

TrustScapes: A Visualisation Tool to Capture Stakeholders' Concerns and Recommendations About Data Protection, Algorithmic Bias, and Online Safety

International Journal of Qualitative Methods

Volume 22: 1–10

© The Author(s) 2023

DOI: 10.1177/16094069231186965

journals.sagepub.com/home/ijq

Sachiyo Ito-Jaeger^{1,2} , Giles Lane³ , Liz Douthwaite^{4,*}, Helena Webb^{4,*}, Menisha Patel⁵, Mat Rawsthorne⁶ , Virginia Portillo⁴, Marina Jirotko⁷, and Elvira Perez Vallejos^{1,2,4}

Abstract

This paper presents a new methodological approach, TrustScapes, an open access tool designed to identify and visualise stakeholders' concerns and policy recommendations on data protection, algorithmic bias, and online safety for a fairer and more trustworthy online world. We first describe how the tool was co-created with young people and other stakeholders through a series of workshops. We then present two sets of TrustScapes focus groups to illustrate how the tool can be used, and the data analysed. The paper then provides the methodological insights, including the strengths of the TrustScapes and the lessons for future research using TrustScapes. A key strength of this method is that it allows people to visualise their ideas and thoughts on the worksheet, using the keywords and sketches provided. The flexibility in the mode of delivery is another strength of the TrustScapes method. The TrustScapes focus groups can be conducted in a relatively short time (1.5–2 hours), either in person or online depending on the participants' needs, geographical locations, and practicality. Our experience with the TrustScapes offers some lessons (related to the data collection and analysis) for researchers who wish to use this method in the future. Finally, we describe how the outcomes from the TrustScapes focus groups should help to inform future policy decisions.

Keywords

trust, stakeholder engagement, co-creation, algorithmic bias, data protection, online safety, responsible research and innovation, patient and public involvement

Introduction

We continue to witness a significant development in digital technologies and artificial intelligence (AI). Artificial intelligence is “a family of techniques where algorithms uncover or learn associations of predictive power from data” (Panch et al., 2019, p. 1). Algorithms are a set of step-by-step instructions that computers follow to solve a problem. Research discussions on the topic of AI and digital technologies oscillate between celebration and fear. On the one hand, AI and digital technologies are seen to create new opportunities and improve efficiency, ranging from preventing fraud within online banking to improving accuracy of cardiovascular risk prediction (Weng et al., 2017). On the other hand, they pose dangers to privacy and safety, such as inappropriate sharing of

¹NIHR Nottingham Biomedical Research Centre, Nottingham, UK

²Faculty of Medicine and Health Sciences, The University of Nottingham, Nottingham, UK

³Proboscis, London, UK

⁴School of Computer Science, The University of Nottingham, Nottingham, UK

⁵King's Business School, King's College London, London, UK

⁶HD Labs, Stockport, UK

⁷Department of Computer Science, University of Oxford, Oxford, UK

*These authors contributed equally.

Data Availability Statement included at the end of the article

Corresponding Author:

Sachiyo Ito-Jaeger, NIHR Nottingham Biomedical Research Centre; Institute of Mental Health, Faculty of Medicine and Health Sciences, University of Nottingham, Innovation Park, Triumph Road, Nottingham NG7 2TU, UK. Email: sachiyo.ito-jaeger@nottingham.ac.uk



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

user's personal data by social networking service companies (Forrest, 2019). Concerns about privacy and fear of data exploitation are major factors influencing untrustworthiness in digital technologies (Adjekum et al., 2018; Ito-Jaeger et al., 2023; Liverpool et al., 2020; Sbaffi & Rowley, 2017). Concerns related to AI includes algorithmic bias, which occurs when "the application of an algorithm compounds existing inequalities in socio-economic status, race, ethnic background, religion, gender, disability or sexual orientation to amplify them and adversely impact equality" in society (Panch et al., 2019, p. 1). For examples, racial bias has been found in the outcomes of algorithms used to predict the likelihood of defendants recommitting crimes in the United States (Angwin et al., 2016).

Due to such concerns, some potentially beneficial technologies are not utilised by the target users. For example, young people and adults with previous experience of depression reported that although they could see the potential benefits of algorithms identifying depression from their social media content, the majority would not consent to their data being used for this purpose, regarding it as exposing and invasive (Ford et al., 2019). They believed that the risks to privacy were more significant (Ford et al., 2019). For users to feel safe to use such new technologies, it is essential to involve end-users through stakeholder engagement to identify their concerns and incorporate their ideas in the technologies to make them safer and more trustworthy.

Stakeholder Engagement

Stakeholder engagement is one of the pillars for responsible research and innovation (RRI) and provides an inclusive methodology to facilitate constructive dialogue that can drive the development of meaningful and relatable digital interventions (Webb et al., 2018). To maximise the impact on the target population, stakeholders, including end-users, must be actively involved in the development, production, implementation, and evaluation of new digital interventions (i.e., co-creation) (Jirotko et al., 2017). In recent years, stakeholder engagement has gained an increasing interest in a wide range of areas, including in digital technologies and in health and social care, and the importance and benefits of involving the public has been highlighted (Cluley et al., 2022; Gagnon et al., 2021; Hugh-Jones et al., 2022; Ito-Jaeger et al., 2021, 2022; Modigh et al., 2021). It is recommended that we should gather input from a diverse representative sample of end-users (Calvo et al., 2020).

Similarly, the multi-stakeholder initiative is a governance strategy that works on the basis that those most affected by an issue or change in a particular field should be involved in the discussion, decision making, and management of the particular issue at hand (Malcolm, 2008). This technique has been frequently implemented in internet governance (e.g., the Internet Governance Caucus) (Malcolm, 2008). The Internet Society described the multi-stakeholder initiative as working best on complex issues (Internet Society, 2016), including:

- a. Decisions that impact a wide and distributed range of people.
- b. Overlapping rights and responsibilities across sectors.
- c. Different forms of expertise needed.
- d. Legitimacy and acceptance of decisions that directly impact implementation.

The issues related to data protection, online safety, and algorithmic bias are such complex issues, and thus, stakeholders must be actively involved in the discussions.

Existing Assessment Tools and Public Engagement Initiatives

A number of assessment tools have been developed to evaluate new technologies, systems, or programs in regard to data protection, online safety, and algorithmic bias. Data Protection Impact Assessment (DPIA) is a process to help data controllers to identify, assess, and reduce the data protection risks of a project (Information Commissioner's Office, 2022; Ivanova, 2020). Other tools have also been created for public agencies and developers to identify and minimise bias in algorithmic decision-making systems (Duarte, 2017; Reisman et al., 2018). However, many of the assessment tools have been created for only specific stakeholders (e.g., data controllers, public agencies, and developers) but not for others (e.g., end-users) (Ayling & Chapman, 2022).

In recent years, an increasing number of public engagement and education initiatives have been developed (e.g., National Institute for Health and Care Research, 2019; Vincent, 2019). For example, in Citizens' Juries, people from a diverse, representative group are invited to participate in a series of workshops (3–5 days), where 'jurors' are asked specific questions (e.g., "What impact will automated decision systems have on broader social structures and interactions?") and consulted by experts in the specific field (Royal Society for the encouragement of Arts Manufactures and Commerce, 2019). Jurors are then asked to draw on the information that they learned to reach a conclusion on the questions. Whilst this method is a strong example of public engagement, it may not be feasible for many to implement due to the time-intensive process. In this paper, we introduce a new methodological approach, TrustScapes, an open access tool designed to identify and visualise stakeholders' concerns and policy recommendations on algorithmic bias, data protection, and online safety for a fairer and more trustworthy online world, that can be carried out in a relatively short time.

Methodology

TrustScapes

TrustScapes is part of the UnBias Fairness Toolkit, designed and created by Giles Lane (GL) with Alice Angus (AA) and Alex Murdoch, as part of the UnBias project funded by the

Engineering and Physical Sciences Research Council (EPSRC) (Koene et al., 2018; Rovatsos et al., 2019). The Fairness Toolkit aims to raise awareness and facilitate a *public civic dialogue* about how algorithms shape online experiences and to reflect on possible changes to address issues of online fairness (Lane et al., 2018). The tools are not just to enable critical thinking, but also to promote *civic thinking* – supporting a more collective approach to imagining the future in contrast to the individual approach that such technologies often lead to.

TrustScapes is a proactive, inclusive RRI tool (Andrews et al., 2022) that aligns with UK Research and Innovation (UKRI) expectations of anticipation, reflexivity, engagement and action (a considered response to mitigate risks) (Stilgoe et al., 2013; UKRI Engineering and Physical Sciences Research Council, 2022). TrustScapes includes a worksheet, keywords, and sketches. All these resources can be retrieved online (<https://unbias.wp.horizon.ac.uk/fairness-toolkit/>).

The TrustScapes worksheet is designed for stakeholders, including end-users, to visualise their concerns about algorithmic bias, data protection, and online safety, and what they would like to see changed to make the online world fairer and more trustworthy. The worksheet captures both their feelings about the current situation and their dreams and ideals for solutions and what the digital world could (or should) be in a dynamic and visual way. Visualization has consistently been proven to be an effective strategy to organise ideas and thoughts (e.g., Holley & Dansereau, 1984; Keller & Tergan, 2005).

Four prompts are included on the worksheet, these are:

1. Describe an experience of online bias, unfairness, or untrustworthiness you have had or are concerned about.
2. Illustrate what is important to you about this experience.
3. How do you think issues are being addressed by companies and authorities?
4. Ideally, what would you like to see done?

The TrustScapes keywords and sketches are provided to inspire participants when they complete the worksheet. The benefits of using visual illustrations to aid people in understanding complex concepts have been widely shown (e.g., Brotherstone et al., 2006; Carney & Levin, 2002). Both keywords and sketches can be printed out on standard office stickers to use on the worksheet, and participants are also encouraged to contribute their own drawings and insights.

These materials were co-created through a series of workshops with young people and other stakeholders in the United Kingdom. Young people were selected as co-creators as they are digital natives who grow up online and use devices as part of their daily lives. In addition to these young people, a range of experts who engage in research, industry, and civil

society were invited to provide comments and feedback on the tools during the development stages. In the following section, we illustrate how the TrustScapes approach was co-created through a series of workshops.

Co-Creators

Young People. Following the approval of the project by the research ethics committee at University of Oxford Social Sciences and Humanities Inter Divisional Research Ethics Committee (Ref R42596/RE001), three groups of young people in the United Kingdom were recruited to participate in the co-creation workshops (Table 1).

Group A was recruited to the project through personal contact with a member of the research team. Groups B and C became involved in the project through pre-established research relationship with the research team. Prior to the first workshop, the young people and their parent/carer were provided with the written information sheet and ‘young person and parent/carer consent form.’ The consent form was signed by both the young people and their parents/carers if they were willing to participate in the workshops. As stated in the consent form, no information about participants was collected or stored. Additionally, in accordance with the consent form, although the workshops were audio recorded and the worksheets were collected, they are permitted to be shared only among the project team members. Thus, in this paper, we focus on the procedure undertaken during the workshops and our reflections on the workshops.

Stakeholders. A range of experts were recruited through personal contacts with members of the research team to comment on several iterations of the toolkit. The experts included academics and researchers from the computer science, information security, and social sciences fields, and experts from the technology industry and from civil society organisations and NGOs.

Procedure

Young People’s Workshops. Each group of young people participated in two workshops (2 hours each). Each workshop was facilitated by the designer (GL) at Proboscis, a non-profit creative studio with expertise in social and cultural engagement and co-facilitated by 2–3 researchers with a background in Human-Computer Interaction or the Social Sciences. The first workshop focused on ‘awareness’ of bias in algorithmic systems, and the second focused on ‘empowerment’ to act when encountering bias in such systems. Stimulus material (Supplemental Material) was sent to each group in advance to provide some context about the issues to be explored in the workshops. This material included some short scenarios and links to online videos exploring the issues. Certificates of Participation were issued to each of the participants in the workshops for them to use as records of achievement.

Table 1. Characteristics of the Schools and Community.

	School/Community	N	Age Range	Characteristics
Group A	Inner London School	20	12–17	All girls school
Group B	Greater London School	12	12–17	Mixed school
Group C	Community Centre	5	16–22	Young women not in employment, education, or training

Young people's workshop 1: This workshop used the activity of creating a large MindMap with the participants to understand what kinds of digital devices and apps they use, what kinds of activities they do online, what they understand by 'algorithm', how it and the data it uses might affect them, and why it might be important for young people to know about this.

Young people's workshop 2: This workshop used the Mind mapping technique from Workshop 1 to develop what a TrustScape might look like. The participants were asked to imagine factors contributing to bias, trust and fairness in the systems they use, and how they might be affected by them. Resonant keywords and suggestions for the image sketches (both visual and descriptive) were gathered as part of the workshop.

Stakeholder's Workshop. The stakeholders participated in a three-and-a-half-hour workshop facilitated by the designer (GL) and co-facilitated by researchers with a background in Human-Computer Interaction and the Social Sciences. The workshop introduced the concept of and early designs for the Fairness Toolkit, including the TrustScapes, to stakeholders and sought insights from them regarding what it could offer them, and how this might be affected.

Reflections on the Co-Creation Workshops

Many of the young people who participated in the co-design workshops expressed a certain fatalism and lack of agency regarding how they use technology which seems to foster a sense of isolation and inability to effect change. This was coupled with a very limited sense of their rights and how the law protects them in their interactions with service providers, institutions and big companies. Unsurprisingly, they often feel that their voice is not listened to, even when they are the targets of some of the most aggressive marketing techniques deployed by the big platforms and retailers online and off. Due to this perceived lack of understanding and agency, and feelings of powerlessness, participants often found it difficult to articulate their feelings about the online world or what could be improved. The co-creation workshops were dynamic and engaging, and participants soon began to open up and offer ideas when they realised there were 'no stupid questions' and that the facilitators really wanted to hear from them. This was aided by the use of creative ways of recording what the participants said (e.g., using stickers and drawings to illustrate what was being said); experiencing this 'opening up' helped

the facilitators to understand what might work for the TrustScapes.

Designing the TrustScapes

The design of the worksheets aimed to address these existing issues, to give people the tools to articulate their concerns and empower them to do so. Additionally, the tool was designed to be completed in a relatively short time (1.5–2 hours) so that it was feasible for many people to use (e.g., in a group activity at a school, youth group, or other context). The tool needed to function independently of the project team so that it could be used by people without our presence or direct facilitation. As such, it was decided to include a worksheet and supporting materials (sketches and keywords) that could be printed as stickers. The worksheet is purposefully simple with prompts rather than leading questions so that it can be used by a wide range of people.

The keywords and sketches were created based on a large number of ideas generated during the engagement workshops. These were refined down into a draft selection by the designer (GL) and the illustrator (AA) and presented to the wider research team at a meeting. Feedback from the team informed the choices for the proposed final selection. The Handbook was written last to accompany the TrustScapes (and the other tools in the Fairness Toolkit) and provide background, guides for using the tools, and further links and information (UnBias, 2018).

Testing the TrustScapes

The tool was tested with experts during a second stakeholder workshop, again facilitated by the designer (GL) and researchers in Human-Computer Interaction and the Social Sciences. This workshop previewed the unfinished toolkit in a near-to-complete stage and sought feedback on the tools from the perspective of stakeholders. The TrustScapes were also tested with two independent groups of young people during workshops (i.e., different from the young people who participated in the co-creation workshops). One group was part of an advisory group for the overall UnBias project, aged between 13 and 17, and the other group was a small group of undergraduates at a University in London. The comments and feedback received were largely positive. Participants understood the instructions easily and were quickly engaged in creating their stories using the keywords and sketches, as well as creating their own drawings. They were also keen to talk

through their experiences afterwards. Anecdotal evidence suggests that the TrustScapes helped the young people to articulate their concerns and gave them confidence to present ideas. Only small changes were recommended, which included changing a few words on the keywords and sketches lists as they were not familiar to some people.

Example TrustScapes Focus Groups

To illustrate how TrustScapes are used, and the data analysed, we describe two sets of TrustScapes focus groups. The first TrustScapes focus groups were conducted in person, whereas the second focus groups were conducted online. Both explored the stakeholders' views and concerns about using AI in healthcare (AI chatbots for the first focus groups, and AI in health apps for the second focus groups). To illustrate how TrustScape focus groups can be held online, we will briefly describe the procedure for conducting online TrustScapes focus groups on Example 2 below.

Example 1: AI chatbots (in-person TrustScapes focus groups)

Chatbots are systems that interact with human users through online messaging. These bots operate through algorithms to give an automated response to a user's message, offering a virtual and instant 'therapy' (Abd-alrazaq et al., 2019). The bots offer daily conversations to monitor mood, advise self-help techniques, guided meditation, and education on the importance of sleep, exercise, and nutrition. This is an appealing method of counselling for many people as it is instantaneous, consistent, and anonymous, which are key reasons why young people go online to seek help with their mental health (Pretorius et al., 2019). The first series of TrustScapes focus groups were held to explore stakeholders' concerns and recommendations about AI chatbots in mental health care in the United Kingdom.

Participants. Following the approval of the study by the research ethics committee of the University of Nottingham Division of Psychiatry & Applied Psychology (Ref 0450), purposive sampling was used with only certain stakeholders invited. The participants included digital mental health service users, computer scientists (students and staff) and medical students. Prior to the focus groups, the participants received the participation information sheet, privacy notice, and consent form and submitted the consent form to confirm their attendance. A total of 20 people participated in the focus groups. Each session lasted for 2 hours with a short break after 1 hour. Recruiting emails were sent to the administrations of the relevant institute and Schools at the university who disseminated the emails around the respective groups. Flyers were also placed around the relevant

buildings of the same university. Data saturation became apparent with major trends clear by the end of the second focus group thus a third would not have been appropriate.

Procedure. An initial icebreaker allowed participants an opportunity to introduce themselves, get to know one another, and create a more open environment for discussion. This was followed by a presentation on the background of AI in mental health to give participants some examples of current interventions and raise a few ethical considerations.

Participants were separated into smaller groups, where each group consisted of two to five participants with representatives from each sector (service users, tech developers and medical students). Each group completed a TrustScape worksheet while group discussion was guided by a facilitator, one for each group (see Figure 1 for an example of completed TrustScapes). Once TrustScapes were completed, a member from each group presented their ideas to the whole group. This was followed by whole group discussion. All discussions were audio recorded and transcribed verbatim for analysis.

Participants understood the instructions easily and completed the task in groups with minimal input from the facilitators. Conversations were lively and the role of the facilitator mainly focused on ensuring all group members contributed equally and that a volunteer was nominated to present back to the whole group. Feedback from participants was very positive, highlighting the opportunity to interact with people from different backgrounds. For example, while medical students appreciated the opportunity to interact with service users and discover their viewpoints, participants with a computer science background appreciated engaging with the issues and concerns that future doctors and end users experience when accessing digital mental health services.

Analysis of TrustScape Worksheets and Transcriptions. The data collected from the TrustScapes worksheets and transcriptions was qualitatively analysed using thematic analysis following a six-step method of analysis, as proposed by Braun and Clarke (2006, 2019). A researcher (1) became familiar with the content by reading and re-reading the transcripts, (2) generated initial codes, (3) searched for themes, (4) reviewed the themes, (5) defined and named themes, and later (6) produced the report. Throughout the process, research meetings were held for debriefing and discussions among the researchers.

Results

Types of Data Collected. The data fell into three main themes and three subthemes for each main theme. Table 2 provides quotes and sketches for a few of the themes and subthemes to illustrate the types of data collected.

The reflections on the in-person TrustScapes focus groups and data analysis from this study are reported in the 'Methodological Insights' section.

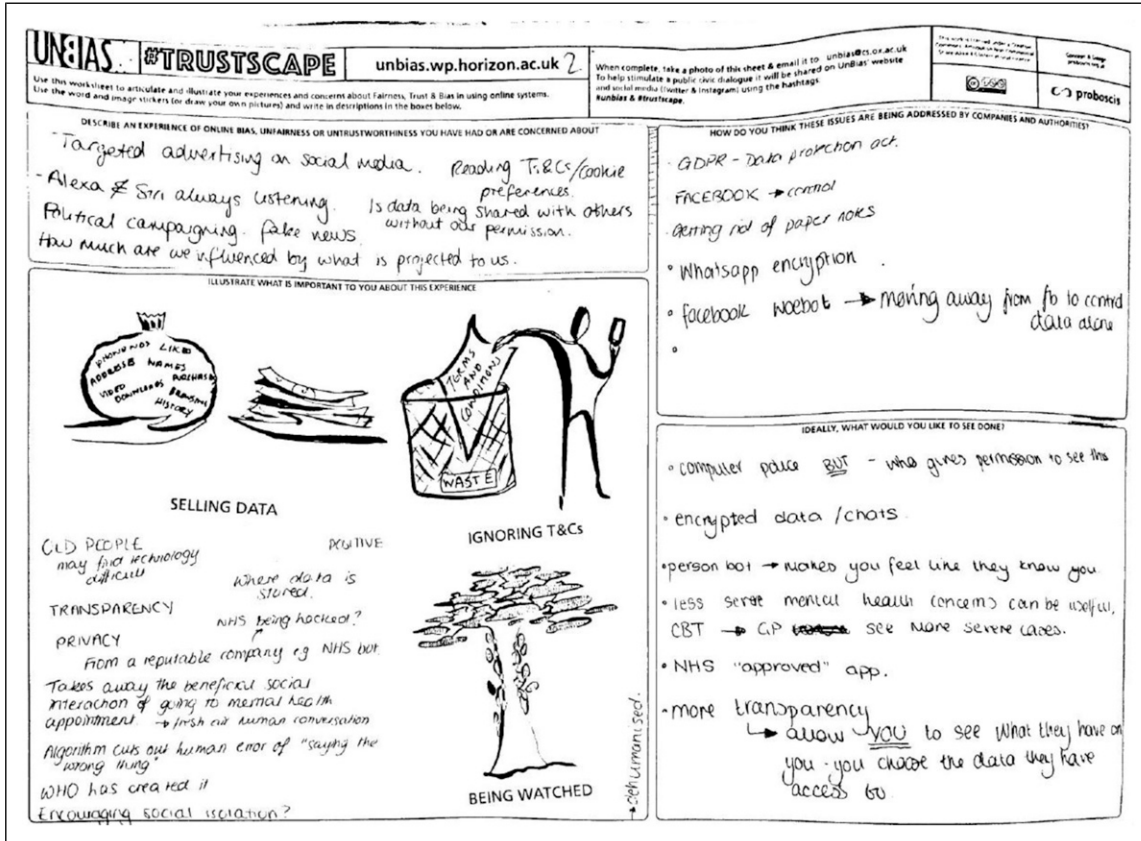


Figure 1. Example of completed TrustScapes.

Example 2: AI in health applications (online TrustScapes focus groups)




To illustrate how TrustScapes can be used online, we briefly report the procedure of online workshops. We also report the reflections on the online workshop in the ‘Methodological Insights’ section. The TrustScape workshops were held to explore health apps reviewers’ and stakeholders’ opinions and recommendations about AI in health apps. More specifically, specialist advisers were asked about the worst things that could happen with AI in health apps and how to avoid them. Research consent was not necessary as the participants acted as specialist advisers providing valuable knowledge and expertise based on their experience, and thus the workshop falls under the category of Patient and Public Involvement (Involve, 2009).

Procedure. The TrustScape workshops were conducted online via Miro (<https://miro.com/>), an online collaborative whiteboard platform, and were facilitated by a service user researcher and a technical specialist. The service user researcher had prior experience facilitating TrustScapes focus groups. A total of 34 people participated in the workshops. The specialist advisers included app reviewers, clinical leads, review clients of digital health technology and apps, regulators including health focused industry bodies, app development advisors,

app users (e.g., clinicians and patients), and members of the technical community (e.g., computing and engineering professional bodies, AI academics). A total of seven workshops were conducted, and each session consisted of an average of six participants and lasted for 2 hours with a short break after 1 hour.

An initial icebreaker was followed by a presentation on the AI in health apps and the importance of including stakeholders in the development of apps. During each workshop, participants collaborated with one another to complete one TrustScape. Participants understood the instructions easily with no need to provide additional clarifications. The conversations were not audio recorded as the facilitators attempted to capture all comments on the Miro sticky notes by encouraging participants to use the keywords and sketches provided and also asking participants to expand on their reason for selection. The online platform enabled both semi-anonymous dialogue and asynchronous access so participants could review and contribute further when convenient for them. Although the original plan was to conduct workshops for each stakeholder type, the feedback from the participants was that they appreciated having a mix of perspectives in the room and it generated many good discussions which then fed into nuanced considerations being captured on the TrustScapes Miro boards.

Table 2. Illustrative Themes and Subthemes Generated from the TrustScapes Focus Groups.

Main Theme	Subtheme	Quotes and Sketches
Human and robot interactivity	Social isolation	 <p>ECHO CHAMBER</p> <p>“It might actually end up being really counterproductive and isolating them from society and not letting them actually get out, live a life and interact, which is one of the better things for a lot of people with depression, is managing –having that assistance in that first step” (mental health nursing student)</p>
Individual safety and data privacy	Data protection	 <p>SELLING DATA</p> <p>“So we were thinking about selling data and how as well as that just being illegal and quite hurtful and wrong in lots of ways, it also is particularly upsetting because it feels like personal things that are emotions are also commodities or something that could be used against you, or something that, you know people would want to buy to manipulate you or to use in some way.” (service user)</p>
	Confidentiality	 <p>BEING WATCHED</p> <p>“Because in the NHS at the moment the only people that have access to medical records is doctors, nurses, healthcare professionals, on a need to know basis. So you can’t just have somebody randomly clicking open your file and reading through all your notes. But it would be particularly concerning if you could just have some office employee, on an admin team halfway round the world, reading through all of your information and all the messages that you sent.” (medical student)</p>

Methodological Insights

Strengths of TrustScapes

A key strength of the TrustScape method is that it allows participants to organise their ideas and thoughts visually in the four separate sections. It has been consistently shown that visualization is effective in organising ideas and thoughts (e.g., [Holley & Dansereau, 1984](#); [Keller & Tergan, 2005](#)). The prompts on the worksheet helped participants to start discussions smoothly, and the keywords and sketched helped them to articulate their concerns. As described above, the keywords and sketches have been co-created with stakeholders with the aim of developing meaningful and more relatable intervention ([Webb et al., 2018](#)). The use of sketches is unique and crucial given that they provide additional information on concepts that may be less familiar to some participants (e.g., filter bubbles, echo chamber). The benefits of using visual illustrations to aid people in understanding complex concepts have been widely studied

(e.g., [Brotherstone et al., 2006](#); [Delp & Jones, 1996](#)). Visual illustrations are helpful materials for everyone, but especially for a younger audience ([Carney & Levin, 2002](#)).

Another strength of the TrustScape method is that it is highly interactive as TrustScapes are completed during group discussions. Previous studies have consistently shown that interactive programmes are more effective than non-interactive, lecture-based programmes ([Bond & Hauf, 2004](#); [Ennett et al., 1994](#)). TrustScapes can be completed in groups with minimum input from a facilitator as the instructions are easy to understand.

The flexibility in the mode of delivery is another strength of the TrustScape method. That is, it can be employed both in person and online. In fact, in our first study, the TrustScape workshops were conducted in person and the worksheets were completed on paper. In the second study and a recently published study ([Andrews et al., 2022](#)), the workshops were conducted online via Miro, and the worksheets were completed using the Miro sticky notes. In another recently

published study, whilst the workshops were conducted online, the participants completed the worksheets on paper and emailed them to the researcher (Ito-Jaeger et al., 2023). Researchers can choose the mode of conducting workshops depending on participants' needs (e.g., accessibility to computer and the internet connection, sufficiency in using a software), geological locations, and practicality. Another strength of the TrustScapes method is that it can be completed in a relatively short period of time (1.5–2 hours) compared to other existing stakeholder engagement methods, which can take multiple days (e.g., Royal Society for the Encouragement of Arts Manufactures and Commerce, 2019).

Lessons for Future Research Using TrustScapes

Our experience with the TrustScapes offers some lessons for researchers who wish to use this method in the future. The first lesson relates to the data collection and analysis. In our first study, we collected and analysed both the audio transcripts of the focus groups discussions and the TrustScapes worksheets. The analysis of transcripts was necessary as some participants provided limited information on the worksheet. The possible reasons include insufficient time to complete the worksheet and feeling that they have provided their ideas verbally. While audio transcripts provide rich data and more context, the analysis of long transcripts is a time-intensive process. In future research, if researchers wish to analyse the data from the worksheets alone, we recommend that they provide enough time for participants to complete the worksheet and make the task of participants as easy as possible by providing stickers with the keywords and sketches printed on. In the second study, participants' ideas were collected only on Miro sticky notes. Whilst this way of collecting data is more efficient and reduces time for analysis, some Miro sticky notes included only a few words and lacked the context. Thus, it is recommended that future researchers encourage participants to provide more context on the Miro sticky notes by writing a sentence, rather than only words.

A second lesson for future research is to provide participants with lists of the keywords and sketches prior to the TrustScape workshop. As the lists are lengthy, it is crucial that participants have enough time to familiarise themselves with the lists in advance. Researchers should also allocate enough time to explain some possibly unfamiliar words and illustrations and the time for participants to ask questions before the TrustScape focus groups start.

Informing Policy

The TrustScapes method provides an opportunity for stakeholders to give opinions and recommendations about algorithms bias, data protection, and online safety. The outcomes from the TrustScape workshops should inform future policy decisions. MetaMap, another tool contained in the Fairness Toolkit, is a worksheet designed for policy

makers, regulators, members of the public sector, researchers, and industry, to respond to the ideas provided by participants in the TrustScapes (Rovatsos et al., 2019). By selecting and incorporating a TrustScape from those shared on the Unbias website (Unbias, 2018), policy makers can respond to the opinions and recommendations provided by the stakeholders. The completed MetaMaps will also be shared online via Twitter (ReEnTrust, 2016) to encourage the further public civic dialogue and to demonstrate the value of participation to people in having their voice listened and replied to, acting as a foundation for building trust and ongoing engagement.

Conclusion

The TrustScape is a valuable method to identify and visualise stakeholders' concerns and policy recommendations on data protection, algorithmic bias, and online safety. One of the strengths of this method is that people can visualise their ideas and thoughts on the worksheet and use the keywords and sketches. The TrustScape workshops and worksheets can also be completed in a relatively short time either in person or online depending on the participants' needs. Stakeholder engagement is one of the pillars of RRI and a key element for developing new interventions in a socially desirable and acceptable way (Jirotko et al., 2017). Thus, it is critical that policy makers, regulators, members of the public sector, researchers, and industry respond to the ideas provided by the stakeholders in the TrustScapes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Engineering and Physical Sciences Research Council (EP/N02785X/1). Sachiyo Ito-Jaeger and Elvira Perez Vallejos acknowledge the financial support of the NIHR Nottingham Biomedical Research Centre.

ORCID iDs

Sachiyo Ito-Jaeger  <https://orcid.org/0000-0001-9664-7797>

Giles Lane  <https://orcid.org/0000-0001-5222-4469>

Mat Rawsthorne  <https://orcid.org/0000-0002-7481-693X>

Data Availability Statement

TrustScape worksheets are openly available from the University of Nottingham data repository at <https://doi.org/10.17639/nott.7313>.

Supplemental Material

Supplemental material for this article is available online

References

- Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, *132*, 103978. <https://doi.org/10.1016/j.ijmedinf.2019.103978>
- Adjekum, A., Blasimme, A., & Vayena, E. (2018). Elements of trust in digital health systems: Scoping review. *Journal of Medical Internet Research*, *20*(12), Article e11254. <https://doi.org/10.2196/11254>
- Andrews, J. A., Rawsthorne, M., Manolescu, C., Burton McFaul, M., French, B., Rye, E., McNaughton, R., Baliouis, M., Smith, S., Biswas, S., Baker, E., Repper, D., Long, Y., Jilani, T., Clos, J., Highton, F., Moghaddam, N., & Malins, S. (2022). Involving psychological therapy stakeholders in responsible research to develop an automated feedback tool: Learnings from the EX-TRAPPOLATE project. *Journal of Responsible Technology*, *11*, 100044. <https://doi.org/10.1016/j.jrt.2022.100044>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias risk assessments in criminal sentencing*. ProPublica.
- Ayling, J., & Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? *AI and Ethics*, *2*(3), 405–429. <https://doi.org/10.1007/s43681-021-00084-x>
- Bond, L. A., & Hauf, A. M. C. (2003). Taking stock and putting stock in primary prevention: Characteristics of effective programs. *The Journal of Primary Prevention*, *24*(3), 199–221. <https://doi.org/10.1023/B:JOPP.0000018051.90165.65>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Brotherstone, H., Miles, A., Robb, K. A., Atkin, W., & Wardle, J. (2006). The impact of illustrations on public understanding of the aim of cancer screening. *Patient Education and Counseling*, *63*(3), 328–335. <https://doi.org/10.1016/j.pec.2006.03.016>
- Calvo, R. A., Peters, D., & Cave, S. (2020). Advancing impact assessment for intelligent systems. *Nature Machine Intelligence*, *2*(2), 89–91. <https://doi.org/10.1038/s42256-020-0151-z>
- Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *14*(1), 5–26. <https://doi.org/10.1023/A:1013176309260>
- Cluley, V., Ziemann, A., Feeley, C., Olander, E. K., Shamah, S., & Stavropoulou, C. (2022). Mapping the role of patient and public involvement during the different stages of healthcare innovation: A scoping review. *Health Expectations: An International Journal of Public Participation in Health Care and Health Policy*, *25*(3), 840–855. <https://doi.org/10.1111/hex.13437>
- Delp, C., & Jones, J. (1996). Communicating information to patients: The use of cartoon illustrations to improve comprehension of instructions. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, *3*(3), 264–270. <https://doi.org/10.1111/j.1553-2712.1996.tb03431.x>
- Duarte, N. (2017). *Digital decisions tool*. <https://cdt.org/insights/digital-decisions-tool/>
- Ennett, S. T., Tobler, N. S., Ringwalt, C. L., & Flewelling, R. L. (1994). How effective is drug abuse resistance education? A meta-analysis of project DARE outcome evaluations. *American Journal of Public Health*, *84*(9), 1394–1401. <https://doi.org/10.2105/AJPH.84.9.1394>
- Ford, E., Curlewis, K., Wongkoblap, A., & Curcin, V. (2019). Public opinions on using social media content to identify users with depression and target mental health care advertising: Mixed methods survey. *JMIR Mental Health*, *6*(11), Article e12942. <https://doi.org/10.2196/12942>
- Forrest, A. (2019). *Facebook data scandal: Social network fined \$5bn over 'inappropriate' sharing of users' personal information*. Independent. <https://www.independent.co.uk/news/world/americas/facebook-data-privacy-scandal-settlement-cambridge-analytica-court-a9003106.html>
- Gagnon, M.-P., Tanchou Dipankui, M., Poder, T. G., Payne-Gagnon, J., Mbemba, G., & Beretta, V. (2021). Patient and public involvement in health technology assessment: Update of a systematic review of international experiences. *International Journal of Technology Assessment in Health Care*, *37*(1), Article e36. <https://doi.org/10.1017/S0266462321000064>
- Holley, C. D., & Dansereau, D. F. (1984). Chapter 1 - the development of spatial learning strategies. In C. D. Holley, & D. F. Dansereau (Eds), *Spatial learning strategies* (pp. 3–19). Academic Press. <https://doi.org/10.1016/B978-0-12-352620-5.50007-2>
- Hugh-Jones, S., Pert, K., Kendal, S., Eltringham, S., Skelton, C., Yaziji, N., & West, R. (2022). Adolescents accept digital mental health support in schools: A co-design and feasibility study of a school-based app for UK adolescents. *Mental Health & Prevention*, *27*, 200241. <https://doi.org/10.1016/j.mhp.2022.200241>
- Information Commissioner's Office. (2022). *Data protection impact assessment*. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/>
- Internet Society. (2016). *Internet governance - why the multi-stakeholder approach works*. <https://www.internetsociety.org/resources/doc/2016/internet-governance-why-the-multistakeholder-approach-works/>
- Involve. (2009). *Patient and public involvement in research and research ethics committee review*. <https://www.invo.org.uk/wp-content/uploads/2011/12/INVOLVENRESfinalStatement310309.pdf>
- Ito-Jaeger, S., Perez Vallejos, E., Curran, T., & Crawford, P. (2022). What's up with everyone? A qualitative study on young people's perceptions of cocreated online animations to promote mental health literacy. *Health Expectations: An International Journal of Public Participation in Health Care and Health Policy*, *25*(4), 1633–1642. <https://doi.org/10.1111/hex.13507>
- Ito-Jaeger, S., Perez Vallejos, E., Curran, T., Spors, V., Long, Y., Liguori, A., Warwick, M., Wilson, M., & Crawford, P. (2022).

- Digital video interventions and mental health literacy among young people: A scoping review. *Journal of Mental Health*, 31(6), 873–883. <https://doi.org/10.1080/09638237.2021.1922642>
- Ito-Jaeger, S., Perez Vallejos, E., Logathasan, S., Curran, T., & Crawford, P. (2023). Young people's trust in cocreated web-based resources to promote mental health literacy: Focus group study. *JMIR Mental Health*, 10, Article e38346. <https://doi.org/10.2196/38346>
- Ivanova, Y. (2020). *The data protection impact assessment as a tool to enforce non-discriminatory AI forthcoming in*. Spring Proceedings of the Annual Privacy Forum.
- Jirotko, M., Grimpe, B., Stahl, B., Eden, G., & Hartswood, M. (2017). Responsible research and innovation in the digital age. *Communications of the ACM*, 60(5), 62–68. <https://doi.org/10.1145/3064940>
- Keller, T., & Tergan, S.-O. (2005). Visualizing knowledge and information: An introduction. In S.-O. Tergan, & T. Keller (Eds), *Knowledge and information visualization: Searching for synergies* (pp. 1–23). Springer Berlin Heidelberg. https://doi.org/10.1007/11510154_1
- Koene, A., Dowthwaite, L., Lane, G., Webb, H., Portillo, V., & Jirotko, M. (2018, August 22, 2018). *UnBias: Emancipating users against algorithmic biases for a trusted digital economy*. KDD. http://www.kdd.org/kdd2018/files/project-showcase/KDD18_paper_1804.pdf
- Lane, G., Angus, A., & Murdoch, A. (2018). *UnBias fairness toolkit*. Proboscis. <https://doi.org/10.5281/zenodo.2667808>
- Liverpool, S., Mota, C. P., Sales, C. M. D., Čuš, A., Carletto, S., Hancheva, C., Sousa, S., Cerón, S. C., Moreno-Peral, P., Pietrabissa, G., Moltrecht, B., Ulberg, R., Ferreira, N., & Edbrooke-Childs, J. (2020). Engaging children and young people in digital mental health interventions: Systematic review of modes of delivery, facilitators, and barriers. *Journal of Medical Internet Research*, 22(6), Article e16317. <https://doi.org/10.2196/16317>
- Malcolm, J. (2008). *Multi-stakeholder governance and the internet governance forum*. Terminus Press.
- Modigh, A., Sampaio, F., Moberg, L., & Fredriksson, M. (2021). The impact of patient and public involvement in health research versus healthcare: A scoping review of reviews. *Health Policy*, 125(9), 1208–1221. <https://doi.org/10.1016/j.healthpol.2021.07.008>
- National Institute for Health and Care Research. (2019). *Involving the public in complex questions around artificial intelligence research*. <https://www.nihr.ac.uk/blog/involving-the-public-in-complex-questions-around-artificial-intelligence-research/12236>
- Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: Implications for health systems. *Journal of Global Health*, 9(2), 010318. <https://doi.org/10.7189/jogh.09.020318>
- Pretorius, C., Chambers, D., Cowan, B., & Coyle, D. (2019). Young people seeking help online for mental health: Cross-sectional survey study. *JMIR Mental Health*, 6(8), Article e13524. <https://doi.org/10.2196/13524>
- ReEnTrust. (2016). *UnBias.Algorithms*. ReEnTrust. https://twitter.com/unbias_algos
- Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic impact assessments: A practical framework for public agency accountability*. <https://ainowinstitute.org/aiareport2018.pdf>
- Rovatsos, M., Mittelstadt, B., & Koene, A. (2019). *Landscape summary: Bias in algorithmic decision-making: What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?* <https://www.gov.uk/government/publications/landscape-summaries-commissioned-by-the-centre-for-data-ethics-and-innovation>
- Royal Society for the Encouragement of Arts Manufactures and Commerce. (2019). *Democratising decisions about technology: A toolkit*.
- Sbaffi, L., & Rowley, J. (2017). Trust and credibility in web-based health information: A review and agenda for future research. *Journal of Medical Internet Research*, 19(6), Article e218. <https://doi.org/10.2196/jmir.7579>
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- UKRI Engineering and Physical Sciences Research Council. (2022, March 31, 2022). *Anciticipate, reflect, engage and act (AREA)*. <https://epsrc.ukri.org/research/framework/area/>
- UnBias. (2018). *UnBias: Fairness toolkit*. <https://unbias.wp.horizon.ac.uk/fairness-toolkit/>
- Unbias. (2018). *Unbias: TrustScapes*. <https://unbias.wp.horizon.ac.uk/category/trustscapes/>
- Vincent, J. (2019). *Finland is making its online AI crash course free to the world*. <https://www.theverge.com/2019/12/18/21027840/online-course-basics-of-ai-finland-free-elements>
- Webb, H., Koene, A., Patel, M., & Vallejos, E. P. (2018). Multi-stakeholder dialogue for policy recommendations on algorithmic fairness. Proceedings of the 9th International Conference on Social Media and Society, Copenhagen, Denmark. <https://doi.org/10.1145/3217804.3217952>
- Weng, S. F., Reys, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*, 12(4), Article e0174944. <https://doi.org/10.1371/journal.pone.0174944>