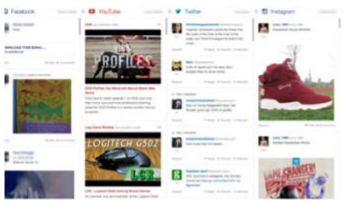
Summary report of outcomes 2nd ReEnTrust Stakeholder Workshop



Workshop date: 6 December 2019
Report publication date: March 2020

Report authors:
Jun Zhao, University of Oxford
Menisha Patel, University of Oxford
Philip Inglesant, University of Oxford

Introduction

This report summarises the outcomes of the Second ReEnTrust Stakeholder workshop, which took place at Mary Ward House in London on Friday 6 December. It was attended by 7 participants and 5 ReEnTrust researchers.

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 second of the series. In our first stakeholder workshop (April 2019), we have explored stakeholders' general conceptualisation of trust, in relation to an algorithmic system. We found several factors key to trust development, including consistency, transparency and fairness of algorithmic results.

In this workshop, our key goal was to gain feedback to our two latest technology prototypes, which were designed to facilitate users' perception of algorithm transparency, fairness and trustworthiness:

An algorithm exploration sandbox tool, which investigates how user groups develop their trust perceptions of algorithms, when being provided with algorithm explanations and an ability to explore algorithmic behaviours by filtering and querying algorithms. Hotel booking was chosen as the recommendation scenario for this prototype, as it was perceived relevant to all user groups and particularly to the target groups 16-25 and 65+.

The trust mediation tool, which aims to enable users to raise trust-related queries and issues related to AI-algorithms and influence the behaviour of the algorithms accordingly. The tool is expected to be used following users' interactions with the aforementioned "algorithm sandbox". The initial prototype was inspired by a `chatbot' style mediation process, facilitating the dialogue between users' express of algorithm trust concerns and structured responses generated by mediation algorithms, with an aim of enhancing users' trust of algorithmic outcomes.

Methodology

The workshop used a combination of focus group discussions and co-design activities to encourage participants' active involvement and share of reflection.

Focus group discussion of algorithmic trust

We were interested to explore the perspective of stakeholders regarding how algorithm explanations may affect their perception of algorithm trust.

The stakeholder participants were given an A3 print-out of the search results returned by two recommendation algorithms supported by our prototype, along with explanations about why these results were returned to them. They were then given 20 minutes to work as a group to discuss these two scenarios regarding how to perceive these explanations may help the character to make choices of hotels or to develop the trust of the search results. At the end of the session, groups then shared their thoughts with the room. All materials are enclosed in Appendix A.

Group discussion and co-design of algorithm mediation tool

The session started with a short presentation of the current design of our trust mediation tool prototype, by using two fictional scenarios, as illustrated below:

- One is a hotel booking scenario, where a fictional character would like to negotiate with the mediation tool to ensure that the hotels recommended to her/him are ranked in a way that is more trustworthy to her/him.
- The other is a news aggregator scenario, where the character tried to negotiate with the mediation tool to ensure that it is not presenting her news feeds by implicitly profiling her as a female.

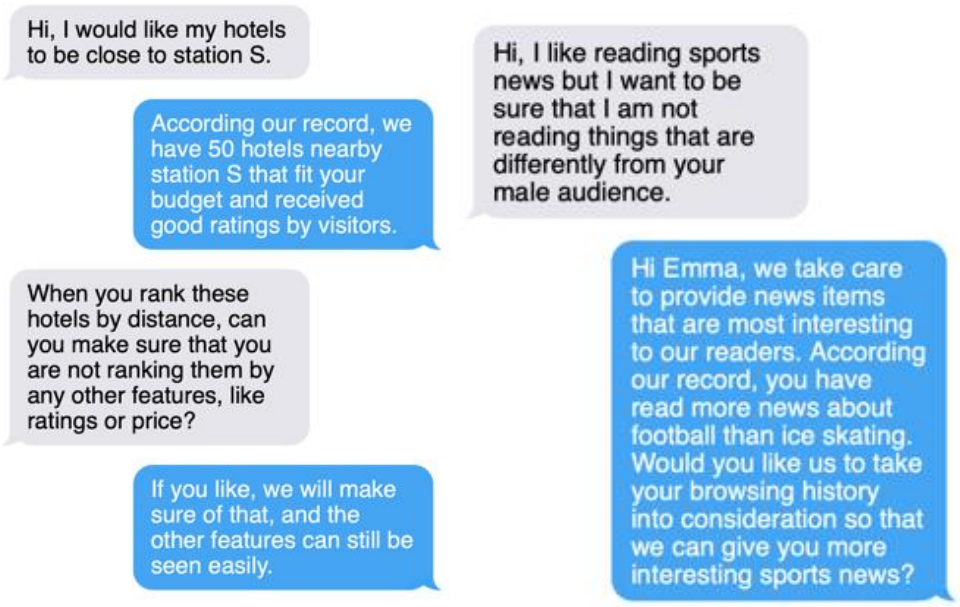


Fig 1: Two fictional scenarios to be supported by the trust mediation tool prototype: a hotel booking mediation scenario (left), and a news feed mediation scenario (right).

Groups were then asked to carry out a role-play activity, by imagining themselves as one of the fictional characters and thinking how they may react to the response of the tool. By the end of the activity, groups were asked to use 'post-it's to express what they liked and disliked of the prototype. After that, participants were invited to work as a group to design a better mediation tool for 20 minutes. They could choose to either improve the existing design in ways that are meaningful to them or propose something completely new. All materials can be found in Appendix B.

Key message summary

Below is a quick summary of key findings from our workshop, based on researchers' notes and the co-design outcome.

1. Our algorithm sandbox prototype was found to provide a novel and useful way for users to learn about algorithms. However, stakeholders suggest that to develop trust of algorithmic outcomes, they require a lot more information about how algorithms function. Examples of such information include more transparency about how the information is obtained by the algorithms in the first place, what other information the algorithm knows about me, or whether cookies are used to give me the personalised results. Participants were also concerned about social engineering transparency, e.g. who may control the algorithms or how the results are ranked, whether commissions are exchanged between the platform and hotels. These broader concerns are valuable inputs to our algorithm trust rebuilding approach and the positioning of the scope of our Algorithm Playground tool.

2. Our choice of hotel booking scenario - though relevant to all user groups- may not necessarily reach levels of trust consideration by users. Stakeholders suggested we consider integrating a higher stakes scenario within the sandbox in order to engender a deliberation of trust-related issues.

3. The Trust Mediation tool was welcomed as a novel approach to help users communicate their algorithmic concerns (such as the range of data used by an algorithm or how results should be filtered). Stakeholders found that the dialogue-like approach offered a flexible way to gain more transparency, awareness and understanding of algorithms. However, they were unsure of the broad scope that the mediation tool was positioned at, which seemed ambitious and would require lots of thinking to ensure the mediation language is appropriate, accurate and trustworthy.

4. The trustworthiness of the mediation approach could introduce unnecessary noise in the process of users' development of trust of algorithms: How could users trust that these mediation tools are being completely transparent and honest? This raised a great demand on how we may design the communication between users and the tool, and how the tool needs to provide a constructive and meaningful suggestion of alternatives or solutions to users' concerns.

5. The mediation tool has also been challenged to address situations where users may not know what they do not know. Participants discussed whether the tool could provide suggestions to users regarding the kind of questions they could ask, or some possible actions they could take. In this way, the mediation tool could act as a great facilitator, particularly for naive users.

Responsible Innovation (RI) is fundamental to the ReEnTrust project. An important part of this approach is to ensure an open and inclusive development approach so that stakeholder concerns are embedded in the processes and outcomes of the research. The feedback from the stakeholders provided valuable input to our tool development from an early stage. It would be exciting to follow this up in ReEnTrust by

investigating how a different algorithmic search scenario may affect people's perception of algorithms, and how people's broader needs for information transparency, which affects their development of trust, could be better facilitated by our next version of the tool prototype.

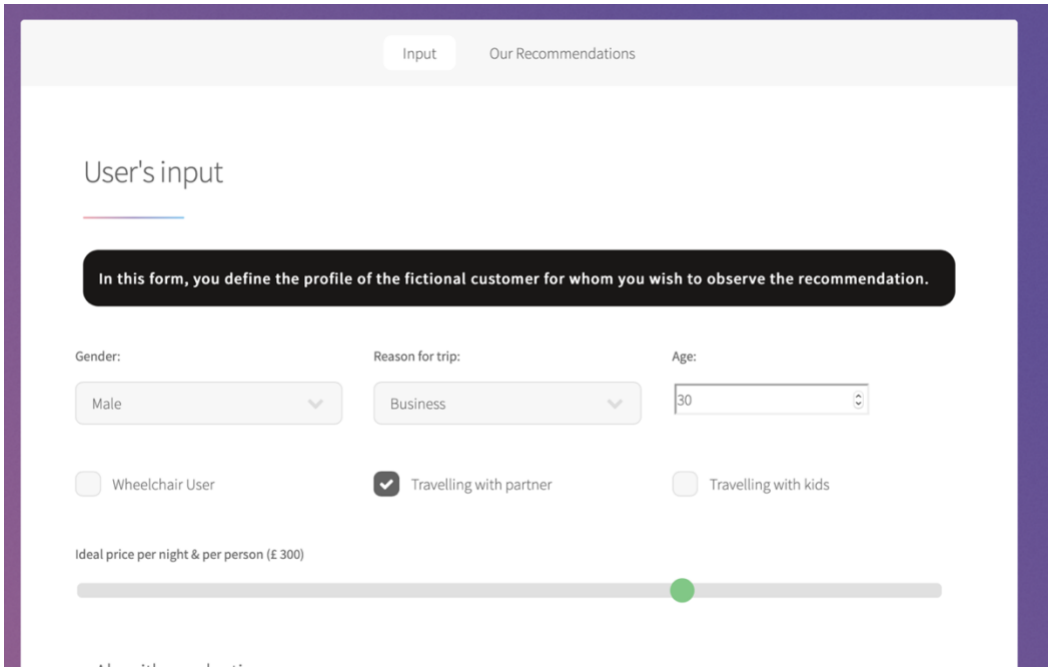
Appendix 1. Algorithm Playground Focus Group

Here is your group task sheet. Based on the hotel booking scenario for a fictional character, Tom, on our Algorithm Playground prototype platform, you will work together as a group, to think about the results returned for Tom and respond to the questions set out towards the end of the task sheet.

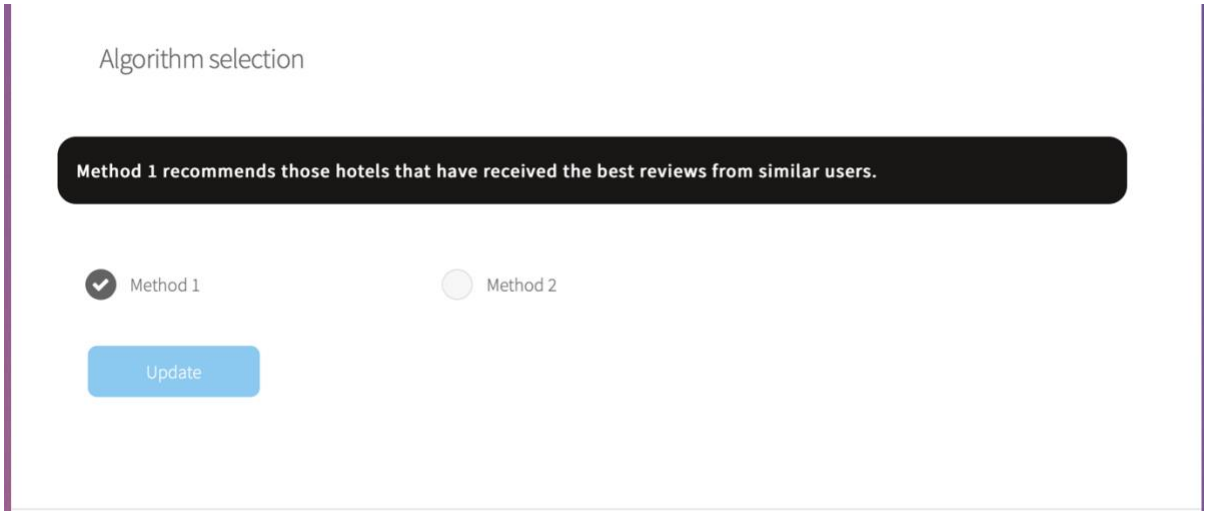
The scenario:

Tom is 30 years old. He is single and travels with his partner during his business trip, with a budget of £300 per night. They would like to have a hotel with good facilities, for example, for them to dine out in the evening.

Here is Tom's search input at our Algorithm Playground platform:

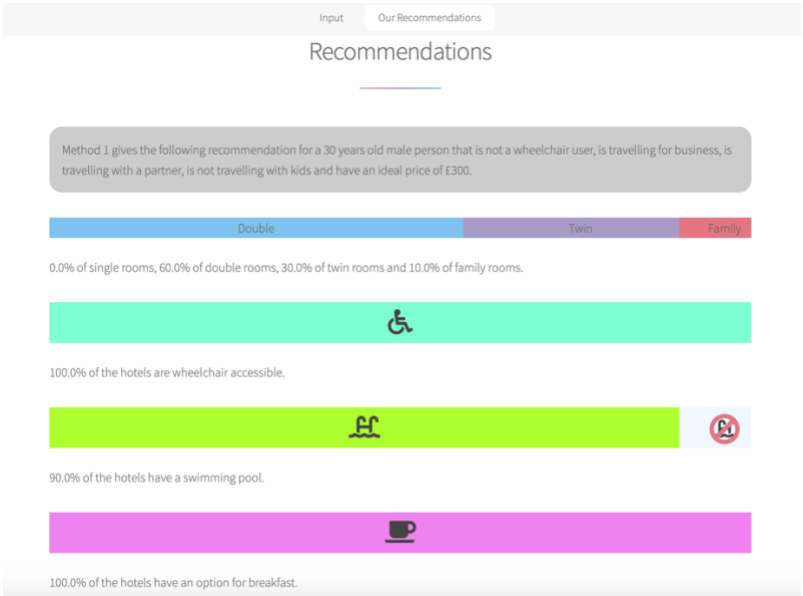


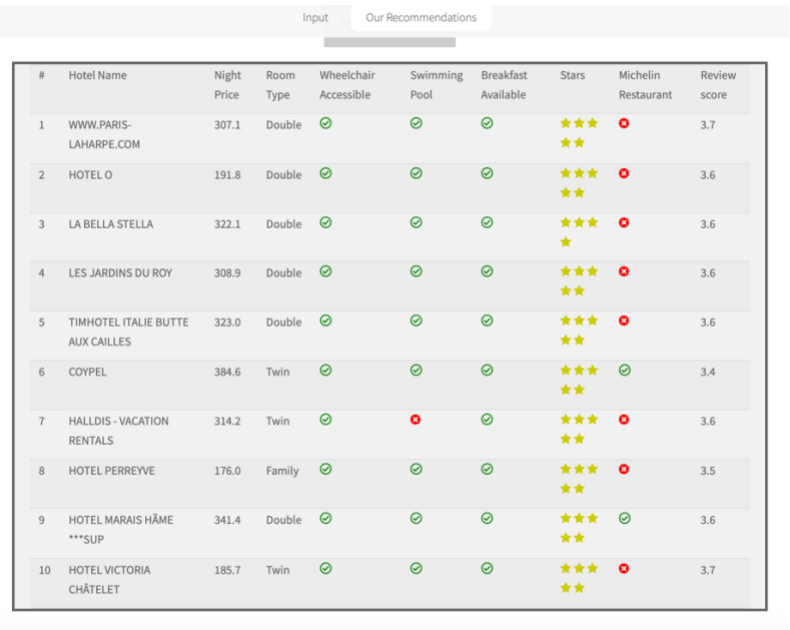
Tom can make a selectin of algorithms: Method 1 recommends hotels that have received best reviews from similar users; while Method 2 recommends hotels that are most similar to the ones similar users have stayed in.



Search results by Method 1

Below show the top 10 hotels returned for Tom from Method 1:
- The majority (6) of the rooms are double room, although 4 of them are not.
- All the hotels provide breakfast and 9/10 of them have a swimming pool.
- The highest ranked hotels appear on the top of his results, even though some of them are above his budget.





#	Hotel Name	Night Price	Room Type	Wheelchair Accessible	Swimming Pool	Breakfast Available	Stars	Michelin Restaurant	Review score
1	WWW.PARIS-LAHARPE.COM	307.1	Double	✓	✓	✓	★★★★	✗	3.7
2	HOTEL O	191.8	Double	✓	✓	✓	★★★★	✗	3.6
3	LA BELLA STELLA	322.1	Double	✓	✓	✓	★★★★	✗	3.6
4	LES JARDINS DU ROY	308.9	Double	✓	✓	✓	★★★★	✗	3.6
5	TIMHOTEL ITALIE BUTTE AUX CAILLES	323.0	Double	✓	✓	✓	★★★★	✗	3.6
6	COYPEL	384.6	Twin	✓	✓	✓	★★★★	✓	3.4
7	HALLDIS - VACATION RENTALS	314.2	Twin	✓	✗	✓	★★★★	✗	3.6
8	HOTEL PERREYVE	176.0	Family	✓	✓	✓	★★★★	✗	3.5
9	HOTEL MARAIS HÃME ***SUP	341.4	Double	✓	✓	✓	★★★★	✓	3.6
10	HOTEL VICTORIA CHÂTELET	185.7	Twin	✓	✓	✓	★★★★	✗	3.7

Search results by Method 2

Below show the top 10 hotels returned for Tom from Method 2:

- The rooms are a mix of single, twin and family rooms.
- All the hotels provide breakfast, although only 2/10 of them have a swimming pool.
- Hotels are ranked probably by how much they are like to the ones stayed by other people similar to Tom.

#	Hotel Name	Night Price	Room Type	Wheelchair Accessible	Swimming Pool	Breakfast Available	Stars	Michelin Restaurant	Review score
1	HOTEL DE VENDOME	107.4	Family	✗	✗	✓	★★	✗	2.2
2	LE TOURNE-BOUCHON	223.5	Twin	✓	✓	✓	★★★★	✗	3.5
3	HOTEL RÉSIDENCE SAINT-CHRISTOPHE PARIS	98.6	Single	✗	✗	✓	★★★	✗	2.5
4	HOTEL DU TRIANGLE D'OR	114.4	Family	✗	✗	✓	★★	✗	2.1
5	HOTEL DE NEUVE	61.6	Single	✓	✗	✓	★★★	✗	2.5
6	LIBERTEL AUSTERLITZ JARDIN DES PLANTES	235.8	Twin	✓	✓	✓	★★★★	✗	3.6
7	HOTEL DE MARSEILLE	106.5	Twin	✗	✗	✓	★★★★	✗	2.6
8	ANGELY	78.3	Single	✓	✗	✓	★★★	✗	2.5
9	PULLMAN PARIS TOUR EIFFEL	204.2	Twin	✗	✗	✓	★★★★	✗	2.4
10	PARIS VISIT SERVICES	82.7	Single	✗	✗	✓	★★★	✗	2.4

Questions:

1. Please discuss Method 1.
   a. What do you think of the algorithmic method that was used to produce outcomes for Tom?
   b. What are your thoughts on the description of the algorithmic method?
   c. What are your thoughts on the outcomes?

2. Please discuss Method 2.
   a. What do you think of the algorithmic method that was used to produce outcomes for Tom?
   b. What are your thoughts on the explanation of the algorithmic method?
   c. What are your thoughts on the outcomes?
3. Which algorithmic method would you prefer to select if you were Tom? and why?
4. What aspects of the tool do you feel facilitate an understanding of the role of algorithms in the filtering of information?
5. What do you think could be added/improved to further facilitate an understanding of the algorithm?
6. How do you think this tool could be used in practice?
7. How useful do you feel a tool like this is for enabling users to understand how algorithms feature in their digital world?

Appendix 2. Co-Design Mediation Tool

Here are two mediation scenarios. Please imagine yourselves to be the fictional character in each scenario, take a look at how the character may interact with our mediation tool, and then response to a few questions at the end of each scenario.

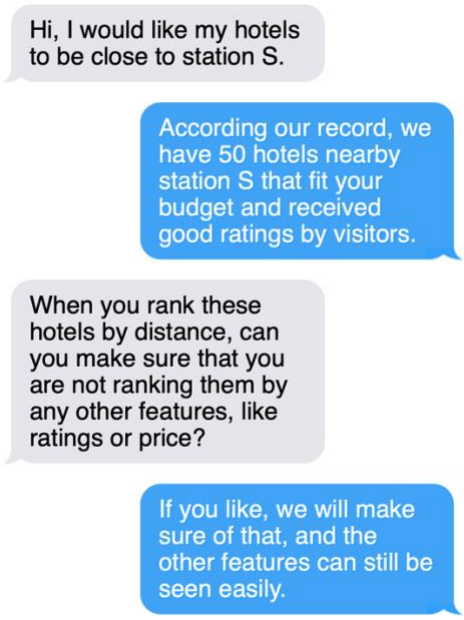
Scenario 1:

1. Jake wants to book a hotel for a leisure trip. He would like the algorithm to search for hotels stayed by people similar to him, and particularly priorities his needs for a convenient location that is close by to a metro station.

2. This preference cannot be expressed when searching for these hotels on the web site, so Jake decided to make use of “Personal Booking Assistant” provided by the hotel booking website.

3. Jake communicates his preferences to the booking assistant (the machine) and wants to make sure that the algorithm is being transparent.

4. He ends up being given 50 hotels sorted by their distance to the station S with increased transparency.



Questions:

1. Do you understand what our tool is trying to do?

2. Does the dialogue make sense to you?

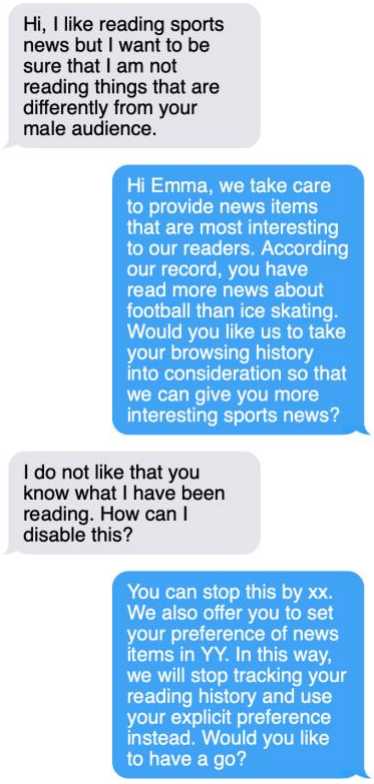
3. What do you think Jake would feel after this conversation with the 'booking assistant'?

4. If you were Jake, are there any other concerns that you would like to negotiate with our 'booking assistant'?

Scenario 2:

1. Emma wants to read about sports news that are interesting to her, but she does not like the newsfeed to give her only things they think a 30-year woman would be interested, so she started a conversation with the 'newsfeed assistant' provided by the website

2. Emma communicates her concern to the assistant (machine), which pointed out her preference conflicting with how the algorithm works and thus proposed her a set of selections.



Questions:

1. Do you understand what our tool is trying to do?

---

2. Does the dialogue make sense to you?

---

3. What do you think Emma would feel after this conversation with the 'newsfeed assistant'?

4. If you were Emma, are there any other concerns that you would like to negotiate with the ‘newsfeed assistant'?