

Imperial College
London



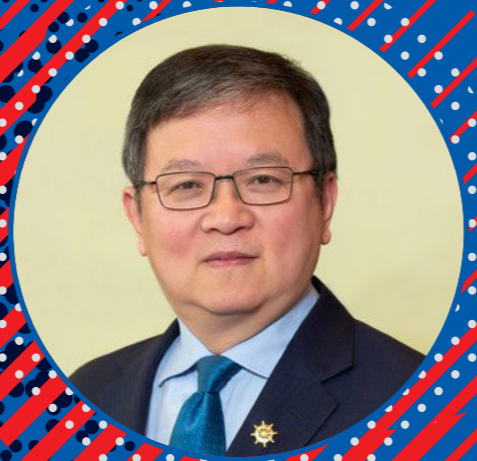
Data Science Institute

imperial.ac.uk/data-science

 @ImperialDSI

Summary

2	Welcome	26	Key Industrial Collaborations
4	The Journey of the Data Science Institute	30	International Academic Collaborations
6	Our Strategy	31	International Conferences
7	Our People and Governance	32	Influencing Policy
8	Our Multidisciplinary Laboratories	33	Outreach
10	Our Lecturers	34	In the Media
11	Our Research Fellows	35	Visitors to the DSI
12	Research from our Academic Fellows		
14	Cutting Edge Research		
17	Our Infrastructure		
18	Multi-Disciplinary Collaborative Projects		
20	Data Observatory		
22	Education		
24	Our Alumni		
25	Supporting Excellence		



Welcome to the Data Science Institute

Big Data is frequently described as heralding a revolution and the transformation of society. How we manage, process and analyse this data to extract knowledge or insight is an important question for academics, businesses and policy makers. Given its powerful theoretical foundations, data science is now becoming an important tool for many other disciplines, from natural sciences to engineering, from medicine to finance.

We have accomplished many achievements since we launched in 2014 through engaging with the College community and advancing data science in many areas of research.

I am proud of the data science research community that has flourished at Imperial College and is driving innovation in data science research and education on a global scale.

Professor Yike Guo → Director of the Data Science Institute



“New technologies such as AI and robotic systems will enhance jobs, changing the type of activity that is done, as technology has done throughout the ages. We will see some jobs disappearing and new jobs appearing.

We have a great advantage as a country in this area, thanks to the excellence of the research base in our universities. That bedrock lets innovation flourish. More and more, innovation requires collaboration across disciplines and the Data Science Institute was

created precisely with the mission to foster collaboration between our academics in different faculties. Together with the other global institutes, it is the shop window for Imperial and the impactful work we are doing in tackling those big challenges.”

Professor Nick Jennings

→ Vice Provost for Research and Enterprise



“Over the past decade, data science has established itself as one of the highest priorities on the national and international agenda. Governments, businesses and educators are fast realising that to progress society, the vast quantities of data must be managed and understood. Data Science is an increasingly important aspect across all disciplines.

I am delighted to endorse the progress and contribution made by the Data Science Institute in its first five years. It has established a solid foundation to enable multi-disciplinary research to grow in collaboration with businesses and international partners. On a personal level, I am strongly supportive of its innovative approach and commitment to ensuring the education and training of the next generation of data scientists and engagement with the wider data science community.

I do hope that you enjoy reading of its achievements.”

Sir Keith O’Nions

→ Chairman of Data Science Institute Advisory Board

Journey of the Data Science Institute

The Data Science Institute (DSI) is a major Imperial College London initiative that brings together Imperial's existing data science activities and expertise, and provides a focus and a catalyst for new partnerships.

The DSI supports multidisciplinary collaborations between the College's academic experts in many disciplines such as healthcare, financial services, climate science, and city infrastructure to create solutions to complex problems. Alongside research, the Institute fosters the next generation of data scientists and engineers by developing a range of postgraduate and executive courses.

In the last 5 years the DSI has grown to include 7 Academic Labs, has attracted over £50m in funding for data science research, technology and infrastructure and has published over 300 papers.

Thanks to its many research collaborations both across College and with a variety of external academic and industrial partners, the DSI is establishing its role as an international hub in data science.

DSI STATISTICS

50+ research projects bids supported	30 collaborative projects across College	300+ papers published	7 Multidisciplinary Labs	>£50m research funding	70% research funding from industry
--	--	---------------------------------	------------------------------------	-------------------------------------	--

Timeline

2014

- April** - Launch of the DSI
- May** - Signing ceremony for the Imperial College-Zhejiang University joint Applied Data Science Lab
- October** - Building works begin at the DSI hub

2016

- January** - BigNeuron Open Analysis Hackathon
- March** - Research collaboration with Jaywing on Cognitive Marketing
- July** - Research collaboration with Thomson Reuters
- October** - Research collaboration with LexisNexis
- November** - Launch of the Business Analytics Lab
- December** - Launch of the Behaviour Analytics Lab

2015

- March** - Memorandum of Understanding is signed with the Research Organization of Information and Systems (ROIS)
- May** - DSI hub opens
- August** - DSI hosts the 2015 International Conference on Brain Informatics & Health
- September** - Research collaboration with Keybridge and CCBI on data economy
- October** - Chinese President Xi Jinping visits the DSI
- October** - Collaboration agreement with Genomics England
- November** - Opening of the KPMG Data Observatory
- November** - Launch of the DSI Strategy 2016–2021
- November** - Launch of the SOCA Lab and Data Assimilation Lab
- November** - China-UK Big Data Summit
- December** - Launch of the Data Economy lab

2017

- February** - Launch of the Algorithmic Society Lab
- February** - Academic Fellow programme opens
- April** - Worley Parsons PhD studentship in data economy established
- June** - HRH The Duke of York visits the DSI
- June** - Signing ceremony for the HNA Research Centre for Future Data Ecosystems

2018

- March** - Launch of the Machine Learning Lab
- April** - £100k as seed funds to 4 projects on Machine Learning
- May** - Beijing Mayor, Mr Cai Qi, visits the DSI
- August** - DSI co-hosts the KDD 2018 conference in London
- September** - MSc in Health Analytics and Machine learning launched in collaboration with the School of Public Health

2019

- February** - UKRI funding awarded to doctoral training programme in AI for healthcare, led by Dr Aldo Faisal
- March** - Launch of the Data Learning Group
- April** - Collaboration initiated with Portuguese research institute INESC TEC
- May** - DSI co-organises the MARBLE conference on Blockchain Economy
- June** - First Workshop on Machine Learning and Data Assimilation for Dynamical Systems
- August** - Launch of Royal Bank of Canada scholarship in machine learning
- September** - Launch of the CSIC Future Digital Ocean Innovation Centre
- November** - Validate AI conference

Our Strategy

Vision

Our vision is to use data to create a better world.

Mission

Our mission is to foster, advance and promote excellence in data science research and education across all domains for the benefit of society.



Strategic Priorities

- 1** To act as a focal point for coordinating data-driven scientific **research** at Imperial through stimulating cross-disciplinary collaboration.
- 2** To train and **educate** the new generation of data scientists.
- 3** To develop data management and analysis **technologies** and services for supporting data driven research.
- 4** To enable the **translation** of data science innovation by close collaboration with partners including industry, and supporting commercialisation.
- 5** To **promote** data science and its applications to the general public and to influence policy makers.

Our People and Governance

ADVISORY BOARD

Sir Keith O'Nions, Chair	Mr Larry Hirst CBE	Sir John O'Reilly
Professor Deborah Ashby, OBE	Professor Jackie Hunter	Professor Paul J Taylor
Professor Yike Guo	Professor Nick Jennings	Mr Paul Taylor

MANAGEMENT COMMITTEE

Faculty of Engineering Dr Aldo Faisal Professor Yike Guo Professor Will Knottenbelt Professor Chris Pain Professor Alessandra Russo	Business School Dr Armand Leroi Dr Mark Kennedy	Faculty of Natural Sciences Professor Niall Adams	Faculty of Medicine Professor Paul Matthews
--	---	--	--

INSTITUTE STAFF

Professor Yike Guo Director	Mr Senaka Fernando Data Observatory systems architect	Ms Fay Miller Research group administrator	Dr James Scott-Brown Research associate in visualizations
Mrs Alice Ashley-Smith Operations manager	Mr Florian Guitton System and infrastructure manager	Ms Diana O'Malley PA and Events Manager	Dr Kai Sun Deputy operations manager
Dr Anna Cupani Stakeholder engagement manager	Ms Ping Huang China partnerships manager	Dr Yves-Alexandre de Montjoye Lecturer	Dr Wenjia Bai Lecturer

RESEARCH FELLOWS AND ASSOCIATES

Dr Rossella Arcucci Research fellow	Dr César Quilodran Casas Research associate	Dr Ovidiu Șerban Research fellow	Dr Shuo Wang Research associate
Dr Chengliang Dai Research associate	Dr K S Sesh Kumar Research fellow	Dr Nguyen Truong Research associate	Dr Xian Yang Research fellow

POSTGRADUATE RESEARCHERS AND RESEARCH ASSISTANTS

Ana-Maria Cretu Postgraduate researcher	Florimond Houssiau Postgraduate researcher	Pierre Richemond Postgraduate researcher	Pan Wang Postgraduate researcher
Ibrahim Emam Research assistant	Shubhan Jain Research assistant	Mihai Suteu Postgraduate researcher	Xiaoyu Zhang Research assistant
Ali Farzaneh Far Postgraduate researcher	Yuanhan Mo (Max) Postgraduate researcher	Arnaud Tournier Postgraduate researcher	Jingqing Zhang Postgraduate researcher
Andrea Gadotti Postgraduate researcher	Philip Nadler Postgraduate researcher		

MULTIDISCIPLINARY RESEARCH LABS

The DSI promotes collaborations through dedicated laboratories bringing together researchers who investigate a common topic from different disciplines. They currently are as follows:	Social and Cultural Analytics Lab Professor Armand Leroi	Data Economy Lab Professor Willian Knottenbelt	Business Analytics Lab Dr Mark Kennedy
	Data Assimilation Lab Professor Chris Pain	Behavioural Analytics Lab Dr Aldo Faisal	Machine Learning Lab Professor Alessandra Russo
	Algorithmic Society Lab Dr Yves-Alexandre de Montjoye		

ACADEMIC FELLOWS

The DSI has a growing community of Academic Fellows from all faculties who share an interest in data science. Through the fellowship programme they get preferential access to the Data Observatory can apply for funding from the DSI and get support in grant applications.

→ Our Multidisciplinary Laboratories

ALGORITHMIC SOCIETY LAB

Lab Director: Dr Yves-Alexandre de Montjoye

Big data dramatically increases our ability to follow and influence the behaviour of individuals and collectives. Together with plenty of opportunities, progress in this area also raises serious questions around privacy and fairness.

The Lab provides leadership, in the UK and beyond, in the safe, anonymous, and ethical use of large-scale behavioural datasets coming from Internet of Things (IoT) devices, mobile phones, credit cards and browsers.



BEHAVIOURAL ANALYTICS LAB

Lab Director: Dr Aldo Faisal

Brain activity, eye tracking, and skeletal movements can now be monitored with an unprecedented level of accuracy. This knowledge will be used to understand and predict human and biological behaviour much more effectively than before.

The Lab taps into this huge pool of data extracting precious information through analytical methods and algorithms to develop innovative technology and understand how human brain works.



BUSINESS ANALYTICS LAB

Lab Director: Dr Mark Kennedy

Established in 2014, the Lab brings together business-focused academics and specialists in data science to do cutting-edge research on how data and artificial intelligence are changing business and society.

It collaborates with companies that have problems, people and datasets available and that want to make the most of the dynamic field of data science, preparing for the change ahead.



DATA ASSIMILATION LAB

Lab Director: Professor Chris Pain

Data assimilation is a methodology that combines measurements and numerical predictions to obtain an accurate representation of a modelled system. It can be applied to a variety of topics such as oceanography and flooding motions, atmosphere and climate changes, urban and indoor flows, oil reservoirs, vehicle design and several industrial processes.

The Lab promotes and leads scientific advances and technological innovations through data assimilation, sensitivity/uncertainty/error analysis, design optimization and control, and computational modelling, simulation and visualisation.



DATA ECONOMY LAB

Lab Director: Professor William Knottenbelt

Large public datasets are nowadays available through government open-source websites. A deluge of data is also collected by private companies and datasets are sold and traded as commodities.

At the Data Economy Lab, researchers are carrying out research into the economic, legal and policy mechanisms underlying the emerging data economy in the UK and worldwide.



MACHINE LEARNING LAB

Lab Director: Professor Alessandra Russo

Machine Learning is a branch of artificial intelligence in which algorithms identify patterns in large amounts of data, build models and take decisions about them, in an iterative process that makes the model more and more robust with minimal human intervention, thus resembling the traditional human learning process.

The Lab promotes and leads scientific advances in this area, developing autonomous decision-making systems that learn from small amounts of data.



SOCIAL AND CULTURAL ANALYTICS LAB

Lab Director: Professor Armand Leroi

The Social and Cultural Analytics (SOCA) Laboratory was launched in November 2015 and brings together Imperial College researchers and colleagues from allied institutions, to investigate how culture and society work.

The Lab members study the evolution of music, the neurobiology of creativity, the dynamics of twitter networks, digital markets and online collaborations, the application of text mining to health and much more.



Our Lecturers

DR YVES-ALEXANDRE DE MONTJOYE

Dr Yves-Alexandre de Montjoye is a lecturer jointly in the Department of Computing and the Data Science Institute, where he heads the Computational Privacy Group. He's a technical expert appointed by Parliament to the Belgian Data Protection Agency and was a Special Adviser to EC Commissioner Vestager in 2018–2019.

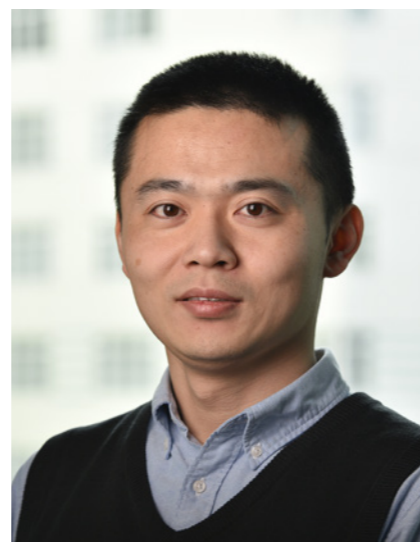


My research aims at understanding how the unicity of human behaviour impacts the privacy of individuals in large-scale metadata datasets. I approach the privacy of metadata (mobile phone, credit cards, or browsing metadata) from two perspectives. First, I developed the concept of unicity to study the risks of re-identification of large-scale metadata datasets. Second, I use machine learning techniques to study what can be inferred from metadata about individuals.

DR WENJIA BAI

Dr Wenjia Bai is a lecturer jointly in the Department of Brain Sciences and the Data Science Institute. His research focuses on developing novel image computing and machine learning algorithms for medical image analysis and applying the algorithms to clinical research.

My research on medical imaging is at the interface between computer science and medicine. It is an area of research which may lead to more accurate diagnosis and treatment planning. The analysis of medical images is largely performed manually and in a time-consuming fashion. Automated image analysis tools have the potential to greatly improve the efficiency of this analysis. In addition, machine learning algorithms may be able to recognise subtle patterns in medical images and healthcare data that are clinically relevant but may not be easily appreciated by humans. I am working on a variety of large-scale medical image datasets of brain, heart, and abdominal organs.



Our Research Fellows

DR JULIO AMADOR

Julio holds a PhD in economics from the University of Essex and has held different research positions, both in the UK and abroad. He is now a research fellow at Imperial Business Analytics, a laboratory held together by Imperial College's Data Science Institute and Business School. His area of expertise is applied machine learning (ML) and his research includes big-data studies of online political participation and applying ML to categorize public opinion and automatically identifying fake news. Julio is currently dedicated to the study of misinformation.

DR ROSSELLA ARCUCCI

Rossella is a research fellow in data analysis. The main topic of her research is data assimilation and/or with machine learning, applied to several case studies. Her areas of expertise are numerical analysis, scientific computing, and the development of methods, algorithms and software for scientific applications on high-performance architectures including parallel and distributed computing. At the DSI, she leads and coordinates the activities of the Data Learning Working Group and supervises MSc students and early career researchers.

DR XIAN YANG

From 2012 to 2018 Xian worked as research associate in the Data Science Institute at Imperial College London. During this period, she took part in many cross-European research projects, such as UBIOPRED (severe asthma subtyping), eTRIKS (knowledge management platform for translational medicine), OPTIMISE (multiple sclerosis disease prognosis and treatment) and iHealth (clinical treatment pathway optimisation). She developed machine learning methods to analyse and construct predictive models from Omics, clinical and survey data. From 2018 to 2019, she worked at Microsoft Research Asia on data intelligence for cloud systems.

She is now a research fellow at the Data Science Institute working mainly in the medical NLP area.

DR OVIDIU ȘERBAN

At the DSI, Ovidiu focuses on modern data science techniques, including large scale visualisation and realtime natural language processing. His research interests and expertise also include machine learning, affective computing, and interactive system design. He works in collaboration with the Institute for Security Science and Technology at Imperial College.

DR K S SESH KUMAR

Sesh is a research fellow in machine learning. He wants to propose scalable solutions to submodular optimisation, Bayesian optimisation and Bayesian nonparametrics. He finished his doctoral studies at ENS/INRIA, Paris under the supervision of Prof. Francis Bach in 2016. He was a post-doctoral researcher in the Kolmogorov Group, IST Austria headed by Professor Vladimir Kolmogorov until October, 2017. He joined Imperial College London in December, 2017 as a Leverhulme Centre for the Future of Intelligence research fellow for the Trust and Transparency project at the Statistical Machine Learning Group, where he was until August 2019.

DR FERNANDO ROSAS

Fernando is a research fellow at Imperial College London, based at the Centre For Psychedelic Research (Department of Medicine), and also affiliated with the Centre for Complexity Science, and the Data Science Institute. His current work is focused on the development of tools to enable a deeper understanding of the interdependencies that can take place in systems composed of many interacting agents. Improving this understanding is paramount for future advances in neuroscience, genetics, network science and many other fields that explore systems composed by many interacting variables. His interest is in the most fundamental and theoretical aspects of this problem, and in the consequences and applications in diverse contexts, related to basic sciences, engineering and arts.

Research from our Academic Fellows

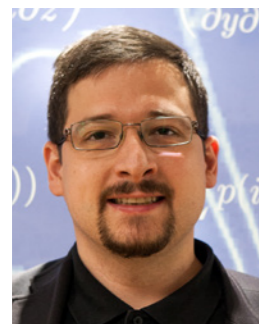
The AI@Imperial network brings together experts across engineering, science, healthcare and business, developing AI methods and systems across a wide range of applications.



In my research lab we study machine learning systems and biological brains: learning from the brain how to advance technology, and vice versa using advanced technology to reverse-engineer the brain – we call it “Neurotechnology”.

We combine methods from computing, physics and engineering with experimental human studies to understand how the brain works and use this understanding to improve patients’ lives.

Dr Aldo Faisal



My goal is to minimise suffering and avoidable deaths in modern healthcare. Clinical expertise in rare diseases is not easily accessible to everyone. Using machine learning, we can provide all doctors with decision support systems that operate in real time.

Dr Bernhard Kainz

HUMAN-IN-THE-LOOP COMPUTING FOR HEALTHCARE

Dr Kainz works to improve medical imaging, thus supporting doctors’ diagnosis of rare diseases. The analysis of biomedical images such as foetal scans is an essential practice during pregnancy screening and monitoring. This is an area where decisions can make the difference between life and death. To support doctors in their medical assessments, he uses machine learning methods. This approach where human interaction is combined with algorithms decision pathways, is referred to as “human-in-the-loop” and is key in clinical decisions.

Access to credit

Innovations in statistical technology, for example in predicting the likelihood to access credit, have raised concerns about the damages of racial profiling to specific groups of people.



Using data on US mortgages, we predicted default using traditional and machine learning models. We found that Black and Hispanic borrowers are disproportionately less likely to gain from the introduction of machine learning. We believe there is much more to be done to understand the impacts of the use of these technologies.

Professor Tarun Ramadorai

Looking at the sky

My research in cosmology is about analysing, interpreting and making sense of cosmological observations, in order to learn more about the properties and nature of dark matter and dark energy. I develop and apply new methods in Bayesian statistics and data analysis, machine learning, deep learning and AI to understand the history and nature of the Universe.

Professor Roberto Trotta



Fighting against antimicrobial resistance

Antimicrobial agents are frequently prescribed according to a one size fits all model yet responses to antimicrobials vary widely. This frequently leads to poor treatment, negative clinical outcomes and excessive antimicrobial use, a key driver of antimicrobial resistance (AMR).

A small electrochemical biosensor with a closed-loop control system has been developed to monitor antimicrobial levels in real-time so that appropriate therapeutic levels are achieved.

Professor Alison Holmes

Cutting Edge Research at the Data Science Institute

Data assimilation is a methodology that combines measurements and numerical predictions to obtain an accurate representation of a modelled system. It can be applied to a variety of research topics such as oceanography and flooding motions, atmosphere and climate changes, urban and indoor flows, oil reservoirs, vehicle design and several industrial processes.



THE MAGIC PROJECT

Buildings and transportation are the biggest culprits for high levels of pollution in our cities. Traditional approaches to urban environmental control rely on heating, ventilation and cooling systems that consume energy and produce toxins and CO₂ in an unsustainable cycle.

Dr Rossella Arcucci, Research Fellow at the DSI says: “This project, run by Imperial College in collaboration with the Universities of Cambridge and of Surrey, develops an integrated suite of tools that use natural ventilation in buildings to reduce demand for energy and ensure air pollutants are diluted below levels that cause adverse health effects.”

INHALE FOR CLEAN AIR

The DSI is part of a large consortium that studies how physical components interact with cells, how this affects human health and how this knowledge can be used to improve air quality.

They have been awarded an EPSRC Physics of Life grant of £2.8 million to study how to mitigate air pollution. The proposal was led by Professors Chris Pain (Department of Earth Science and Engineering and Director of the Data Assimilation Lab at the DSI) and Fan Chung from the National Heart and Lung Institute (NHLI)

Machine Learning is a branch of artificial intelligence in which algorithms identify patterns in large amounts of data, build models and take decisions about them. This iterative process makes the model more and more robust with minimal human intervention, resembling the human learning process.

BREXIT AT THE BORDER

Dr Ke Han and his team simulated motorways queues in the hypothesis of longer border checks in Dover after Brexit. This meant measuring queuing patterns during the day and studying their interaction with the local traffic. He explains: “We found that just 2 extra minutes spent on each vehicle at the border could more triple the existing queues and at peak times Kent could see nearly five hours of traffic delays”.



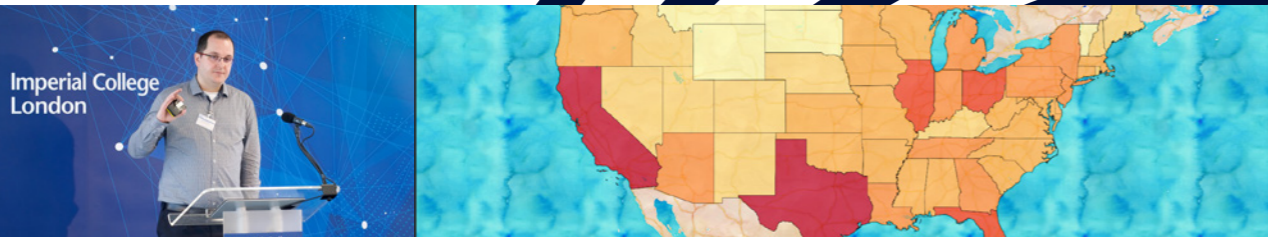
SEED FUND BEARING FRUIT

In May 2018 the DSI awarded a seed fund to 4 research projects in the area of Machine Learning. The projects awarded covered research in medicine (Professor Simon Schultz), air quality (Dr Ke Han), natural language processing (Dr Marc Deisenroth) and experiment modelling (Dr Ruth Misener). This funding consolidated existing collaborations and supported further funding applications to national and international bodies.

THE OPAL PROJECT

OPAL is a joint project between MIT and Imperial College London with telecommunications partners Orange and Telefonica. It consists of an open suite of software and algorithms that give access to statistical information extracted from anonymised, secured and formatted data. The data can be used by national statistical offices, civil society organisations, media partners and other bodies to make informed decisions and efficient interventions.

Such interventions are successful if the data can be correlated to key indicators such as poverty, literacy, population density, social cohesion, in an accurate, timely and reliable manner. Dr Yves-Alexandre de Montjoye, leading the project for Imperial College, says: " We want to provide information that has statistic relevance while ensuring that privacy and fairness are maintained.



THE SENTINEL PROJECT

The extensive use of social media platforms in recent years makes it possible for public health bodies to detect disease outbreaks earlier than traditional methods would allow, thus acting more promptly against pandemics. The SENTINEL project looked at Twitter for clues that people may be experiencing an illness. By scanning millions of tweets from US users, researchers trained a software to classify them as health-related or not. "The word *cold* could mean that someone has a cold or that it's cold outside – explains Dr Ovidiu Şerban – so we designed a classifier that, through machine learning, teaches itself to distinguish what is a health context and what isn't". It all started with human annotators looking at entire sentences and classifying them. "We fed these to the computer which looked for language patterns in how we distinguished the health tweets from the non-health tweets. It then applied this to new tweets".

To test how good the system is compared to humans, the researchers looked at the level of agreement in classifying the tweets between two people and used it as benchmark. "Our human annotators agreed with each other between 84% and 89% of the time when selecting whether tweets were health-related or not. Our neural network is currently around 85% accurate, so it's almost as reliable as us!" says Dr Şerban.

With increased accuracy, it may be possible in the future to monitor in real time how and where outbreaks happen and act accordingly.

→ Our Infrastructure

Society is witnessing an explosion in the level of data being generated, much of it with the potential to improve the way we live our lives. To help realise this aim the DSI promotes big data scientific research and applications by building and integrating a private cloud platform for researchers across faculties. Industry leaders in IT equipment such as Huawei have also donated state of the art IT infrastructure to the institute. Today, the DSI manages petabyte-scale data on hundreds of containers supported by a superfast 100Gb/s network backbone.



Multi-Disciplinary Collaborative Projects

TACKLING BRAIN CANCER

Gliomas are a range of devastating and progressive brain tumours affecting around 25,000 people each year in Europe and responsible for the majority of deaths from primary brain tumours. Nine research institutes and three private sector organizations have joined in the project AiPBAND (An Integrated Platform for Developing Brain Cancer Diagnostics Techniques). Its mission is to improve technologies for the early diagnosis of brain tumours using molecular biomarkers in the blood. The project has been funded under Horizon 2020, the EU's research and innovation program as a Marie Curie Innovative Training Network (MSCA-ITN). The DSI focuses on the creation of a big sensing intelligent warehouse for translational and clinical diagnosis study.

DIGITAL FINGERPRINTS FOR NEURODEGENERATIVE DISEASES

The DSI is part of a Europe-wide consortium of 47 partners from universities, companies and research institutes which has received funding from the European Commission to study neuro-degenerative diseases. The 5-year project called IDEA FAST will look at digital endpoints for fatigue and sleep disturbances. This granular data collected from patients will provide more objective, sensitive and reliable measures of the severity and impact of these symptoms on patients with neurodegenerative disorders (NDD) and immune-mediated inflammatory diseases (IMID). A better understanding of the mechanisms behind the insurgence of such diseases will improve the efficiency of clinical trials and the quality of treatment for patients.

ASTHMA BIOMARKERS

U-BIOPRED or Unbiased biomarkers for the prediction of respiratory disease outcomes, is a €20m research project founded by the Innovative Medicines Initiative (IMI). The project consists of longitudinal clinical studies on adult and paediatric subject cohorts, as well as associated translational studies using animal, in vitro and in silico models of asthma. The goal of the project is the identification of novel biomarkers of severe asthma acquired by both invasive and noninvasive sample collection techniques, and enhance the understanding of asthma aetiopathogenesis leading to the discovery of novel therapeutic targets.

DATA SCIENCE FOR LOW-CARBON FUTURE

The DSI is part of the EPSRC-funded project PREdictive Modelling with Quantification of Uncertainty for MultiphasE Systems (PREMIERE), a multidisciplinary collaboration between Imperial, the University of Birmingham, University of Cambridge, UCL, and the Alan Turing Institute. The project received £6.5m funding from EPSRC. The PREMIERE project will help to deploy energy systems that will make a significant contribution to generating low-carbon power. This will support the UK in meeting its net-zero emissions pledge by 2050.

DIGITAL CITIES

Digital City Exchange was a five-year Digital Economy multi-disciplinary research programme at Imperial College London. Researchers explored ways to digitally link utilities and services within a city, for new technical and business opportunities. Developments in pervasive sensing, large-scale modelling, new analytical and optimisation techniques and web services technologies all offer a new wave of opportunities. The programme of research focuses on harnessing digital systems to combine and repurpose city data, ultimately transforming the planning and use of cities. The DSI contributed to the creation of an advanced platform for building services based on city data collected by various sensors.


Data Observatory

The Institute's Data Observatory (DO) was one of the first and largest visualisation suites in Europe. It provides a multi-dimensional and immersive environment to analyse large and complex data sets.

The DO is used for data exploration projects (Mars imaging, Runkeeper, SENTINEL, tranSMART amongst others), teaching courses and as a showcase space for current research projects. The Observatory is also connected to an immersive Microsoft Kinect perception environment, made of multiple sensors which detect people's movements in the room and track their gaze. This information can be used as a predictor for viewers' interests to different section of the demo.


The whole infrastructure is powered by a new Open Source project, called Open Visualisation Environment (OVE), which allows us to create immersive experiences for Large Scale Visualisation Systems. Such a system has various applications, from real-time visual analytics, such as detection of traffic patterns using Transport For London's public cameras, to an immersive projection of the Map of the Universe, the largest interactive visualisation of the known universe. All these projects involve multi-disciplinary research groups working on various topics from digital currencies (Bitcoins demo, in collaboration with Professor Knottenbelt), to transportation management (Shanghai Metro Analysis in collaboration with Dr Anderson), from policy analysis (Chinese migration in collaboration with Zhejiang University) to public health interventions (Anti Microbial Resistance in collaboration with Dr Céire Costelloe). They all rely on the team's expertise in data visualisation, artificial intelligence, user interaction and large scale distributed system design.

DO SPECIFICATIONS


64 
screens


32 
workstations

132M 
pixel display environment

313° 
panoramic view

PEOPLE STATISTICS

30+ 
student projects

180+ 
MSc students taught
on data science
modules

30+ 
Exec Ed sessions

5 
public affairs
meetings

150+ 
delegations

RESEARCH DATA

40+
visualisations in
the Observatory

Open Visualisation Environment
Development of a Data Platform
for Large Scale Visualisation

Sanjeev Gupta, Professor in the Department of Earth Science and Engineering, heavily involved in the NASA *Curiosity* rover mission routinely uses the Data Observatory for teaching undergraduate courses. "The Data Observatory is a unique environment because it allows students to interact with high-resolution visualisations in a way that would not be possible on a small screen. This is particularly true for space exploration."



Education

The Data Science Institute supports the broader College educational mission with a variety of courses on Big Data and Artificial Intelligence for students in various departments, from Design Engineering to the Business School.

Together with the School of Public Health, the DSI delivers the MSc in Health Data Analytics and Machine Learning, a one-year full-time course that gives students a solid background in analysing and managing health data.

The DSI also contributes to the MRes in Biomedical Research (Data Science), also within the Faculty of Medicine amongst several other courses.



DOCTORAL TRAINING CENTRES

In October 2019, the new Centre for Doctoral Training in *AI for Healthcare* was launched. This doctoral programme funded by UKRI (United Kingdom Research and Innovation) will train up to 100 doctoral researchers over the next 5 years at the interface between medicine and computer science. Dr Aldo Faisal, director of the centre, says: "Our goal is to improve patients' care and well being. This unique training will allow us to grasp the incredible opportunities that artificial intelligence offers to both patients and healthcare professionals."

The DSI is also jointly setting up a doctoral training programme with the University of Southampton and several industrial partners on *AI for Future Society*. Professor Dame Wendy Hall, from the University of Southampton, says: "AI is having a profound impact on our society, today and tomorrow. Thanks to this collaboration with the DSI at Imperial College, we will train the new generation of AI researchers to be the future leaders in on an AI-empowered world."

Bespoke Winter and Summer Courses

The Data Science Institute offers bespoke winter or summer courses for partner universities around the world.

They are designed for undergraduate students with an interest in data science and AI, allowing them to expand their knowledge outside traditional term time.

The programme uses group discussions, workshops, off-campus excursions and guest speakers to enrich, enhance and develop business knowledge and practical skills in a fun and supportive learning environment.



Executive Education

The Data Science Institute works in close collaboration with the Business School's Executive Education team to deliver a number of open and custom programmes for business executives from Vodafone, Strongways and Naspers.

Many of these programmes take place in our Data Observatory. Open courses currently provided include Digital Transformation Strategy and AI and Machine Learning in Finance.

The DSI also developed customised programmes for partners such as Barclays and Prudential.



→ Our Alumni

Axel Oehmichen (DSI postgraduate researcher)

Axel co-founded the start-up Secretarium where he acts as Chief Data Science Officer. He is applying his knowledge in distributed computing and privacy to build a confidential computing platform to simplify the setup and integration of privacy-respecting applications in finance and life science.

May Yong (DSI Postgraduate researcher)

May is a Senior Research Software Engineer at The Alan Turing Institute. She is building tools to minimise data ambiguity so that data is interpreted and used the way the data collectors intended.

Dan McGinn (DSI research assistant)

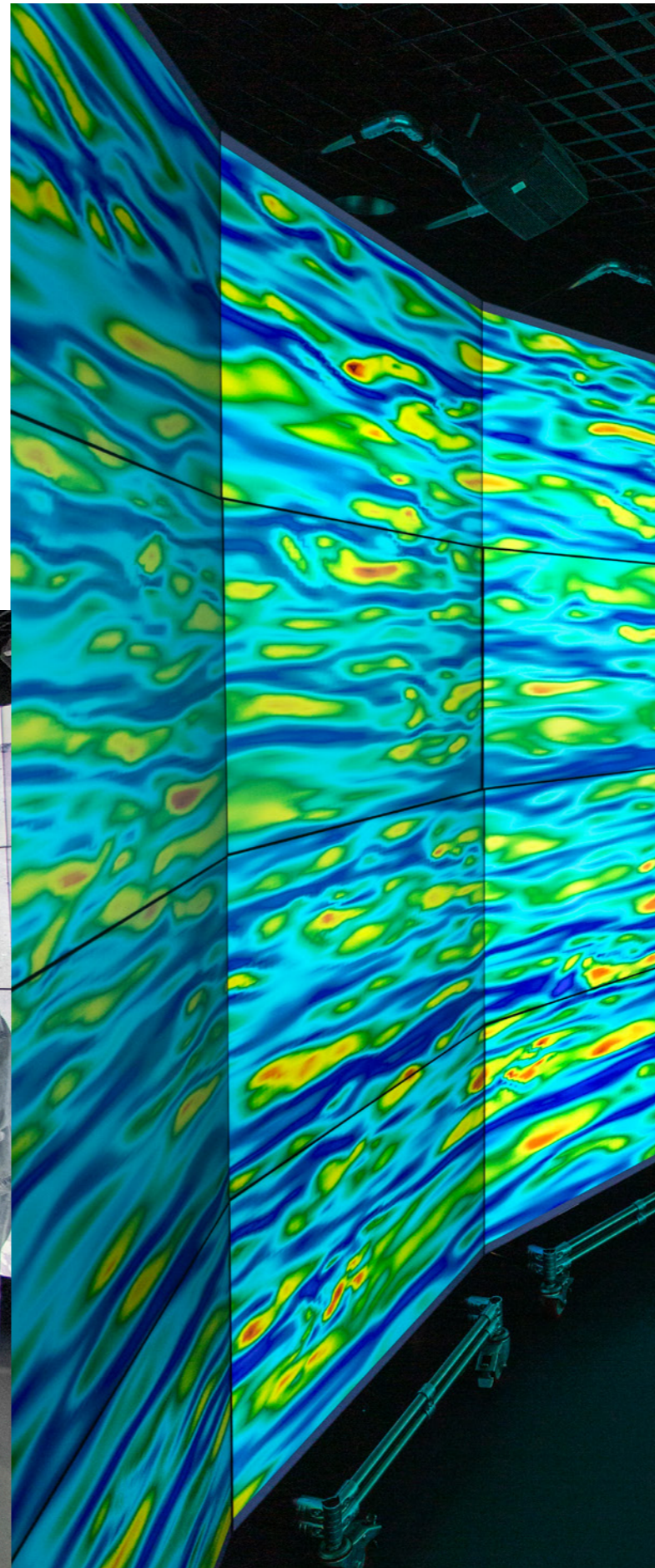
Dan is the co-founder Catalact, a company that uses data science techniques to detect patterns and anomalies in crypto blockchains to understand the behaviours behind market flows and identify new trends in real time.

Hao Dong (DSI Postgraduate researcher)

Hao is currently an Assistant Professor at Peking University. His research is at the intersection of deep learning, computer vision and graphics. He is interested in unsupervised learning and lifelong learning.



→ Supporting Excellence beyond Research



DATA SPARKS

Together with the Imperial Business School, and leveraging some of the world's best academics, industry experts and technologies, the DSI supports the Data Spark scheme, an agile and innovative research and development platform that enables business leaders to experiment and test new ideas.

Through a student placement scheme, companies supplied data that constituted the basis of 6-week exploratory student projects with support from an academic mentor with speciality in the problem domain. These projects are designed to support longer-term research and were visualised in the Data Observatory (secondary research project).

STUDENTS ENTREPRENEUR

Imperial students Annika Monari and Alan Vey co-founded Aventus – a blockchain-based event ticketing solution that practically eliminates fraud and unregulated touting.

Aventus launched the token sale of its AventusCoin (AVT) - "the Bitcoin of ticketing" – in 2017. They sold out within seven minutes, receiving over 60,000 Ethereum – then equivalent to around \$20m – from investors.

BEST THESIS IN DATA SCIENCE

In 2015 the DSI awarded its Best Thesis Prize, worth £3000, to Melissa Turcotte, for her PhD in Statistics at Imperial titled: Anomaly Detection in Dynamic Network. The Best Thesis Prize, sponsored by IBM, recognised an outstanding PhD thesis completed within the fields of data science or data-driven scientific research at Imperial College London.

Key Industrial Collaborations

The DSI has strong relationships with a huge variety of industrial partners. They contribute to fund research projects, provide exchange opportunities for our students and accelerate technology transfer. Some of the companies we have been working with are reported below:

- China Construction Bank Corporation International (CCBI): £2m donation and a joint Lab on data economy
- China Shipbuilding Industry Corporation (CSIC): £3m research funding for a joint Lab on digital ocean
- Genomics England : research funding for building translational medicine research infrastructure
- GlaxoSmithKline (GSK): research funding to support personalised medicine research
- HNA Group: research collaboration in blockchain infrastructure
- Huawei: research collaboration in data science innovation
- Jaywing: PhD studentships in neuromarketing
- KPMG: donation for building Data Observatory and collaboration in building the Business Analytics Lab and Data Sparks
- LexisNexis: PhD studentship in text mining technology
- Royal Bank of Canada (RBC): PhD studentships in financial text analytics
- Thomson Reuters, now Refinitiv: research fellow in text-based big data analysis
- Worley Parsons: PhD studentship in data economy



TEXT ANALYSIS TECHNOLOGY FOR LEGAL INDUSTRY

More and more companies are aware of the importance of quantitative analysis tools to support decision making. LexisNexis is a global provider of legal, regulatory and business information and analytics. They supported the work done at the DSI by Jingqing Zhang, a postgraduate researcher studying text mining, multimodality, deep learning and data science.

WORKFORCE OF THE FUTURE

Royal Bank of Canada funded a project that looked at data from their workforce over the years. This research collaboration allowed the company to get insight about career progression, requalification and changes in workforce tasks. After identifying hurdles and biases in promotion and retention of staff, the company could improve their practices and be more strategic in the allocation of resources.

DATA ECONOMY MODELLING

The Australian engineering company WorleyParsons, with expertise in procurement and construction, co-funded Philip Nadler's PhD research in the area of social and economic impacts of data science. His research is at the intersection of economics and computer science. He is using data science technology, such as high-performance computing and statistical inference, to develop economic models that are more robust and accurate in their predictions.

PEOPLE FLOW

MAKES AND LADDERS IN A LARGE BANK

YE 2012

SERVICE

YE 2013

LEAVERS

YE 2014

LEAVERS

NEW PLATFORM FOR TREATING MULTIPLE SCLEROSIS

OPTIMISE is a collaboration between Imperial College London and the biopharmaceutical company Biogen Idec, who have a long-standing commitment to developing therapies for people with Multiple Sclerosis (MS). By comprehensively capturing and managing data in ways that can be implemented at a low cost and a large scale, the project allowed researchers to better monitor outcomes and evaluate new treatments. The Data Science Institute developed a custom-made software platform used by OPTIMISE to store, curate and analyse data.

DATA ECONOMY RESEARCH

Chinese Construction Bank International is a company that provides a range of execution and clearing services for base metal contracts on the London Metal Exchange. In 2015 they contributed to the creation of the DSI's Data Economy Lab, with a £2m donation. Their contribution supported research into the economic, legal and policy mechanisms required for the emerging Data Economy in the UK and worldwide. Working with colleagues in economics and social science, the DSI is pioneering research in the key areas of open data business model, digital money and digital service exchanges.

FUTURE DIGITAL OCEAN

In 2019 the Data Science Institute signed an agreement with the China Shipbuilding Industry Corporation for a 5-year collaboration. The £3million investment will focus on technologies and platforms for machine learning in areas of relevance for the shipping industry such as ocean data (Dr Rossella Arcucci) and blockchain for logistics (Dr Mark Kennedy).

The collaboration will be led jointly by Professor Yike Guo and Professor Julie McCann and include academics working on distributed deep learning (Dr Wenjia Bai), integrated data management (Dr Thomas Heinis) and ocean data assimilation (Professor Chris Pain).

BUSINESS ANALYTICS

KPMG has been one of the key partners of the Data Science Institute since its inception and supported the creation of the Data Observatory, and the Imperial Business Analytics Centre. The Centre is led by Dr Mark Kennedy and researchers at Imperial College Business School as part of the College's Data Science Institute, which is developing new data science methods and technologies and supporting world class data-driven research. In the past 5 years, the KPMG Centre developed innovative approaches, analytical methods and tools for big data, giving UK businesses the opportunity to solve complex issues, such as enabling banks to predict fraud or helping retailers better understand consumer behaviour.

International Academic Collaborations

IMPERIAL COLLEGE-ZHEJIANG UNIVERSITY JOINT APPLIED DATA SCIENCE LAB

The joint Lab for Applied Data Science is a collaboration between the Data Science Institute and Zhejiang University (ZJU), China. Its aim is to combine the expertise and research activities in data-driven scientific research and innovation at ZJU and Imperial to build a world-class Lab. Future applications include bioengineering, brain science, energy systems, meteorology and materials science. The Lab acts as the base to coordinate the growing collaboration between the two institutions and hosts seconded students, researchers and professors.

Over the past 5 years, it has welcomed 8 PhD students and postdoctoral researchers working in the Lab. Dr Wei Xiang is currently a visiting researcher from the Modern Industrial Design Institute, College of Computer Science and Technology, at Zhejiang University. He works on a stereoscopic visualization program, analysing cognitive effect of stereoscopic images to generate content for scientific and 3D data visualization using our unique “bare eyed 3D visual wall” facility.



INESC TEC

The Data Science Institute has initiated a collaboration with the Portuguese research institute INESC on Artificial Intelligence and its applications with a series of events. In 2019 a delegation of twelve scientists from the Portuguese research institute INESC TEC came to London to attend a workshop on “AI and its applications” co-organized and hosted by the Data Science Institute. Researchers from both institutions had a chance to present their work, answer questions about the biggest challenges in the sector and discuss potential collaborations. The main areas where the collaboration will focus are medical imaging, infrastructure maintenance, social media analytics and data privacy.

In July 2019, a delegation of UK researchers in AI and data science flew to Lisbon to take part in a panel discussion on the future of AI during the Portuguese Science Week, where the UK was special guest. Researchers of both countries discussed how AI and data science will shape our societies in the next 20 years and how this strategic collaboration will foster research in this area.

International Conferences

BIH 2015

The DSI led the organisation of the 2015 International Conference on Brain Informatics and Health (BIH 2015), which was held at the Royal Geographical Society from 30th August to 2nd September, 2015. Co-Chaired by Professors Yike Guo and Karl Friston, BIH 2015 provided a premier forum that brought together researchers and practitioners from neuroscience, cognitive science, computer science, data science and neuroimaging technologies with the purpose of exploring the fundamental roles, interactions as well as practical impacts of Brain Informatics.

KDD 2018

The 24th SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2018) took place in ExCeL London on August 19-23 2018. Started in 1989, KDD is the oldest and largest data mining conference worldwide. KDD 2018 conference was the largest KDD conference ever, with more than 3400 participants from 99 countries and 1588 submissions. The DSI played a significant role in KDD 2018, from the proposal phase to the development of the conference programme, to the management of event logistics, attracting researchers and practitioners in data science, data mining, knowledge discovery, large-scale data analytics, and big data.

During the KDD conference, a one-day round-table co-organized by the DSI and the Alan Turing Institute, brought together the Directors and leading academics of major data science institutes around the world to discuss about the state of affairs and the future of the discipline. The participants discussed the definition and role of data science and the strengths and challenges of their institutes.



MARBLE 2019

Mathematical Research for Blockchain Economy (MARBLE) is a new international conference series initiated by the DSI. It aims to provide a high-profile platform for mathematicians, computer scientists and economists from both academia and industry to present latest advances and innovations in key theories and models of blockchain.

The first MARBLE conference was held at Santorini Greece, from 6th to 10th May 2019. MARBLE 2019 focused on the mathematics and economics behind blockchain, seeking to bridge the gap between practice and theory. Professor Yike Guo (General Chair), Professor William Knottenbelt (Programme Chair), Dr Kai Sun (Organising Chair), Diana O'Malley (Finance Chair), and members of the Centre for Cryptocurrency Research and Engineering have contributed to the success of the conference.



→ Outreach



ROYAL SUPPORT TO IMPROVE MENTAL HEALTH

Imperial experts from the DSI and the Department of Mathematics, along with representatives from Heads Together, analysed YouGov data on emerging trends in relation to attitudes to mental health.

Heads Together is a mental health initiative spearheaded by The Duke and Duchess of Cambridge. It aims to tackle the stigma associated with mental health and inspire people to talk more openly about mental wellbeing.

The team of Imperial researchers who analysed the data included Professor Yike Guo and Drs Xian Yang and David Birch from the DSI. Ms Sarah Jones, a research postgraduate from the Institute for Global Health Innovation at Imperial and an advisor to the Royal Foundation, also assisted, along with Professor Michael Crawford, from the Department of Brain Sciences, and Dr Reza Drikvandi, from the Department of Mathematics. Around 14,000 responses from 6 surveys were analysed by the team. This was presented to their Royal Patrons at a private viewing.

IMPERIAL LATES



Imperial Lates give you the chance to bring out your inner child and have fun with our research, as well as debate the major issues of the day with our scientists and engineers. Thanks to its unique visualisation facility, the DSI is best placed to accommodate interactive visits and display large maps and their changes in real time. We showed the turbulence around an engine magnified and in slow motion, what we know at the moment about astronomical objects, thanks to the Map of the Universe, how diseases spread and how contagion can be monitored and more.

→ Influencing Policy

The Forum is Imperial's new programme to connect researchers with policy-makers and discover new thinking on global challenges.

In a workshop on AI and the future of work, researchers presented their work to civil servants on different aspects of AI in the workplace. The Forum has previously engaged policy-makers in workshops on other key areas of Imperial research, including vaccines, air quality, data privacy and AI's healthcare applications, and has further workshops planned for the next academic year.



Evidence Week

Evidence Week in Parliament is an initiative of Sense about Science – the independent charity that promotes the public interest in sound science and evidence – in collaboration with the House of Commons Library, POST (the Parliamentary Office of Science and Technology) and the House of Commons Science and Technology Committee. Yves-Alexandre de Montjoye and Andrea Gadotti from the Algorithmic Society Lab at the DSI, represented Imperial at an event that took place last June at the House of Commons, where they explained to MPs and committee members the importance of adequate policies around data anonymization.

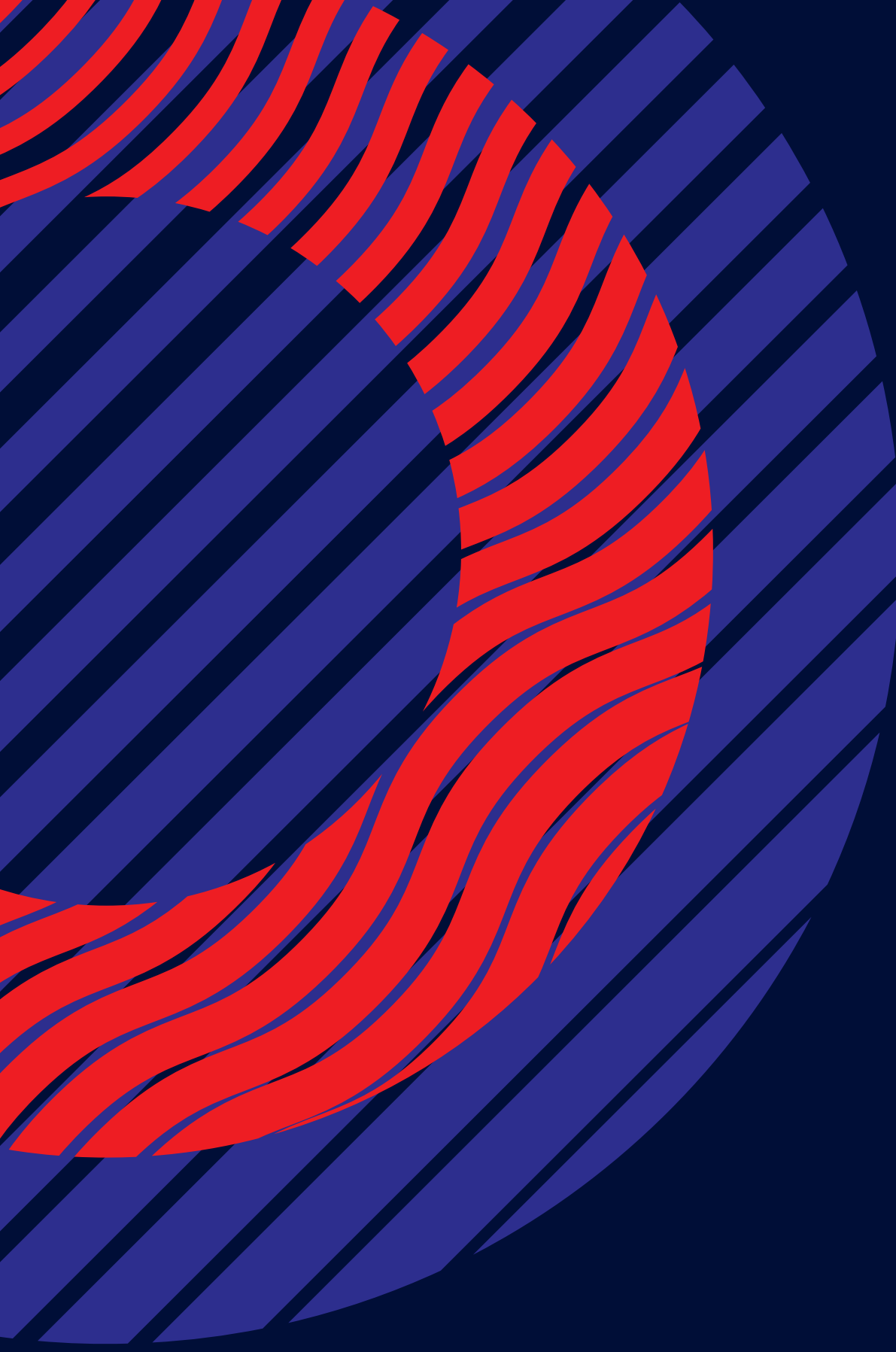


In the Media

- "The Secret Science of Pop" is a BBC programme featuring Professor Armand Leroi's work displayed in the DO; January 2015
- Motherboard: Europe's largest Data Observatory turns Big Data into Big Images; November 2015
- BBC Click: Bitcoin Visualisation; December 2015
- The Gadget Show: Imperial College's Futurists being interviewed; February 2016
- Sir Charlie Bean speaking at the DSI about "Future of UK economic statistics set out at Imperial; March 2016
- The Drum's; May 2016
- A feature in the I; May 2016
- An ITN Productions piece of video content that has been distributed across a number of outlets including AOL, MSN, Yahoo, Irish Independent and Huffington Post
- Panorama: Gambling (Professor David Nutt); June 2016
- Opinion piece on Advertising Week Europe and The Huffington Post online; September 2016
- Professor Sanjeev Gupta from the Department of Earth Science and Engineering was filmed at the DSI Observatory by the BBC 4 TV programme the Sky at Night; April 2018
- "Professor dreams of big-data future" China Daily UK – interview with Yike Guo; June 2018
- "Major forum focuses on high-quality development" China Daily – Article reporting Yike Guo's comments on China being one of the major regions for the development of artificial intelligence in the world; November 2018
- "Princess Eugenie goes back to university - details" Hello! – February 2019
- Forbes: Where Is The Best Place To Create A Startup In Europe?; April 2019
- "The Progress 1000: London's Most Influential People 2019" Evening Standard - Imperial academics and alumni featured in the list included Dr Ling Ge, former Visiting Fellow at Imperial's Data Science Institute
- "Changes to EU antitrust enforcement on Big Tech urged" Financial Times article quoting Yves-Alexandre de Montjoye, April 2019

Selected list of visitors to the DSI

- Aug-15 Meng Jianzhu, Secretary of Central Political and Legal Affairs Commission, China
- Oct-15 President of China Xi Jinping and First Lady Peng Liyuan
- Oct-15 George Osborne, Former Chancellor of the Exchequer, UK
- Jan-16 Professor Ana Mari Cauce, President of the University of Washington, US
- Jun-16 Lu Hao, Governor of Heilongjiang, China
- May-17 Jan Tore Sanner, Norway's Minister for Local Government and Modernisation
- Jun-17 HRH Prince Andrew Duke of York
- Jun-17 Koji Tsuruoka, Japanese Ambassador to the UK
- Jul-17 Matt Hancock MP, Secretary of State for Health and Social Care, UK
- Oct-17 HRH Prince William Duke of Cambridge
- Dec-17 Professor Chen Xu, Chairperson of Tsinghua University, China
- Feb-18 Albert II Prince of Monaco
- Mar-18 Teo Chee Hean, Deputy Prime Minister of Singapore
- Mar-18 Kersti Kaljulaid, President of Estonia
- Mar-18 Dr Grigory V. Trubnikov, Russian Science Minister
- Apr-18 Professor Muhammad Yunus, Nobel prize winner
- Apr-18 Professor Manuel Heitor, Portugal's Minister of Science, Technology and Higher Education
- Apr-18 Shen Yiqin, Governor of Guizhou, China
- May-18 Cai Qi, Communist Party Secretary of Beijing
- May-18 Michael Gove MP, Secretary of State for Environment, Food and Rural Affairs, UK
- May-18 Thomas Stelzer, Governor of Upper Austria
- Nov-18 Karen Bradley MP, Former Secretary of State for Digital, Culture, Media and Sport, UK
- Dec-18 Professor Wu Zhaohui, President of Zhejiang University, China
- Feb-19 Princess Eugenie
- Jun-19 Jack Dorsey, CEO and co-founder of Square, CEO and founder of Twitter
- Jul-19 Wang Zhigang, China's Minister of Science



⇄ Imperial.ac.uk/data-science

⇄ [@ImperialDSI](https://twitter.com/ImperialDSI)