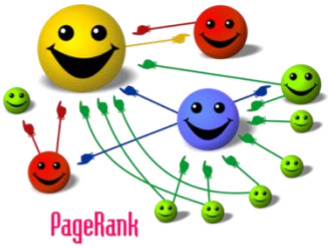
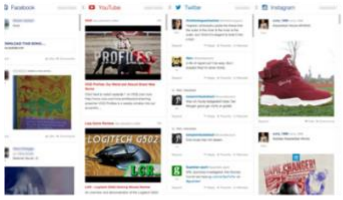
Summary report of outcomes 1st ReEnTrust Stakeholder Workshop --- Trust in Algorithmic Decision-Making





Workshop date: 15 April 2019
Report publication date: 6 November 2019

Report authors:
Jun Zhao, University of Oxford
Menisha Patel, University of Oxford
Ansgar Koene, University of Nottingham

Executive summary

On April 15th 2019 the first ReEnTrust project stakeholder engagement workshop took place at the Digital Catapult Centre in London, UK. The workshop brought together participants from academia, education, NGOs and enterprises to discuss trust in relation to algorithmic decision-making services and platforms. As the first stakeholder workshop of the series, we aimed to achieve an initial understanding from the various stakeholders regarding how trust is conceptualised by these organisation and how it is established and maintained for themselves or for their users. Following on a plenary discussion around anticipated drivers and inhibitors of trust in algorithmic online services, we used two use case studies about online shopping and online hotel booking to dive deeper into this conceptualisation of trust.

Our key findings include:

- Trust is commonly recognised as fundamental in the context of algorithmic decision-making, and the emerging big data and computational power is raising new challenges to our existing notion of trust.
-
- Trust in relation to algorithms needs to be considered in a broader context by considering the various other elements constitute the Internet and the digital services/systems that algorithms operate on.

- Consistent with literature from sociology and psychology, our stakeholders have pointed out that trust is an emotional and psychological state of an individual. Like most emotional states, trust can hardly be expressed as a binary yes or no, but often as a matter of degree.

- The emotional state of feeling trust is closely related to a subjective perception of the trustworthiness of a piece of information or platform. This trustworthiness is a multifaceted concept with many components, such as transparency, privacy, reliability, reputation, appearance and behaviour of the website, all filtered through the user's previous experience.

- The gap of digital literacy observed in most Internet users could make their recognition and establishment of trust in relation to algorithmic decision-making much more challenging, and it is a critical issue that we need to bear in mind when designing trust-enabling technologies.

These findings are critical to our ReEnTrust project, to recognise the complex socio-economic context in relation to algorithmic decision-making, when designing new technology solutions to facilitate users' establishment of trust. In the next stakeholder workshop, we would like to inspire more focused discussions with stakeholders regarding the role of different trust components in helping them negotiate trust in algorithms, and how technologies may help them mediate these different components more effectively. This would give us deeper insights regarding how a responsible trust (re-)building digital system should look like.

Introduction

This report summarises the outcomes of the first ReEnTrust project stakeholder engagement workshop that was held at the Digital Catapult Centre in London, UK, on April 15th, 2019. These workshops aim to bring together individuals from a range of professional backgrounds who are likely to have differing perspectives on issues related to ***trust in algorithmic decision-making***. The workshops are opportunities to share perspectives and seek answers to key project questions such as:

- How do commercial and non-commercial organisations establish and maintain user trust in the algorithmic systems they use?

- How do these organisations themselves establish trust in their own systems?

- How is trust conceptualised by these organisations?

The ReEnTrust project strongly emphasises co-creation in the design of the technologies and policy recommendation we will develop. The stakeholder workshops are one format for us to achieve this and we aim to host three such workshops during the project. This workshop is the first of the series and is aimed to explore stakeholders' perspectives about how trust is formed or lost by users when interacting with existing platforms involving algorithmic decision-making (such as hotel booking, loan applications, digital policing, etc) and what factors would help users to negotiate or establish trust.

For more information on the workshops, see also the call for participation in Annex 1.

Methodology

To facilitate an open discussion all stakeholder engagement activities are run under Chatham House rule – meaning that views expressed can be reported back elsewhere but that individual names and affiliations cannot, unless explicit consent is given. The outcomes described in this report combine the data obtained from the audio recordings and the notes made during the discussions. Prior to publication, the report was circulated for approval and/or amendment by all participants.

Since this was the first of our workshops, taking place at an early stage of the project, the primary aim was to gain a broader understanding of the range of issues and perspectives that the various stakeholders bring to the topic of Trust regarding algorithmically mediated online services. To this end, the discussion was anchored by two case studies (see Annex 2): online shopping and holiday booking.

The workshop started with an opening presentation about the ReEnTrust project and the goal of the stakeholder workshop. Then the floor was open to all participants, for a short group discussion about their understanding of trust and experience of trust in algorithmic services in their professional settings. We then had a plenary discussion regarding the following two questions:

1. What do you think about trust in algorithmic systems?
2. What is your experience of trust in algorithmic services in your professional settings?

After this, we took a short break and the room was broken up into two focus groups, to discuss the two case studies. Finally, the workshop concluded by a plenary reflection of their perception of algorithmic trust and suggestions on how to possibly mitigate these issues.

The workshop was attended by 9 participants (+ 3 facilitators) representing three SMEs/industry, two consultancy agencies, two academic institutions/groups, one NGOs/not-for-profits, and one governmental regulatory agency.

Background

The ReEnTrust project

Interaction with online Web-based platforms is becoming an essential part of people's everyday lives. This allows data-driven computer algorithms to exert a massive influence on society, which can be the search results we receive, news we read, products we are recommended, or bank loans we are permitted with. There is growing attention regarding how the computational algorithms are being designed and used in a way that may introduce biases or unfair decisions upon us, or manipulate the information we consume. This has led to a serious breakdown of trust in our society --- users do not know when to trust the outcomes of algorithmic processes and, consequently, the platforms that use them. As trust is a key component of the Digital Economy where algorithmic decisions affect citizens' everyday lives, this is a significant issue that requires addressing.

The ReEnTrust project (https://reentrust.org/) aims to address this issue of 'trust in algorithms, artificial intelligence and automation' by seeking to identify ***fundamental principles for algorithm design*** and ***technological tools*** informed by stakeholder engagement for better understanding of the core principles of the user-platform trust dynamic and providing innovative socio-technical solutions to support it.

Focusing on AI algorithms and large scale platforms used by the general public, our research questions include: What are user expectations and requirements regarding the (re)building of trust in algorithmic systems? Is it possible to create technological solutions that (re)build trust by embedding values in recommendation, prediction, and information filtering algorithms and allowing for a productive debate

on algorithm design between all stakeholders? To what extent can user trust be (re)gained through technological solutions and what further trust (re)building mechanisms might be necessary and appropriate, including policy, regulation, and education?

Stakeholder engagement activities

As part of an RRI (Responsible Research and Innovation) approach, we will carry out interviews, surveys and workshops with stakeholders to gain stakeholder perspectives on trust breakdown and regain in relation to algorithm-driven online platforms. We invite stakeholders from academia, education, government/regulatory oversight organisations, civil society organisations, media, industry and entrepreneurs to join us in exploring the requirements for trust rebuilding and identifying future best practices for responsible and trustworthy innovations. The activities will take place over two years (2018 – 2020) and will seek to identify relevant perspectives and concerns as well as provide feedback on our project activities. They will also produce:

- Policy guidelines for engendering trust in the design, development, and use of algorithms.
- Evidence and documentation describing the RRI methodology developed through ReEnTrust work.
- A portfolio of real-world case studies to inform responsible practice.

Key Findings: Perception of Trust

Through this section, we outline key emergent themes from our opening group discussion concerning the conceptualisation of trust, and participants' experiences of trust in relation to technologies.

We summarise our highly insightful multi-stakeholder discussion into three key areas of considerations:

1. **Broad trust-related issues**: comments in relation to trust and the technological context in which it is situated;
2. **Unpacking of trust:** reflections on what is key if we are to conceptualise trust in relation to the use of algorithms/technologies;
3. **Reflections on matters compounding the breakdown of trust:** issues which stakeholders felt may undermine trust in algorithms/technologies.

We now consider each of these key areas in turn. It should be noted that whilst we did frame the workshop in relation to the use of algorithms, there were times where stakeholders considered trust in relation to technologies in a more general sense.

Broad trust-related Issues

In this subsection, we outline key themes which emerged in relation to trust and the technological context in which it is situated.

Trust is important and fundamental in the context of algorithmic decision-making

Our stakeholder group was drawn from a different array of professional backgrounds, and often had nuanced perspectives on how we can conceptualise trust etc. However, what remained clear and aligned between them is that trust was seen as both an important and fundamental consideration in this context of algorithmic decision-making. This was especially reflected in the terminology used in the discussion. For example, ways in which trust was articulated include being considered as a "massive deal" [p2] or "really important" [p1].

Trust is embedded across multiple layers of the Internet *including* in relation to algorithms

Although our plenary discussion was initialised around trust in algorithms, stakeholders would often refer to trust being situated across multiple layers of an online *ecosystem,* rather than an isolated consideration focused on algorithms. "Ecosystem" has been used to refer to the interdependency of different aspects of the Internet, for example, "the technical stuff", "an app", "the TCP IP" and "algorithms" [p5]. There were suggestions of the need to look at the different parts of the *ecosystem* when we consider trust in relation to algorithms. It was also suggested that the end users may have more trust in some layers of the ecosystem than others, such as more trust in "apps" than in the "higher levels of the network" [p4].

The emergence of the new technological context of algorithms raises new thinking about trust

Through discussion, participants would point out that of course, the notion of trust in itself and in relation to technologies is not new, despite being very relevant to the use of algorithms. For example, a participant mentioned that issues of trust already exist in relation to big data, and additionally values like privacy and security are "basics that people expect" [P2] irrespective of algorithmic decision-making. However, what emerged is that algorithms introduce new trust-related problems, and that these problems could be extensions of/exacerbate existing issues that are already prevalent in the digital space. This is in particular given the scalability and extent of data used; the keenness of different industries to embrace it; and the differential knowledge held within industries and varying levels of regulation in relation to its (potential) use.

Unpacking of trust

During the session, we were also able to understand the different ways in which participants conceived of trust as a concept. These views were often framed in regard to technologies in a more general sense than algorithms themselves. Though we were unable to probe these perceptions in-depth, we can at least begin to unpack how trust is considered by different stakeholders; and what key dimensions may constitute trust.

Trust is an emotional relationship between users and technologies

Trust was alluded to as akin to established notions surrounding human-human relationships such as "emotional safety" [p1]. It was suggested that similar emotion-oriented considerations, though not "exactly the same" [p1] could be translated to the human-technology relationships. It was pointed out that this kind of translation is already present if we consider how characteristics we attribute to people we trust may also inform brand loyalty [p9].

Trust emerges from consistency and reliability in the behaviour of technologies

It was suggested that it is important for technologies to demonstrate "consistency of behaviour" [p1]- similar to our preferences for predictability related comfort in our human interaction. Ensuring that such expected behaviour can be maintained and that technologies are reliable- making sure that- for example, "the computers and robots don't go crazy" [p2]- is an important feature of trust.

Transparency is important to maintain trust in technologies

The visibility of "what these things are doing" and "how are you making decisions' [p2] was seen as an important contributor to trust in technologies. An algorithm-specific example of understanding the outcome of a mortgage application was raised by a participant, where the "computer says no" [p2], rejecting an application. The participant relayed how it would be important for the recipient of this decision to understand the decision-making rationale underlying this algorithmic decision-making outcome.

Fairness is important in engendering trust in algorithms

There was a recognition of some of the biases and potentially serious discrimination that may occur when algorithms are embedded in decision-making processes. Thus, fairness to ensure non-discrimination in algorithmic decision-making was seen as important for engendering trust. For instance, the example of a service provider being able to demonstrate that a decision "is fair across its customer base" [p2] was used to illustrate the importance of this point.

The perception of trust should be extended to trustworthiness

The differentiation between the notions of "trust" and "trustworthiness" [p4] emerged during the workshop. It was raised that for most organisations or even governments, gaining and sustaining public trust is an essential component of their successful operation. However, it may be that the underlying intentions of eliciting such trust are not necessarily conducive to the public good. Instead, being *trustworthy* (seen as: being worthy of trust) was pertained to be a more appropriate attribution, especially when we consider the complexity and pervasiveness of technologies, such as the use of algorithms. It was felt that in this framing of "trustworthy" we are ensuring that there are positive intentions underlying the use of algorithms.

Trustworthiness is a multi-dimensional concept

There was an awareness that trustworthiness is in itself is a multi-faceted and overarching concept, constituted of different components. Components such as "transparency and scrutiny", "robustness and resilience" and "privacy and liability" [p9] were identified.

Reflections on matters compounding a breakdown of trust

During our discussion participants also deliberated more deeply on the issues that may undermine trust in algorithms/technologies.

The underlying technology infrastructure and ecosystem may contribute to a trust breakdown

There were concerns from participants over the collective harnessing of technologies and processes which have been developed distinctly in their own right. It was suggested that since the development of the multiple layers and aspects comprising the Internet has been fragmented over time, unanticipated problems "which weren't originally envisaged" in relation to individual parts [p5] may emerge when these varying components come together. The emergence of potentially problematic and unforeseen issues was seen to likely impact the trust users have in technologies.

There has been a lack of responsible innovation

It was suggested that trust has been undermined given that there has been a lack of foresight and responsibility in the ongoing development of the Internet. This has allowed potentially harmful innovations to emerge without adequate consideration of the end-users and their safety online. This point was articulated as the Internet and its application space "running as a research space in public" [p5]. The implications of the way things have so far developed were seen as potentially dangerous on a global level.

Users have a lack of awareness about the implications of technologies

The impact of digital literacy on trust emerged during the discussion. It was suggested that most people are unable to understand what all the aspects of the "ecosystem" are- including more technical aspects such as algorithms [p4]. There were concerns surrounding these varying levels of digital literacy, and

how they may have an impact on user's abilities to scrutinise their online space and the use of algorithms in order to determine their trust in these processes. Instead of being able to critique processes algorithms may seem like "magic" [p4] and people may not understand "their internet is not the same as the internet of their friends, or of their colleagues" [p4]. Such exemplars were used to convey the quite problematic implications of a lack of user awareness in relation to a fair consideration of the trustworthiness of algorithms.

An open challenge raised to address this gap in awareness was to find ways to increase digital literacy, so people are able to decide in an informed manner if they trust what they experience online.

Case Study Discussions

Case Study: Online Shopping

Here are key points from the online shopping case study (see appendix 2) discussions. The statements below are not direct quotes.

1. Reviews and trust

- If people can't trust whether reviews are genuine, then the values of these reviews are questionable.
- Platforms create an illusion that reviews are normalised, independent, or statistically valid, but they are really not.
- At the same time, people (consumers) are often misguided and believe that these reviews are independent assessments. People often do not understand how these reviews are generated, and would misplace trust on a '5-star' review.
- Review stimulation takes place after shopping, such as consumers receiving an email prompting them to leave a review and consumers often leave reviews if they liked or terribly disliked the products. This, as well as fake reviews, both contribute to the trust issues of online reviews.
- The fundamental question is: when people lose trust, is it the platform or the reviews?

2. Personal issues with trust in algorithmic systems from the stakeholders

- When using a platform with known flaws or issues, additional elements can be useful to help us be critical, such as reviews from accredited people, or being extra careful with promotions.
- People often read between the lines of reviews, just like how they would treat opinions from friends, who may not exactly be an expert of a product but provides their opinions nonetheless.
- The stake in the decisions can lead to different decision-making processes. People would do more reviews for items of higher value and also be more likely to go to 'traditional' big brands for such items.

- There are lots of hidden promotions online and their relationships with big tech or big search engines are not always clear. Online shopping algorithms are not just filtering for you, but 'promoting' to you.
-

3. Struggles that current consumers face

- People of lower educational level or people of the older generation would particularly struggle to 'read between the lines' or understand how the technology works
- Understanding and distrust of algorithms does not impact our trust of platforms; some of us do not trust the platforms from the beginning.

4. Issues recognised by stakeholders

- It's not just algorithms, but the data sources going into the algorithms that may have caused the problems.
- Big brands seem not to enforce any quality reviews of contributed content (such as users' reviews)
- Who should take the regulatory role and who should be regulated is unclear and can be challenging.

5. The proposed focus for future

- Focus on enhancing uses' ability to establish the trustworthiness of reviews or information online.
- Focus on education and users' digital literacy development.
- Better regulations are required, but unclear how best to achieve this.

Case Study: Holiday Bookings

Existing displays of our lack of trust in relation to online holiday booking:

- The fact that we compare sites to check information: this seems to be a common mechanism which we adopt- and shows that we do have concerns about the information we see on holiday booking sites;
- The existence and use of consumer reviews to try and validate information about hotels- shows that there is a concern amongst users about trust;
- The fact that many interactions with hotels are now mediated by platforms rather than talking to hotels themselves – this has created the problem of trust because there is a layer of asymmetry;
- There is a growing awareness of pressure selling techniques- participants discussed their own experiences of these with one another.

Concerns about the credibility of information on online sites:

- There was concern that information regarding the availability and quality of hotels was not entirely true and undermined by pressure selling techniques: such as scarcity notifications- e.g. “last one remaining”; or the misleading notifications such as “50 people looking at this hotel” which may not necessarily be true at all- or may not represent people looking to stay in a hotel for the same time-period;
- This causes a lack of trust given an asymmetry of information on the consumer side- it was pointed out that hotels as well as platforms may be responsible for some of the misinformation (such as paying people for false reviews; or paying extra to game the algorithms and appear higher in rankings).
- There were concerns that reviews, though instantiated to help verify information- could be questionable: such as payments by hotels for fake reviews; incentives such as free rooms to influencers to provide good reviews; or only collecting good reviews- such as offering only passengers who have been on a timely flight the opportunity to review a flight booking platform.
- Profiling of users was also a concern: the potential of information being altered according to a user’s location or perceived wealth (according to the device they were using to browse) were raised.

A breakdown in trust would not necessarily lead to a site not being used:

- There were a number of reasons raised as to why there would be a continued use of a site despite awareness and experience of pressure selling, and a related lack of trust:
- “They all do it”- given the commonality of such pressure-selling or other mechanisms used by all platforms and hotels to sell their services- then there would be no reason to look elsewhere;
- Convenience was seen as a key driver- particularly for people who are busy and travel for work or pleasure- and also in terms of going through a process of going through the setting up of another account on another platform- entering card details etc.
- The opportunity versus cost: if consumers feel they are getting a good deal, then this might outweigh their using another site.
- It was also suggested that we may have varying levels of trust- given that we do provide these sites with highly sensitive information such as our bank details and personal information must mean there is some extent of extant trust;
- The value of the product being used to the user was also seen as a feature in whether they would use a site deemed potentially untrustworthy. Goods/services perceived to be of lower consequence or value to the user were more likely to be purchased without an in-depth consideration of the mechanisms of the site.

Solutions to trust in online holiday booking platforms

- Education of users was seen as a good way to raise awareness regarding the practices of holiday booking platforms. It is seen as a means to overcome difficulties in regulating and enforcing unfair selling practices and information representation;
- Indeed, enforcing regulation on global sites was seen as hugely challenging and unlikely- particularly given the proprietary nature of commercial algorithms;

- Consumer-driven “protective” sites were also seen as a way in which information about some of the problematic elements of comparison sites could be raised for a more informed consumer base. It was suggested that there are already some sites that attempt to do this.

Appendix 1. Call for Participation

REENTRUST
Rebuilding and Enhancing Trust in Algorithms

1st ReEnTrust Stakeholders workshop

Date/Time
Date(s) - 15/04/2019
12:00 pm - 4:30 pm

Location
Digital Catapult

First ReEnTrust stakeholders workshop, focusing on the topic of trust in algorithmic mediated online services; policy and practice, technical matters, and understandings.

During this workshop we will start by providing an introduction into the ways in which the ReEnTrust project is investigating trust in algorithmic services and possible tools for trust mediation. We will then hand out a number of case studies on loss of trust. The intent is for these to serve as discussion starters to that we can hear from you about your concerns regarding trust in algorithmic services and your thoughts on methods how to manage this trust.

We are inviting participants from a wide range of disciplines and sectors in order to gain a broad perspective on the issues and to hopefully provide you with a valuable opportunity to gain fresh insights. Participants will come from:

- **academia** from multiple disciplines (e.g. Engineering, Computer Science, Internet Research, Law, Business Schools, Schools of Education and Science Technology & Society studies)
- **public sector** organizations (e.g. European Commission, Nottingham City Council)
- **private sector** organizations (e.g. Bibblio, Samsung and Amazon)
- **third sector** organizations (e.g. Doteveryone, 5Rights, Mental Health Foundation, Polka Theatre)

Where: The workshops will take place at the Digital Catapult in London (101 Euston Road, NW1 2RA, London).

When: April 15th, 2019 from 12:15am to 5pm.

Workshop schedule:

- 12:00 – 12:45 Lunch
- 12:45 – 13:00 Welcome and Introduction to the Workshop
- 13:00 – 13:20 Introduction to ReEnTrust Project & Case Study Task
- 13:20 – 14:20 Case Study Group Work
- 14:20 – 14:30 Break (Tea & Coffee)
- 14:30 – 15:30 Report Back & Group Discussion
- 15:30 – 16:15 Plenary Discussion
- 16:15 – 16:30 Summary

Privacy/confidentiality and data protection

All the workshops will be audio recorded and transcribed. This is in order to facilitate our analysis and ensure that we capture all the detail of what is discussed. We will remove or pseudonymise the names of participating individuals and organisations as well as other potentially identifying details. We will not reveal the identities of any participants (except at the workshops themselves) unless we are given explicit permission to do so. We will also ask all participants to observe the Chatham House rule – meaning that views expressed can be reported back elsewhere but that individual names and affiliations cannot.

A short report on the outcomes of the workshop will be circulate among all the ReEnTrust stakeholder panel members (including those that could not attend the workshop) for comments, feedback and corrections.

Appendix 2. Case Studies

Case Study 1: Online shopping

Which? is an independent consumer organisation in the UK[1]. It has recently been investigating customer reviews posted on popular shopping platforms such as Amazon.com. Which? estimates that up to 97% of adults use online reviews when researching a product to buy. It also cites estimates from the Competition and Markets Authority that £23bn a year of UK consumer spending is influenced by online customer reviews.

October 2018 Which? report evidence of wide scale posting of fake reviews on online shopping sites[2]. These are reviews which do not represent the genuine views of the poster or refer to a genuine purchasing experience.

Which? investigators identified several fake review 'factories' on Facebook – called Amazon Deals Group and Amazon UK Reviewers amongst other names. On these groups sellers offer their products for free or at a discount in return for positive reviews online. In a typical process, a group member buys the item through Amazon – thereby ensuring it registers as a verified purchase – and posts a positive review on the site with, for instance, a five-star rating. If the sellers are satisfied with the review, they send the group member a refund for the item.

November 2018 Which? report that in the past 2 years at least 30,000 customer reviews have been removed from Amazon[3]. They had been posted in 16 popular categories of technology and home products, mostly in relation to unknown brands. Which suggest the reviews have likely been removed as they were fake. Which? advises consumers to be wary of: 1) buying certain categories of products such as headphones and cameras; 2) buying from unknown brands and 3) relying solely on customer ratings, without looking into the text of reviews more deeply.

January 2019 Which? report a further practice that might manipulate product reviews[4]. Amazon allows sellers to group together similar products – for instance that only differ in terms of colour or size – so that the same reviews for them appear across the whole range. Adding in artificial variations can inflate the number of positive reviews and removing certain variations can hide negative reviews.

1 https://www.which.co.uk/
2 https://www.which.co.uk/news/2018/10/the-facts-about-fake-reviews/
3 https://www.which.co.uk/news/2018/11/thousands-of-fake-customer-reviews-removed-from-amazon/
4 https://www.which.co.uk/news/2019/01/which-uncovers-fake-five-star-reviews-flooding-amazon/

Which? also provide consumer advice as part of their mandate. Their website includes a page on how to spot fake reviews[5]. Their tips are:

1) Inspect the comments in reviews not just the star ratings
2) Look for suspicious language that appears more like an advertisement than genuine opinion
3) Check the dates of reviews – as fake reviews of a product are often posted in batches at the same time
4) Look at the posters' other reviews – if they post a large number of 5 star reviews, they are likely being paid to do so.
5) Pay attention to mid-range reviews – as they are more likely to be honest
6) Watch out for poster 'bias' – such as if they say they have bought the product but not actually used it yet.
7) Check the purchase has been verified
8) Check what else the reviewer has bought (if that information is available) – is there a consistent pattern?
9) Use an online tool to help check for fake reviews. These include Fakespot[6] and ReviewMeta[7]

Questions for discussion

1. To what extent does the presence of fake reviews on online shopping sites create problems that might undermine consumer trust?

2. How might awareness of fake reviews undermine consumer trust and what consequences might this have?

3. In your own experience, have you come across similar issues in the algorithmic systems, which may have undermined end users' trust? How?

4. How useful are the tips for spotting fake reviews suggested by Which? Are there some further tips consumers could follow?

5. Following their investigation, Which? received the following response from Amazon. To what extent is this response satisfactory?

> *Our goal is to make reviews as useful as possible for customers. We do not permit reviews in exchange for compensation of any kind, including payment. Customers and sellers must follow our review guidelines, and those that don't will be subject to action including potential termination of their account'.*

6. Writing fake reviews under a false identity is a crime under Italian law and can lead to prison sentences[8]. Is this an appropriate way to deal with the problem? What kinds of mechanism do you think could enhance users' trust in reviews in online shopping systems

5 https://www.which.co.uk/consumer-rights/advice/how-to-spot-a-fake-review
6 https://www.fakespot.com/
7 https://reviewmeta.com/
8 https://www.theguardian.com/world/2018/sep/12/man-jailed-italy-selling-fake-tripadvisor-reviews-promo-salento

Case Study 2: Hotel booking sites

October 2017 The Competition and Markets Authority, which promotes competition for the benefit of UK consumers, opens an investigation[9] into online hotel booking sites. These are sites such as booking.com etc. that allow customers to compare prices across a number of hotels. The CMA notes that up to 70% of people booking hotels use these sites and their investigation follows concerns that some of their practices might breach consumer law[10]. The CMA announcement states:

> *The CMA is concerned about the clarity, accuracy and presentation of information on sites, which could mislead people, stop them finding the best deal and potentially break consumer law.*
>
> *Its investigation will examine several practices, including:*
>
> - *Search results: how hotels are ranked after a customer has entered their search requirements, for example to what extent search results are influenced by other factors that may be less relevant to the customer's requirements, such as the amount of commission a hotel pays the site.*
> - *Pressure selling: whether claims about how many people are looking at the same room, how many rooms may be left, or how long a price is available, create a false impression of room availability or rush customers into making a booking decision.*
> - *Discount claims: whether the discount claims made on sites offer a fair comparison for customers – for example, the claim could be based on a higher price that was only available for a brief period, or not relevant to the customer's search criteria, for example comparing a higher weekend room rate with the weekday rate for which the customer has searched.*
> - *Hidden charges: the extent to which sites include all costs in the price they first show customers or whether people are later faced with unexpected fees, such as taxes or booking fees.*

The CMA announces it has written to companies across the sector to understand more about their practices. It also calls for customers who use the sites and hotels that advertise with them to share their experiences.

June 2018 The CMA launches enforcement action against a number of hotel booking sites[11]

> *The CMA will be requiring the sites to take action to address its concerns, where they are believed to be breaking consumer protection law. It can either secure legally binding commitments from those involved to change their business practices or, if necessary, take them to court.*

9 https://www.gov.uk/cma-cases/online-hotel-booking
10 https://www.gov.uk/cma-cases/digital-comparison-tools-market-study
11 https://www.gov.uk/government/news/cma-launches-enforcement-action-against-hotel-booking-sites

February 2019 The CMA announces that six of the sites subject to enforcement action have voluntarily agreed to make changes[12]. The sites are: Booking.com, Agoda, Hotels.com, ebookers, and trivago. The agreed changes are:

> *Search results: making it clearer how hotels are ranked after a customer has entered their search requirements, for example telling people when search results have been affected by the amount of commission a hotel pays the site.*
>
> *Pressure selling: not giving a false impression of the availability or popularity of a hotel or rushing customers into making a booking decision based on incomplete information. For example, when highlighting that other customers are looking at the same hotel as you, making it clear they may be searching for different dates. The CMA also saw examples of some sites strategically placing sold out hotels within search results to put pressure on people to book more quickly. Sites have now committed not to do this.*
>
> *Discount claims: being clearer about discounts and only promoting deals that are actually available at that time. Examples of misleading discount claims may include comparisons with a higher price that was not relevant to the customer's search criteria. For example, some sites were comparing a higher weekend room rate with a weekday rate or comparing the price of a luxury suite with a standard room.*
>
> *Hidden charges: displaying all compulsory charges such as taxes, booking or resort fees in the headline price. Sites can still break that price down, but the total amount the customer has to pay should always be shown upfront.*

All changes will need to be implemented by September 2019 and the CMA will monitor compliance. The CMA also states that it will write to other hotel booking sites to set out clear expectations for how they should be complying with consumer protection law.

Questions for discussion

7. To what extent do you think the four practices listed by the CMA could have created problems that might undermine consumers' trust on these algorithm-based systems?

8. How might awareness of these practices undermine consumer trust and what consequences do you think this might this have?

9. Have you come across similar issues in the algorithmic systems commonly used in your professional settings, which may have undermined end users' trust?

10. Do you think enforcement action is the only mechanism that exists to prevent or limit these practices? If you were given the opportunity to resolve the issues discussed above, what other mechanisms, if any, would you put in place?

12 https://www.gov.uk/government/news/hotel-booking-sites-to-make-major-changes-after-cma-probe

Appendix 3: Information Sheet

INFORMATION SHEET FOR PARTICIPANTS

REENTRUST STAKEHOLDER ENGAGEMENT WORKSHOPS

[CUREC reference number CS_C1A_19_015]

You are being invited to take part in a research study. Participation is open to all members of the ReEnTrust project stakeholder panel. This sheet gives you information about the study to help you decide if you would like to take part.

We would like to invite you to contribute to our ongoing ReEnTrust research study by attending one of more of our stakeholder engagement workshops. These workshops involve professionals from various fields relevant to the project. They are discussion-based sessions in which participants consider issues around trust between users and algorithms/systems. The workshops provide an opportunity for stakeholders to put forward their perspectives and shape our project outcomes as we seek to identify mechanisms that can foster trust between users and algorithms/systems.

The workshops will be audio-recorded and transcribed. They might also incorporate short pre and post workshop questionnaires. We intend to quote extracts from these data in our project publications and outputs. However, we will take great care to anonymise all our data and store it carefully.

If you are interested in taking part please read the rest of this information sheet, which provides further details of the study. You can also speak to a member of the research team if you have any questions. You can take your time to decide whether you would like to take part in the study.

What is the ReEnTrust project about?

Algorithmic systems and automated processes for decision-making are now prevalent on online platforms and across our lives. As interaction on online Web-based platforms is becoming an essential part of people's everyday lives and data-driven AI algorithms are starting to exert a massive influence on society, we are experiencing significant tensions in user perspectives regarding how these algorithms are used on

the Web. These tensions result in a breakdown of trust: users do not know when to trust the outcomes of algorithmic processes and, consequently, the platforms that use them. Our project explores new technological opportunities for platforms to regain user trust and aims to identify how this may be achieved in ways that are user-driven and responsible.

The ReEnTrust project is developing experimental online tools that allow users to evaluate and critique algorithms used by online platforms. Alongside this we are conducting extensive engagement with users, stakeholders, and platform service providers to produce an improved understanding of what makes AI algorithms trustable. We will also develop policy recommendations and requirements for technological solutions, plus assessment criteria for the inclusion of trust relationships in the development of algorithmically mediated systems and a methodology for deriving a "trust index" for online platforms that allows users to assess the trustability of platforms easily.

What are the stakeholder engagement workshops about?
The ReEnTrust stakeholder panel forms a central part of our project engagement. Relevant professionals from academia, education, government/regulatory oversight organisations, civil society organisations, media, industry and entrepreneurs have joined our stakeholder panel to receive information about our activities. We now invite all our panellists to take part in a series of stakeholder workshops. In these meetings panellists will share their perspectives and experiences in order to accomplish productive and wide-ranging discussions on our project themes. Four workshops will take place over the timespan of the project and will centre on discussion-based activities that cover topics such as: controversies arising from the contemporary prevalence of algorithmic systems online and offline; instances of trust breakdown between users and systems; algorithmic governance; and mechanisms to foster trust between users, algorithms and systems. These workshops will play a central role in helping us to fulfil our project aims.

Why have I been chosen?
Participation is open to all members of the ReEnTrust stakeholder panel.

Do I have to take part?
No. Your participation is totally voluntary. You can also choose to attend as many or as few workshops as you prefer.

If you do take part in a workshop, you can withdraw during from participation at any point and you do not have to give a reason for doing so. You can also withdraw after a workshop – meaning that we will remove all data relating to you – for up to six weeks after each workshop. If you withdraw, your contributions to the sessions will not be transcribed and will not be used as data. Any questionnaire responses will be destroyed. No reference will be made to any information you provided in connection to the workshop when we report our study findings.

What will happen to me if I take part? If you agree to take part, you will be invited to attend up to four stakeholder workshops. These will take place in easy-to-reach locations – most often central London. It may also be possible to attend via teleconference. You will be asked to participate in activities that might include some or all of the following:

- completion of a short pre workshop questionnaire
- small group and plenary group discussions on contemporary controversies involving the use of algorithms, algorithmic governance, and trust
- small group and plenary group discussions of case studies of the breakdown of trust between users and algorithmic systems
- provision of feedback on ReEnTrust project activities, including comments on designs for our online tools and policy guidelines.
- completion of a short post workshop questionnaire

The first set of workshops will take place in April 2019, and will then run at 4-6 month intervals until the end of the project in late 2020. Each workshop will be a half-day event, typically taking place over three to four hours. Refreshments will be provided. There will also be opportunities for participants to network with other attendees.

The workshops will be audio recorded and transcribed. If participants agree, we will also take photos during the workshops to place on our project website.

What are the benefits and risks of the study?

We hope that our participants will find the workshops interesting and enjoy having an opportunity to put forward their professional perspectives in order to shape our project outputs. The workshops also provide an opportunity for participants to network with others in relevant fields. Being involved poses no specific risks other than the risks involved in undertaking ordinary everyday activities. The discussions will not cover personal or highly sensitive issues so participants are also unlikely to find them uncomfortable. We also understand that participants cannot always reveal certain information about their working practices, employers etc. Participants are welcome to decline to talk about the specifics of their own work as they feel is appropriate.

No formal preparation will be necessary for the workshops and they are designed to avoid placing a heavy time burden on participants. But in case attendees would like time to think through the issues involved beforehand, we will email round the main topics to be discussed at least a week before each event takes place.

Privacy/confidentiality and data protection

We are highly sensitive to the need to maintain strict procedures around the privacy and anonymity of our research participants. Unless you give us specific permission to do otherwise, we will not tell anyone beyond the ReEnTrust project team and those attending on the day that you have taken part. We will also ask all participants to observe the Chatham House rule during the workshop – meaning that views expressed can be reported back elsewhere but that individual names and affiliations cannot.

All the workshops will be audio-recorded and transcribed. This in order to facilitate our analysis and ensure that we capture all the detail of what is discussed. The transcripts will be prepared by members of the project team or a specialist transcription service. We will keep a 'master copy' of the full transcript

including participant names etc. but for day-to-day analysis we will use anonymised versions. We will assign each participant a study number and pseudonym and use these instead of real names on the anonymised transcripts. We will also remove or pseudonymise the names of participants' organisations as well as other potentially identifying details in the transcripts. We will link questionnaire responses to participants so that we can connect them to comments made during the workshops. Similarly, we will keep master copies of questionnaire responses but anonymise all other copies by using participant study numbers and pseudonyms. If participants provide explicit consent, we will take photographs during the workshops and place them on our website. We will use extracts from the workshop transcripts and questionnaire responses in our project journal articles, conference presentations and other outputs; these will all be anonymised.

All data and project documentation will be kept securely. Electronic data will be password protected and securely stored. It will only be shared with members of the project team and we will use secure mechanisms for doing so. Printed information, including signed consent forms, will be stored in secure, locked, university spaces. The University of Oxford is the data controller with respect to your personal data and all your data collected in the study will be handled in accordance with their their requirements – these incorporate your rights under the General Data Protection Regulation.

Once the project and project write-up is complete all backup copies of recordings and other data will be destroyed. However due to University of Oxford regulations and EPSRC funding requirements we will store the original data securely for a minimum of 10 further years after publication. Our funding body, the EPSRC, is committed to the responsible sharing of anonymised research data amongst university research professionals. For this reason, anonymised data may be placed within a data repository that can be accessed by other researchers. Access is only possible under strict conditions that ensure good ethical practice. No personal details will be placed within the repository. It is possible for you to consent to take part in the study but refuse permission for your data to be placed in a repository.

What will happen to the results of the study?

As part of this project we will produce publications in academic journals and University publications (including webpages) and will present our findings at conferences and other events. These activities might include the use of short data excerpts such as transcript extracts and questionnaire responses. All data excerpts will be anonymised in the ways described above.

Who is organising and paying for the study?

The ReEnTrust project is funded by the Engineering and Physical Sciences Research Council. It is a collaboration between the Universities of Oxford, Nottingham and Edinburgh. The project is led by Professor Marina Jirotka in the Department of Computer Science at the University of Oxford. EPSRC reference EP/R03363/1.

Who has reviewed the study?

This project has been reviewed by, and received ethics clearance through, the University of Oxford Computer Science Departmental Research Ethics Committee.

Who should I contact for further information?
If you have any questions or require more information about this study, please contact project researcher Jun Zhao.

Dr Jun Zhao
Department of Computer Science
University of Oxford
Parks Road
Oxford
OX1 3QD

01865 273875
jun.zhao@cs.ox.ac.uk

Who do I contact if I have a concern about the study or I wish to complain?
If you have a concern about any aspect of this project, please speak to our project leader at the University of Oxford:

- Professor Marina Jirotka marina.jirotka@cs.ox.ac.uk or 01865 270550.

Professor Jirotka will do her best to answer your query. Your concern should be acknowledged within 10 working days and you will be given an indication of how she intends to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the chair of the Research Ethics Committee at the University of Oxford (using the contact details below) who will seek to resolve the matter in a reasonably expeditious manner:

Chair
University of Oxford Computer Science Departmental Research Ethics Committee.
Department of Computer Science
University of Oxford
Parks Road
Oxford
OX1 3QD

ethics@cs.ox.ac.uk

Thank you for reading this information sheet and for considering taking part in this research. Please feel free to ask any questions if you would like further details of the study.