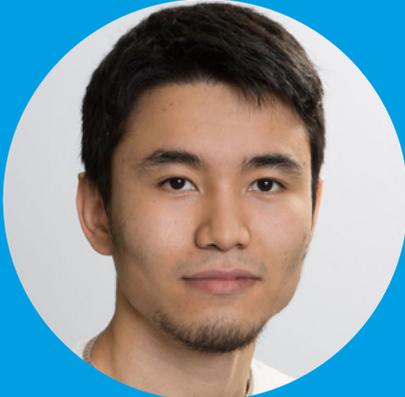


# The Alan Turing Institute



---

Doctoral Students  
2016 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 inaugural 2016 doctoral cohort, but also of their contribution to the success of the Turing as a whole.

The students have been inspiring in the research they have undertaken, and prolific in their output. They have produced over 50 papers; attended more than 50 conferences, workshops, and symposia; and developed work with industrial partners including Intel, Apple and NASA. 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 businesses to apply their skills in tackling real-world problems.

The Alan Turing Institute is built on a strong legacy of pioneering work in theoretical and applied mathematics, engineering and computing, blended with increasingly diverse contributions from academic disciplines, that help make data science a truly multi-disciplinary endeavour. At a time when we are expanding our role as a driving force in data science and AI communities, we are pleased to be able to reflect on how the 2016 doctoral cohort has contributed to our work.

It is testimony to the excellence of the Institute, including this cohort, that we have grown so much in the years they have been studying with us. Their contributions have often been highlighted as key markers of our collective success.

These are successes that have been achieved in a pioneering spirit, with the students working alongside the Turing's business teams to create the collaborative environment, training provision, and research culture suitable for a national institute, which subsequent cohorts have built upon.

We regret that the end of their time with us was interrupted by the global pandemic in 2020. However, we very much hope that as the Turing opens again as a vibrant place to meet and share ideas, we can welcome the 2016 cohort back to celebrate passed vivas, learn about new projects, and develop new opportunities with the leaders of the future.

**Ben Murton**  
Head of Skills

As an Institute we are confident that our paths will cross in the future, as we expect this new generation of leaders to continue to tackle the cutting-edge research challenges and make ground-breaking contributions to real-world issues.

We are proud of your achievements and wish you the very best in the years to come.

**Adrian Smith**  
Institute Director and Chief Executive



## Alex Bird, University of Edinburgh



### Thesis title

Multi-task Dynamical Systems

### Supervisor

Chris K I Williams

### Main outcomes of my research

My work explored how time series models can provide customised fit/predictions at an individual level. The eventual approach we proposed is a general purpose method, which makes use of unsupervised learning to determine a manifold of response types which can be used to tailor the model. By using unsupervised techniques, we can provide customisation without the requirement of highly informative exogenous information. The approach was investigated in the context of linear dynamical systems and recurrent neural network models.

### Impact of my work

The problem of handling individual entities (persons, organisations, products, equities, etc.) in time series data is ubiquitous, hence there is wide applicability of the work. But perhaps the most interesting application during my PhD was attempting to predict patient response to an anaesthetic agent at an individual level. Together with some natural extensions, this work could form the basis of automating steady-state drug infusion in anaesthesia.

### Being a Turing Doctoral Student

The Turing exposed me to a far wider range of ideas than would have been possible in a traditional PhD programme. It was great to be able to socialise and learn from those with backgrounds in maths, stats, computer science, engineering, social sciences, etc. This enabled me to look at my work from a great many perspectives and satisfy my curiosity about other disciplines. Well, at least to an extent!

### Plans following my PhD completion

I'm currently working with some of my Turing friends/colleagues on a start-up venture. Our aim is to facilitate improved access and collaboration within the open source ecosystem and enable developers to access financial benefits from their contributions.

I have very much enjoyed my time in academia, and leaving for industry reminds me a little of leaving home. I would be delighted to return in the future.

### Research areas

Applications  
(Machine learning)

Time series

Unsupervised learning

### Papers & publications

Bird, A. and Williams, C.K.I. (2020). Customizing Sequence Generation with Multi-Task Dynamical Systems. arXiv preprint arXiv:1607.06450.

### Workshops & conferences

Bird, A., Williams, C. K. I. and Hawthorne, C. 2019. Multi-Task Time Series Analysis applied to Drug Response Modelling. In: *22nd International Conference on Artificial Intelligence and Statistics, 16-18 April 2019, Okinawa, Japan.*

**Note:** Shouto Yonekura (UCL), Daniel Wilson-Nunn (University of Warwick), Vidhi Lalchand (University of Cambridge), Odysseas Sclavounis (University of Oxford) were also Doctoral Students.

## David Butler, University of Edinburgh



### Research areas

Cryptography  
(Privacy & trust)      Verification

### Papers & publications

Butler, D., *et al.* Formalising Sigma-Protocols and Commitment Schemes using CryptHOL. IACR Cryptol. ePrint Arch. Upcoming in *Journal of Automated Reasoning*.

Butler, D. and Aspinall, D. (2019). Multi-Party Computation. *Archive of Formal Proofs*. [www.isa-afp.org/entries/Multi\\_Party\\_Computation.html](http://www.isa-afp.org/entries/Multi_Party_Computation.html).

Butler, D. and Lochbihler, A. (2019). Sigma-Protocols and Commitment Schemes. *Archive of Formal Proofs*. [www.isa-afp.org/entries/Sigma\\_Commit\\_Crypto.html](http://www.isa-afp.org/entries/Sigma_Commit_Crypto.html).

### Selected workshops & conferences

Butler, D., Aspinall, D. and Gascón, A. 2020. Formalising  $\Sigma$ -Protocols and Commitment Schemes using CryptHOL. In: *3rd ZKProof Workshop, 20 April – 21 May 2020, online*.

Butler, D., Aspinall, D. and Gascón, A. 2020. Formalising Oblivious Transfer in the Semi-Honest and Malicious Model in CryptHOL. In: *Certified Programs and Proofs (CPP), 20-21 January 2020, Louisiana, US*.

Butler, D., Aspinall, D. and Gascón, A. 2019. Formalising Oblivious Transfer in the Semi-Honest and Malicious Model in CryptHOL. In: *Foundations of Computer Security (FCS), 24 June 2019, New Jersey, US*.

### Thesis title

Formalising Cryptography using CryptHOL

### Supervisors

David Aspinall, Adrià Gascón

### Main outcomes of my research

In my work I have constructed frameworks for reasoning formally about a number of fundamental cryptographic constructions – namely multi-party computation, commitment schemes and Sigma protocols. These frameworks allow for machine-checked security proofs of these constructions, increasing confidence of the correctness of the proofs.

### Impact of my work

I provided others with a rigorous platform upon which to formally consider fundamental cryptographic primitives. During our work we were able to point out an inconsistency in the definition of Sigma protocols which results in incorrect proofs for some more complex constructions. In particular we were able to point to the correct definition.

### Being a Turing Doctoral Student

It gave me access to an interesting network of people I otherwise would not have been able to meet. This is particularly useful when considering options post-PhD.

## Corinne Cath-Speth, University of Oxford



### Research areas

Computing networks      Ethics

Data science of govt. & politics

### Selected papers & publications

Cath, C. The Technology We Choose to Create: Human Rights Advocacy in the Internet Engineering Task Force. Under peer review at *Telecommunications Policy*.

Rincón, C., Keyes, O. and Cath, C. (2021). Speaking from Experience: Trans/Non-Binary Requirements for Voice-Activated AI. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW1).

Cath, C. (2018). Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376 (2133).

### Selected workshops & conferences

Cath, C., *et al.* 2020. Leap of FATE: Human Rights as a Complementary Framework for AI Policy and Practice. In: *2020 Conference on Fairness, Accountability, and Transparency, 27-30 January 2020, Barcelona, Spain*. New York: Association for Computing Machinery.

### Books

Knodel, M., Uhlig, U., ten Oever, N. and Cath, C. (2021). *How the Internet Really Works*. No Starch Press.

### Thesis title

Changing Minds and Machines: Human Rights Advocacy in Internet Governance

### Supervisors

Victoria Nash, Gina Neff

### Main outcomes of my research

Outcomes included 12 peer-reviewed publications, four book chapters, one book (co-published), 10+ media articles, three workshops and conferences organised (at the Turing, Harvard and Oxford) and an 80,000+ word thesis on the politics of internet infrastructure and civil society participation therein.

### Impact of my work

Furthering understanding of how internet governance cultures influence the politics of technology; strengthening the participation of civil society in technical spaces; and providing policy processes with on-the-ground insights from ethnographic research.

### Being a Turing Doctoral Student

It allowed me the time and space to explore what it means to be an engaged academic and find novel avenues for translating knowledge back to concrete processes of advocacy and policy-making.

### Plans following my PhD completion

I hope to continue to do ethnographic research and work in the policy space of internet infrastructure.

## Edward Chuah, University of Warwick



### Thesis title

Features Correlation-based Workflows for HPC Systems Diagnosis

### Supervisor

Arshad Jhumka

### Main outcomes of my research

Modern-day data centres and high performance computing (HPC) systems are comprised of complex combinations of networks, processors, storage systems and operating systems. Recent research has demonstrated the value and significance of combining system failure logs with resource utilisation data for system failure diagnosis (and error detection). However, the massive amount of data that large HPC systems generate presents a significant challenge in processing the data effectively for system failure diagnosis. My PhD research addresses the challenge by developing a new framework for failure diagnosis. The framework uses resource usage data and system logs in its analyses. I evaluated multiple feature extraction methods and correlation algorithms and implemented two failure diagnostics workflows. The first diagnostics workflow identified error cases that occurred frequently. The second diagnostics workflow identified error cases that are rare.

### Impact of my work

My PhD research has produced a set of recommendations on what one should look for in the resource use data and system logs. For example, in the resource use data, correlations of network data frame errors and network packets transmitted can be used to detect network problems. In the system logs, correlations of DNS lookup failure and FTP software status messages can be used to identify a configuration error in the DNS software.

### Being a Turing Doctoral Student

In 2016, I received an offer to complete my PhD studies at The Alan Turing Institute. I knew this was an opportunity I could not miss. I am grateful for the strong support the Institute has provided, not only for a very prestigious and competitive Turing Doctoral studentship but for access to new collaborators in industry and academia.

### Plans following my PhD completion

I hope to use the knowledge I gained during my PhD studies to develop a workflow for failure prediction. The field I am interested in is fault tolerance with emphasis on systems diagnosis and failure prediction using statistics and data mining.

### Papers & publications

Chuah, E., *et al.* (2019). Towards Comprehensive Dependability-Driven Resource Use and Message Log-Analysis for HPC Systems Diagnosis. *Journal of Parallel and Distributed Computing (JPDC)*, 132C.

### Workshops & conferences

Chuah, E., *et al.* 2019. Using Resource Use Data and System Logs for HPC System Error Propagation and Recovery Diagnosis. In: *17th IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA)*, 16-18 December 2019, Xiamen, China.

Chuah, E., *et al.* 2017. Enabling Dependability-Driven Resource Use and Message Log-Analysis for Cluster System Diagnosis. In: *24th IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC)*, 18-21 December 2017, Jaipur, India.

Chuah, E., *et al.* 2016. Using Message Logs and Resource Use Data for Cluster Failure Diagnosis. In: *23rd IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC)*, 19-22 December 2016, Hyderabad, India.

## Andreas Grammenos, University of Cambridge



### Thesis title

Federated Dimensionality Reduction

### Supervisors

Cecilia Mascolo, Jon Crowcroft

### Main outcomes of my research

The main outcome of my research was the introduction of a federated scheme to do dimensionality reduction using principal component analysis (PCA) while being tolerant to missing values, as well as providing differential privacy.

### Impact of my work

PCA is a ubiquitous and fundamental tool in data analysis and machine learning pipelines and also has important societal applications, such as poverty measurement.

Computing PCA on large-scale data is not only challenging from the computational point of view, but also from the public policy point of view. Indeed, new regulations around data ownership and privacy, like GDPR, have imposed restrictions in data collection and storage.

My contributions enable large-scale decentralised computation of PCA in settings where each compute node – be it large (servers), thin (mobile phones), or super-thin (cryptocurrency blocks) – contributes in an independent and asynchronous way to the training of a global model, while ensuring the ownership and privacy of the data.

### Being a Turing Doctoral Student

Being a Turing student was one of the most rewarding experiences in my life. The environment is fantastic and is highly recommended for anyone that wants to do cutting-edge research. However, the most important bit is that it allowed me to foster collaborations with disciplines that would be impossible to facilitate had I been just in Cambridge. These vibrant and interdisciplinary collaborations not only expanded my research horizons but also increased both the breadth and scope of my research.

Finally, for a PhD program, the overall student support was extensive both in terms of resources available but also in terms of actual support from the student services, which was a pleasant surprise.

### Plans following my PhD completion

Next steps are unclear but will definitely include research activities, be it in a corporate or academic setting; ideally, both!

### Selected research areas

Complexity (Algorithms)	Data structures
Distributed (Algorithms)	Numerical (Algorithms)

### Papers & publications

Grammenos, A., *et al.* (2020). Federated Principal Component Analysis. Pre-print <https://arxiv.org/abs/1907.08059>.

Grammenos, A., *et al.* (2019). MOSES: A Streaming Algorithm for Linear Dimensionality Reduction. *Transactions of Pattern Analysis & Machine Intelligence*, 42 (11).

Grammenos, A., *et al.* (2018). Efficient, privacy aware federated model sharing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2 (1).

### Workshops & conferences

Grammenos, A., *et al.* 2020. Dissecting the Workload of a Major Adult Video Portal, Tyson. In: *Passive and Active Measurement (PAM)*, 30-31 March 2020, online.

Grammenos, A., *et al.* 2019. Efficient, privacy aware federated model sharing. In: *UK Mobile, Wearable and Ubiquitous Systems Research Symposium (MobiUK)*, 1-2 July 2019, Oxford, UK.

Grammenos, A., *et al.* 2018. Online pattern discovery in distributed, high-dimensional, streaming data under the YOLO principle. In: *EuroSys Doctoral Workshop*, April 23 2018, Porto, Portugal.

## Javad Hosseini, University of Edinburgh



### Thesis title

Unsupervised Learning of Relational Entailment Graphs from Text

### Supervisor

Mark Steedman

### Main outcomes of my research

My research is in the area of natural language processing and machine learning. I have developed methods to automatically process large amounts of news data and extract entailment graphs between events in the text. For example, if we know that someone has visited a location, we can predict that that person has also arrived in that location and will leave that location.

The basic entailment graphs are sparse and noisy. I proposed applying global soft constraints on them which improved the graphs significantly. I showed that the tasks of entailment graph induction and knowledge graph completion are complementary. I introduced a new problem of contextual link prediction and proposed a model for it. I showed that by grounding events in context and using large pre-trained language models (such as BERT), we can get further improvements on the entailment graphs.

### Impact of my work

Personal assistants such as Apple, Siri or Google Assistant are now very good at handling queries, such as checking the weather or knowing the birthplace of celebrities. They try to find the answers to many of our questions by searching the web. However, in many cases they struggle at providing a definite answer because there are many ways of asking the same question, while the answers to only a few of them are explicitly stated over the web. We are usually left with snippets that might or might not contain the answer to our questions.

Google Assistant does not currently (Sep 2020) correctly answer the question "Has Zidane played in a Champions League Final?"; however, if we change the question to "Has Zidane scored in a Champions League Final?", it can find a good answer. I have developed models to learn entailments between natural language relations, e.g. knowing that when an "athlete scores in a game" entails that the "athlete has played in the game". The entailments can be used to improve question and answering systems.

### Being a Turing Doctoral Student

I was lucky to be among the first Turing doctoral cohort. The Institute helped me not only in supporting me in my own research but also in providing me with a great community and in making me familiar with wider research. I did my experiments on Microsoft Azure machines and I am grateful that the Institute generously allowed me to use the Azure credits necessary for my research. Without Azure, I would not have been able to conduct my research well and develop the large-scale models that I proposed as part of my work. The Turing has a very friendly environment with many social events. We had access to training from the first days which was great. I am grateful that the Turing supported us attending conferences so that we could present our findings and become familiar with ongoing research.

### Plans following my PhD completion

My plan is to work as a researcher in an industrial position in the fields of natural language processing and machine learning.

### Research areas

Natural language processing

Deep learning

### Papers & publications

Hosseini, M. J., *et al.* Learning Typed Entailment Graphs with Global Soft Constraints. *Transactions of the Association for Computational Linguistics*, 6.

Hosseini, M. J., *et al.* Open-Domain Contextual Link Prediction to Learn Entailment Graphs. Under review.

### Workshops & conferences

Hosseini, M. J., *et al.* 2019. Duality of Link Prediction and Entailment Graph Induction [poster]. In: *57th Annual Meeting of the Association for Computational Linguistics*, 28 July - 2 August 2019, Florence, Italy.

Hosseini, M. J., *et al.* 2019. Learning Typed Entailment Graphs with Global Soft Constraints. In: *57th Annual Meeting of the Association for Computational Linguistics*, 28 July - 2 August 2019, Florence, Italy.

## Katherine Helen Oliver, University of Cambridge



### Thesis title

Obstacles to Wearable Computing

### Supervisors

Jon Crowcroft, Richard Mortier

### Main outcomes of my research

I used a novel method of participatory design fiction at the inspiration phase of a research through design of everyday connected wearables. What I learned is that the participants wanted something from wearable technology that can't be bought on the mass market: self-expression, creativity and face-to-face connections with others.

### Impact of my work

Inspired by the creativity of the participants, I started a prototype before the pandemic, incorporating more functionality and integration of personal data. I have also been exploring collaborations with experts in privacy-preserved federated machine learning to develop a more context-aware device while improving on the privacy of the system. An OEM consultancy in the States has shown a lot of interest, and I look forward to seeing how they apply this maker-based method in an industrial setting.

### Being a Turing Doctoral Student

At the Turing you can meet the leading experts in data science all under one roof. You are certain to connect with like-minded people who see what you're getting at and, at the same time, have perspectives you would never have thought of.

The generous financial support enabled me to focus on my work without stressing over my living conditions, so I feel much more fortunate than the stereotypical PhD student.

### Plans following my PhD completion

I am currently researching Russian underground forums at the Cybercrime Centre in the Security Group at Cambridge. I am also developing my research into the user-centred privacy and security of wearable computing.

### Research areas

Human computer interface

### Papers & publications

Oliver, H. (2019). Design fiction for real-world connected wearables. *5th ACM Workshop on Wearable Systems and Applications (WearSys'19)*, 21 June 2019, Seoul, Republic of Korea. New York: Association for Computing Machinery. DOI: 10.1145/3325424.3329664.

Oliver, H. (2019). Participatory design fiction for innovation in everyday wearable IoT systems. *4th ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI 2019)*, 15-18 April 2019, Montreal, Canada. New York: Association for Computing Machinery. DOI: 10.1145/3302505.3312597.

### Selected workshops & conferences

Oliver, H., and Mortier, R. 2021. How not to be seen - privacy and security considerations in the design of everyday wearable technology. Extended abstract. Conditionally accepted for *Competitive Advantage in the Digital Economy 2021*, 2-3 June 2021, Venice, Italy.

Oliver, H. 2020. Participatory Design Fiction for Wearables II: The Sequel. In: *UK Systems Research Challenges Virtual Workshop*, 16 December 2020, online.

Oliver, H. 2020. Design Fiction as a Method for Innovation in Everyday Wearable IoT Systems. In: *Wearable and Ubiquitous Systems Research Symposium*, 2 July 2019, Oxford, UK.

## Adilet Otemissov, University of Oxford



### Thesis title

Dimensionality reduction techniques for global optimisation

### Supervisor

Coralia Cartis

### Main outcomes of my research

While optimisation techniques are widely used in machine learning, can machine learning techniques be used in optimisation? Inspired by this question, my research investigates the effectiveness of a particular machine learning technique called 'dimensionality reduction' applied to a difficult optimisation problem. The work is mostly theoretical with novel applications of tools from machine learning, random matrix theory and conic integral geometry in optimisation. Through understanding geometry and merging random search techniques with dimensionality reduction, my work develops algorithmic frameworks that are more scalable than existing generic global optimisation methods.

### Impact of my work

In my work, I optimise a particular class of functions called 'functions with low effective dimensionality'; these appear in applications mainly related to parameter studies such as neural networks, climate modelling, complex engineering and physical simulations. Along with the practical applications, my work presents new ways of analysing random methods in global optimisation and opens up new avenues of research in this direction.

### Being a Turing Doctoral Student

I found the Turing to be a vibrant community passionate about data science and artificial intelligence. Seminars and talks at the Turing made me appreciate benefits as well as potential harm of the fast-developing field of data science in our everyday lives.

While I was based at the Turing during the first year of my PhD, the Institute provided me with training in a wide range of data science subjects including machine learning theory, optimisation, programming and more. The Turing created an exceptional platform/environment for networking of people from different areas of research from fields as remote as topology and ethics.

I thank the Turing for providing me with a great opportunity to be a part of its endeavour and allowing me to pursue a PhD within its walls. I established (hopefully) long-lasting connections and found new friends who cleared my eyes to look at life anew.

### Plans following my PhD completion

I am planning to pursue a career in academia. I am considering doing a post doc after my PhD either (probably) in Europe or back home in Kazakhstan. The fields I am interested doing research in include (global) optimisation, specifically the use of random methods and functional transformation techniques, and some topics in random matrix theory and matrix completion.

### Research areas

Nonlinear programming

Probability

Stochastic optimisation

### Papers & publications

Cartis, C. and Otemissov, A. (2020). A dimensionality reduction technique for unconstrained global optimization of functions with low effective dimensionality. Manuscript submitted for publication.

### Workshops & conferences

17th EUROPT Workshop on Advances in Continuous Optimization, 28-29 June 2019, Glasgow, UK.

28th Biennial Numerical Analysis Conference, 25-28 June 2019, Glasgow, UK.

23rd International Symposium on Mathematical Programming, 1-6 July 2018, Bordeaux, France.

6th IMA Conference on Numerical LA and Optimization, 27-28 June 2018, Birmingham, UK.

SIAM Conference on Optimization, 22-25 May 2017, Vancouver, Canada.

## Nicole Peinelt, University of Warwick



### Thesis title

Detecting Semantic Similarity: Biases, Evaluation and Models

### Supervisor

Maria Liakata

### Main outcomes of my research

My doctoral research has focused on automatically predicting semantic similarity between two short English texts. Based on an analysis of existing semantic similarity detection datasets, I found that the majority contained significant levels of lexical overlap bias between text pairs. This bias can be exploited by simple baselines to achieve seemingly high performance on traditional evaluation metrics, without any deep language understanding. Therefore, I proposed more challenging metrics which take lexical overlap into account during evaluation. On this basis, I developed new deep learning models which are less reliant on direct text overlap and instead harness other sources of information, such as topic models and linguistically-informed word representations.

### Impact of my work

The impact of my work is twofold. The first part of my work contributes to a growing body of research within natural language processing (NLP) which seeks to highlight existing biases and find ways to address them. The wider goal of this research is to develop more resilient NLP models which can be safely employed in real-world settings. The second part of my work contributes to recent deep learning methods, which can be applied to a wide range of NLP problems without the need for time-consuming manual feature-engineering. For example, the developed semantic similarity detection models can be used to identify paraphrases, as well as to select the correct answer to a question given a collection of possible answers.

### Being a Turing Doctoral Student

Being at the Turing allowed me to expand my network and horizon by meeting bright people from many different fields. The Institute offered a vibrant atmosphere with frequent insightful talks and inspiring events. As a Turing student, I had the opportunity to actively contribute to this environment, for example by organising reading groups for the NLP interest group. These weekly meetings brought together NLP experts from across the Institute and partner universities to discuss recent publications and developments in our field. As organiser, I had the unique opportunity to directly invite and host speakers. Moreover, my research benefitted immensely from the access to high-end computing platforms which enabled me to conduct large-scale experiments. The Institute's support further allowed me to attend international summer schools and present my research at top-tier NLP conferences (e.g. ACL).

### Research areas

Deep learning

Supervised learning

Natural language processing (NLP)

Neural networks

### Papers & publications

Peinelt, N., Rei, M. and Liakata, M. (2020). GiBERT – Introducing Linguistic Knowledge into BERT through a Lightweight Gated Injection Method. arXiv preprint: 2010.12532.

Peinelt, N., Nguyen, D. and Liakata, M. (2020). tBERT – Topic Models and BERT Joining Forces for Semantic Similarity Detection. *Proceedings of the 58th Conference of the Association for Computational Linguistics*.

Peinelt, N., Nguyen, D. and Liakata, M. (2020). Better Early than Late: Fusing Topics with Word Embeddings for Neural Question Paraphrase Identification. arXiv preprint: 2007.11314.

Peinelt, N., Liakata, M. and Nguyen, D. (2019). Aiming beyond the Obvious – Identifying Non-Obvious Cases in Semantic Similarity Datasets. *Proceedings of the 57th Conference of the Association for Computational Linguistics*.

Peinelt, N., Liakata, M. and Hsieh, S. (2017). Classifier Guesser – A Context-based Classifier Prediction System for Chinese Language Learners. *Proceedings of IJCNLP 2017 (System Demonstrations)*.

# The Alan Turing Institute

