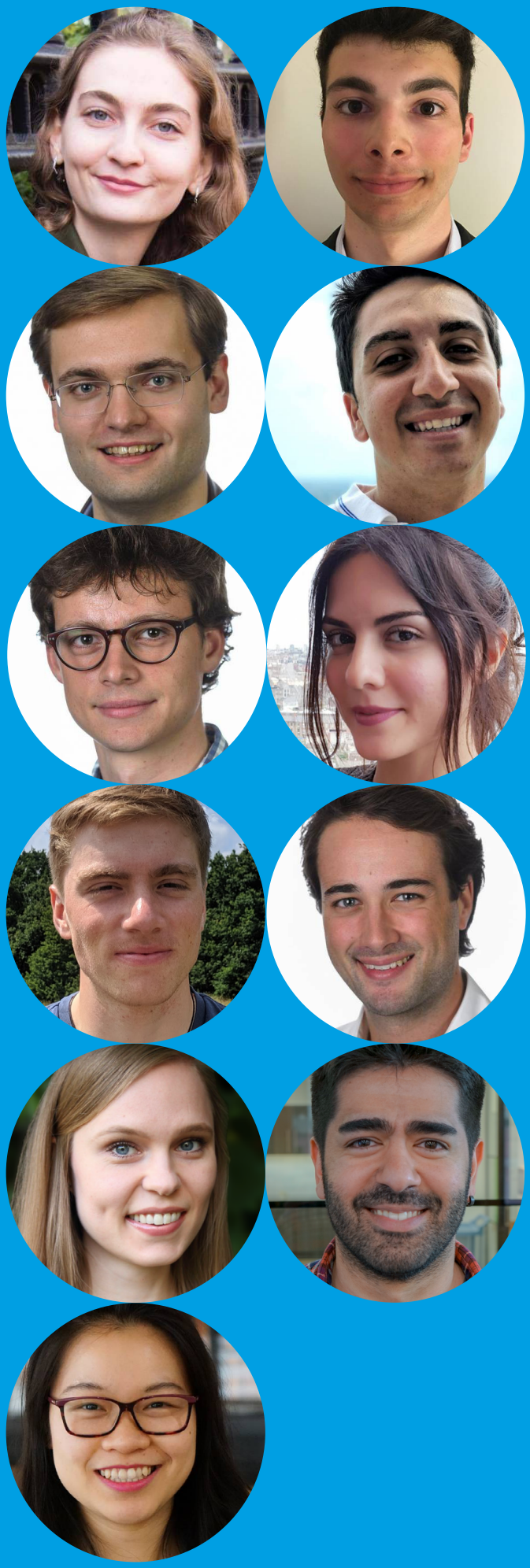


The Alan Turing Institute

Doctoral Students
2017 Yearbook

Foreword

Since its foundation in 2015, one of the institutional goals of the Turing has been training new generations of data science and AI leaders. This yearbook is therefore not only a celebration of the individual success of our 2017 doctoral cohort, but also of their contribution to the success of the Institute as a whole.

This cohort has continued the trend of Turing doctoral students being both inspiring and prolific in their research output. They have produced over 50 research papers; attended more than 45 conferences, workshops, and symposia; and developed work with industrial partners including G-Research, HSBC and Facebook. Since graduation that impressive trajectory has continued, with members of the cohort going on to work in leading companies, taking the next steps on the academic career path and founding their businesses to apply their skills in tackling real-world problems.

Our second cohort of students reflected the broadening diversity of research interests at the Turing and the ongoing strength of the interplay with our university partners.

The students visited the Institute regularly and meeting them and discussing their work was a personal highlight of my first year at the Institute. We learned so much from this cohort about how a national institute could foster the ambition these young leaders have; from restructuring training programmes to facilitating internship opportunities. Their openness to give constructive feedback continued to improve the doctoral scheme and inform how the Turing could serve students at a national level.

We had all hoped that the pandemic may have resolved enough for us to meet back together before the 2017 cohort moved on. Even though a large event was not possible, the cohort continued to feel connected to the Institute, as we celebrated successes and the emerging of new research opportunities. We aimed to train the next generation of leaders, and this cohort has made us proud and confident that we have met our goal.

Dr Ben Murton
Head of Skills

Students

It has been quite a journey for the past three and half years, but we are glad that we have walked the path together. The memory of us all meeting for the first time during the induction week is still vivid, just like those days we spent attending classes and listening to talks in Enigma as well as walking to Kerb food market for Friday lunches. We feel very lucky to be part of one of the Institute's cohorts of doctoral students and grateful to have got to know you all during this experience.

Our desks are no longer part of shared clusters in the British Library and for many months we've not been able to sit down at the Turing kitchen for a quick coffee and chat. While PhDs are rarely linear journeys, the pandemic made sure that ours required some additional navigation. But as the Institute had to close its doors, we relocated our desks and chairs closer to our home universities or home cities, set up our screens and keyboards, and stuck to our research plans or revised them as needed. It's been truly admirable to witness a gradually strengthening stream of news about paper acceptances in conferences and journals, and to share in the joy of achievements such as internships, teaching jobs, plans for future destinations, and other happy life events amongst our cohort.

Many of you have moved away from London as time has gone by and we are no longer the fresh faces at the Turing. We hope that wherever you are, you have found this PhD a rewarding journey and treasure the memories you've made over the past years. Even as time continues to pass, we hope that you'll keep cultivating some of that curiosity, wonderment, and tenacity that we know permeated Alan Turing's lifework. We miss you all and wish you the very best for your future endeavours.

Looking forward to seeing you soon again,

Helen and Sanna

Note: Alex Campbell, Alex Mansbridge, Nicolas Anastassacos, Prateek Gupta*, Florentin Goyens and Amartya Sanyal were also Doctoral Students.

*Due to intermission of studies, Prateek's yearbook entry will be included in a future edition

Albesë Demjaha, UCL

Research areas

- Cognitive science
- Management science
- Research methods
- Social psychology
- Modelling (Statistical methods & theory)

Selected papers & publications

Demjaha, A., Caulfield, T. and Pym, D. 2 Fast 2 Secure: A Case Study of Post-Breach Security Changes. 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).

Demjaha, A., Parkin, S., and Pym, D. You've left me no choices: Security economics to inform behaviour intervention support in organizations. STAST 2019: Socio-Technical Aspects in Security and Trust (pp 66-86).

Demjaha, A., Parkin, S. and Pym, D. Economics–driven behaviour intervention support in organizations. 8th International Symposium 'From Data to Models and Back (DataMod 2019).

Thesis title

Modelling and co-design of security policy for cultural and behavioural aspects in organisations

Supervisors

David Pym, Angela Sasse, Simon Parkin

Main outcomes of my research

My research focuses on the behavioural and cultural aspects of information security. More specifically, it explores the cultural components which may impact compliance with security policy in organisations. My work takes an interdisciplinary approach and considers knowledge from behavioural economics which can help contextualise the decision-making process of employees when making security choices. The thesis methodology is based on co-design principles in that it attempts to co-design security together with the relevant stakeholders, whether that is with co-researchers from other fields (such as modellers), or security managers in organisations.

Impact of my work

Although it has been a long time since the human factor has been identified as critical in the field of information security, challenges with security compliance persist. My work provides in-depth insights into the challenges that employees face with security in their organisations and provides recommendations that consider a multitude of important factors - such as the context of bounded decision-making and other behavioural and cultural aspects. My proposed methodology provides an iterative process of co-design, engaging researchers, modellers and stakeholders. It creates a shared understanding of the system to be modelled, identifies the purpose of the model, and ensures that the specified model is aligned with the needs of the stakeholders and fits within the limitations created by data availability.

Being a Turing Doctoral Student

The Turing was a second home outside the university. It was a place full of diversity, research, excellence, growth, and most importantly, understanding. Being a Turing Doctoral student allowed me to feel empowered and motivated to continuously grow and improve my research. It also provided a fantastic network of people which led to great conversations and collaboration opportunities. The Turing constantly provided opportunities for improvement, through various training, talks, and the chance to showcase your research in front of the researchers as well as the Business Team. Although I will be sad to leave the vibrant Turing office, I will always be a part of the community. I will take with me many important lessons and connections with so many wonderful people.

Plans following my PhD completion

I am switching to industry to work in the Cyber Security area, as I have recently been employed at KPMG Norway as a consultant in the Cyber & Security department.

Alexis Bellot, University of Cambridge

Research areas

- Applications (Machine learning)
- Causality
- High dimensional inference
- Modelling (Statistical methods & theory)

Workshops & conferences

NeurIPS 2019 (December 2019, Vancouver, Canada).

Thesis title

Causal Inference and Hypothesis Testing with Heterogeneous Medical Data

Supervisor

Mihaela van der Schaar

Main outcomes of my research

I developed a range of methods for causal inference and hypothesis testing with applications in the medical domain which frequently encounter heterogeneous data, biased in a myriad of ways and of unconventional presentation including irregularly-sampled time series data and high-dimensional data.

Impact of my work

A number of new problems can be tackled with the methods I developed. For instance, hypothesis testing was typically confined to data samples with the same number of observations and noise patterns, but I showed that a modification of existing tests lends itself to this pattern of observation and that hypothesis tests may be consistently applied.

Being a Turing Doctoral Student

It allowed me to meet extraordinary people, many of whom I hope to remain in touch with for a long time. They have influenced my research and, more importantly, my overall experience which trickles down to many aspects of my life. The highlights of my time at the Turing were the people and the kitchen space. The skills I learned through being a Turing PhD student were learning to speak about my research in many different levels of detail given the wide-ranging backgrounds of people at the Turing. I believe this to be a super important skill for any researcher to have an impact in the world.

Plans following my PhD completion

I will start a postdoc at Columbia University in the Department of Computer Science with Professor Elias Bareinboim.

Andrew Burnie, UCL

Research areas

- Applications (Machine learning)
- Complexity (Algorithms)
- Human computer interface
- Neural networks
- Natural language processing
- Systems theory
- Social media
- Causality
- Time series
- Non-parametric & semi-parametric methods

Selected papers & publications

Burnie, A., Burnie, J. and Henderson, A. 'Developing a Cryptocurrency Assessment Framework: Function over Form'. Ledger, 3, July 2018.

Burnie, A. 'Exploring the Interconnectedness of Cryptocurrencies using Correlation Networks'. In Cryptocurrency Research Conference 2018. (Anglia Ruskin University, Cambridge, UK, May 2018).

Burnie, A., Henderson, A., and Burnie, J. Putting Names to Things: Reconciling Cryptocurrency Heterogeneity and Regulatory Continuity. *Journal of International Banking and Financial Law (JIBFL)*, 33(2): 83-86, February 2018.

Thesis title

Processing Social Media Text for the Quantamental Analyses of Cryptoasset Time Series

Supervisors

Emine Yilmaz, Tomaso Aste

Main outcomes of my research

Quantitative analysis of social media text and cryptoasset time series is combined with the qualitative identification of fundamentals to increase our knowledge of what determines the value of cryptoassets across time. The outcomes are a new cryptoasset classification and three new natural language processing techniques. These new techniques find words associated with the phase of declining bitcoin prices in 2017-18; apply neural networks (word2vec) to extract topics; and apply a framework to identify plausible causes of phasic shifts in the ether and bitcoin price series. This shows the one-off effect of regulatory bans on bitcoin and the recurring effects of rival innovations on ether price, consistent with bitcoin providing a form of money and Ethereum providing a platform for developing applications.

Impact of my work

My research creates a user-friendly cryptoasset classification based on token functionality that has been published in the peer-reviewed journal Ledger and which formed part of the written evidence submitted to the UK Parliament Digital Currencies Inquiry to inform public policy on cryptoassets, in conjunction with Eversheds Sutherland (International) LLP. My research further quantitatively assesses social media discussion forums to identify what events and concerns are associated with major shifts between different phases of price, necessitating new natural language processing methodologies that recognise the need for non-parametric analyses to quantitatively examine discussion forums. This moves the debate from previous analyses of volume and sentiment to associating changes in price with specific events and concerns. This resulted in publications in Royal Society Open Science, SIGIR and Frontiers in Blockchain.

Being a Turing Doctoral Student

As a Turing Doctoral student, I had access to a multi-disciplinary network of experts, first-rate facilities, an excellent support team, and financial sponsorship. I found the Turing classes invaluable in that they capitalised on the variety of expertise available at the institute and provided invaluable insights into different areas of AI and data science research. These talks helped to inform where my PhD research could be taken and how I could develop my skillset to get there. The Alan Turing Institute is a true centre of excellence in the data science and AI field and it has been a great pleasure working here.

Plans following my PhD completion

I am an expert in machine learning, AI and natural language processing with extensive experience extracting actionable insights from a variety of data types. I intend to take the expertise that I developed throughout the PhD and during my previous commercial experience to add value for businesses and organisations trying to make sense of their data.

Francesco Cosentino, University of Oxford

Research areas

Numerical (Algorithms)
Numerical analysis
Calculus & analysis
Dynamical systems & differential equations
Dynamic/static (Mathematical modelling)
Time series
Stochastic (Mathematical modelling)
Monte Carlo methods
Probability
Simulation
Stochastic optimisation
Uncertainty quantification

Papers & publications

Cosentino, F., Oberhauser, H. and Abate, A. 'A Randomized Algorithm to Reduce the Support of Discrete Measures', NeurIPS 2020, arXiv:2006.01757.

Cosentino, F., Oberhauser, H. and Abate, A. 'Carathéodory Sampling for Stochastic Gradient Descent', arXiv e-prints, arXiv:2006.01819v2.

Cosentino, F., Oberhauser, H. and Abate, A. 'Grid-Free Computation of Probabilistic Safety with Malliavin Calculus', arXiv preprint, arXiv:2104.14691.

Thesis title

Formal Verification meets Stochastic Analysis

Supervisors

Harald Oberhauser, Alessandro Abate

Main outcomes of my research

The main outcomes of my research have been to introduce a new randomised algorithm to solve the recombination problem, i.e. finding a reduced measure guaranteed by the Tchakaloff and Carathéodory Theorems, and to present a new way to compute probabilistic safety regions in Formal Verification without fixing a grid in the space.
In particular, the first algorithm has been applied to accelerate optimisation procedures in the big-data framework, and to reduce the number of states necessary to approximate an software-defined everything (SDE) using discrete space Markov Chains.

Impact of my work

Formal Verification is strictly related to concepts of dynamical systems safety, and often numerical techniques must be used to compute the likelihoods of interest. Extend these studies to systems represented by means of SDEs has been the first goal. In general, my research can be applied to various tasks, where the aim is to evaluate desired quantities relatively to stochastic systems.

Being a Turing Doctoral Student

Being a Turing Doctoral student allowed me to interact with people far from my domain of research, extending my personal points of view. Having access to both the Turing and my university was excellent and I really enjoyed it. The interdisciplinarity of the Turing enabled me to extend my network and my research interests, and it boosted my interest in machine learning. The seminars were helpful to introduce me to interesting topics, and probably being a Turing member was essential. It is also relevant to remark the financial and technical support provided by the Turing, which was great.

Julien Vaes, University of Oxford

Research areas

Numerical (Algorithms)
Convex programming
Operations research
Game theory
Numerical analysis
Stochastic optimisation

Papers & publications

Vaes, J. and Hauser, R. (2021). Optimal Trade Execution with Uncertain Volume Target, *Journal of Computational Finance*.

Workshops & conferences

NFORMS Annual Meeting 2019 (October 2019, Seattle, USA).

ICCOPT 2019 - International Conference on Continuous Optimization (August 2019, Berlin, Germany).

ICOEP - International Conference on Optimization and Equilibrium Problems (July 2019, Dresden, Germany).

INFORMS Annual Meeting 2018 (November 2018, Phoenix, USA).

Thesis title

Pricing power future contracts under risk: a game theoretical approach

Supervisor

Raphael Hauser

Main outcomes of my research

My thesis is articulated in two different parts:

1) First I have a theoretical part where I analyse generalised Nash equilibriums when players have non-differentiable payoff functions. I then study the uniqueness and stability of the equilibrium. Then I study the uniqueness of an equilibrium when players are risk averse.

2) In a second step, I apply this theory to trading under uncertainty.

Impact of my work

The impact of my work is twofold: some theory about game theory and then models in finance.

Being a Turing Doctoral Student

It allowed me to be well supported/advised. It helped me also to establish contact with people in a similar field as mine. It also helped me to make some friends.

Plans following my PhD completion

I am starting a postdoc at UCL.

Lina Gerontogianni, University of Cambridge

Research areas

- Applications (Machine learning)
- Unsupervised learning
- Monte Carlo methods
- Deterministic (Mathematical modelling)
- Non-parametric & semi-parametric methods
- Modelling (Statistical methods & theory)
- Probability
- High dimensional inference
- Simulation

Workshops & conferences

Medical Genetics (October 2018).

Greek Stochastics (August 2019, Corfu, Greece).

University of Kent, School of Mathematics, Statistics and Actuarial Science (November 2019, Canterbury, UK).

Thesis title

Variational Mixture Models for non-Gaussian observations: Applications to molecular data

Supervisor

Leonardo Bottolo

Main outcomes of my research

My doctoral research provides speedy model-based clustering tools for DNA methylation data of continuous and discrete type. DNA methylation is an epigenetic modification that represses gene expression when attached in a promoter region. Aberrant levels of this modification may work as an indicator of cancer development or other serious health conditions.

In our research, we conduct analyses on real blood samples of normally and artificially conceived neonates - some of them diagnosed with a rare developmental disorder - to reveal their hidden group structure, based solely on the methylated pattern of their genome (methylome). Our clustering algorithms successfully determine the group of unhealthy neonates, while discover healthy groups of artificially conceived newborns with irregular methylation in imprinting regions (parent-of-origin expressed genes), implying potential development of congenital disorders, such as Beckwith-Wiedemann syndrome and Silver-Russell syndrome.

Impact of my work

The demand for determining the hidden groups of non-normally distributed DNA methylation data, such as betaintensities derived from array-based platforms, or methylated counts from Bisulfite Sequencing techniques, motivated us to build ready-to-use clustering packages for a wide range of non-Gaussian data scenarios. These model packages, are also able to supply the user with the important methylated genomic regions per group. In a nutshell, our research contributes in facilitating the data analysis of epigenetic biological processes, like DNA methylation, by providing a plethora of tools to choose from according to the nature of the genetic data.

Being a Turing Doctoral Student

Being an Alan Turing doctoral student is a decision I would definitely make again and recommend to anyone who is seeking to undertake a PhD programme in data science. The exposure to this interdisciplinary environment, in conjunction with the endless support from the Turing faculty, render the Turing an ideal place to get started with your research career.

Plans following my PhD completion

I have already secured a job position at the Francis Crick Institute as a Senior Biostatistician, starting end of September 2021.

Michael Murray, University of Oxford

Research areas

- Compression (Algorithms)
- Numerical (Algorithms)
- Information theory (Applied mathematics)
- Deep learning
- Asymptotic (Statistical methods & theory)
- Probability

Papers & publications

Murray, M., Abrol, V. and Tanner, J. (2022). 'Activation function design for deep networks: linearity and effective initialisation', *Applied and Computational Harmonic Analysis*, vol.59, 117-154.

Murray, M. (2021). 'From matrix factorisation to signal propagation in deep learning: algorithms and guarantees'.

Murray, M. and Tanner, J. (2020). 'Encoder blind combinatorial compressed sensing'. arXiv:2004.05094.

Murray, M. and Tanner, J. (2018). 'Towards an understanding of CNNs: analysing the recovery of activation pathways via Deep Convolutional Sparse Coding'. arXiv:1806.09888.

Thesis title

From Matrix Factorisation to Signal Propagation: Algorithms and Guarantees

Supervisor

Jared Tanner

Main outcomes of my research

First, an algorithm for solving blind-decoder multimeasurement combinatorial compressed sensing, which has optimal sample complexity up to logarithmic factors, and second, an analysis of the role of the activation function deployed in a deep neural network on signal propagation in both the forward and backward pass. With regards to the second the key takeaway is that using an activation function with a scalable linear region around the origin can be used to avoid certain problems at initialisation.

Impact of my work

The first contribution of my DPhil has applications in sketching, coding and compressed sensing in contexts where the encoder matrix is unknown or too large to transmit and store in regards to the decoder. The second contribution provides principles for activation function selection in deep neural networks.

Being a Turing Doctoral Student

It allowed me to gain awareness of the wider machine learning, AI and data science communities and the interdisciplinary nature of these research areas. It also gave me access to great resources and events early on. The highlights of my time at the Turing were meeting and making some great friends!

Plans following my PhD completion

In September I will be starting an Assistant Adjunct Professorship at UCLA.

Nikolas Kuhlen, University of Cambridge

Research areas

Applications (Machine learning)

Natural language processing

Unsupervised learning

Management science

Research methods

Information theory (Statistical methods & theory)

Modelling (Statistical methods & theory)

Papers & publications

Kuhlen, N. and Preston, A. (2020). News Entropy. Available at SSRN: https://ssrn.com/abstract=3820449 or http://dx.doi.org/10.2139/ssrn.3820449.

Carvalho, V. M., Draca, M. and Kuhlen, N. (2021). Exploration and Exploitation in US Technological Change.

Kuhlen, N. (2021). Endogenous Technology Space. Available at SSRN: https://ssrn.com/abstract=3899948 or http://dx.doi.org/10.2139/ssrn.3899948.

Selected workshops & conferences

Cambridge-UCL Applied Micro PhD Seminar (June 2021).

Economics and Data Science Seminar, (March 2021, Zurich, Switzerland)

Monash-Warwick-Zurich Text-as-Data Conference - Q&A Session (February 2021).

UCLA-Warwick Machine Learning Seminar (November 2020).

Thesis title

Essays on Probabilistic Machine Learning for Economics

Supervisors

Vasco M. Carvalho, Mingli Chen, Chenlei Leng

Main outcomes of my research

My PhD thesis consists of three essays that explore the use of probabilistic machine learning techniques in combination with information-theoretic concepts to answer economic questions. Over the past years, economists have started applying machine learning methods to a wide range of topics. Probabilistic methods in the context of unsupervised learning represent one particular modelling approach at the intersection of computer science and statistics. While widely used in applied statistics, these models, however, do not necessarily provide relevant and interpretable outputs from an economist's perspective. In my research, I appeal to information-theoretic methods to summarise the probabilistic information inferred from such models and construct economically meaningful measures.

Impact of my work

My research builds on a theoretically well-defined framework that allows to empirically measure previously abstract concepts such as policy uncertainty and exploration patterns in innovation from unstructured data. For instance, one of my papers introduces the concept of 'news entropy' to characterise the relationship between news coverage and the economy. Intuitively, news entropy decreases as the news focus on a smaller set of pressing topics. We observe that news entropy exhibits clear negative spikes close to important economic, financial, and political events and find that decreases are associated with two key features: an increase in uncertainty measures and a macroeconomic contraction.

Being a Turing Doctoral Student

The Turing provided a fantastic environment for me to carry out my interdisciplinary research and I was fortunate to be part of a wonderful doctoral cohort and the wider Turing community. Being a Turing Doctoral student allowed me to pursue a research agenda that is not very standard in my field and pushes the boundaries of applied machine learning in economics. I hugely benefited from being surrounded by a diverse group of experts in their respective areas and the resulting cross-pollination of ideas. In addition, being a Turing student not only provided me with freedom in terms of my actual research but also regarding all the elements that were crucial to the practical implementation of this undertaking. Specifically, over the course of the PhD, I spent my time across three different universities, switched PhD programmes halfway through and participated in many workshops and conferences. Without the support by the Turing, both financially and – more importantly – by the amazing academic services team, I would have not been able to do my research and potentially not even reach the end of the PhD. Lastly, I was also able to initiate an industry corporation with HSBC's Big Data Lab through the Economic Data Science programme which was fundamental to the development of my technical skills involved in working with genuinely large, real-world data sets.

Plans following my PhD completion

I will join the next cohort of Entrepeneur First.

Sanna Ojanperä, University of Oxford

Selected research areas

Graph theory

Ethics

Research methods

Data science of government & politics

Neural networks

Management science

Monte Carlo methods

Social networks

Causality

Social media

Selected papers & publications

Fitzgerald, J., Ojanperä, S. and O'Clery, N. (2021). "Is academia becoming more localised? The growth of regional knowledge networks within international research collaboration." Applied Network Science, 6(1), 1-27.

Katta, S., Ferrari, F., Ojanperä, S., Salem, N., Taduri, P., Neerukonda, M., and Graham, M. (2021). Skills for the Planetary Labour Market: Indian Workers in the Platform Economy. Oxford, UK; Manchester, UK; Cape Town, South Africa; Bangalore, India: Fairwork Foundation.

Bertolini, A., Graham, M., Neerukonda, M., Ojanperä, S., Parthasarathy, B., Srinivasan, J., Taduri, P. and Ustek-Spilda, F. Forthcoming in "Platformising Informality, One Gig at a Time" In A. Suriti, and U. Huws (Eds.), Platformization and Informality: Pathways of Change, Alteration, and Transformation, London: Palgrave Macmillan.

Ojanperä, S. (2021). "Tackling the inevitable AI-induced rise in global inequality." The New Statesman/Spotlight, September 17, 2021.

Thesis title

New Forms of Work in the Emerging Digital Economy

Supervisors

Felix Reed-Tsochas, Neave O'Clery, Mark Graham

Main outcomes of my research

The world of work is undergoing rapid changes all around the world. While the actors and processes involved are multiple and complex, the emergence of digital labour platforms as sites to access paid work is a central development affecting job seekers and those looking to contract workers around the world. My research investigates the patterns of work carried out through digital labour platforms. Adopting an evolutionary perspective and applying methods from data science and network science, I focus on the skill landscape of platform labour and the job trajectories of platform workers with a special interest in their geographical underpinnings.

Impact of my work

The many changes related to the future of work are a salient research area, as they directly and indirectly impact individuals, economies, and governments around the world. Among the key policy questions is the growing need to ensure that education, training, and skills development reflect the current and future demands of workplaces. My research contributes to the understanding of digital labour platforms as sites to access job opportunities, the prevalence and value of skills and skill combinations, and the job trajectories of platform workers in particular in terms of skill development. While these platforms constitute a limited part of the larger digital economy, my research will be able to address the developmental impacts of strategies such as digital policies, which place an emphasis on nurturing a market for digital goods and services with the aim of increasing productivity and prosperity.

Being a Turing Doctoral Student

It has been a wonderful experience to find a second academic home at the Institute during my doctoral studies. I have discovered new methods, research approaches, and research questions through my interactions with fellow students and Turing researchers, which have enriched my own research agenda tremendously. Through organizing and attending various events and workshops based at or facilitated by the Institute, I got to collaborate with stakeholders beyond academia, including in the government, NGOs, and the private sector. These experiences have been very valuable for me as beyond generating academic contributions, I aim for my research to generate practical and policy contributions.

While I have spent countless hours in early morning and late night trains between Oxford and London over these past years, it has been a huge privilege to be part of this vibrant and closely-knitted community. I will always treasure my experience as a Turing doctoral student and look forward to many future collaborations with my excellent Turing peers and colleagues!

Plans following my PhD completion

Upon graduation, my plan is to pursue an academic career. I aim to apply for assistant lectureships or post-doc positions where they are available in a topic closely connected to my doctoral research and wider research agenda. I also intend to apply for funding to conduct independent research projects as an early career researcher.

Taha Yusuf Ceritli, University of Edinburgh

Research areas

Applications (Machine learning)

Supervised learning

Papers & publications

Ceritli, T., Williams, C. K. I. and Geddes, J. (2020). ptype: probabilistic type inference. Data Mining and Knowledge Discovery, 34(3):870-904.

Thesis title

Probabilistic Type Inference for the Construction of Data Dictionaries

Supervisors

Chris Williams, Charles Sutton

Main outcomes of my research

My main goal was to become an independent researcher before I arrived at the Turing. I believe that I have improved my skills (both analytical and soft skills) during my PhD. Perhaps the biggest opportunity for me has been an interdisciplinary project at the Turing on semi-automating data science, where I collaborated with other researchers and data scientists.

Impact of my work

I have contributed to the development of robust type inference methods for tabular data. These methods have been shown to perform better then existing methods on real-world datasets. We have also made our work publicly available as a Python package that can be used by the practitioners.

Being a Turing Doctoral Student

I think that the Turing provides a unique environment for academics by bringing them closer with data scientists, who perhaps have more experience in dealing with the challenges in real-world datasets than academics. I found this stimulating for research.

Plans following my PhD completion

I will start working as a postdoc at the University of Oxford on machine learning for healthcare in the coming months.

Zhenzheng (Helen) Hu, UCL

Selected research areas

Applications (Machine learning)

Supervised learning

Unsupervised learning

Estimation theory

Monte Carlo methods

Modelling (Statistical methods & theory)

Non-parametric & semi-parametric methods

Papers & publications

Hu, Z., Kosmidis, I. and Manalopoulou, I. (2021). 'Modelling misclassification in binary studies' (In preparation to submit to Statistical Science).

Thesis title

Dirichlet process probit misclassification mixtures model for misclassified binary data

Supervisors

Ioanna Manolopoulou, Ioannis Kosmidis

Main outcomes of my research

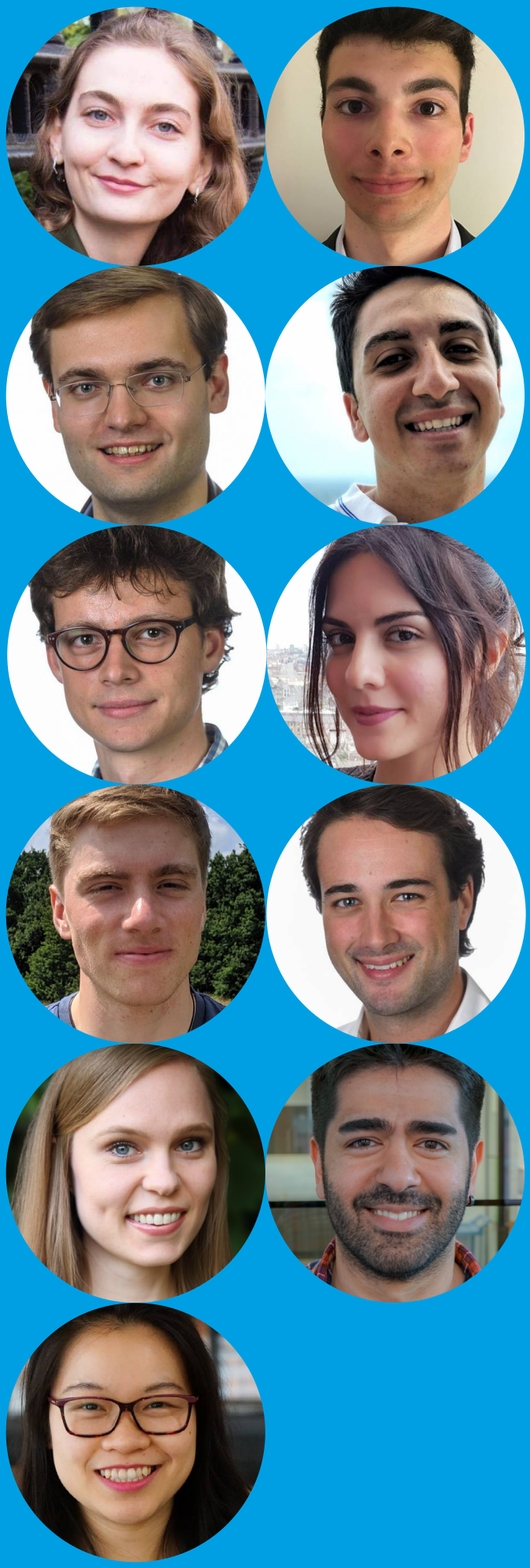
My research addresses a common challenge in binary studies: observations which are potentially mislabelled but whose mislabelling does not occur randomly. I proposed a flexible Bayesian method that models misclassification without assuming constant misclassification probability nor the availability of validation data.

Impact of my work

My research provides the first systematic review of current methods for dealing with misclassification. This could benefit researchers and practitioners facing the problem of misclassification at hand. My research also proposes a flexible method that models misclassification without assuming constant misclassification probability nor the availability of validation data. The proposed method can be applied to any binary data prone to contamination in the responses regardless of application area.

Being a Turing Doctoral Student

Being a Turing student made my graduate studies much more enjoyable that it would have been. It allowed to learn many interesting areas that is not directly related to my PhD and meet many people who I have formed good friendship with. Most importantly, I felt being part of the Turing community for the past four years and felt supported. This was important to my growth both academically and personally. Without it, my PhD would have been much more depressing and frustrating.



turing.ac.uk
@turinginst