TRUST IN AI TRACK

UN AI FOR GOOD SUMMIT 2018

Geneva, 15 – 17 May 2018

UNIVERSITY OF CAMBRIDGE

CFI LEVERHULME CENTRE FOR THE FUTURE OF INTELLIGENCE

THE ROYAL SOCIETY

LEVERHULME TRUST

Trust in AI

Artificial intelligence (AI) has the potential to dramatically accelerate the pace at which the United Nations' Sustainable Development Goals (SDGs) can be achieved. Maximising AI's potential for good will depend on building and earning trust in AI, in several dimensions. This track will focus on three dimensions of trust. Developers of AI solutions must earn the trust of communities to which such solutions are offered. AI developers and others working for beneficial AI must trust each other, across cultural, national and corporate boundaries. And AI systems themselves must be demonstrably trustworthy.

This track will explore these three dimensions of trust under three Themes:

A. *Building trust for beneficial AI – trust by stakeholder Communities*
B. *Building trust for beneficial AI – trust across boundaries*
C. *Building trust for beneficial AI – trustworthy systems*

On Day 2 we will present three projects outlines in each Theme. Participants will be invited to refine the projects, to propose potential research, impact, and funding partners, and to join projects as collaborators. In addition, on Day 3, we will present a vision for an international collaborative organisation – TrustFactory.ai – to launch further projects of this kind, designed to engineer trust for beneficial AI.

Track Hashtags: #AIforGood #TrustFactoryAI

Twitter handles: @ITU, @LeverhulmeCFI, @royalsociety

Programme – Trust in AI Track

Day 2

9:00 – 9:15	Introduction: Framing of the Trust in AI track Huw Price (CFI), Stephen Cave (CFI), Francesca Rossi (IBM, CFI) and Claire Craig (RS)
9:15 – 10.30	Presentation/discussion of Theme A projects (25 mins each incl. Q&A) Chair: Stephen Cave (CFI)

Building trust for beneficial AI – trust by stakeholder communities

1. Dr Becky Inkster (Department of Psychiatry, Cambridge): Building better care connections: establishing trust networks in AI mental healthcare
2. Dr Dina Machuve (Nelson Mandela African Institute of Science and Technology and Technical Committee Member for Data Science Africa): Assessing and Building Trust in AI for East African Farmers: A Poultry App for Good
3. Irakli Beridze (United Nations Interregional Crime and Justice Research Institute, UNICRI): Building Trust in AI – Mitigating the Effects of AI-induced Automation on Social Stability in Developing Countries & Transition Economies

10:30 – 11:00	**Coffee break**
11:00 – 12:30	Presentation/discussion of Theme B projects (30 mins each incl. Q&A) Chair: Claire Craig (RS)

Building trust for beneficial AI – trust across boundaries

1. Prof LIU Zhe (Peking University, Beijing): Cross-cultural comparisons for trust in AI
2. Dr Kanta Dihal (Leverhulme Centre for the Future of Intelligence, Cambridge): Global AI Narratives
3. Prof David Danks (Carnegie Mellon University): Cross-national comparisons of AI development and regulation strategies – the case of autonomous vehicles

12:30 – 13:30	**Lunch**
13:30 – 15:00	Presentation/discussion of Theme C projects (30 mins each incl. Q&A) Chair: Francesca Rossi (IBM, CFI) *Building trust for beneficial AI – trustworthy systems* 1. Dr Jess Whittlestone (Leverhulme Centre for the Future of Intelligence, Cambridge): Bridging the policy-technical gap for trustworthy AI 2. Dr Rumman Chowdury (Accenture): Trustworthy data: creating and curating a repository for diverse datasets 3. Dr Krishna Gummadi (Max Planck Institute, Saarbrücken) and Dr. Adrian Weller (Leverhulme Centre for the Future of Intelligence, Cambridge, and Alan Turing Institute, London): Cross-cultural perspectives on the meaning of 'fairness' in algorithmic decision making
15:00 – 15:30	**Coffee Break**
15:30 – 16:30	Panel Discussion – *Trust in AI: Opportunities and Challenges* Chair: Huw Price (CFI) Speakers: Prof Hagit Messer-Yaron, World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), UNESCO, 'Trust in AI by educating engineers to ethically aligned design'; Elena Tomuta, Comprehensive Nuclear-Test-Ban Treaty Organization, 'Trustworthy AI Systems: Lessons Learned from an Arms Control Application'; Joe Westby, Amnesty International, 'A human rights framework for trustworthy and accountable AI'
16:30 – 17:30	Project breakout discussions Chair: Stephen Cave (CFI) Attendees will choose which project's discussion group to join.
17:30 – 17:45	Wrap Up
17:45 - 18:15	(Core team only) Next steps

Theme A.1: Building better care connections: establishing 'trust networks' in AI mental healthcare

The Challenge

Depression is the leading worldwide cause of disability and ill health 1 and by 2030 mental health is predicted to be the leading global disease burden. The rapidly emerging AI mental healthcare field is hoping to address these issues; however, serious obstacles to care still remain, especially for disadvantaged, vulnerable populations with mental health problems. Knowing who to trust and where to turn during a mental health crisis can be more of a disorienting maze than a sign-posted roadmap. Family? Friends? Community? Local authorities? Government? Social Media platforms? App developers? Third parties? Healthcare systems? Psychiatrists? Other care professionals? Academics? Advertisers? Non-Profits?

The Solution

In order to move forward, we must first identify where trust has broken down and where it still exists. This must be considered across a wide network of different stakeholders' perspectives. Only then can we start to create new practice models of AI mental healthcare that are fundamentally built on trust, thereby creating a transparent vision across sectors.

Proposed Methods (1 Year Plan Outline)

Stage 1: Scoping Report on "Trust Networks" Setting: Trust Hackathon Event & Follow-up Roundtable Meeting, London, UK (17th July 2018)	• 'Trust' is the core theme of the 2018 Digital Innovation for Mental Health Conference (17-18th July, London, UK). At this conference, there will be a **Trust Hackathon Event** where a range of mental health tech start-ups will engage with individuals from diverse sectors and life perspectives in order to 'hack' issues on trust, privacy, policy, discrimination, efficacy, culture, design etc. • Follow the hackathon, a **Roundtable Meeting** will take place to translate discussions into a scoping report called 'Trust Networks'. The aim of this report will be to identify all relevant stakeholders and evaluate the degree of 'trust' that may or may not exist amongst each of these complex relationships.

Stage 2: Scoping Survey & Analysis Setting/time: TBD	• Contributors and survey developers will use the outcomes generated by the "Trust Networks" scoping report to help identify key questions centred around trust in AI mental healthcare from multiple perspectives, which will be used to inform a global survey. Analysis and dissemination will follow.
Stage 3: Meeting to establish a new practice model framework for AI mental health founded on trust Setting/time: TBD	• The aim of this meeting is to establish a new practice model framework in a way that is centred around repairing and amplifying trust amongst different stakeholders and across sectors (e.g., bridging online-offline experiences, giving people the power to build their communities, combining AI, social media and pioneering methods in social psychiatry etc.).

Confirmed Public Partners

- **The Lancet Psychiatry** (Editor, Niall Boyce)
- **All-Party Parliamentary Group, UK, Young People & Social Media** (Chair, Chris Elmore MP)
- ***AI & Society: Knowledge, Culture and Communication*** (Editor, Karamjit S. Gill)
- **Department of Global Health and Social Medicine**, Harvard University, USA (Prof Vikram Patel, Dr John Nasland and colleagues
- **It's OK To Talk, India** (Non-Profit Organisation & Research Institute)
- **Royal Society of Public Health, UK** (CEO, Shirley Cramer)
- **Digital Innovation in Mental Health Conference, UK** (Creator, Becky Inkster)
- **Positive Computing Group, Australia** - ARC Future Fellow, Professor and Director of the Wellbeing Technology Lab (University of Sydney)

Contact

Becky Inkster, bi212@medschl.cam.ac.uk

References:

1. "Depression: let's talk" says WHO, as depression tops list of causes of ill health. News Release, Media Centre, World Health Organisation. Article released 30.03.2017; accessed 16.04.2018:

http://www.who.int/mediacentre/news/releases/2017/world-health-day/en/

Theme A.2: Building Trust in AI for East African Farmers: A Poultry App for Good

The Challenge

The availability of agricultural data is challenging in developing countries. In East Africa, the case includes poultry farmers, most of which operate on a small scale; they don't have a systematic way to collect, analyse and store information and thus they do not see the value of data for running their enterprises. The farmers are mainly organized on social media groups, such as WhatsApp and Facebook, to address their needs for information on poultry farming practices and to advertise their products to consumers. It requires building trust among poultry farmers so that data can be collected properly and helpful information can be disseminated.

The Solution

This project will assess the causes behind the lack of trust focusing on the case of poultry farmers, a group who has been identified by researchers as having the most to gain from AI. It will also start to build that trust with an app designed to address the specific concerns and needs of the farmers. With the help of AI, farmers would be able to better predict crop and livestock health which would increase the production of food for the entire country.

The Proposed Approach

Aim: The goal of this project will be to assess the challenges for building trust in AI among farmers in East Africa and to propose of a solution to help increase food production.

Methods: First, there will be a survey to understand the causes behind the lack of trust in AI. Then, with the information gained from an initial survey, AI technologies will be built as part of a hackathon at the Data Science Africa Conference in November. Afterwards, the app will be implemented and a post survey will be conducted to understand if trust has shifted.

Outcomes: The outputs and response rate of the pre and post surveys on trust will be initial outcomes of this work. Other measures include the overall use of the application by the farmers and any increase of food production that the app helps to provide. We will track and monitor both things with the data collected from the app.

Timetable

- Summer 2018 - Survey is created with the help of experts be able to access trust and other problem areas accurately. Survey is released to farmers.

- Fall 2018 - Survey results are collected and analysed. Results to be presented at the Data Science Africa Conference in November 2018. Initial app will be built by African
- technologists and researchers attending these workshops.

- Winter 2018 - App in development.

- Spring 2019 - Beta use of app by farmers.

- Summer 2019 - Post survey will be released and initial results of app usage and food production will be published. Further improvements to app will be made.

Current Partner

Data Science Africa

Contact

Dina Machuve (PhD) dmachive@gmail.com, and Ezinne Nwankwo enwankwo17@gmail.com

The Challenge

Artificial Intelligence (AI) has undergone a renaissance, edging its way from the realms of science fiction into the very functioning of society. Growing computational power and an increasing abundance of data have been at the core of this, significantly improving AI capabilities and broadening the range of its real-world application. Technology giants and Governments alike are also increasingly investing in AI, in the hopes of enhancing efficiency, optimising resource allocation, reducing costs, and creating new revenue opportunities.

Masked behind the benefits of AI however, are the unexplored consequences of AI-enabled automation, which threaten to undermine trust and belief in AI through widespread displacement of workers and usher in social instability through new waves of migration and increased crime rates. Although automation will be global, the impact will not be felt evenly. Developing countries and economies in transition will bear the brunt of disruption, as traditional labour-cost advantages are swiftly undercut. Moreover, the solutions of developed countries to diversify economic bases or reskill workforces may be more challenging in these countries.

The Solution

The project will seek to more precisely understand the impact of AI-induced automation on developing countries and economies in transition from a social stability perspective, focusing on migration flows, crime rates and security. Relying on this, tailored support will be provided to a pilot country(-ies) to develop a roadmap of actions to mitigate potential negative impact. Mitigating the negative effects will, in turn, contribute to building trust and belief in AI as a positive agent for change. This is also in line with Sustainable Development Goals 8, 10 and 16.

The Proposed Approach

1. Identify and liaise with points of contact in a select sample of countries throughout Eastern and Southern Europe; the Caucasus; Central, Eastern, and Western Africa; Central Asia; and South-East Asia (*months* 1-3).
2. Carry out a general (country/region-neutral) AI impact assessment, focusing on migration, crime and security (*months* 1-6).
3. Identify a pilot country(-ies) for an in-depth national impact assessment (months 6-12).
4. Organise workshop(s) with national authorities and concerned entities in the selected country(-ies) to validate the results of the assessment (*month* 12).
5. Design a roadmap of actions to mitigate the impact of AI, such as awareness-raising events or the review, development or updating of policies (*months* 12-20).
6. Organise a final awareness-raising workshop(s) for policy-makers to encourage political support for the implementation the roadmap (*months* 20-24).

Timetable
The project will be implemented according to the above proposed approach over 24 months.

Current Partners
The project will be implemented by the United Nations Interregional Crime and Justice Research Institute (UNICRI) together with the Leverhulme Centre for the Future of Intelligence (CFI) at the University of Cambridge, and World Economic Forum (WEF). Other organizations and entities will also be invited to participate in project activities as implementation proceeds.

Contact
Irakli Beridze, UNICRI, irakli.beridze.un.org

Theme B.1: Dimensions of trust for beneficial AI – a cross-cultural study

The Challenge

The task of developing AI in the service of the SDGs and other global benefits requires relationships of trust of several kinds: (i) The technology and its developers and advocates must be trusted by potential users; (ii) AI technologists and governments must trust each other to work together on issues of regulation and governance; (iii) AI researchers, on the one hand, and scholars from the humanities and social sciences, on the other, must trust and respect each other's expertise, to work together on the impacts and challenges of AI. However, in all of these cases there are cross-cultural variations in the way issues of trust are perceived and discussed. These variations are potential obstacles for global cooperation for beneficial AI.

The Solution

This project aims to turn these obstacles into opportunities. It is the pilot stage for a larger multi-year cross-cultural investigation of the dimensions of trust for beneficial AI. The full study will greatly improve our understanding of the opportunities and challenges for building trust in these dimensions, and of their regional variations. It will also build cross-cultural trust and understanding between the growing networks of scholars working on the impacts of AI. The initial 12 month pilot phase will lay the foundations for the full project.

The Proposed Approach

Aim: This pilot phase project aims (a) to build a strong group of leading international experts (academics, technologists, and policy-makers) and organisations (e.g., national Academies) committed to working together on a longer project; and (b) to develop a roadmap for discussions and research on the three identified dimensions of trust (i, ii, iii), under a cross-cultural umbrella.

Method: The project will hold two scoping workshops within 12 months, one in the UK and one in China. These workshops will develop a roadmap for the longer project, and produce a report provisionally entitled *Cross-cultural Perspectives on Trust in AI – Opportunities for Impact.*

Outcomes: The outcomes of the 12 month pilot phase will be the establishment of the project team and roadmap for the multi-year project, and the report just described. The latter will also serve as a standalone document, indicating important research directions for other groups working on trust in AI, especially with a cross-cultural focus.

Timetable

- Fall 2018: Workshop 1 (provisionally Cambridge)
- Spring 2019: Workshop 2 (provisionally Beijing); draft report.

Current Partners

Department of Philosophy & the (planned) Center for Philosophy and the Future of Intelligence, Peking University, Beijing. Professor LIU Zhe.

Leverhulme Centre for the Future of Intelligence (CFI), University of Cambridge, UK. Prof Huw Price, Dame Onora O'Neill, Dr Yang Liu

Contact

Huw Price, hp331@cam.ac.uk

Theme B.2: Global AI Narratives

The Challenge

The impact of AI will be truly global, and managing it for the benefit of all will require international, cross-cultural collaboration. But different cultures see AI through very different lenses. Diverse religious, linguistic, philosophical, literary and cinematic traditions have led to diverging conceptions of intelligent machines. To build trust across cultures we must understand these different ways of seeing what AI can and should be.

The Solution

We will mobilise scholars around the world to gather and present their cultures' narratives of what AI is and the role it should play in our lives. Research on these narratives will be shared in a series of workshops, and later in widely available academic publications and public interventions.

The Proposed Approach

Aim: The AI Narratives project, launched in 2017 by CFI and the Royal Society, has begun the work of collating and analysing the AI narratives prevalent in North America and Europe. At this summit, we are seeking to build academic partnerships around the world with whom to explore how narrative traditions in other regions have shaped both popular hopes and fears for AI, and how this influenced the local development and implementation of technology.

Methods: We will bring together experts in a series of workshops, each in a different region, and disseminate their respective findings to each other and to our North American and European partners.

Outcomes: The outputs of these workshops would be collected into reports and other media, and will eventually culminate in an edited book to be published with a top academic publisher in 2021.

Timetable

- Summer 2018: workshop 1 - East Asia
- Fall 2018: workshop 2 - Africa
- Winter 2018: workshop 3 - South America
- Spring 2019: workshop 4 - South Asia
  - Each of these workshops would be followed by a report
- Early Summer 2019: Report to *UN AI For Good Global Summit 2019*
- [2021: Publish edited book on Global AI Narratives]

Current Partners

Leverhulme Centre for the Future of Intelligence (CFI), University of Cambridge, UK. Dr Kanta Dihal, Dr Stephen Cave.

The Royal Society, London, UK. Dr Claire Craig, Lindsay Taylor.
Waseda University, Tokyo, Japan. Professor Toshie Takahashi.

Contact

Dr Kanta Dihal, ksd38@cam.ac.uk

Theme B.3: Cross-national comparisons of AI development and regulation strategies – the case of autonomous vehicles

The Challenge

Successful development of AI & robotic systems that can benefit the social good in diverse cultures & countries requires many different types & instances of trust. One key challenge is for developers to be able to trust, and thereby learn from, one another's experiences & practices as they build technologies that will be deployed cross-culturally and cross-nationally. The development of that trust depends partly on understanding the relevant differences: How do different countries regulate a technology? How do different cultures engage or interact with it? This project will focus on SDG 11 (sustainable cities and communities) using the specific technology of autonomous vehicles (AVs), and more specifically the interactions between AVs and pedestrians. These interactions exhibit clear national and cultural differences (and involve some of the most vulnerable members of the urban environment) and inspire questions related to the legal and cultural constraints, guidelines, & rules, that can, or should, apply for AVs, particularly with regards to pedestrian interactions. Furthermore, how can this information be used to build inter-developer trust to speed ethical development and deployment of AVs that are suitable for China, USA, Singapore, the Netherlands, and so on?

The Solution

The end goal of this project is to propose concrete implementation strategies that facilitate trust between AVs and pedestrians nearby. To arrive at this final stage we will first conduct a systematic survey (perhaps via a conference) of different regulatory systems & cultural expectations for autonomous vehicles, with a focus on acceptable or legal interactions between AVs and pedestrians (particularly, interactions that could develop or impair trust). We will then examine potential ways to increase trust among all relevant parties, whether between developers; between the public and developers; or between the public and the AVs.

The Proposed Approach

Aim: Use autonomous vehicles as a case study to explore cross-national & cross-cultural differences in expectations and interactions when interacting with autonomous systems.
Methods: Survey of regulatory systems, "best practices," and current research for AVs. The primary work of the survey would occur through in-person workshops, with targeted follow-up.
Outcomes: A report outlining the results of the survey, including critical analysis, and identification of key challenges for developers working across national & cultural boundaries.

Timetable

- 2-3 workshops, evenly distributed across the coming months, located in cities with current partners (including possibly Pittsburgh, Singapore, and/or Beijing)

Current Partners

Carnegie Mellon University, USA. Prof David Danks

Tencent Research Institute, China. TBD

Land Transport Authority, Singapore. Wee Shann

TU Delft, Netherlands. Prof Aimee van Wynsberghe

Catelijne Muller, member EESC, President of the Permanent Study Group on Artificial Intelligence

Contact

David Danks (email: ddanks@cmu.edu)

Theme C.1: Bridging the technical-policy gap for trustworthy AI

The Challenge

Government use of AI could hugely improve public services, but only if we are able to ensure the trustworthiness of the systems being used. Assessing the potential risks and benefits of a given application of AI is challenging, particularly given that those assessing the use of AI systems in government often lack technical expertise. Without a detailed understanding of how an AI system works, governments risk either trusting the output of an AI system too much (with potentially harmful consequences), or too little (failing to make the most of AI's potential).

The Solution

We will focus on helping decision-makers to communicate better with technical developers to understand and identify potential sources of bias, error or negative consequences for a given AI system, as well as better understanding where AI can do most good in public services.

The Proposed Approach

Aims: To improve communication between decision-makers in government deploying AI systems and those with a deep technical understanding of those systems, in order to ensure governments are well-equipped to assess the trustworthiness of AI systems.
Methods: A series of workshops that bring together a combination of senior decision-makers and technical experts, to better understand: (a) what information decision-makers feel they lack/need, (b) what information it's possible for technical experts to provide, and (c) common mistakes/miscommunications between the two groups.
Outcomes: We will produce reports based on each of these workshops. A final output of this project might be a set of questions decision-makers can ask of technical experts to help them assess trustworthiness, and guidance for developers about the kinds of answers that are likely to be helpful.

Timetable

- Summer 2018: workshop 1 - government-focused: what info do governments lack/ need?
- Autumn 2018: workshop 2 - developer-focused: what info can be provided?
- Winter 2018: workshop 3 - combined: sources of miscommunication?
- Early 2019: development of framework and final report

Current Partners

Leverhulme Centre for the Future of Intelligence (CFI), University of Cambridge, UK. Dr Jess Whittlestone

Potential partners: Trustworthy Technologies Initiative (University of Cambridge), Dr. Laura James. ASI Data Science, Dr. Marc Warner.

Contact

Jess Whittlestone, jlw84@cam.ac.uk

Theme C.2: Trustworthy data: creating and curating a repository for diverse datasets

The Challenge

Key to the recent success of AI technologies has been the rapid increase in available data on which they can be trained to carry out new tasks or functions. While enabling a range of new products and services, these new uses of data also pose challenges in data management and raise questions about how to support the development of AI while managing fairness and combating bias in the implementation of AI in data.

Data reflect the society within which they were created. Therefore, even 'correct' data can be biased, as they reflect institutionalised cultural and social inequality. When such datasets are used to train AI systems, the resulting algorithms may inherit these biases. Recent years have seen a number of examples of the negative consequences of such bias, whether through image recognition systems failing for some users, recommender systems failing some potential job applicants, or concerns about predictive analytics in the criminal justice system.

Could new approaches to data sharing and curation help manage bias in data science and AI systems?

The Solution

There are already examples of repositories which maintain datasets in a way that is intended for data science use and the development of machine learning methods, such as the UCI Machine Learning archive.

A fairness-focussed data repository that maintains datasets intended for use in data science – in which data is categorised, cleaned, and appropriately labelled – could help those developing AI systems to manage bias in the data. Datasets maintained in the system would be accompanied by descriptions that explain which population the dataset is representative of, and potentially additional descriptors about bias or sensitive variables. UN partners could source geographically and topically diverse datasets.

The goal of this repository would be to deliver a set of datasets that the data science public could confidently use for training algorithms. It could also produce a methodology for 'vetting' datasets for bias.

The Proposed Approach

Aim: Create a publicly accessible repository of data for various data science & AI needs that is contextually appropriate and 'unbiased'.

Method: Curate and collect data from governments, organizations, companies for different types of data science and AI models. Develop a methodology for determining bias (data bias and societal bias).

Outcome: Online public repository.

Timetable

Curate datasets: 3-6 months

Develop framework for understanding bias and apply to datasets: 3 months

Create repository and organize data appropriately: 3 months

Current Partners

Accenture

Turing Institute

Contact

Rumman Chowdhury, Accenture, rumman.chowdhury@accenture.com

Sebastian Vollmer, University of Warwick, svollmer@turing.ac.uk

Theme C.3: Investigating cultural perspectives on fairness in algorithmic decision making

The Challenge

Algorithms are increasingly deployed in applications that affect human lives in important ways, such as making decisions related to hiring or criminal sentencing. Accordingly, there is a growing call for these systems to be fair. But notions of fairness can vary significantly across countries and cultures, and over time.

The Solution

We will provide templates to enable scholars around the world to gather and present their cultures' perspectives in a consistent way, which will facilitate direct comparison and a deeper understanding of the underlying drivers of fairness notions. Research will be shared in academic publications and presented in workshops and conferences, enabling real-world application.

The Proposed Approach

Following initial work in [1], we will develop templates to allow ourselves and partners across countries and cultures to gather a diverse mix of opinions in a standardized format. Our approach investigates opinions on the fairness of using specific features when making algorithmic decision recommendations, and extends this analysis to understand underlying latent views about feature characteristics, such as volitionality, relevance and causal influence. Results can be incorporated into real-world decision systems, as shown in [2]. In future work, the process of gathering opinions could be repeated to see how views are shifting, and whether they are moving together or apart.

[1] N. Grgic-Hlaca, E. Redmiles, K. P. Gummadi and A. Weller. Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In the Web Conference (WWW), 2018.

[2] N. Grgic-Hlaca, M. Zafar, K. P. Gummadi and A. Weller. Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning. In the Association for the Advancement of Artificial Intelligence conference (AAAI), 2018.

Timetable

- Late 2018: Workshop / conference - *Relating Fairness Perceptions to Political Views and Demographic Factors*
- Spring 2019: Workshop / conference - *Cross-Cultural Study of Fairness Perceptions*
- Early Summer 2019: Report to *UN AI For Good Global Summit 2019*
- [2020: Journal paper - *Human Perceptions of Algorithmic Fairness*, incorporating previous studies and exploring temporal effects on perceptions of fairness]

Current Partners

Max Planck Institute for Software Systems, Saarbrucken, Germany. Prof Krishna Gummadi, Nina Grgic-Hlaca.

University of Maryland, USA. Elissa Redmiles.

Leverhulme Centre for the Future of Intelligence (CFI), University of Cambridge, UK. Dr Adrian Weller.

Contact

Prof Krishna Gummadi, gummadi@mpi-sws.org, Dr Adrian Weller, aw665@cam.ac.uk

TEAM

Row 1: Toshie Takahashi, Gaenor Moore, Susan Gowans, Adrian Weller, Irakli Beridze, Becky Inkster. **Row 2:** Huw Price (with Bede), Dina Machuve, Stephen Cave, Rumman Chowdhury, Kanta Dihal, Yang Liu. **Row 3**: Jess Montgomery, David Danks, Krishna Gummadi, Kay Firth-Butterfield, Liu Zhe, Francesca Rossi. **Row 4:** Charlotte Stix, Claire Craig, Seán Ó hÉigeartaigh, Rafael Calvo, Jess Whittlestone, Ezinne Nwankwo, Laura James.

Notes

Notes

Notes