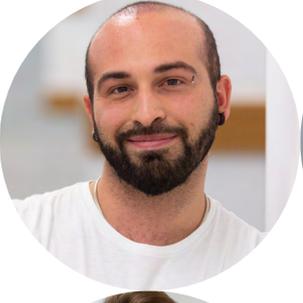
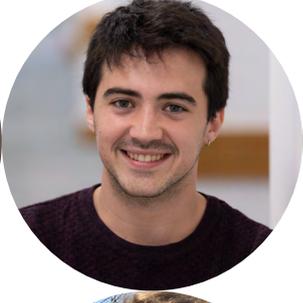


The Alan Turing Institute



Turing Enrichment
Scheme
2019/20 Yearbook

Foreword

Thank you. In this most strange year I am determined to start with “thank you” as this is a message of celebration and a recognition of achievement of our 2019/20 Enrichment Scheme cohort. This year has been our biggest, and I also think our best, in terms of realising the vision we have developed for the programme of bringing together a diverse group of doctoral students, connected by their desire to lead the way in data science and AI.

This cohort threw themselves into life at the Turing, and then continued to build a community and found ways to innovate in the face of the unparalleled challenges presented by the pandemic. Before we left the office, I was so pleased to see students having lunch together, hearing about events and activities that were being organised and recognising how involved this cohort was in life at the Turing.

I was amazed at your resilience and patience when a new way of working was needed, and humbled that so many of you continued to engage remotely.

I hope that we will continue to hear from you as you complete your doctoral studies, and you will consider the Turing part of your extended community. In all you do, go forward with those foundations of ethical and responsible approaches in data science and AI that we hope you have seen in action at the Turing, as you become the next generation of leaders.

Dr Ben Murton

Head of Professional Development and Academic Programmes

Everyone has the tendency to say that their year was the most challenging. Falling into the cliché of saying it myself, I think we all agree that a world pandemic trumps most things. We started our Enrichment program not even imagining that by the end of six months we'd all be in a lockdown, that the Fridays in the pub would be replaced by virtual calls, or that the coffee we'd have next week would be postponed indefinitely. But the truth is that we stuck through it, and whilst we were going through such a difficult and unexpected time, we are managing to get out on the other side. The Turing will forever remain in our minds as

The enrichment scheme is currently one of the most coveted academic placements in the UK. The number of applications far exceeds that of accepted applicants. It seems therefore appropriate to characterise the former group of individuals as some of the brightest and most promising minds in the country. Some were lucky enough to spend the best part of six months in the Institute while others less so. As you skim through the pages of this book, you will meet each talent, learn about their research, their motives and the outcomes of this great experiment. The result is a welcome glimpse into a most enjoyable and

not only the unforgettable place where we had a placement during our PhDs, but the Institute that helped us grow and stayed with us in difficult times. Our cohort did not have one last get together, but when normality sets, we will meet and toast (hopefully not at a social distance) to the fact that not only were we colleagues, but pandemic mates! I really enjoyed being your rep and I sincerely wish you have managed to overcome the outstanding circumstances we have faced.

Beatriz Costa Gomez

Student Representative Enrichment

fruitful past. But amongst what is left untold, is a story of friendship. Of lunchtime meals filled with laughter and plans not unfolded. Of stone circle aspirations and ideas sprung together. Of doctoral angst halted over two sizzling cups of coffee. These moments, they will endure.

Pedro Pinto Da Silva

Student Representative Enrichment

Students



Note: George Elder (QMUL), Solon Karapanagiotis (University of Cambridge), Alessandro Ragano (University College Dublin), Risa Ueno (University of Cambridge) and Eilish Gibson (University of Oxford) were also 2019/20 Enrichment students.

Victoria Carr, King's College London



Project title

Genomic surveillance of the spread of antimicrobial resistance

Supervisors

David Moyes, Peter Mullany

Main outcomes of my research

Microbes that are resistant to antimicrobials can spread across human populations. These microbes contain genes that code for antimicrobial resistance, which can also spread between microbes. A resistance gene from the genome of resistant microbe can be copied/cut and pasted into another genome of a non-resistant microbe, potentially making the microbe resistant.

The transfer of genes between microbes happens continually but in different ways. Many of these genes are part of "mobile genetic elements" which enable the resistance genes to be copied or cut from one genome and pasted into another genome. The challenge is to discover what mobile genetic elements are associated with antimicrobial resistance genes and how prevalent these are.

Impact of my work

A knowledge of how antimicrobial resistance genes are connected to mobile genetic elements would help researchers predict the emergence of new resistant pathogens before epidemics.

Being a Turing Enrichment student

It allowed me to access to the Turing's cutting-edge computational resources has been essential for my research. Also being part of the Turing community has given me an unique opportunity to explore new avenues of research beyond my discipline and receive valuable feedback from experts and mentors.

"An unique opportunity to explore new avenues of research beyond my discipline."

Laura Carter, University of Essex



Project title

Gender stereotyping in automated systems in UK child protection services

Supervisors

Lorna McGregor, Roisin Ryan-Flood

Main outcomes of my research

I am looking at the ways in which data and algorithms are used in the UK's child protection services, to classify and/or predict outcomes for families. My research uses queer theory and feminist analysis to interrogate the use of categorisation in these systems.

It also applies a human rights analysis to assess whether these systems risk violating Article 5(a) of the Convention on the Elimination of All Forms of Discrimination Against Women, which requires states to eliminate practices based on gender stereotypes.

Impact of my work

My research aims to analyse the impact of data-driven systems on the families whose lives are subject to scrutiny by child protection services in the UK, and to identify individuals and families who risk being harmed by these systems, and in particular by the use of stereotypes and categorisations.

My work contributes to the growing body of work in critical algorithm studies, and questions the use of algorithmic systems that may have negative impacts on people's lives.

Being a Turing Enrichment student

As an Enrichment student, I was able to be part of a different community from my home university and faculty, which exposed me to different ideas and ways of thinking about my research. I'm grateful for the many discussions I've had with students, staff and fellows, all of which have broadened and strengthened my work.

I've also enjoyed joining the formal and informal discussion groups within the Institute, particularly the Digital Humanities discussion group, as well as contributing to The Turing Way project. It's also been great to contribute to the work of The Alan Turing Institute. Being part of the Gender and LGBTI+ working group has been an enjoyable way to contribute to the community.

I've also benefited from being the student liaison on the Ethics Advisory Group, which has allowed me to work with the group that navigates the ethical responsibilities of working in an emerging and fast-moving field, and supports other researchers to do the same.

"It exposed me to different ideas and ways of thinking about my research."

Michael Casey, University of Southampton



Project title

Information theory for single-cell biology

Supervisors

Ben MacArthur, Ruben Sanchez-Garcia

Main outcomes of my research

Single-cell gene expression experiments are rich sources of data for many biological phenomena. However, the resulting data is complex and high-dimensional. It is, as such, difficult to know a priori how useful a given data set is. During my time at the Turing, I have worked on developing an information-theoretic framework for assessing how much information is available from such data sets, providing a general measure of data set utility.

Impact of my work

There are increasingly large gene expression data sets being produced, particularly as part of the Human Cell Atlas project, with increasingly complex models. Without robust evaluation measures, we cannot make full use of these data sets, potentially losing out on novel biological discoveries.

Being a Turing Enrichment student

Being at the Turing has increased my confidence as a data scientist. Being surrounded by talented people and talking to them about their work has been an inspiring experience. On a more practical level, the bringing together of people from across disciplines has been great, as I've been able to learn more about techniques beyond those used in my field. In particular, a conversation over coffee led me to the application of information theory in economics, something I would have never otherwise looked at, that ended up inspiring much of my current approach.

“Being at the Turing has increased my confidence as a data scientist.”

Alicia Cork, University of Exeter



Project title

Studying social identities online

Supervisors

Miriam Koschate-Reis, Richard Everson, Mark Levine

Main outcomes of my research

The main outcomes of my research are a new methodology for incorporating social psychological aspects into current data science methodologies. As it stands, much research using human data ignores the psychology of the human and presents each individual as equal. My research provides a way of testing offline social psychological theories in the online realm. Since being at the Turing, I have broadened my understanding of how psychology can be applied in various different aspects of computational social sciences.

Impact of my work

The impact of my work both enhances the tools currently used by data scientists and offers psychologists a new way to understand and study online human behaviour, whilst also looking to understand how psychology can be practically applied online to aid in the prevention and detection of criminal network involvement. More specifically, with this research project, I have been working directly with the National Crime Agency to better understand how we can use NLP techniques to learn more about the group membership of those who engage in online cryptomarkets.

Being a Turing Enrichment student

Being a Turing student has enabled me to gain a greater appreciation for how the social sciences can be best applied to optimise data science research. I have gained an understanding of the current pitfalls of relying solely on a technical and mathematical approach to making sense of data, and thus have a new-found appreciation for the use of social theory within this domain.

Moreover, being an Enrichment student at the Turing has given me the opportunity to get involved in projects outside of my immediate research. I have started a few different collaborations that I believe will change my career trajectory and this never would have been possible without the support of the Turing network.

In addition to providing me with a bigger picture of how the social sciences and data sciences can work together, being a Turing student has also enabled me to develop my fine-grained technical skills. Coming from a non-data science background, the workshops and seminars provided by the Turing have proven fundamental to my understanding of best practices within data science research.

“The Turing has given me the opportunity to get involved in projects outside of my immediate research.”

Beatriz Costa Gomes, University of Manchester



Project title

Quantitative description of microtubule networks in neurodegenerative diseases: software development and image analysis

Supervisors

Andreas Prokop, Matthias Heil

Main outcomes of my research

Creating a software to analyse images with as little user input as possible. I built a pipeline to analyse and classify the neuronal images also using CNNs.

Impact of my work

The ability to quantitatively analyse phenotypes in different neurodegenerative diseases, using *Drosophila* neurons.

Being a Turing Enrichment student

It has allowed me the chance to learn new methods to analyse the images, which was a whole new chapter in my thesis. It also allowed me to meet people that I wouldn't have otherwise and some of which I now consider close friends. I also had the chance to participate and organise a lot of activities, from stand-up comedy to an artificial intelligence Christmas workshop.

“It allowed me to meet people that I wouldn't have otherwise and some of which I now consider close friends.”

Akira Endo, London School of Hygiene & Tropical Medicine



Project title

Inference framework for school-based infectious disease surveillance in the presence of multiple-layers of mixing

Supervisors

Sebastian Funk, Adam Kucharski, Paul Fine

Main outcomes of my research

During my time at the Turing, I explored the possibility of an inference framework for infectious disease transmission data on a large-scale social network: schools and households in particular. To incorporate the flexibility of machine learning methods into the existing infectious disease modelling framework, I developed a fast computation algorithm that can handle complex outbreak data with a network structure. I will further extend this algorithm and study the school-household transmission dynamics of influenza to identify potential determinants of the disease spread.

As COVID-19 outbreak began to unfold during the Enrichment Scheme, I was also involved in multiple COVID-related research projects. I wrote a short paper quantifying the potential role of superspreading events in driving the spread of COVID-19.

Impact of my work

As directly transmitted infectious diseases mainly spread by conversational contacts, it is important to study how diseases spread in the network of social contacts. It has been known that school children play key roles in the spread of the disease in schools and households. The dataset of influenza available for analysis included more than 10,000 students and their family members, which is one of the largest datasets on the disease spread over the school-household network structure. Exploring this dataset with flexible models will help us obtain clues to better epidemic control strategies.

Being engaged in COVID-19 research was one of the greatest opportunities for me to directly contribute to society with my expertise. The superspreading paper attracted wide attention not only from academics but also from media, as preventing superspreading events is a crucial element for post-lockdown exit strategies.

Being a Turing Enrichment student

The Enrichment Scheme was a great opportunity to connect with researchers using similar methods and tools for different goals. Also being in the middle of my PhD program, spending time in a new research environment with new people helped me to refresh my mind and to reframe academic life.

As a student from an interdisciplinary field, talking to other students and researchers with stronger backgrounds in data science was a precious and exciting experience. Although after the COVID-19 outbreak it was not as easy to casually chat with colleagues, the active interaction with the Turing people was the best part of the Enrichment Scheme.

“The active interaction with the Turing people was the best part.”

Marcel Gehrung, University of Cambridge



Project title

Triage-driven diagnosis for early detection of oesophageal cancer

Supervisors

Florian Markowetz, Rebecca Fitzgerald

Main outcomes of my research

Cytosponge technology has been developed and supported by scientists at the University of Cambridge from conception through to the ongoing implementation pilot. The Cytosponge is a cell collection device that consists of an encapsulated sponge on a string, which enables sampling of cells from the oesophagus. This cell sample can be screened for cancer-associated changes. The analysis of Cytosponge microscopy slides is comparable to finding a needle in a haystack. It involves repetitive tasks, resulting in insufficient time for thorough review of difficult cases. I developed a semi-automated triage approach to assist the pathology analysis of Cytosponge samples. This increases accuracy and decreases screening time, increasing throughput and enabling pathologists to focus on difficult cases.

Impact of my work

Late-stage cancer diagnosis can dramatically impair a patient's quality and length of life. For cancer of the oesophagus, fewer than 2 in 10 patients survive longer than 5 years. However, if this cancer can be treated at an early stage, then over 8 in 10 patients survive beyond 5 years. Existing methods to detect this early stage are limited to endoscopy, which is expensive, invasive, and unpleasant. The COVID-19 pandemic has paused several key healthcare services, including endoscopy, an aerosol-generating procedure, resulting in an extensive backlog of patients. There is a clear need for non-invasive approaches for Barrett's Oesophagus diagnosis. The analysis of oesophageal cell samples collected using the Cytosponge device benefits patients who can be examined in an office-based, nurse-led clinic in primary or secondary care. My semi-automated approach substantially increases the throughput of screening processes, making the technology available to more patients. I am now driving further development and adoption of the technology through Cytel.

Being a Turing Enrichment student

The process of analysing Cytosponge samples is resource-intensive and mentally rigorous, even for an experienced pathologist. With the Turing's support, I developed and validated a novel machine learning model which stratifies patients into high- and low-confidence diagnostic results. During the Enrichment placement, I was able to modify the model to be generally applicable to more sample types. I explored the implications of triage models on potential pathologist workload reduction and which heuristics have to be determined for other analysis pathways in order to apply similar approaches.

The environment in London offered great opportunities for interdisciplinary exchange between staff and students. It provided office space which I used for focused work and meetings with collaborators. The Turing community, through various channels of communication and excellent organisation, is of tremendous value for data science researchers like me and I thoroughly enjoyed it.

“The environment in London offered great opportunities for interdisciplinary exchange between staff and students.”

Katriona Goldmann, Queen Mary University of London



Project title

Stratified medicine in autoimmune diseases

Supervisors

Myles Lewis, Michael Barnes

Main outcomes of my research

My research focuses on integrating genomic, transcriptomic and clinical data from clinical trials in autoimmune diseases. Using statistical and machine learning methods, I look to predict specific phenotypes and drivers of disease in order to move towards personalised or stratified medicine and improve patient response to treatment. I also focus on creating interactive computational tools to analyse big data online. This allows me to automate research pipelines, as well as visualise and share biological results with the wider research community.

Impact of my work

Currently many autoimmune disease treatments are prescribed through trial-and-error until clinicians find a plan that works. This results in patients being prescribed ineffective treatment which can potentially have adverse side-effects. The main goal of my research is to identify sub-groups of patients with distinct mechanisms of disease, or particular treatment responses.

This stratified approach allows us to develop a deeper understanding of these groups which can lead to the identification of new drug targets and treatment strategies. This will ultimately allow clinicians to recognise treatments that are most effective for particular groups of patients and ensure they receive the optimal treatment plans.

Being a Turing Enrichment student

During my time as an Enrichment student I was lucky enough to study alongside a fantastically diverse and enthusiastic cohort of students. Working as part of the Turing's collaborative environment has allowed me to think about my own research with new perspectives from different disciplines. This experience has already begun to push my research in directions that would have seemed daunting a year ago.

Attending the various lectures and workshops at the Turing has also allowed me to further my programming and machine learning skills. In particular, the research engineering course gave me the confidence and know-how to publish my own open-source software following The Turing Way guidelines to ensure it was both accessible and reproducible.

“I was lucky enough to study alongside a fantastically diverse and enthusiastic cohort of students.”

Ferran Gonzalez Hernandez, UCL



Project title

Unlocking pharmacokinetic literature with natural language processing

Supervisors

Joseph Standing, Frank Kloprogge, Waty Lilaonitkul, Juha Iso-Sipilä

Main outcomes of my research

During my time at the Turing, I worked on developing Natural Language Processing (NLP) pipelines to recognise specific biomedical entities in scientific articles and efficiently extract pharmacological information from those articles. With support from the software engineering team, I deployed a machine learning classifier on Azure to filter scientific literature in real-time.

Impact of my work

The short-term impact of my work is to facilitate the availability of comprehensive and up-to-date pharmacological data by leveraging NLP and machine learning approaches to process biomedical text. In the long-term, I hope that this research can accelerate preclinical drug development by generating more comprehensive and up-to-date biomedical databases.

Being a Turing Enrichment student

The main highlight of being at the Turing has been an amazing and diverse community that I found every day at the office. Being surrounded by remarkably passionate and driven people allowed me to continuously learn from other colleagues and explore different fields. From a training perspective, the Enrichment Scheme really helped me to improve my software engineering skills and my ability to effectively contribute to open-source projects.

“An amazing and diverse community.”

Daniele Guariso, University of Sussex



Project title

Towards a certification of financial algorithms

Supervisor

Omar A Guerrero

Main outcomes of my research

In this project, we explore the linkage between government spending and the indicators of Sustainable Development Goals (SDGs) set by the United Nations. That is, we try to predict the movement of development indicators from information about changes in budgetary allocations. This is a challenging problem given all the potential interactions between hundreds of indicators and government programmes.

During my time at the Turing, I tried to address this issue by developing an alternative framework based on Artificial Neural Networks (ANNs). This approach provides a flexible architecture to discover those policy issues where indicators are sensitive to budgetary changes.

Impact of my work

This research takes advantage of the global agenda for fiscal transparency, an important movement that has been pushing governments to publish large datasets of public spending and, in some cases, to link them to the SDGs. What we want to achieve with the project, is to provide an innovative tool that can assist policy makers to gauge the actual impact of public spending when framing their development plans. The lessons that can be obtained from our approach can be useful to help treasuries around the world in making the best use of their resources in order to achieve the SDGs.

Being a Turing Enrichment student

At the Turing, I had the chance of being part of a truly diverse community and engage with researchers working on completely different domains. I learnt from the expertise of the other Enrichment students in their specific fields; their different perspective when tackling problems. During the Enrichment Scheme, I widened my set of methodological skills thanks to the several training opportunities offered, developing new heuristics to address my research.

The Enrichment Scheme has been an incredible opportunity to start new multidisciplinary collaborations both with other PhD students and fellows of the Turing. All of this wouldn't have been possible in the “disciplinary narrowness” of my university department. After this experience, I would advise any PhD student to spend some time at the Turing. If you want to broaden your perspectives as a researcher, this is the place to be.

“If you want to broaden your perspectives as a researcher, this is the place to be.”

Man Luo, University of Warwick



Project title

Rebalancing fast expanding EV sharing systems with multi-agent deep reinforcement learning

Supervisors

Hongkai Wen, Hakan Ferhatosmanoglu

Main outcomes of my research

During this year at the Turing, I have been working in multi-agent reinforcement learning applied to electric vehicle sharing systems. In this research, we study the incentive-based rebalancing for fast expanding EV sharing systems. We model the rebalancing task as a Multi-Agent Reinforcement Learning (MARL) problem, which directly takes the range and charging properties of the EVs into account.

We propose a novel approach of policy optimisation with action cascading, which is able to work with the expansion dynamics in the system and solve the formulated MARL. To stabilise training and improve data efficiency, we also propose a regularised reward function to train the policy networks.

Impact of my work

The rebalancing problem of shared e-mobility considered in this project is one of the key research topics to create more efficient and effective operations of these systems. This has received extensive interests from both academia and industry, and the positive effects are significant, especially in urban cities where the current traffic system and real estate market are already reaching their limits. In that sense, shared e-mobility systems with the capability of balancing themselves are a promising solution, which leads to more environmentally friendly cities and significant decrease in private transportation, creating more housing and public spaces. This work has been published by IJCAI 2020.

Being a Turing Enrichment student

The Turing Enrichment Scheme offered a unique opportunity to support my development. From the financial and technical perspective, they provided excellent which really facilitated my research. In terms of academia, the placement enabled me to work with the world-leading Turing researchers in the data science community, which is extremely valuable to my research. The generous support and opportunities of training such as events, seminars and workshops at Turing has also helped me to gain new techniques and knowledge during the placement.

“The Turing Enrichment Scheme offered a unique opportunity to support my development.”

Sally O'Brien, University of Cambridge



Project title

Data and information management for multi-hazard analysis in critical infrastructure organisations

Supervisor

Kristen MacAskill

Main outcomes of my research

My research aims to develop a process to assess the extent to which data quality and information fragmentation in CI organisations is impacting the ability to effectively manage risk exposure to multiple hazards.

Impact of my work

Industry 4.0 is enabling CI organisations to transform the way their systems are being managed. Organisations are collecting and processing more data than ever before in the hope that they can gain new and useful insight into how they can operate, maintain and protect their systems and assets more effectively.

However, the true benefit of these advancements are often far from being realised in CI organisations because of fundamental data management issues within the organisations. My research focuses on better understanding data quality and management problems in CI organisations and developing opportunities for overcoming these issues in practice.

Being a Turing Enrichment student

It allowed me to engage, network and collaborate with a wide range of students, academics and industry partners. This hugely benefited my research as I was able to explore 'out of the box' ideas and options that were developed in group and individual discussions. Even informal chats with colleagues over a cup of coffee lead to some great research developments.

Being a Turing Enrichment student also provided me with an active and supportive environment for conducting my research in. The Turing research groups, teams and facilities were brilliant for conducting research, brainstorming ideas and receiving feedback.

“Being a Turing Enrichment student allowed me to engage, network and collaborate.”

Tugce Oruc, University of Birmingham



Project title

Understanding the working mechanism of antibiotic production via computational approaches

Supervisors

Peter Winn, Christopher Thomas

Main outcomes of my research

My research focuses on understanding the working mechanism of large biomolecular complexes producing chemicals that can be used as antibiotics. For that purpose, I am using deep neural networks and statistical analyses. Thanks to the casual communications with the people from the Turing, I found the opportunity to better understand the mathematical background of the approaches I have been using, besides learning variety of research areas and their application that I was not aware. Additionally, thanks to the courses I took at the Turing, I learned how to convert my scripts into reliable and convenient software.

Impact of my work

Understanding the structure and the working mechanism of the antibiotic-producing biomolecular complexes is a critical step to design more successful experiments to develop new drug candidates. The approaches I have been improved, and the tool I am working on to develop can be used to work on similar large biomolecular complexes promising to generate more drug candidates for the fight against antibiotic resistance.

Being a Turing Enrichment student

It allowed me to meet many people from different educational backgrounds which provided me the opportunity to learn many application areas of machine learning and statistical approaches apart from my topic. Additionally, being in a non-biology oriented environment forced me to explain my work in different ways which not only improved my communication skills but also made me gain additional perspectives to my work. The trainings and courses I took improved my skills and knowledge, and the seminars and lectures I attended provided me to follow recent advancements in the artificial intelligence and data science area in general.

“The trainings and courses I took improved my skills and knowledge.”

Feergus Pendlebury, King’s College London and Royal Holloway, University of London



Project title

The limitations of machine learning for security

Supervisors

Lorenzo Cavallaro, Johannes Kinder, Kenneth Paterson

Main outcomes of my research

My research focuses on the challenges of deploying machine learning (ML) systems in hostile environments. While ML has shown to be effective in lab settings there are a host of complications that arise when used in the wild: adaptive attackers that seek to evade and poison classifiers; an evolving environment that results in concept drift; and poor explainability that limits the use of ML as part of a larger analysis pipeline.

I’ve demonstrated that previous work has vastly overestimated the performance of ML classifiers in security – including how several complex state-of-the-art methods can be outperformed with much simpler alternatives, and shown that large-scale attacks against malware detectors are a realistic threat by generating evasive malware using automated software transplantation.

Impact of my work

The main impact of my work is in improving security pipelines as well as making them more robust against attackers. There are a wide variety of detection tasks in security that can benefit from machine learning: malware detection, network intrusion detection, spam filtering, authorship attribution, website fingerprinting, tackling social network abuse, etc.

However, while classifiers show great promise in these areas, they often begin to fail after deployment as adversaries adapt to the new model – a reaction that can be sudden and extreme and is difficult to simulate in lab experiments. By understanding these adversarial processes and the performance degradation they induce it will become easier to design new methods that can be relied upon in the real world.

Being a Turing Enrichment student

Being involved with the Turing has been a real joy; to be able to engage and collaborate with such a vibrant and varied community of researchers has been one of the highlights of my PhD so far. Working alongside people from many different disciplines has really expanded my knowledge of different data science techniques and methodologies, which has also enabled me to expand the kinds of research topics that I feel comfortable working on.

“A vibrant and varied community of researchers has been one of the highlights of my PhD so far.”

Ignacio Perez Pozuelo, University of Cambridge



Project title

Digital phenotyping through multimodal, unobtrusive sensing

Supervisors

Nick Wareham, Soren Brage, Cecilia Mascolo

Main outcomes of my research

I have worked on a number of projects grouped into three research streams. First, I've worked on the use of wearable devices to make sleep stage and sleep quality predictions, yielding a number of publications and conference presentations. Second, I've done research related to cardiorespiratory fitness inferences using multimodal wearable sensors. This stream of work comprised a number of sub-projects: (1) a traditional epidemiological study, exploring associations between resting heart rate and cardiorespiratory fitness; (2) a study applying established machine learning tools in large-scale populations and (3) a study leveraging large quantities of multimodal data through self-supervised learning. All are currently under revision for publication. I also initiated a multidisciplinary collaboration with colleagues from the Turing, Oxford and Cambridge to work on an ethics piece related to data governance principles for digital phenotyping technologies.

Impact of my work

- **Research/new discoveries:** Overall, my research explores the development of new methods for human activity recognition, focusing on sleep and fitness inferences, by leveraging the multimodal nature of modern wearable and mobile sensors and machine learning. Through this I've highlighted the strengths and constraints of applying these approaches to large-scale populations.
- **Policy and data governance:** Through a multidisciplinary working group, we developed a set of recommendations for digital phenotyping data governance and submitted it for publication.
- **Open/reproducible science:** One of the major challenges in my research is that most of the tools are developed by commercial enterprises which keep them private with little clarity about the validity and robustness of the inferences made. Most open-source resources are device- and task-specific, which constrains their use. By creating an open-source Python library I have aimed to address these issues.

Being a Turing Enrichment student

The Institute is a great place to learn and grow as a data scientist through the courses and training it offers. Encouraged by this training and with the support from the research engineering group, I ventured into developing my first Python package, HypnosPy, which will be released soon. My time at the Turing has, without a doubt, been the most stimulating and productive research period of my PhD. I was devastated it was cut short due to the pandemic. The Turing fosters interdisciplinary collaboration, through the vast array of talks and seminars and in the nature of the space itself. I established a collaboration with a colleague who works on the intersection of ethics and policy after meeting over coffee (I never tired of the Turing coffee machines). I attended a seminar where I re-connected with an old friend and was able to link up one of his research projects with my home group. My PhD has evolved beyond what I could have imagined when I first started my time at the Turing.

“My PhD has evolved beyond what I could have imagined when I first started my time at the Turing.”

Pedro Pinto da Silva, Newcastle University



Project title

Identification and impact assessment of recurring traffic bottlenecks using ANPR technology

Supervisors

Stephen McGough, Matthew Forshaw, Phil Blythe

Main outcomes of my research

Following my first conversations at the Institute, I realised that the goals of my project needed to be updated and better defined to generate truly useful outcomes for traffic management. This continuous and fruitful exchange of ideas, which can only happen in places like the Turing, pushed me to better specify my goals and clearly articulate the impact of my research. Consequently, I was able to perfect my pitch to new collaborators with whom ongoing partnerships continue to bear fruits, even as I reach the final stages of my degree.

Impact of my work

Before local authorities implement interventions to address traffic congestion, they develop funding proposals that need to be supported by clear evidence of their impact. In the case of recurring traffic bottlenecks, we demonstrate that Automatic Number Plate Recognition (ANPR) technology can be used effectively to identify new cases and quantify their impact. By combining estimates of excess delay and features of road importance, we propose a prioritisation mechanism that allows local authorities to rank identified bottlenecks by their impact and prioritise corrective measures accordingly.

Being a Turing Enrichment student

My experience at the Turing was shaped by the people I met during my stay. Getting to know and work with my fellow students, every single one a promising talent, Enrichment or doctoral, was surely the highlight of my time at the Institute. I had the opportunity to learn about their projects, propose new ideas and receive great constructive feedback in return. As one of their representatives I organised activities to promote student engagement, such as the Turing student poster showcase, and bonding outside the Institute, like after-work Fridays and Christmas dinner.

I'm ecstatic that many work relationships and friendships that started at the Institute will outlast our time there, even in the face of a pandemic. I also had the chance to learn from several leading academics that I met through Turing-led initiatives, such as the mentorship programme, the research engineering clinic and the tools, practices and system seminars to name a few. New collaborations were born from unscheduled conversations at tea point. Overall, my Enrichment experience will always be remembered as one of the high points of my academic career.

“My Enrichment experience will always be remembered as one of the high points of my academic career.”

Mousam Rai, University of Warwick



Project title

Track/shower characterisation using deep learning techniques

Supervisor

John Marshall

Main outcomes of my research

A significant improvement to overall performance of current pattern recognition software with a focus on developing novel track and shower characterisation methods, using machine learning and deep learning approaches, and improving on the current sensitivity measurement of proton decay at the Deep Underground Neutrino Experiment Far Detector.

Impact of my work

A highlight is the implementation of my work in Deep Underground Neutrino Experiment, which is an international collaboration with 184 different institutions across 31 different countries contributing to the experiment, with the UK being the second largest contributor after the US. My work will be crucial in improving proton decay sensitivity measurement and will help test the validity of various supersymmetric and non-supersymmetric grand unified theory.

Being a Turing Enrichment student

It allowed me the opportunity to connect with other students and discuss our projects. This has been very helpful in terms of understanding all the various types of research that The Alan Turing Institute supports. Talking with fellow students about problems in your area and asking for their opinions allowed me to gain a new perspective on problems and helped solve issues. The research staff have been very helpful too, as they are approachable and willing to invest time in you and your project to ensure that you received well-informed and educated advice.

Apart from the academic assistance, the Turing has also helped me understand social issues such as inclusive work environments and workplace etiquette, as they have regular training and workshops that aim to shed light on to these subject matters.

“It allowed me the opportunity to connect with other students and discuss our projects.”

Lorenzo Rimella, University of Bristol



Project title

High-dimensional hidden Markov models

Supervisor

Nick Whiteley

Main outcomes of my research

The main outcomes of my research at the Turing are two papers: (1) “Dynamic Bayesian Neural Networks” and (2) “Inference in Stochastic Epidemic Models via Multinomial Approximations”. The former combines Bayesian neural networks with hidden Markov models, which was one of the aims at the beginning of the placement at the Turing. The latter proposes an approximate Forward-Backward algorithm for compartmental models via multinomial approximation, which raised up during the lockdown. In the near future, I see myself keeping exploring the role of hidden Markov models in high-dimensional statistics and machine learning.

Impact of my work

In general my research has multiple fields of application. (1) was applied to predict the next frame in a waving flag video. There are multiple extension of such work and we are planning to explore the weather forecasting field. Finally, (2) has been applied to epidemic models and in particular to predict the initial stage (the reproduction number) of COVID-19 in Wuhan. This is unfortunately something that closely touched all of us, and we hope that a clearer understanding of the evolution of a disease could help in preventing a wide spread.

Being a Turing Enrichment student

It allowed me to work in a dynamic fast-paced environment, which was particularly stimulating and raised my research to the next level. I had access to multiple computational resources and so I had the chance to speed up my research by running jobs on multiple HPC. Moreover, even during the lockdown the Turing proposed useful alternatives to keep the researchers involved and busy.

For example, they allowed Enrichment students to attend online courses, and in my case I had the chance to get multiple certificates to add to my CV. However, the most valuable thing is the feeling of being part of a huge community of great researchers, which not only made me more confident, but also gave me the possibility to be engaged with interesting events. Last but not least, I talked with other researchers about my work and I got useful feedback, which in some cases also evolved in collaborations.

“The most valuable thing is the feeling of being part of a huge community of great researchers.”

Seth Sharp, University of Exeter



Project title

Polygenic prediction of autoimmune disease using machine learning

Supervisors

Richard Oram, Michael Weedon, Michael Inouye

Main outcomes of my research

Machine learning using genetic markers in the region of the genome responsible for immune system function (human leukocyte antigen) has helped strengthen our ability to predict autoimmune disease onset, such as type 1 diabetes and coeliac disease. Using machine learning we have managed to build prediction models that are equally, if not, more accurate than existing genetic tests.

Human immune system genes are incredibly complex and the data available from studies vast. At the Turing I have worked to adapt machine learning and deep learning algorithms to immune system genetics and determine the best method to model this complexity.

Impact of my work

My models can be applied with inexpensive genetic testing and shows promise as a test that may one day be available in clinic. A follow-up study has shown combining my genetic testing with typical clinical features can identify young people who will go on to develop type 1 diabetes extremely accurately.

The ability to accurately and inexpensively identify those at high genetic risk of autoimmunity has enabled a number of population level screening studies due to begin shortly. By accurately identifying babies that will develop a disease we can recruit into important studies such as prevention drug trials.

Being a Turing Enrichment student

The Enrichment Scheme enabled me to develop a collaboration with a group led by a Turing Fellow which has proved very valuable in my research. This has opened doors both for continued collaborative research and for my research career. Coming from a clinically-focused group it was fantastic to engage with those in more technical research and has enabled a cross-disciplinary collaboration which will continue past my Enrichment placement.

The Institute was also a great place for personal development as the diversity of researchers and projects they support has exposed me to research far outside of my field that I would never of otherwise come across. This has been a great opportunity for learning and has helped inform my future in research after my PhD.

“It was fantastic to engage with those in more technical research.”

Aditi Shenvi, University of Warwick



Project title

Dynamic graphical modelling for public health and policing

Supervisor

Jim Q Smith

Main outcomes of my research

My research is concerned with developing statistical models which have a graphical interface, that can express contextual conditional independences between the components within the graph topology. I have developed methodologies within this framework to enable analysis and inference for dynamic systems with components that evolve both regular and irregularly in time. At my home university, I focussed on applications in public health. As part of a Turing Defence and Security project I had the opportunity to use graphical models for modelling criminal collaborations. Also, I have started a new project (during lockdown!) on mixture models with Turing Fellow, Silvia Liverani.

Impact of my work

The Turing defence project I worked on has the potential to have a real impact in terms of how monitoring and resource allocation are carried out for policing suspected individuals. Additionally, the model classes I have developed during my PhD are applicable to a wide range of real-world systems. I have applied them to analysing treatment interventions in public health problems such as early epilepsy treatment and falls support for the elderly. However, these model classes basically provide a flexible framework and can also be applied to other domains.

Being a Turing Enrichment student

Although I only spent 2.5 months at the Turing office due to the lockdown, my time there and the support I have received from the Turing community during lockdown have made my Enrichment placement very memorable. There was a wide range of academic and non-academic talks, seminars, training and resources available. I had the chance to interact with PhD students and other researchers coming from various disciplines with a wide range of educational backgrounds who were all, in some way, working on data science and AI. This naturally led to conversations and discussions which were intellectually stimulating and helped me broaden my perspective.

In particular, it made me appreciate the value of finding a decent solution first and then worrying about perfecting it – which is not always how more theoretical science departments function. I also really appreciate the Turing’s focus on making research reproducible! One of the main highlights for me was the research software engineering course and the general programming support. Lastly, I also appreciate how the Turing adapted to the lockdown situation through virtual mentorship meetings, and provided us with a training budget and extended access to the IT resources!

“Conversations and discussions which were intellectually stimulating and helped me broaden my perspective.”

Vasilis Stavrinides, UCL



Project title

The microenvironmental contexture of the MRI-characterised prostate

Supervisors

Mark Emberton, Caroline M Moore, Hayley Whitaker

Main outcomes of my research

MRI is currently recommended before biopsy for the detection of cancer in the prostate. My research has so far demonstrated that prostate MRI lesions caused by cancer are radiologically and clinically different from those caused by benign causes, such as inflammation. These differences can be exploited in order to help some men with very low prostate cancer risk to avoid biopsy altogether. Furthermore, we have indicated a roadmap for the monitoring of low-risk prostate tumours through regular MRI rather than conventional biopsy. This approach improves patient compliance and reduces the number of biopsies needed to detect progression to aggressive disease.

Impact of my work

If taken further, the work I completed can help some men undergoing tests for suspected cancer to avoid unnecessary biopsies and their complications. Also, it could improve the active surveillance of small prostate tumours that do not necessarily need immediate treatment. Both these small but important contributions were published recently in prestigious clinical journals and presented in major international conferences.

Being a Turing Enrichment student

The greatest benefit of my Enrichment Scheme was the formation of new collaborations with experts from other universities. This allowed me to get a closer look into the research currently performed in statistics, mathematics and computer science and think about how ideas from these disciplines can be applied to solve clinical problems. I also made several new friends that I hope to keep for life.

Together with another Enrichment student from the MRC Biostatistics Unit in Cambridge, I completed a decision curve analysis of models that could allow men with suspected prostate cancer avoid biopsy through the use of MRI. This analysis formed a part of a paper now published in a prestigious clinical journal.

I am currently working on the molecular dynamics of prostate cancer with other Turing students, while I am actively engaged in a community of clinicians with quantitative research interests. The experience at the Turing also increased my confidence in my own statistical knowledge and ability.

“I also made several new friends that I hope to keep for life.”

Obi Thompson Sargoni, UCL



Project title

Exploring the impact of autonomous vehicles on pedestrian experience in urban spaces through agent-based simulation

Supervisors

Ed Manley, Jack Stilgoe

Main outcomes of my research

Creating a novel framework for modelling pedestrian road crossing behaviour that combines quantitative models of decision-making from psychology with models of pedestrian movement in geographic spaces.

Impact of my work

To help broaden the space of possible autonomous impacts by incorporating effects of vehicle traffic on pedestrian route choice and journey times and by considering the effect of changes in driver and pedestrian behaviour. I hope this will help transport planners be more sensitive to the ways the transport system might change in response to the arrival of autonomous vehicles.

Being a Turing Enrichment student

It allowed me to learn more about reinforcement learning, neural networks, and helped to improve my coding skills. These have been very useful skills to develop and I hope to apply them to my research.

“It allowed me to learn more about reinforcement learning.”

Anthony Tuckwell, University of Warwick



Project title

NLP and machine learning for cultural economics

Supervisors

Daniel Sgroi, Elliott Ash, Thomas Hills

Main outcomes of my research

While at the Turing I've been working on a project that applies NLP to investigate the links between politics and religion in the US using a database of religious sermons and congressional speeches. One of the things we've been looking at is whether the use of religious language might be driven by motivations to get re-elected. Indeed, we've found that senators increase their use of religious language at the end of their election cycle, and that this relationship only holds after the senate chamber was televised.

I've also been working on a project which applies machine learning to predict the happiness of chart music overtime, which can then be compared to survey-based measures of happiness to test whether popular music has the potential to encode collective emotion in a society.

Impact of my work

NLP and machine learning are very much at the forefront of economic research, with an increasing number of economic researchers making use of these methods. It's been exciting to have the opportunity to use them for my own research, which we'll be looking to publish in general-interest and science journals.

Being a Turing Enrichment student

Being at the Turing has helped my research enormously. The project about happiness and music would not have happened unless I had been to the Turing: I met two fantastic collaborators (Emmanouil Benetos and Alessandro Ragano) who do work in the field of music informatics who have been invaluable in helping to get the project up and running. For the project about politics and religion in the US, I received extremely helpful feedback from a number of fellow students and Turing researchers and I learnt a lot about NLP in the process! I also participated in great workshops on deep learning and CNNs.

Most of all though, the open and friendly environment was very conducive to interacting with and learning from others; I met a whole range of extremely kind and interesting people, and hope to keep in touch with all of them.

“The open and friendly environment was very conducive to interacting with and learning from others.”

Sam Van Stroud, UCL



Project title

Monitoring ecosystem resilience

Supervisors

Nick Barlow, Camila Rangel Smith

Main outcomes of my research

To help Nick and Camila in the design and creation of a Python package, pyveg. pyveg uses remote sensing data openly available on Google Earth Engine to monitor the resilience of dryland ecosystems around the world. The package has been written with research software engineering best practices in mind, in collaboration with a team of ecologists at the University of Exeter. pyveg contains functionality to interface with, and download data from, Google Earth Engine, and also to run a suite of time series analysis on the data. Results from different locations around the world can then be aggregated to form a macroscopic picture of the health of dryland ecosystems across the world.

Impact of my work

Monitoring the health environment is crucial if we are to drive policy changes and make targeted actions to reduce our impact on the global ecosystem. Dryland ecosystems are, in particular, important to understand, as they are generally thought to be near a potential “tipping point”. Tipping points are bifurcation points in the state space of the ecosystem. In the case of semi-arid dryland ecosystems, reaching a tipping point could mean irreversible and catastrophic collapse of the ecosystem to a permanently arid state. We want to be able to predict ahead of time when these collapses will occur, in order to better prepare for them.

Being a Turing Enrichment student

Being a Turing Enrichment student allowed me to get real hands-on experience of being both a research software engineer, and also a data scientist. I gained an invaluable insight into the process of designing and building a medium-scale piece of research engineering software. Through the Turing, I was also exposed to a variety of talks, events, and seminars on a huge variety of topics. From open democracy and citizen participation, to bitcoin and state of the art techniques in graph learning. I met researchers from a similarly wide variety of disciplines tackling exciting problems, from applying old-school logic programming techniques to discovering novel therapeutics for cancer patients, to adversarial information security.

“I met researchers from a similarly wide variety of disciplines tackling exciting problems.”

Jin Wang, University of Exeter



Project title

Deep learning based system optimisation and maintenance for cloud and multi-access edge computing

Supervisors

Geyong Min, Jia Hu

Main outcomes of my research

Applying cutting-edge deep reinforcement learning (DRL) technologies for optimisation and intelligent maintenance of computing and communication systems has attracted huge recent research attentions in both academia and industry. My PhD research lies in using DRL to improve the quality of services, energy efficiency, and reliability of cloud computing and multi-access edge computing systems. So far, I have designed and implemented novel DRL-based methods that can significantly reduce the running latency and energy consumption of an MEC system. These researches have been published/submitted to journals and conferences such as IEEE communication magazines, IEEE Transactions on Parallel and Distributed Systems, IEEE Globecom.

Impact of my work

The optimisation and maintenance of large-scale computing systems are intractable. Most current solutions are based on heuristic/ approximate algorithms. However, those methods rely heavily on expert knowledge, once the scenario changes, considerable human efforts, and expertise are required to tune the heuristics or mathematical models to adapt those methods to new scenarios, which is time-consuming and sometimes unrealistic. My research aims at adapting learning-based methods for system optimisations (improve the resource utilisation and energy efficiency), which will alleviate the requirements of expert knowledge and let the system learn optimal optimisation strategies based on the data.

Being a Turing Enrichment student

The Enrichment placement provided a perfect opportunity for me to build collaborations with researchers and other students from a broad range of disciplines, which benefited me a lot. The multiplicity of research topics at the Turing really broadened my perspective and inspired me to contribute to exciting research in different fields.

Through talking and discussion to those brilliant minds, I got lots of good ideas for solving problems in my research. Specifically, I feel really lucky to have met a PhD student who has statistic backgrounds and works on hidden Markov models in a seminar. We discussed a lot about our research, and I found his research can help me solve a problem that plagued me for a long time. Overall, I really enjoyed life at the Institute, I have met lots of good friends and learned things outside of my special areas.

“The multiplicity of research topics at the Turing really broadened my perspective.”

Tiejun Wei, Queen Mary University of London



Project title

Exploring the impact of autonomous vehicles on pedestrian experience in urban spaces through agent-based simulation

Supervisor

Christopher Duffy

Main outcomes of my research

I gained enormous amount of knowledge on deep learning and other statistical techniques which helped me a lot in my day-to-day research. We successfully applied graph convolution network (GCN) to our structural analysis on light-harvesting complex II (LHCII) system. This will greatly help us understand the causality of pigment-pigment coupling and their structural change.

Impact of my work

This methodology is quite universal and can possibly be applied to drug discovery field. As we used scalar label to structural data, this can be easily mutated to biological affinity label and give pharmaceutical science a new way to discover drug leads.

Being a Turing Enrichment student

One of the greatest things about being at Turing is that you meet a lot of people. There are experts from different universities and prestigious companies. They give free talks and seminars frequently and this really broadens your insight into data science and inspires you to apply the most-recent approach to your own project.

It also allowed me to do research in an open-office space among other PhDs, most of them being Enrichment students as well. This allowed me to see what others do in their area and learn how they approach a problem.

“One of the greatest things about being at Turing is that you meet a lot of people.”

Lizhi Zhang, University of Bath



Project title

Realistic constraints for data-driven network analysis

Supervisor

Tiago P Peixoto

Main outcomes of my research

During my Enrichment Scheme, I worked on developing statistical models that describe the generative process of network data. The outcome is an algorithm can be used to reveal community structure in large-scale networks. We managed to come up with a model being focused on assortative structure (vertices are locally densely connected) of networks. We will follow this direction to develop models with other realistic features. That includes but not limited to evolution of networks, multiplayer structure, etc.

Impact of my work

We developed a variant of the general stochastic block models focusing on assortative structure. Under the Bayesian modelling framework, our model provides a tool to test the existence of assortative structure, by which many non-statistical approaches fail to justify. Our method has the advantage of coping with underfitting and overfitting problem in one go.

The result of our work should be beneficial to anyone interested in analysing the structure of networks. A typical use of community detection algorithm is to identify functional components in biological networks.

Being a Turing Enrichment student

The Enrichment Scheme at the Turing has been great studying experience. The best thing about the Institute is that it provides a research environment with no disciplinary boundary. Throughout my time at the Turing, I had been surrounded by smart minds coming from diverse backgrounds.

Daily interaction with others in the Institute allows me to learn interesting problems in different fields. In addition, the experience also helped me hone the soft skills needed for a research career. The Institute organised a series of excellent events bringing students and fellows together. There are many opportunities to form new collaborations and that is the most unique thing about the Institute.

“I had been surrounded by smart minds coming from diverse backgrounds.”

Konstantinos Zinelis, Imperial College London



Project title

Spray formation in non-Newtonian fluids

Supervisor

Omar Matar

Main outcomes of my research

The application of AI and particularly machine learning techniques are currently leading to significant advancements in the field of computational fluid dynamics. Hence, in the current work we investigate the integration of data-driven methods to multi-physics models for predicting the spray formation in complex fluids. Classification approaches are implemented to predict the different flow regimes of a viscoelastic jet, accounting for how the interplay among the acting forces, such as capillarity and elasticity, determines the flow behaviour and the formation of droplets. Last but not least, machine learning algorithms are being developed for optimising the population balance equation to capture properly the effect of elasticity during breakage and coalescence processes.

Impact of my work

The data-driven modelling of atomisation processes can really expand the current capabilities of computational fluid dynamics models, by reducing considerably their significantly high computational cost, as well as exploiting the great number of both already available experimental and simulation data. In this project, we present how these data would contribute to answer still open questions about how elasticity alerts the dynamics and mainly the droplet size distributions in sprays of real fluids. Data-driven modelling can also enhance the predictability of existing fluid mechanics models, optimising at the same time the experimental set-up needed to investigate further the outcome of these highly complex phenomena.

Being a Turing Enrichment student

Being an Enrichment student at The Alan Turing Institute primarily introduced me to the fabulous and impactful world of big data, and I became aware of how a field, which is strongly different from my own research area, would be the right tool to tackle the challenges of my PhD project.

The great availability of various seminars, workshops and students' talks, from the fundamentals of data science to specific AI applications, ensures a well-planned route for acquiring all the skills and knowledge which are necessary today in the modern academic and professional environment.

What inspired me most was the multi-disciplinary and collaborative culture of the Institute, where I was privileged to interact daily with powerfully bright minds from various research fields, in being exposed to their stimulating, innovative and creative ideas. Last but not least, the Enrichment and doctoral students cohort was really supportive, communicative and friendly, making my Turing experience greatly enjoyable and sincerely unforgettable.

“I was privileged to interact daily with powerfully bright minds from various research fields.”



turing.ac.uk
@turinginst