doctoral Cohort), Ilan Price (2018 Doctoral Cohort) and Christian Blex (2018 Doctoral Cohort) are included in this yearbook due to a later submission date. Sedar Olmez (University of Leeds) is a 2018 Doctoral student funded by The Alan Turing Institute and ESRC for 4.5 years (2018/19 to 2022/23).

Peter Carr (University of Birmingham) was also a 2019 Doctoral student. Both Peter and Cécile de Bézenac (University of Leeds) are 2019 Doctoral students funded by The Alan Turing Institute and ESRC for 4.5 years (2019/20 to 2023/24) and on submission of the yearbook entry were on their 3rd year of PhD."

Students

Over the course of the past three and a half years, we have undertaken a significant journey together. The memories of our initial meeting during the induction week are still with us, as are the days spent attending classes in Enigma together and the daily lunches. We consider ourselves fortunate to be doctoral students within the Institute and express gratitude for the opportunity to get to know all of you during this experience.

Completing a PhD at The Alan Turing Institute offers a unique experience. Working with individuals from various academic disciplines, including statistics, computer science, linguistics, philosophy and psychology; is crucial due to the interdisciplinary nature of data science and AI. The networking benefits extend beyond academia, allowing interactions with colleagues from software companies, policymakers, and venture capitalists - all essential stakeholders in the AI ecosystem. The Institute effectively facilitates connections among these entities, enriching our PhD experience.

The highlight of the program is undoubtedly the cohort itself. Despite our diverse backgrounds, we quickly formed connections through regular lunches and social events. We appreciate the involvement of the Academic Services team in enhancing our Turing experience.

The pandemic posed challenges, making it harder to maintain these bonds. The vibrant Turing office space, surpassing previous academic environments, was sorely missed. Nonetheless, we are grateful for the opportunities provided by the Turing and the doors it has opened for our future upon completion of the doctoral programme.

The pandemic further complicated the non-linear nature of PhDs, making many adjustments necessary. We have been unable to gather at the Turing kitchen for informal conversations over coffee. As the Institute closed its doors, we relocated our desks and chairs closer to our respective home universities or cities, adapted our research plans accordingly, and continued our work.

Nevertheless, it has been amazing to witness an increasing number of paper acceptances in conferences and journals, as well as share in the achievements of internships, teaching positions, and future plans among our cohort. As time has passed, many of you have moved away from London, and we are no longer newcomers at the Turing. We hope that wherever you are, you have found this PhD journey rewarding and cherish the memories created over the years. We encourage you to maintain the curiosity, wonder, and perseverance that characterised Alan Turing's life work.

We miss all of you and wish you the best in your future endeavors. We look forward to reconnecting with you in the near future.

Giulia Occhini and Joe Early
(2019 Doctoral Student Representatives)

Augustinas Sukys, University of Edinburgh

Research areas

- Dynamical systems & differential equations
- Neural networks
- Nonlinear dynamics
- Applications (Machine learning)
- Stochastic (Mathematical modelling);

Selected publications & papers

Published papers:

1. J. Holehouse, A. Sukys, and R. Grima, “Stochastic time-dependent enzyme kinetics: Closed-form solution and transient bimodality”, The Journal of Chemical Physics 153, 164113 (2020).

2. A. Sukys and R. Grima, “MomentClosure.jl: automated moment closure approximations

in Julia”, Bioinformatics 38, 289–290 (2021).

3. A. Sukys, K. Öcal, and R. Grima, “Approximating solutions of the Chemical Master equation using neural networks”, iScience 25, 105010 (2022).

Presented a research poster in:

4. The 20th conference on Computational Methods in Systems Biology (CMSB 2022).

5. The 12th European Conference on Mathematical and Theoretical Biology (ECMTB 2022).

Thesis title

Approximation and inference methods for stochastic systems biology

Supervisors

Ramon Grima

Main outcomes of my research

My research is focused on developing and implementing novel methods to approximate the chemical master equation, which is commonly used to model biochemical systems such as gene regulatory networks, where stochastic fluctuations can have a significant effect on the system dynamics. Over the course of my PhD, I worked on deriving approximate closed-form solutions for enzyme kinetic mechanisms, built a software tool for automated moment closure approximations and proposed a neural network-based approach for surrogate modelling of chemical reaction networks.

Impact of my work

The chemical master equation in most cases cannot be solved analytically and its exact simulation can be very computationally expensive. The approximation methods we have developed can be used to greatly reduce this computational cost and model stochastic biochemical systems much more efficiently, enabling otherwise infeasible parameter exploration studies and parameter inference from experimental data.

Being a Turing Doctoral Student

I really enjoyed the dynamic multidisciplinary environment that the Turing offers, which is great for making connections and seeking collaborations. I had a chance to meet many fellow Turing doctoral students and researchers that have been a constant source of support throughout my studies. I also appreciate the freedom I had to learn a great deal about diverse research topics and pursue my own interests that helped me to develop as a researcher along the way.

Plans following my PhD completion

I am yet to decide between academia and industry.

Cecile de Bezenac, University of Leeds

Research areas

- Multi-agent reasoning
- Causality
- Agent-based modelling
- Modelling (Statistical methods & theory)
- Graph theory
- Simulation
- Research methods
- Spatial analytics
- Social networks
- Uncertainty quantification

Thesis title

Uncertainty in spatial causal inference: an Agent Based Modelling approach

Supervisors

Alison Heppenstall, Ed Manley, Peter Tennant

Main outcomes of my research

In my research, I am interested in better understanding the challenges of spatial causal inference and evaluating the methods that are used to identify causal structures in spatial contexts. I use an Agent Based Modeling approach to quantify the uncertainty generated by the spatial and social assumptions underlying many observational studies. My aim is more generally to develop a framework for thinking about space and causality.

Being a Turing Doctoral Student

The Turing is an incredible community of researchers, with a clear ambition to act on the many fronts of interdisciplinary research. Being a Turing Doctoral Student has therefore meant, for a little while at least, being part of this movement and seeing it evolve rapidly. With the realisation that a PhD is very often an isolating experience, the purpose of the Doctoral Programme was to create a broad but cohesive group of students and equip them with the resources to become talented researchers. While lockdown has largely impacted this road-map, I feel that I have indeed integrated a group of curious and motivated students that have helped me keep going through the PhD.

Plans following my PhD completion

Although it is difficult to plan ahead, until the PhD is completed, I look forward to integrating a community of researchers and working with a team on research interests that are genuinely shared. The constant exchange of ideas and the knowledge of advancing on projects for more than a degree is something I have been aiming for and what has brought me to academic research.

Christian Blex, University of Oxford

Research areas

- Complexity (Algorithms)
- Data science of government & politics
- Multi-agent systems
- Research methods
- Applications (Machine learning)
- Social media
- Agent-based modelling
- Social networks
- Graph theory

Selected publications & papers

1. Paper in the Journal of Mathematical Sociology with Taha Yasseri "Algorithmic Bias cannot stop fragmentation in homophilic networks"

2. Paper in Journal of Computational Social Science with Taha Yasseri, Patrick Gildersleve, Rachel Dinh "Computational Courtship: Understanding the evolution of online dating through large-scale data analysis"

Thesis title

Causal inference methods for supporting, understanding, and improving decision-making

Supervisor

Scott Hale

Main outcomes of my research

Polarisation on social media is a quasi-natural outcome given network structures and social mechanisms, such as homophily. In that way social network are self-polarising entities, independently from aspects such as algorithmic exacerbation. This includes phenomena such as self-fragmentation or inverted spiral of silences via reverse agenda-setting. Self-fragmentation is essentially a mathematical limit of homophilic networks, whilst an inverted spiral of silence makes majorities self-silence through a variety of factors and extremists more emboldened thus distorting the visible online opinion distribution. These phenomena are obviously strengthened under current algorithmic regimes, which encourage more homophilic attachments and information exchange.

Impact of my work

My work shows that social media networks are self-polarising, even in the absence of algorithmic regimes currently in place. This casts a radically different light on the naive outlook that social media networks are beneficial areas of democratic information exchange. Given their self-polarising nature they are inherently destabilising for public discourse, irrespective of algorithmic interventions. In fact, it may imply a responsibility for policy makers and social media platforms to explore ways to mitigate underlying tendencies on top of reversing current exacerbating regimes. Furthermore, this work has shed new light on ideas such as reverse agenda-setting and proposed a polarisation model including an interaction between traditional and social media leading to an inversion of the famous spiral of silence. Moreover, it has given a mathematical proof of the Schelling model in a network setting, thus giving a strong example of self-fragmentation of social networks.

Being a Turing Doctoral Student

To be independent and free in my research. This included to be free of considerations and interests of supervisors, funding councils etc. and just explore my interests, especially after the focus of my work changed slightly during the course of my PhD. Furthermore, the Turing allowed me to work truly interdisciplinary and not be shackled by disciplinary boundaries or orthodoxies. The latter would have been hard to obtain if not part of such an interdisciplinary environment. The Turing was also extremely helpful during the pandemic and when I had a strong change of path in my career and was extremely accommodating. The combination of doctoral with other external studies would not have been otherwise possible. Finally, the prestige of the Turing and the access to world-class researchers allowed me to enter academia at an extremely high level.

Plans following my PhD completion

I am currently the assistant conductor at the Berlin Philharmonic Orchestra and Gustav Mahler Jugendorchester and will further pursue this career.

Dominic Danks, University of Birmingham

Research areas

- Neural networks
- Deep learning
- Applications (Machine learning)
- Unsupervised learning

Selected publications & papers

1. Dominic Danks & Christopher Yau, BasisDeVAE: Interpretable Simultaneous Dimensionality Reduction and Feature-Level Clustering with Derivative-Based Variational Autoencoders, Proceedings of the 38th International Conference on Machine Learning, PMLR 139:2410-2420 (ICML 2021)

2. Dominic Danks & Christopher Yau, Derivative-based Neural Modelling of Cumulative Distribution Functions for Survival Analysis, The 25th International Conference on Artificial Intelligence and Statistics (AISTATS 2022)

3. Fabian Falck, Christopher Williams, Dominic Danks, George Deligiannidis, Christopher Yau, Christopher C. Holmes, Arnaud Doucet & Matthew Willetts, A Multi-Resolution Framework for U-Nets with Applications to Hierarchical VAEs, The Thirty-Sixth Annual Conference on Neural Information Processing Systems (NeurIPS 2022)

Continued at **turing.ac.uk/people/former-doctoral-students/dominic-danks**

Thesis title

Constraint-aware machine learning methods for biomedical data analysis

Supervisor

Christopher Yau, Alastair Denniston, Andrew Beggs

Main outcomes of my research

During my PhD, I developed a number of machine learning approaches to certain biomedical problems, in particular pseudotemporal analysis and time-to-event modelling. In the area of pseudotemporal analysis, I introduced a model, BasisDeVAE (ICML 2021), which performs simultaneous learning of pseudotemporal profiles from cross-sectional data and assignment of those pseudotemporal profiles to interpretable clusters. In the area of survival analysis, I developed neural network based approaches (e.g. DeSurv, AISTATS 2022) capable of flexibly modelling survival data directly without the restrictions or assumptions inherent to many current models and without the need to necessarily work with the hazard function.

Impact of my work

During my PhD, I developed a number of machine learning approaches to certain biomedical problems, in particular pseudotemporal analysis and time-to-event modelling. In the area of pseudotemporal analysis, I introduced a model, BasisDeVAE (ICML 2021), which performs simultaneous learning of pseudotemporal profiles from cross-sectional data and assignment of those pseudotemporal profiles to interpretable clusters. In the area of survival analysis, I developed neural network based approaches (e.g. DeSurv, AISTATS 2022) capable of flexibly modelling survival data directly without the restrictions or assumptions inherent to many current models and without the need to necessarily work with the hazard function.

Being a Turing Doctoral Student

Being a part of the Turing community as a Doctoral Student provided me with numerous valuable experiences and opportunities which I feel fortunate to have had. Firstly, it placed me at the heart of the UK data science scene, allowing easy access to talks, workshops and wider events held at or in conjunction with the Turing, many of which covered topics that I would perhaps otherwise never have engaged with. In addition, it allowed me to easily network with a wide range of researchers working in a variety of areas of data science and its application areas. For example, a chance meeting with a UCL urologist at a Turing health interest group in my first year led to a collaborative paper publication a year later, and a Turing-associated event led to me collaborating with a number of Turing and external colleagues on work which became a successful NeurIPS submission! Without the Turing network, it would have been much more difficult, if not impossible, to engage with such rewarding projects. Finally, the Turing's generous allowances allowed me to engage with a wide array of training and conference opportunities which I may otherwise not have been able to be a part of.

Plans following my PhD completion

I am now working as a Data Scientist within BT Group.

Georgia Tomova, University of Leeds

Research areas

- Graph theory
- Research methods
- Causality
- Simulation

Selected publications & papers

'Published papers:

1. Tomova GD, Gilthorpe MS, Tennant PWG. Theory and performance of substitution models for estimating relative causal effects in nutritional epidemiology. 2022. Am J Clin Nutr 116(5):1379-88.

2. Tomova GD, Arnold KF, Gilthorpe MS, Tennant PWG. Adjustment for energy intake in nutritional research: a causal inference perspective. 2022. Am J Clin Nutr 115(1):189-98.

3. Arnold KF, Gilthorpe MS, Alwan NA, Heppenstall AJ, Tomova GD, McKee M, Tennant PWG. 2022. Estimating the effects of lockdown timing on COVID-19 cases and deaths in England: A counterfactual modelling study. PLoS ONE 17(4):e0263432.

4. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, Harrison WJ, Keeble C, Ranker LR, Textor J, Tomova GD, Gilthorpe MS, Ellison GTH. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. 2021. Int J Epidemiol 50(2):620-32.

Continued at **turing.ac.uk/people/doctoral-students/georgia-tomova**

Thesis title

Using directed acyclic graphs and data simulations to understand causal effects in nutritional epidemiology involving compositional data

Supervisors

Peter Tennant, Michelle Morris

Main outcomes of my research

I used causal diagrams and data simulations to examine different approaches to analysing and interpreting compositional data in the field of nutritional epidemiology. In the context of energy intake adjustment, my research involved placing different adjustment methods in a causal framework, exploring their robustness to specific forms of confounding, and proposing a flexible approach that can be used to robustly estimate a variety of potential estimands. In the context of food substitution modelling, I explored the performance of the different modelling approaches, in particular when the component variables are not in the same units. Finally, I examined the performance of regression-based modelling approaches in the presence of non-linear relationships, and how they compare to established 'gold standard' methods.

Impact of my work

Although the interpretation of different models for energy intake adjustment and food substitution have been discussed previously, my work was the first to explicitly place these in a formal causal framework using directed acyclic graphs (DAGs). I was the first to examine the performance of the different approaches in the presence of confounding by common causes of diet. I also showed that the common practice of using different units of analysis in food substitution models does not respect the compositional relationship between the variables, and may produce uninterpretable estimates. I also showed that, as long as any non-linearity is correctly modelled, a DAG-informed regression-based approach is suitable for the modelling of compositional data in epidemiology.

Being a Turing Doctoral Student

Most importantly, to conduct my research; I am beyond grateful for the funding provided by the Turing. Through this, I have been able to present my research and attend various conferences around the world, which have all made my experience so much more enriching and enjoyable. Being a doctoral student at the Turing allowed me to co-establish and co-lead the Causal Inference Interest Group and be part of the wider community in that field. Through our seminar series, I met and networked with people from different scientific backgrounds that I might not have otherwise had a chance to, which was undoubtedly so valuable.

Plans following my PhD completion

I would like to stay in academia and will be looking for suitable positions in the field of epidemiology.

Giulia Occhini, University of Bristol

Research areas

- Safe and ethical
- Causality
- Applications (Machine learning)
- Natural language processing
- Data science of government & politics
- Spatial analytics

Selected publications & papers

None of my papers or talks were facilitated directly by being at the Turing, but of course by being on a Turing studentship all my scientific output during my PhD was indirectly supported by the Turing.

Thesis title

What, who, where and why: NLP, spatial econometrics and web-archival data for the study of the digital economy

Supervisors

Emmanouil Tranos, Levi John Wolf

Main outcomes of my research

My research develops novel methodological pipelines for the study of digitally mediated economic activities currently not captured by traditional statistical surveys. The problem mostly arises due to statistical surveys' static nature, resulting in a disparity between a country's industrial structure and the way in which this structure is reflected in government industry classification systems. By using unconventional data sources such as web archives, together with techniques from NLP and spatial econometrics, we show improved speed and accuracy in the classification of digital economic activities. Such improvements help us quantitatively answer a range of socio-economic questions relating to economic evolution and inequalities within digital entrepreneurship.

Impact of my work

Providing new, up-to-speed ways to classify and monitor the economy is a long standing issue within governmental statistical agencies. The methodological frameworks and data pipelines introduced within my thesis attempt to address these issues, while at the same time providing concrete applied examples of how this new classification framework can help understanding participation in the digital economy, and whether traditional patterns of demographic and spatial inequalities replicate in this economic field.

Being a Turing Doctoral Student

As a doctoral student at Turing, I had the opportunity to meet with many talented people with varying research interests. This experience provided me with a valuable understanding of the broader AI community and its interdisciplinary nature. Access to Turing's computational resources greatly aided my research, particularly as many of my projects relied heavily on parallel computing. In addition, active engagement with the Research Software Engineering community significantly contributed to my growth as a data scientist.

Plans following my PhD completion

While finishing my PhD, I have been employed part-time as a Research Software Engineer within my University and I am planning to continue down this route. I am also currently qualifying as a yoga instructor and hoping to find a way to carry on these two paths at the same time.

Ilan Price, University of Oxford

Research areas

- Neural networks
- Deep learning

Selected publications & papers

1. Price, Ilan, and Jared Tanner. "Dense for the price of sparse: Improved performance of sparsely initialized networks via a subspace offset." International Conference on Machine Learning. PMLR, 2021. (Jul 18 -24 2021)

2. Trajectory growth lower bounds for random sparse deep ReLU networks. Price, Ilan, and Jared Tanner. "Trajectory growth lower bounds for random sparse deep ReLU networks." 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2021. (Dec 13 - 15, 2021)

3. Price, Ilan, and Jared Tanner. "Improved Projection Learning for Lower Dimensional Feature Maps." ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023. (Upcoming: Jun 4 - 9 2023)

Thesis title

Sparse and low-dimensional deep learning

Supervisor

Jared Tanner

Main outcomes of my research

My research results touch on four areas: Mathematical theory concerning sparse random DNNs; an architecture for training extremely sparse DNNs with minimal trainable parameters; a technique for shrinking the memory footprint of trained CNNs by compressing their feature maps and foldding this compression into the architecture; and techniques for generating and training DNNs with sparse activations.

Impact of my work

To advance research on resource efficient deep learning, moving towards training and using deep networks using less computational time, memory, and energy costs. This has been done through the publication of this work in popular publication venues in the field.

Being a Turing Doctoral Student

A DPhil experience above and beyond what I could have had without my Turing affiliation, and for which I am very grateful. It provided generous funding to attend conferences and summer/winter schools which advanced my knowledge, network, and research impact, as well as giving me the opportunity to give invited talks, with the associated opportunities to build relationships and future collaborations with hosts and others. The London office was a great place to work, albeit intermittently. My affiliation combined with the network of Turing fellows made reaching out to top people in the field, which I did once or twice, more accessible and less intimidating. The Academic Services team was also consistently excellent and helpful with the little roadblocks/admin issues that tend to otherwise cause DPhil students stress.

Plans following my PhD completion

I will be working as a research scientist at DeepMind.

Jiaee Cheong, University of Cambridge

Research areas

- Artificial intelligence
- Computer vision
- Applications (Machine learning)
- Natural language processing
- Speech recognition
- Ethics
- Causality

Selected publications & papers

1. Towards Gender Fairness for Mental Health Prediction. 2023. J. Cheong, S. Kalkan and H. Gunes. To appear as part of the Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23).

2. Towards Causal Replay for Knowledge Rehearsal in Continual Learning. 2023. N. Churamani, J. Cheong, S. Kalkan and H. Gunes. Accepted for oral presentation at AAAI'23 Bridge Program. To appear as part of the Proceedings of Machine Learning Research.

3. Causal Structure Learning of Bias for Fair Affect Recognition. 2023. J. Cheong, S. Kalkan and H. Gunes. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops, 340–349.

4. Causal Fairness for Affective Computing. 2022. J. Cheong, S. Kalkan and H. Gunes. Accepted at NeurIPS 2022 AFCP Workshop.

5. Counterfactual Fairness for Facial Expression Recognition. 2022. J. Cheong, S. Kalkan and H. Gunes. To appear as part of the ECCV 2022 Workshop Proceedings.

6. The Hitchhiker's Guide to Bias and Fairness in Facial Affective Signal Processing: Overview and Techniques. 2021. J. Cheong, S. Kalkan and H. Gunes. IEEE Signal Processing Magazine. Volume: 38, Issue: 6, November 2021.

Thesis title

Causality and Fairness for Affect Recognition

Supervisors

Professor Gunes

Main outcomes of my research

Causality and fairness are important aspects for affect detection given the high-stakes involved in affect detection use-cases. Affect detection encompasses a range of tasks such as mental health detection, facial expression recognition and mood disorder analysis. Given the intricate relations between the different variables involved, a causal understanding of the problem is integral towards developing fairer solutions.

Impact of my work

Its contribution towards Artificial Intelligence (AI) for Social Good (SG). Given the increasing ubiquity and application of machine-learning based algorithms, the problem of bias in these systems is now becoming an even greater source of concern. Research on bias and fairness in machine learning for affect recognition for mental health has been relatively sparse till date. Thus my research sheds light on the root cause of bias in affect recognition and suggests solutions to address them.

Being a Turing Doctoral Student

To expand my horizons by getting acquainted with a diverse range of talented researchers, develop the important skills needed to be a good researcher and feel fully supported throughout the entire learning and growing process.

Plans following my PhD completion

I may pursue post-doctorate research to further my research in this area.

Joseph Early, University of Southampton

Research areas

- Artificial intelligence
- Neural networks
- Applications (Machine learning)
- Reinforcement learning
- Computer vision
- Semi-supervised learning
- Deep learning

Selected publications & papers

1. Joseph Early, Christine Evers, Sarvapali Ramchurn. "Model Agnostic Interpretability for Multiple Instance Learning". ICLR 2022

2. Joseph Early, Tom Bewley, Christine Evers, Sarvapali Ramchurn. "Non-Markovian Reward Modelling from Trajectory Labels via Interpretable Multiple Instance Learning". NeurIPS 2022

3. Joseph Early, Ying-Jung Deweese, Christine Evers, Sarvapali Ramchurn. "Scene-to-Patch Earth Observation: Multiple Instance Learning for Land Cover Classification". Tackling Climate Change with Machine Learning: Workshop at NeurIPS 2022

4. Gregory Everett, Ryan Beal, Tim Matthews, Joseph Early, Timothy Norman, Sarvapali Ramchurn. "Inferring Player Location in Sports Matches: Multi-Agent Spatial Imputation from Limited Observations". AAMAS 2023

5. Keri Grieman, Joseph Early. "A Risk-based Approach to AI Regulation: System Categorisation and Explainable AI Practices". SCRIPTed Journal 2023.

6. Saqib Rahman, Joseph Early, Matt De Vries, Megan Lloyd, Ben Grace, Gopal Ramchurn, Timothy Underwood. "Predicting response to neoadjuvant therapy using image capture from diagnostic biopsies of oesophageal adenocarcinoma". European Journal of Surgical Oncology 2022

Thesis title

Interpretable Multiple Instance Learning: Theory, Methods, and Applications

Supervisor

Gopal Ramchurn, Christine Evers

Main outcomes of my research

My research has investigated how the decision-making process of machine learning models can be better understood. I have specifically focused on multiple instance learning, where models learn from inexact labels, and developed a range of interpretability techniques for use in computer vision, reinforcement learning, and time series analysis.

Impact of my work

During my PhD, I have published both methodological and applied papers at major machine learning conferences such as NeurIPS and ICLR. I have also collaborated on projects involving cancer research, climate change, sports analytics, and AI regulation. In these projects, I worked alongside academics from the University of Southampton, University of Bristol, Queen Mary University, Georgia Tech, and The Alan Turing Institute.

Being a Turing Doctoral Student

As a Turing Doctoral student, I have been able to collaborate on a wide range of projects and talk to people from outside my own research area. This made my PhD a much richer experience. I was also able to co-found an interest group at the Turing.

Plans following my PhD completion

Industry/startup.

Keri Grieman, Queen Mary University of London

Research areas

- Artificial intelligence
- Robotics
- Safe and ethical
- Applications (Machine learning)
- Data science of government & politics

Selected publications & papers

1. Grieman, Early ‘Regulation of Artificial Intelligence: System categorisation and corresponding explainability practices’ (2023) SCRIPTed Journal of Law, Technology & Society 20:1 56-88 <https://script-ed.org/archive/volume-20/issue-1/>

2. Presentation at the 'Business of Automated Mobility Forum' 2021

3. Reed, Grieman, Early, ‘Non-Asimov Explanations Regulating AI Through Transparency’ (2021) Nordic Yearbook of Law and Informatics <eprints.soton.ac.uk/455135/>

4. Grieman, Early 'Regulation of AI and Corresponding Explainability Practices' Poster at AI UK 2020

5. Grieman, Early 'Regulation of AI and Corresponding Explainability Practices' Centre for Commercial Law Studies Staff Seminar 2020

6. Presentation at 'MOVE: the future of mobility' 2020

Thesis title

Regulation of AI Causing Accidental Death

Supervisors

Chris Reed

Main outcomes of my research

In this thesis I examined how to regulate artificial intelligence (AI) causing accidental death, addressing both liability and regulatory structure. This included an examination of the fundamental aspects of fault and liability; an analysis of how technical tools can be used to bridge the techno-legal gap in regulation of AI; and an evaluation of how the UK, EU, US, and Singapore are currently regulating AI as a bloc and in specific sectors. The thesis ultimately provides a jurisdiction-agnostic blueprint for regulation of AI causing accidental death that can be integrated with existing governance mechanisms.

Impact of my work

While the technical field of AI has raced ahead, regulation and legal structures have not kept pace. Given that AI offers desirable advancement in many fields, regulation that balances both innovation and safety is critical. Regulation has so far not addressed in any fundamental way what liability with regard to AI looks. Providing a full regulatory structure for high risk, high reward AI not only provides desirable stability and predictability to industry, but also takes steps to build and earn societal trust. This work also provides key steps in responding to AI incidents in terms of both liability and investigation.

Being a Turing Doctoral Student

Being a Turing doctoral student provided me the support and structure to examine in depth a critical and fascinating area of emerging regulation and law. My discussions with Turing researchers and my fellow students were invaluable both in growing my understanding of the field, and in making sure that the ideas I proposed were grounded in technical workability. Being at the Turing also allowed me to learn from a variety of lectures, discussions, presentations, and simply chats around the coffee machine – all valuable learning opportunities.

Plans following my PhD completion

I will be continuing in a research position on law and regulation of AI and robotics.

Léo Gorman, University of Bristol

Research areas

- Artificial intelligence
- Neural networks
- Applications (Machine learning)
- Reinforcement learning
- Computer vision
- Semi-supervised learning
- Deep learning

Selected publications & papers

Papers

1. Gorman, L. et al. What’s Stopping Knowledge Synthesis? A Systematic Review of Recent Practices in Research on Smallholder Diversity. Front. Sustain. Food Syst. 5, 395 (2021).

2. submitted to Nature Scientific Data) Gorman, L. et al. The Rural Household Multiple Indicator Survey, 50,396 Observations from Farming households in 35 countries (2023)

Conferences

3. Leveraging the Power of Small Scale Data Collection Efforts, Ninth International Conference on Agricultural Statistics (ICAS IX) Washington D.C.

Thesis title

Leveraging the Power of Small-Scale Data Collection Efforts in Agricultural Research for Development

Supervisor

Professor Andrew Dowsey, Dr Jim Hammond, Dr Christopher Woods, Professor William Browne, Professor Mark Eisler

Main outcomes of my research

My thesis is made up of three main bodies of work: 1) improving our understanding of how data is used to improve our understanding of smallholder farmers in Lower- and Middle-Income countries; 2) developing tools to encourage development organizations to collect of interoperable and comparable data on smallholder farmers; 3) developing procedures to analyze these datasets, accounting for potential biases.

Impact of my work

The software I developed to compile a dataset of over 50,000 household surveys from 115 organisations in 35 countries. The data produced by this software has been used development organizations and researchers to better understand the heterogeneity of smallholders, the impact of interventions, and how the gendered control of resources can affect food-security and nutrition. The procedures I have developed, I hope, will help researchers deal with this complex dataset in a transparent way.

Being a Turing Doctoral Student

Working at Turing has allowed me to work with people from different domains with a diverse skill-set. My work was more on the applied end of the spectrum, and I felt lucky to be able to see work from those developing and testing cutting edge methods.

Plans following my PhD completion

I will be a Data Science Specialist at the University of Bristol's Jean Golding Institute.

Limor Gultchin, University of Oxford

Research areas

- Safe and ethical
- Supervised learning
- Neural networks
- Deep learning
- Applications (Machine learning)
- Causality

Selected publications & papers

Papers/conferences:

1. AISTATS 2020:
Gultchin, L., Kusner, M., Kanade, V. & Silva, R.. (2020). Differentiable Causal Backdoor Discovery. Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research 108:3970-3979

2. ICML 2021:
Gultchin, L., Watson, D., Kusner, M. & Silva, R.. (2021). Operationalizing Complex Causes: A Pragmatic View of Mediation. Proceedings of the 38th International Conference on Machine Learning, in Proceedings of Machine Learning Research 139:3875-3885

3. Mastouri, A., Zhu, Y., Gultchin, L., Korba, A., Silva, R., Kusner, M., Gretton, A. & Muandet, K.. (2021). Proximal Causal Learning with Kernels: Two-Stage Estimation and Moment Restriction. Proceedings of the 38th International Conference on Machine Learning, in Proceedings of Machine Learning Research 139:7512-7523

Continued at **turing.ac.uk/people/doctoral-students/limor-gultchin**

Thesis title

Causal and Trustworthy Machine Learning

Supervisors

Matt Kusner, Varun Kanade

Main outcomes of my research

Various novel causally-inspired ML methods, most of which resulted in publications in top venues in the field.

Impact of my work

Enhance our understanding of ML-causal inference intersections, and more trustworthy ML systems, with contributions in interpretability, fairness and invariant predictions.

Being a Turing Doctoral Student

To explore research, to collaborate with others in the ML community in the UK, and to travel between my home institution (Oxford) and the Turing (London) easily and freely.

Plans following my PhD completion

I started a data consultancy, and am working closely with a Private Equity Client in the US. May also be interested in future research opportunities.

Martyn Fyles, University of Manchester

Research areas

- Artificial intelligence
- Neural networks
- Applications (Machine learning)
- Reinforcement learning
- Computer vision
- Semi-supervised learning
- Deep learning

Selected publications & papers

Papers:

1. "Using a household-structured branching process to analyse contact tracing in the SARS-CoV-2 pandemic" - M. Fyles et. al.

2. "Diversity of symptom phenotypes in SARS-CoV-2 community infections observed in multiple large datasets" M. Fyles et. al.

3. "Inferring the relationship between viral load and infectiousness in SARS-CoV-2 cases using contact tracing data" M. Fyles et. al.

4. "Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example" C. Overton et. al.

5. "The role of regular asymptomatic testing in reducing the impact of a COVID-19 wave" M.E.P Silva et. al.

6. "Public perceptions and interactions with UK COVID-19 Test, Trace and Isolate policies, and implications for pandemic infectious disease modelling" G.C. Marshall et. al.

Conferences:

7. "On the relationship between viral load and infectiousness"
- talk, Epidemics 8, poster, EMCTB

Thesis title

Interactions between heterogeneity and interventions for network epidemics

Supervisor

Thomas House, Ian Hall, Lorenzo Pellis

Main outcomes of my research

The overall focus of my research was in developing models for interventions of epidemics. Naturally, a lot of this work was focussed upon the COVID-19 epidemic that was ongoing at the time. In particular, I attempted to quantify various sources of heterogeneity in epidemics, such as super-spreading behaviour, or the clustering of symptoms. I then studied the impact that these sources of heterogeneity had on interventions. For example, large amounts of superspreading can make contact tracing especially effective if it designed with superspreading in mind. Overall I produced three papers, that cover; detailed contact tracing models, co-occurrence patterns in symptoms, and methods for evaluating the efficacy of lateral flow tests.

Impact of my work

As much of my work was performed with the COVID-19 pandemic ongoing, I had the opportunity to present work to SPI-M - the scientific pandemic influenza modelling subgroup of SAGE (scientific advisory group for emergencies). This allowed me to contribute the scientific guidance and response to the COVID-19 pandemic. In addition, I was able to: develop detailed insights into how contact tracing systems are best designed; discovered different symptom co-occurrence patterns in specific age groups; and developed the evidence based for lateral flow tests, which has led to the proposed use of multiplex testing in hospitals.

Being a Turing Doctoral Student

Being a Turing Doctoral student allowed me to develop and improve my skillset in mathematics, data science and software engineering by providing me with opportunities and training that I would not have otherwise had. This meant that I was fully equipped to work at a rapid pace, while also keeping the software to a high standard when this was required to respond to queries from SPI-M/SAGE.

Plans following my PhD completion

I have a job offer from the UK Health Security Agency as a senior infectious disease modeller.

Oliver Hamelijnck, University of Warwick

Research areas

- Supervised learning
- Non-parametric & semi-parametric methods
- Modelling (Statistical methods & theory)
- Time series
- Spatial analytics
- Uncertainty quantification

Selected publications & papers

1. 'Spatio-temporal variational Gaussian processes': Oliver Hamelijnck*, William J. Wilkinson*, Niki A. Loppi, Arno Solin, Theodoros Damoulas, Conference on Neural Information Processing Systems (NeurIPs), 2021

2. 'Transforming Gaussian processes with normalizing flows': Juan Maronās*, Oliver Hamelijnck*, Jeremias Knoblauch, Theodoros Damoulas, International Conference on Artificial Intelligence and Statistics (AISTATS) 2021

3. 'Non-separable Non-stationary random fields': Kangrui Wang, Oliver Hamelijnck, Theodoros Damoulas, Mark Steel', The International Conference on Machine Learning, 2020

4. 'Multi-resolution Multi-task Gaussian Processes', Oliver Hamelijnck and Theodoros Damoulas and Kangrui Wang and Mark Girolami', Conference on Neural Information Processing Systems (NeurIPs), 2019

Thesis title

Scalable Bayesian Inference for Spatio-Temporal Gaussian Processes

Supervisors

Theodoros Damoulas

Main outcomes of my research

Throughout my time at Turing I have published 4 papers at top machine learning conferences. In my research, I have focused on Bayesian methodologies to incorporate inductive biases and prior constraints in spatio-temporal models, while ensuring computational scalability. To achieve this, I have developed Gaussian process-based models that adhere to constraints such as positivity and boundedness, as well as those inspired by physics, and can accommodate observations from diverse sources. Throughout, I have focused on computational efficiency and have introduced a variational framework that enables these models to scale linearly in the temporal dimension by leveraging Kalman smoothing algorithms.

Impact of my work

Gaussian processes are state-of-the-art in spatial statistics. However, many limitations restrict their application to real-world problems. In my work, I have shown how such models can be effectively applied to a wide range of settings and allow for the incorporation of prior knowledge, all whilst tackling the cubic computational cost that hinders most applications. A direct impact of my work has been within Turing's London Air Quality Project, where my models sit at the heart of the system and provide 48 hours forecasts of air pollution across London.

Being a Turing Doctoral Student

My time at Turing has been an enriching experience. The Turing office enhanced my collaborations and being surrounded by experts in the field was inspiring. I had the freedom to work on what I wanted, which was a great driver for my research.

Plans following my PhD completion

I am looking to start a post-doc.

Prateek Gupta, University of Oxford

Research areas

- Multi-agent system
- Information theory (Applied mathematics)
- Operations research
- Applications (Machine learning)
- Artificial intelligence
- Deterministic (Mathematical modelling)
- Multi-agent reasoning
- Simulation
- Neural networks

Selected publications & papers

1. Proactive Contact Tracing, PLOS Digital Health, 2023

2. Prateek Gupta, Tegan Maharaj, Nasim Rahaman, Martin Weiss, et al.

3. Lookback for Learning to Branch, TMLR, 2022

4. Prateek Gupta, Elias Khalil, Didier Chetélat, Maxime Gasse, Yoshua Bengio, Andrea Lodi, M. Pawan Kumar

5. Predicting Infectiousness for Proactive Contact Tracing, ICLR, 2021 (Top 15%)

6. Yoshua Bengio, Prateek Gupta, Tegan Maharaj, Nasim Rahaman, Martin Weiss, et al.

7. Hybrid Models for Learning to Branch, NeurIPS, 2020

8. Prateek Gupta, Maxime Gasse, Elias B. Khalil, M. Pawan Kumar, Andrea Lodi, Yoshua Bengio

Continued at **turing.ac.uk/people/doctoral-students/prateek-gupta**

Thesis title

Imitation Learning for Combinatorial Optimization and Contact Tracing

Supervisor

M. Pawan Kumar, Andrea Lodi, Yoshua Bengio

Main outcomes of my research

I worked on two different applications of imitation learning, specifically, developing a framework to imitate an expert-yet-practically-infeasible heuristic for branching in discrete optimization solvers. These problems arise quite commonly in real-world scenarios such as inventory management or mail delivery systems. The other application is to develop a contact tracing framework for effectively managing pandemics. I undertook this project at the onset of the pandemic, which involved understanding the downsides of the current contact tracing framework, developing a better framework through the use of an agent-based model (ABM), and investigating how AI can be used to enhance the framework.

Impact of my work

The work on developing a superior framework to imitate branching heuristics has its application in the mixed-integer programming solver community. Private firms and open-sourced communities are considering the findings of this work to incorporate into the discrete optimization solvers.

Finally, the work on contact tracing was not directed to be a publishing exercise. We worked closely with the Québec government to deliver a better contact tracing application. The source code for this application, which we called COVI, based on the framework we developed, is open-sourced. However, when the government decided to use no app due to privacy concerns in the wake of the first wave of the pandemic, we decided to publish our findings to share our ideas with a broader academic community and justify our time spent on the project. This work was awarded third place at the 10th UK-wide Doctoral Researcher Award 2021.

Being a Turing Doctoral Student

Being a Turing Doctoral Student allowed me the flexibility of choosing projects and researchers I can work with and customise my experience by choosing to work at various institutes. I was placed at the Turing for my first year, following which I spent several years at the University of Oxford and the Montréal Institute of Learning Algorithms (Mila). Thus, I was able to learn from researchers at various institutes, and form collaborations, all throughout driven by the desire to do quality research to tackle societal problems, a value at the heart of the Turing's mission.

The easy access to the researchers at Turing was also useful when dealing with the ups and downs of doing a PhD.

Plans following my PhD completion

I plan to join Max-Planck Institute in Berlin to conduct research on deep learning and creativity.

Sami Alabed, University of Cambridge

Research areas

- Artificial intelligence
- Non-parametric & semi-parametric methods
- Modelling (Statistical methods & theory)
- Time series
- Spatial analytics
- Uncertainty quantification

Selected publications & papers

1. Alabed, Sami, and Eiko Yoneki. "High-dimensional bayesian optimization with multi-task learning for rocksdb." In Proceedings of the 1st Workshop on Machine Learning and Systems, pp. 111-119. 2021.

2. Alabed, Sami, and Eiko Yoneki. "BoGraph: structured bayesian optimization from logs for expensive systems with many parameters." In Proceedings of the 2nd European Workshop on Machine Learning and Systems, pp. 45-53. 2022.

3. Data Study Group team. (2021, October 11). Data Study Group Final Report: CatsAi. Zenodo. https://doi.org/10.5281/zenodo.5562660

Thesis title

A framework for automated structured Bayesian optimization in computer systems

Supervisors

Eiko Yoneki

Main outcomes of my research

A framework that allows optimising computer systems quickly - reducing energy consumption and costs while improving their speed. It simplifies the process of integrating expert knowledge in models that provide structures to iterative optimisers by using logs and system blueprints to guide the optimiser and model the system better.

Impact of my work

Can be used with any system design or system designing tool - for example, I have a case study on exploring the design space of a chip, leading to finding chips with configurations that uses less energy while providing better speed - these chips can be used in autonomous vehicles, phones, etc. The work can be also applied on databases and reduce the cost of using large databases.

Being a Turing Doctoral Student

Connect with other students who are working on very interesting domain problems and learn about the wider impact of AI on society. I had several training opportunities that improved the quality of my research and research engineering skills. Have an independent advisor to my PhD progress.

Plans following my PhD completion

I am currently working full time as a Research scientist at Google DeepMind. Working on an automated compiler for optimising large-scale neural network models. This combines both the skill I acquired in both my masters and PhD.

Sara Masarone, Queen Mary University of London

Research areas

- Artificial intelligence
- Supervised learning
- Neural networks
- Unsupervised learning
- Applications (Machine learning)
- Deep learning
- Pattern recognition

Selected publications & papers

1. Hernandez G*, Masarone S*, Ross J. Pott J., Brohi K, Barnes M, Pennington D. Development of a proteomic indicator of injury severity (2023). (In progress)

2. NG S, Masarone S, Watson D, Barnes M. The benefits and pitfalls of machine learning for biomarker discovery. Cell and Tissue research. (2023) (Under review)

Workshops:

Masarone S. Representation Learning to Effectively Integrate and Interpret Omics Data. NeurIPS AI for Science workshop (2022).

Conferences and talks:

Selected talk in the ML track at Cold Spring Harbor - Biological Data Science meeting, NY (2022): "Representation learning to integrate and interpret omics data"

Continued at **turing.ac.uk/people/doctoral-students/sara-masarone**

Thesis title

Understanding Trauma: a Journey through Single and Multi-Omics

Supervisor

Professor Daniel Pennington, Professor Michael Barnes

Main outcomes of my research

This thesis explores the applications of machine learning in precision medicine, demonstrating its potential benefits in clinical practice and calling for further development to improve patient survival and trauma treatment. Specifically, this work demonstrates the use of proteomics data collected at hospital admission to objectively identify and predict injury severity in trauma patients. The result is a predictive framework that validates the utility of molecular data (omics) in complementing existing clinical methods. This work also employs unsupervised methods to stratify patients according to different responses to similar injuries, highlighting the presence of different patient subtypes. Furthermore, I demonstrates the ability to predict long-term outcomes using omics data.

Finally, we compare simpler predictive models with graph neural networks to evaluate the advantages and disadvantages of both in clinical settings. Overall, this work underscores the impact of developing machine learning applications in precision medicine and emphasises the importance of developing tailored treatments for different patient groups.

Impact of my work

Despite trauma being one of the leading causes of death worldwide, our understanding of the molecular physiology of this condition remains limited. Furthermore, most injury severity indicators used to categorize patients at admission are based on anatomical-based scores and fail to consider the genome or immune response of these patients.

To address this knowledge gap, this project aimed to develop a proteomic indicator of severity that could be used alongside existing methods to better categorise patients at admission and understand the main drivers of adverse outcomes. This work demonstrates the utility of these methods and identifies the main biological pathways responsible for adverse outcomes in trauma patients.

Being a Turing Doctoral Student

The Doctoral Programme provided me with the opportunity to work with very interesting people and gain exposure to the extensive scope of applied data science projects throughout the UK. Throughout the programme, I met many brilliant individuals across a broad range of fields, some of which were closely related to my own and others vastly different.

Plans following my PhD completion

I'd like to work in the biotech space/R&D.

Sedar Olmez, University of Leeds

Research areas

- Artificial intelligence
- Applications (Machine learning)
- Evolution & adaptation
- Deep learning
- Multi-agent reasoning
- Reinforcement learning
- Neural networks
- Agent-based modelling
- Data science of government & politics
- Stochastic (Mathematical modelling)
- Simulation
- Spatial analytics

Selected papers & workshops

1. Learning Complex Spatial Behaviours in ABM: An Experimental Observational Study, arXiv: Sedar Olmez, Alison Heppenstall, Dan Birks: https://arxiv.org/abs/2201.01099

2. An Agent-Based Model of Heterogeneous Driver Behaviour and Its Impact on Energy Consumption and Costs in Urban Space, Energies: Sedar Olmez, Jason Thompson, Ellie Marfleet, Keiran Suchak, Alison Heppenstall, Ed Manley, Annabel Whipp, Rajith Vidanaarachchi: https://www.mdpi.com/1996-1073/15/11/4031

3. Exploring the Impact of Driver Adherence to Speed Limits and the Interdependence of Roadside Collisions in an Urban Environment: An Agent-Based Modelling Approach, Applied Sciences: Sedar Olmez, Liam Douglas-Mann, Ed Manley, Keiran Suchak, Alison Heppenstall, Dan Birks, Annabel Whipp: https://www.mdpi.com/2076-3417/11/12/5336

Thesis title

The Emergence of Complex Behaviours in Agent-Based Models using Reinforcement Learning

Supervisors

Professor Alison Heppenstall, Dr Dan Birks, Dr Jiaqi Ge

Main outcomes of my research

The main findings from this thesis are split into three parts. Firstly, agents can learn intelligent behaviours from their environment. Secondly, learnt behaviours were in agreement with theoretical and empirical findings from published literature. Thirdly, agents can adapt to previously unbeknown situations and perform relatively well. Overall, this thesis demonstrates that when integrated with agent-based models, neurologically inspired decision-making algorithms can enhance models by introducing learning and adaptability, making these models better placed to support complex real-world decision-making.

Impact of my work

This thesis has contributed to agent-based modelling by demonstrating how novel RL algorithms can enhance the decision-making accuracy of ABMs by introducing learning and adaptability. The proposed hybrid-ABMs were tested on three distinct complex phenomena, and the findings indicate that the RL agents can learn intelligent behaviours from their environment, adapt to previously unbeknown situations, and perform relatively well. The contributions of this research have practical implications for decision-makers in various domains by providing more accurate and adaptable decision-making models. Furthermore, this research highlights the potential of RL algorithms for enhancing the predictive capabilities of ABMs and opens up new avenues for future studies in this area.

Being a Turing Doctoral Student

The Turing Doctoral Studentship was a unique opportunity that only a handful of people would ever receive. Therefore, every second I spent at the Turing was treasured dearly. From my first day at the interview to the last day, I can say, it has been one of the most eye-opening, insightful and mentally draining experiences. I had the luxury of spending time at The Alan Turing Institute and the University of Leeds, the resources I had access to were more than enough. The admin staff at the Turing were efficient and very helpful, so highly skilled that I sometimes mistook them for academics in their own rights. I was able to converse with like-minded scholars and academic veterans whom I would never have had the chance to meet outside of the The Alan Turing Institute due to their busy schedules. Lastly, the Turing Internship Network allowed me to undertake a 6-month industry placement working on exciting projects, while taking a break from the PhD.

Plans following my PhD completion

I have a job lined up as a Senior Researcher in Social Digital Twins at Fujitsu.

Continued at turing.ac.uk/people/doctoral-students/sedar-olmez

Szymon Walkowiak, University College London

Research areas

- Multi-agent system
- Information theory (Applied mathematics)
- Operations research
- Applications (Machine learning)
- Artificial intelligence
- Deterministic (Mathematical modelling)
- Multi-agent reasoning
- Simulation
- Neural networks

Selected publications & papers

Selected publications, papers, achivements and awards can be found at: turing.ac.uk/people/doctoral-students/szymon-walkowiak

Thesis title

Detecting spatial disorientation in large scale wayfinding data - Towards AI-based diagnosis of early-onset dementia

Supervisor

Professor Ed Manley, Professor Hugo Spiers

Main outcomes of my research

Firstly, my PhD research resulted in estimating benchmarks of normal (i.e. “healthy”) navigation across the population and identifying cultural and geographic differences in wayfinding using large scale data obtained from a mobile application video game. This part of the research also helped to validate a number of theories (which were previously not tested on large, global data) related to the effect of individual differences such as sex or age on wayfinding abilities. Secondly, I designed an algorithm which allows detection of spatial disorientation in real-time and defined spatiotemporal features which can be used to predict disorientation based on behavioural characteristics of wayfinding trajectories e.g. accelerometric features, tortuosity, angular body/head direction, path-to-boundary distances etc.

Impact of my work

My research has exerted a substantial impact on our understanding of the contribution of socio-demographic factors such as age and sex on wayfinding abilities and navigational performance at global scale. It has also redefined the way we measure spatial disorientation of navigating agents in various environments by utilising metrics and features which were previously poorly understood or not defined at all. Finally, my research was able to find the link between the characteristics of wayfinding trajectories and spatial disorientation which may serve as an early indicator of developing dementia symptoms.

Being a Turing Doctoral Student

Being a Turing Doctoral student allowed me to develop my data scientist’s skillset to include methods I wasn’t previously exposed to, collaborate with a large number of researchers from diverse academic fields, enhance my soft skills by learning to communicate with other fellow academics and write drafts of research papers. As a Turing Doctoral student I also had access to first class research facilities and a large collection of academic resources including all leading journals, books, and computing facilities. Finally, I was part of the energetic and diverse community of fellow PhD students and leading AI researchers in the UK.

Plans following my PhD completion

I've recently accepted an unconditional offer for the position of the Lead Data Scientist at the Ministry of Justice (Data Science Hub). I'm currently undergoing pre-appointment clearance checks. If I'm unsuccessful, I will continue working as a data science/AI management consultant, which I did before I started PhD at Turing.

Takuo Matsubara, Newcastle University

Research areas

- Asymptotic (Statistical methods & theory)
- Modelling (Statistical methods & theory)
- Estimation theory
- Uncertainty quantification

Selected publications & papers

Published Paper

1. Mastubara, T., Knoblauch, J., Briol, F-X., Oates, C. J. (2022) Robust Generalised Bayesian Inference for Intractable Likelihoods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 84(3):997-1022.

2. Mastubara, T., Oates, C. J., Briol, F-X. (2021) The Ridgelet Prior: A Covariance Function Approach to Prior Specification for Bayesian Neural Networks. Journal of Machine Learning Research 22:1-57.

Paper in Revision or Submission

1. Mastubara, T., Knoblauch, J., Briol, F-X., Oates, C. J. (2023) Generalised Bayesian Inference for Discrete Intractable Likelihoods. In Revision.

2. Mastubara, T., Mudd, R., Tax N., Guy, I. (2023) TCE: A Test-Based Approach to Measuring Calibration Error. In Submission.

Award

1. Best Paper Award, ASA 2022

2. Best Paper Award, ISBA 2021

3. Postgraduate Research Prise, Newcastle University

Thesis title

Bridging the Gap Between Modelling and Computation in Bayesian Statistics

Supervisors

Chris Oates

Main outcomes of my research

Models that involve intractable normalising constants represent a major computational challenge to statistical inference, since the computation of intractable normalising constants requires numerical integration of complex functions over large or possibly infinite sets, which can be impractical. In particular, Bayesian inference for intractable models demands a specially tailored algorithm to bypass evaluation of two nested intractable normalising constants originating from posterior and model simultaneously. My research addressed this computational challenge through the development of a novel generalised Bayesian inference approach built on a Stein discrepancy, called SD-Bayes, with its methodological, theoretical, and computational foundations provided.

Impact of my work

Models that involve intractable normalising constants are ubiquitous in today's "computer" age where the complexity of statistical modelling has been increased more than ever before. Such intractable models appear in many important applications including, random graph models and energy-based models versatile in machine learning. My research opened up a novel efficient approach to performing Bayesian uncertainty quantification for such complex models. In particular, this is the first generalised Bayesian methodology considered in the context of intractable models. Our methodology enables highly efficient and robust uncertainty quantification in the presence of intractable normalising constants, circumventing the associated computational challenge.

Being a Turing Doctoral Student

Being a Turing Doctoral student allowed me to open up the possibilities for my research. Their generous financial support allowed me to concentrate my full focus on my PhD research and participate in a sufficient number of fruitful research activities taken place during my PhD study. I was also privileged to organise a seminar series of the Data-Centric Engineering programme during 2020, where I was fortunate to get acquainted with experienced researchers in the course of the seminar organisation. The Alan Turing Institute offers an accessible environment for me that facilitates to have a frequent discussion about my research themes with collaborators and other researchers in the same field. Their comfortable desk spaces and computational resources assisted me in propelling my research projects consistently. The convenient physical location of the Institute was helpful for me to participate in useful opportunities for my career development. Finally, my affiliation with the Institute as the UK's national institute of data science and artificial intelligence (AI) aided me in promoting myself and gaining fortunate opportunities to develop my knowledge.

Plans following my PhD completion

I will start my position of Willian Gordon Seggie Brown Research Fellow at the University of Edinburgh for the next three years from September 2023, working on foundational facets of Bayesian inference.

Victoria (Torty) Sivill, University of Bristol

Research areas

- Game theory
- Safe and ethical
- Artificial intelligence
- Time series

Selected publications & papers

1. Sivill, Torty, and Peter Flach. "LIMESegment: Meaningful, Realistic Time Series Explanations." International Conference on Artificial Intelligence and Statistics. PMLR, 2022.

2. Torty Sivill, Vanja Ljevar, James Goulding, Anya Skatova. "What Can Transactional Data Reveal About the Prevalence of Menstrual Pain in England?" Digital Footprints Conference 2023

3. Delivered presentation `Explaining Explainable AI' at the Jean Golding Institute Data Showcase 2022, LV Insurance, University of Bristol Faculty of Law

Thesis title

Explainable Time Series: Exploring Post-hoc Local Explanations from a Multidisciplinary Perspective

Supervisor

Peter Flach

Main outcomes of my research

Given an AI system, which perhaps predicts the probability of lung cancer from several lifestyle factors, a post-hoc local explanation is one of the form `this patient was classified as at risk of lung cancer because their blood pressure is in the 95th percentile', i.e. it explains an AI system's output for an individual in terms of their input features. Post-hoc local explanations have been widely successful for explaining image and tabular data yet explaining time series data has been relatively under-explored. My thesis bridges the interpretability gap between existing ways of generating post-hoc local explanations for image and tabular data to time series data. I bring together game theoretic, causal and statistical ideas to develop four methods for explaining time series, uniting counterfactuals, differential attribution, functional decomposition and frequency domain analysis.

Impact of my work

Black-box models, which describe an algorithm unintelligible to humans, have historically been linked with discriminatory, biased and malevolent behaviour and have thus developed a reputation as the villain of AI. Despite the scepticism, black-box models are increasingly ubiquitous in society, particularly following the advent of large language models it is clear these complex AI systems are here to stay. Explaining and auditing their behaviour has therefore never been so important. This is the underlying motivation of my thesis and although the discipline of Explainable AI is still in its infancy I hope that my PhD, in its own limited capacity, has motivated the importance and potential of explanations in a world where the complexity of AI systems is only going to increase.

Being a Turing Doctoral Student

Being a Doctoral student at the Turing has allowed me to engage in so many activities that, had I not been part of this community, would never have even dreamed of. Sadly, I think the norm as a PhD student is an independent research focused, isolated existence. In contrast, the Turing exposed me to the truly inter-disciplinary, multi-faceted world of AI research. Through Data Study Groups, podcasts, workshops, conferences, it is these experiences that have allowed me to realise that there is huge potential for impactful research outside the traditional confines of academia.

Vitaly Zankin, University of Manchester

Research areas

- Numerical analysis
- Ensemble (Mathematical modelling)
- Neural networks
- Estimation theory
- Deep learning
- Monte Carlo methods
- Supervised learning
- Modelling (Statistical methods & theory)
- Simulation
- Uncertainty quantification

Selected publications & papers

1. Clement Etienam, Kody Law, Sara Wade, and Vitaly Zankin. "Fast Deep Mixtures of Gaussian Process Experts". arXiv:2006.13309 (2022)

2. Sebastian De Pascuale, Vitaly Zankin, Jeremy D. Lore, Ben Russo, Paul Laiu, Birdy Phathanapirom, Steven L. Brunton, J. Nathan Kutz. "Uncertainty Quantification for Model Predictive Control of Tokamak Plasma Boundary Simulations with SOLPS-ITER". 64th Annual Meeting of the APS Division of Plasma Physics (2022)

3. Vitaly Zankin, Kody Law. "Sparse Online Variational Bayesian Inference". SIAM Conference on Uncertainty Quantification (2022)

4. Kody Law, Vitaly Zankin. "Sparse Online Variational Bayesian Regression". SIAM/ASA Journal on Uncertainty Quantification (2022)

5. Vitaly Zankin, Kody Law. "Sparse Online Inference with UQ for Regression and Inverse Problems". SIAM Conference on Computational Science and Engineering (2021)

Thesis title

Static and Online Surrogate Modelling with Uncertainty Quantification

Supervisors

Kody Law

Main outcomes of my research

The main outcomes of my research project revolve around developing scalable, efficient, and flexible modelling techniques in Bayesian inference and machine learning. In one study, we proposed an economical and scalable alternative to exact Bayesian inference in a regression problem with sparsity-promoting priors, which successfully handles larger datasets while maintaining comparable performance in variable selection and uncertainty quantification. In another study, we developed a method combining deep neural networks and mixtures of sparse Gaussian processes experts, providing a flexible, robust model that effectively handles complex data behaviours while maintaining superior performance in terms of accuracy and uncertainty quantification. Finally, we combined Bayesian inference with sparse system identification for model evaluation and selection in the context of tokamak plasma boundary dynamics simulation.

Impact of my work

The impact of my work lies in advancing the methods and efficacy of variational Bayesian inference. We have developed a scalable and cost-effective approximate Bayesian inference approach, which can handle large datasets more efficiently. The developed mixture of experts method, which combines deep neural networks and sparse Gaussian processes, delivers robust, flexible, and efficient supervised learning modeling. Through the open-source software and case studies, I hope to empower researchers to construct improved, comprehensible models, reflecting the methodology established in my thesis.

Being a Turing Doctoral Student

Being a Turing Doctoral student granted me the privilege of collaborating with an interdisciplinary network of leading experts and passionate peers from around the UK. This diverse environment broadened my perspective and enriched my research. My time at the Turing enabled me to undertake a challenging research project of my own choosing, this was complemented by access to unparalleled computational resources, training, and conferences, further enhancing my growth as a researcher.

Moreover, the community at the Turing, with its friendly and open culture, played a significant role during the demanding times of my PhD journey. The substantial financial support offered provided a strong foundation, enabling me to focus exclusively on my research without additional burdens. Turing also paved the way for opportunities beyond academia, broadening my understanding of the tech sector and the application of AI in a commercial setting. All these factors have substantially enriched my doctoral studies and paved the way for my future professional journey.

Plans following my PhD completion

Following the completion of my PhD, I plan to leverage my skills in quantitative finance. Simultaneously, I aspire to maintain ties with academia, potentially through part-time teaching or research collaborations.

Xiaoqing Chen, University of Exeter

Research areas

- Deep learning
- Stochastic (Mathematical modelling)
- Graph theory
- Probability
- Modelling (Statistical methods & theory)
- Spatial analytics
- Monte Carlo methods

Selected publications & papers

1. Chen, X., Bannister, R., Shaddick, G., & Zidek, J. V. (2023). On the Stochasticity of Reanalysis Outputs of 4D-Var. arXiv preprint arXiv:2304.03648.

Thesis title

Theory And Application of Highly Multivariate High-dimensional Spatial Stochastic Process

Supervisor

Gavin Shaddick

Main outcomes of my research

1. Rigorously analysed/proved the stochastic property of the data set;
2. Derived a mix graphical spatial model for highly multivariate and high-dimensional spatial data meanwhile account for asymmetry in the cross-covariance matrix;
3. Rigorously compared with another method;
4. Connect the developed method with modern AI technologies;
5 Quantified the health impact of of relative environmental and climatological problems.

Impact of my work

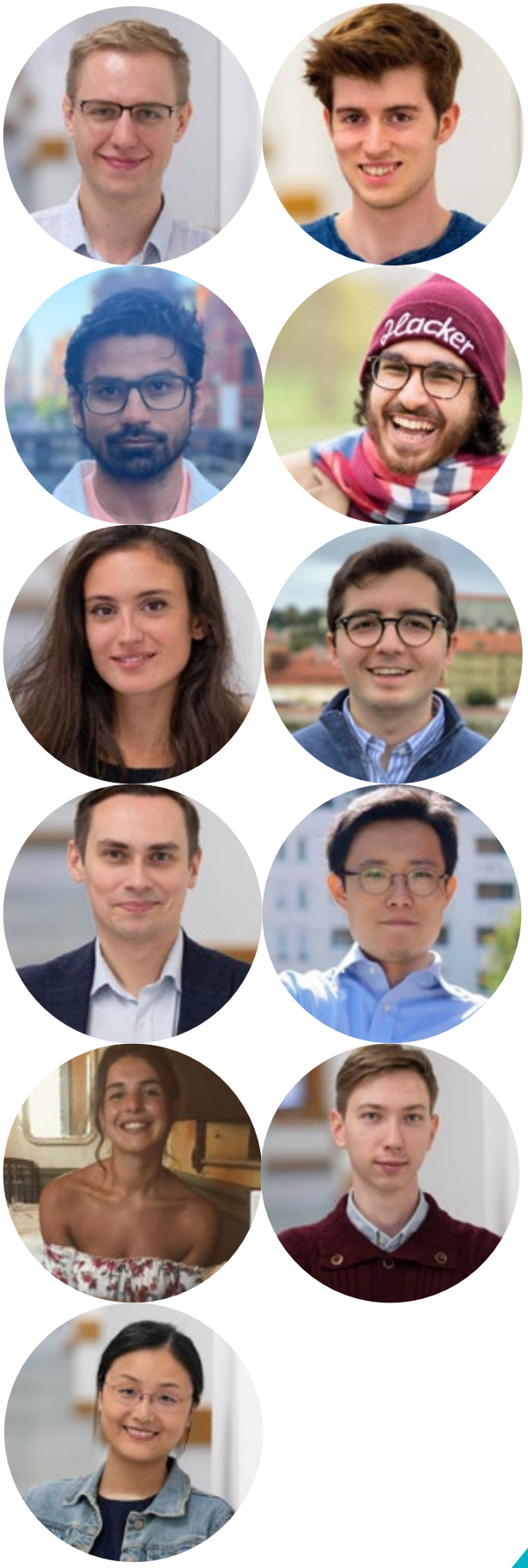
1. The proven stochastic property of the data set can be a solid stepping stone for other spatio-temporal modellers and environmental researchers;
2. The developed modelling framework contributes to the new method to deal with highly multivariate high-dimensional spatial data set, esp. being able to address the asymmetry property;
3. The health impact gives a clear guide to decision-makers;
4. The connection with modern AI technologies empowers the ability of the developed model to address more challenging data set that comes from the real world.

Being a Turing Doctoral Student

1. To have relatively easier communications with external 3rd parties for research needs;
2. Received various support to assist my research;
3. Provide me with plenty of opportunities to learn the world outside of my research domain.

Plans following my PhD completion

I will probably keep doing research-related work.



turing.ac.uk
@turinginst