Journal of Responsible Technology 19 (2024) 100093

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

Research Article

A call to action: Designing a more transparent online world for children and young people

Virginia Portillo [a,*], Liz Dowthwaite [a], Helen Creswick [a], Elvira Pérez Vallejos [a,b], Carolyn Ten Holter [c], Ansgar Koene [a,d], Marina Jirotka [c], Jun Zhao [c]

[a] School of Computer Science, Horizon Digital Economy Research Institute, University of Nottingham, UK
[b] Faculty of Medicine and Health Sciences, University of Nottingham, UK
[c] Department of Computer Science, University of Oxford, UK
[d] Ernst & Young, Brussels, Belgium

ARTICLE INFO

Keywords:
Children and young people
Algorithmic transparency
Digital technologies
Responsible innovation
Qualitative research

ABSTRACT

This paper reports on a qualitative research study that explored the practical and emotional experiences of young people aged 13–17 using algorithmically-mediated online platforms. It demonstrates an RI-based methodology for responsible two-way dialogue with the public, through listening to young people's needs and responding to their concerns. Participants discussed in detail how online algorithms work, enabling the young people to reflect, question, and develop their own critiques on issues related to the use of internet technologies. The paper closes with action areas from the young people for a fairer, usefully transparent and more responsible online environment. These actions include a desire to be informed about what data (both personal and situational) is collected and how, and who uses it and why, and policy recommendations for meaningful algorithmic transparency and accountability. Finally, participants claimed that whilst transparency is an important first principle, they also need more control over how platforms use the information they collect from users, including more regulation to ensure transparency is both meaningful and sustained.

Introduction

It is estimated that globally one third of internet users are children below the age of 18 (Livingstone et al., 2015). In some countries, under-15s are as likely to use the Internet as adults over 25 (UNICEF, 2017). In the UK, 97 % of 12-15 year olds have their own mobile phone, and to go online 94 % use a mobile phone, 54 % a tablet and 63 % a laptop (Ofcom, 2022), exposing them to the opportunities but also the risks that can be encountered online (Hepburn & Christie, 2019; Livingstone et al., 2017; Smahel et al., 2020). Many researchers, non-profits, educators, governments, policy makers and regulators are now focused on the need to create safe online environments for children[1].

The UK government has taken steps in this area. It published a House of Commons report: *'Impact of social media and screen-use on young people's health'* (House of Commons, 2019), and the Information Commissioner's Office (ICO) issued the Age Appropriate Design Code (ICO, 2020). The Code, which was approved by the UK Government and came into force in 2021, follows the principles of the General Data Protection Regulation (GDPR) (European Union, 2016) to ensure online services and apps that are likely to be accessed by children in the UK appropriately safeguard children's personal data. Finally, the UK has now passed legislation in the form of the Online Safety Act (Online Safety Act 2023, 2023). The Act provides for a statutory duty of care for internet companies in all cases where content "is likely to be accessed by children", to ensure children and young people are safer when online. These moves have been supported by organisations that lobby for children's needs and rights (5Rights Foundation, 2019; Barnardo's, 2019; Bentley et al., 2019; Children's Commissioner for England, 2018; Livingstone et al., 2024; OECD, 2020). Children can be particularly at risk from problematic data-collection practices, as their data may be collected inferentially and without meaningful consent (Stoilova et al., 2021).

"Transparency" is a crucial element to ensure regulation and accountability of tech companies (House of Commons, 2019). However,

* Corresponding author.
E-mail address: Virginia.Portillo@nottingham.ac.uk (V. Portillo).

[1] We refer to 'children' as persons under 18 and young people as 10-24 year olds, in line with the terminology used by the UN, UNICEF and the WHO (UNDESA, n. d.)

https://doi.org/10.1016/j.jrt.2024.100093

Available online 23 August 2024

it is not clear what transparency means to internet users, how they feel it can help them, or whether it affects their choices in any way (Dogruel et al., 2020). A report from the All-Party Parliamentary Group on Data Analytics suggested that *"all …organisations [that interact with citizens] should conform to the highest standards of transparency and engagement as their 'licence to provide public services'"* (Tindale & Muirhead, 2019, p. 4). The present study investigated how transparency can be made meaningful to young people, and how they would like to see it manifested in their online world. The study focused on algorithmic transparency, as algorithms are frequently 'invisible' to users but can substantively shape their online experience (Fouquaert & Mechant, 2021).

This paper describes how citizens aged 13 - 17 perceive and experience algorithms in their everyday online worlds; what type of information they want to know when using algorithmically-mediated services; and presents children and young people's needs when using the Internet as a call for action for meaningful algorithmic transparency and accountability. Its contribution is to provide data for an evidence-base to feed into the design of digital technologies aimed at young people, to inform policymakers, to ensure that young people's needs and rights are valued and respected, and to call for technologies to be developed responsibly. Results show that engaging with and educating citizens about the online world enables them to reflect, question, and develop their own ideas on how key challenges related to internet technologies could be approached for a fair and more transparent online world. It also provides evidence for education, policymaking, and regulatory sectors on young people's experiences and recommendations for a more accountable and meaningfully transparent online world.

Algorithmic transparency

Algorithms are so widely used in automated systems, online platforms, websites, and apps, that citizens have become accustomed to algorithms making all kinds of recommendations to them (Garfinkel et al., 2017). Although this might be convenient for users, algorithms can also make decisions about individuals that can unfairly impact people's lives, for example in relation to education, employment, healthcare, and access to credit and criminal justice (Angwin et al., 2016; Federal Trade Commission, 2016).

Algorithmic outputs are often described as 'opaque' as users rarely have a *'concrete sense of how and why*', and the *'inputs themselves may be also unknown or known only partially'* (Burrell, 2016, p. 1). Explaining algorithms and algorithmic decisions – algorithmic transparency – can facilitate accountability (Lepri et al., 2018). The US's Electronic Privacy Information Center declares that algorithmic transparency, a key indicator of internet universality, should be a fundamental human right (EPIC, 2022).

Since May 2018, online platforms providing services to citizens in the EU have to comply with GDPR (European Union, 2016) when collecting users' personal information. The GDPR and the California Consumer Privacy Act of 2018 (California Constitution, 2018), aim to protect users' privacy rights, and balance information and power asymmetries within the Digital Economy. However, more transparency on the functionality of algorithms and accessible information about the potential impact on people's online experiences (such as inferences made about them without their knowledge) should also be made available to users (Wachter & Mittelstadt, 2019). Under the GDPR, the "right to explanation" allows a user to ask for an explanation of an algorithmic decision that was made about them. However, the law does not specify what information is actually required to explain algorithmic decisions to users (Goodman & Flaxman, 2017).

A study by the European Parliament states that *'releasing the source code of an algorithmic system would often not provide meaningful transparency'* (Koene et al., 2019, p II). Explaining exactly how outcomes are reached is technically challenging, but 'meaningful transparency into the behaviour of computing systems is feasible and can provide important benefits' (p. II). Also, providing citizens with algorithmic literacy about the main concepts involved in algorithmic answers, helps participation *'in the public dialogue about the use of these systems'* (p. 69). Explainability is a mechanism to enhance algorithmic transparency (Lepri et al., 2018) and includes how decisions are made, who trains the system, and what data was used in algorithm-based recommendations.

Algorithmic transparency is difficult to generalise across different systems, however, accountability is also difficult to assess (Pasquale, 2015). Demands for algorithmic transparency have to be specific about what is to be disclosed and to whom (de Laat, 2018). Ananny & Crawford (2018) stressed the importance of transparency as understanding the 'behaviour' of algorithmic systems – i.e. the reasons for a decision - rather than looking inside a system.

Users' experiences and preferences

The potential benefits and drawbacks of 'personalisation' algorithms have been widely discussed (Gauch et al., 2007; Oulasvirta & Blom, 2008; CDEI, 2020). Although the 'filter bubble' hypothesis – that refining information according to a user's interest risks placing them in a limited 'bubble' that reinforces one-dimensionality (Pariser, 2011) – has been widely questioned (Dubois & Blank, 2018), it is still possible for personalisation to compromise democratic access to information, diversity of perspectives, and thereby damage high-quality practice and governance. These effects can be reinforced by the fact that users are frequently unaware of algorithmic-decision making and filtration taking place online (Eslami et al., 2015).

Nevertheless, users frequently develop strategies to negotiate everyday problems encountered in algorithmic systems (Eiband et al., 2019) – they may gain some understanding of how an algorithm is working through frequent interaction with a system, by getting unexpected information or outputs (DeVito et al., 2017), or by detecting bias in the system (Eslami et al., 2017).

However, adult users have demanded that technology companies show, in an accessible way, how services work and how decisions pertaining to them are made (doteveryone, 2018a, 2018b). A follow up study indicated that 72 % of adults were aware that their personal information can be 'rented' or sold to third parties; however, their understanding of how organisations actually used their data was limited (doteveryone, 2020). Adults' concerns about online data collection and its privacy implications have been widely reported (Hoofnagle et al., 2010; Fiesler & Hallinan, 2018). These concerns are often associated with issues of information asymmetry between users and online systems regarding personal data practices (Brunton & Nissenbaum, 2015; Bodo et al., 2018). It has been shown that providing algorithmic transparency impacts on people's attitudes and intentions depending on people's engagement with and personal benefit from the algorithm (Eslami et al., 2019).

People's attitudes toward algorithmic decision-making by online systems can also be highly context-specific (Pasquale, 2015). For example users may find personalised recommendations about leisure or socialising acceptable, but not political content (Pew Research Center, 2018). Although Awad & Krishnan (2006) found that online customers who valued information transparency were also less willing to be profiled for personalised offerings, other research has suggested that explanations to users can increase acceptance of the systems (Herlocker et al., 2000) and the recommendations they receive (Cramer et al., 2008). This suggests that transparency could improve users' trust towards algorithmic information (Lee et al., 2015).

Many of these strategies and understandings – so fundamental to a citizen's relationship with the information they access online - are not as available to children and young people, who may have less comprehension of how data is gathered, analysed, and utilised. Adolescents have expressed privacy concerns on how social media and other online platforms manage their personal information (Boyd & Marwick, 2011; Pangrazio & Selwyn, 2017; Perez Vallejos et al., 2017; Silva et al., 2017; Creswick et al., 2019; Livingstone et al., 2019; Dowthwaite et al., 2020).

However, we identify a gap here, as there is little empirical evidence on how young people experience algorithmic functions when actually using online systems and platforms, a lack of analysis on their understanding of how algorithms work, and a shortage of engagement with their concerns and possible recommendations for a more transparent digital environment.

Transparency and digital understanding

Digital literacy can be defined as 'the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills' (ALA, 2022). Digital understanding, on the other hand, is knowing the *impacts* of using digital technologies (doteveryone, 2018b). This can include 'algorithmic literacy' – an understanding of the processes that may be at play behind the scenes and the use of critical evaluation to assess the credibility of the information (Swart, 2021). Therefore, improving both digital literacy and digital understanding is *'essential to rebalance the power between public and technology and allow people to question the way technologies shape society'* (doteveryone, 2018b, p 22). The 5Rights framework[2] stresses digital literacy as a 'digital right', and they champion the need to promote digital understanding, including more education on digital and data literacy in schools.

There have also been demands for better digital education in schools from elsewhere (Hepburn & Christie, 2019). In the UK the ICO has created lesson plans for schools to educate on data literacy, which include how to keep personal and private information safe, as well as information on digital rights (ICO, 2022). Other educational resources have been created outside of schools to allow children, young people, and the general public, to take control of their digital literacy and understanding[3]. Exposure to tools that make the data-collection and sharing behaviours of smartphone apps clearer and more transparent, have been shown to increase users' confidence in their choices (Van Kleek et al., 2017).

However, the aspect of digital literacy that has drawn attention in the regulatory and policy making sector is largely related to children's online safety (e.g., cyberbullying and grooming). This does not focus on other issues of online safety, or teach young people *'to protect their online privacy, encourage control over disclosure practices, and consider the potential commercial and non-commercial uses of their information'* (Bryce & Klang, 2009, p. 1). This paper brings to the forefront the voices of children and young people, detailing their recommendations demanding more choices and control over their online data and meaningful digital transparency.

Responsible innovation approach to the research

A thorough description of the history, interpretation, use, and current challenges of Responsible Research and Innovation (RRI) and Responsible Innovation (RI) can be found in (Owen & Pansera, 2019), where the authors describe some of the challenges that RI has encountered, and the distinction between the policy-led RRI, versus the academic foundation of RI. Within the UK the RI AREA - Anticipate, Reflect, Engage, Act - framework remains the most commonly-used approach. Throughout this study, therefore, we applied the principles of RI as operationalised in the AREA framework (Stilgoe et al., 2013; Owen, 2014; Jirotka et al., 2017).

In line with these RI principles, the research opened up the discussion in the spirit of RI "Engagement" to include the voices of young people. The study gave young people a participatory voice to contribute their experiences to the debate on the responsible design and governance of digital technologies. Key elements of this study also included "Reflection" on the choice of methodology to use, i.e., Youth Juries, that allowed participants to share ideas, discuss, reflect, and develop critical thinking. We also reflected on the need to use age-appropriate educational materials that could best elicit discussion about the impact of algorithms in participant's lives and co-created those resources with young people. Based on these reflections we "Acted" to create an advisory group with young people recruited from the first wave of this study, who volunteered and were keen both to learn more and to contribute to raising awareness. Engagement with the advisory group was key for the study as they co-designed the second wave's workshop activities and educational resources with the researchers. The research design was thus highly participatory. During this study we engaged with multiple stakeholders and acted on young people's concerns to additionally create educational resources to be used beyond the study timeline and provided research evidence to inform policy making/regulation in the UK (Horizon Digital Economy Research Institute & UnBias, 2018).

Methodology

This paper reports results from a study that gave children and young people a space to share their experiences when online, to explore their concerns, and for them to provide recommendations for change. This was done through a focus group method known as Youth Juries, which consists of youth-led discussion, debate, and interactive sessions. This paper draws on results from two waves of juries with a total of 25 sessions. The study was approved by the Ethics Review Board at the lead author's institution.

Participants

Recruitment took place through newsletters and emails to schools from the Midlands in the UK, and the sessions were conducted at either the participants' schools or at the University. All participants were thanked with a high street shopping voucher, a certificate of participation, and a leaflet providing more information about online tracking and other issues. In Wave 1, a total of 14 two-hour juries were conducted with 144 participants aged between 12 and 23, with a mean age of 15 years old. 50.7 % were males; only 4 participants were aged 18 or over. In Wave 2, a total of 11 two-hour juries were conducted with 116 participants aged between 13 and 18, with a mean age of 14.5 years. 59.5 % were males (Table 1). The average number of participants per session was 10.

Procedure

The Youth Jury method is empirically proven to be an effective way to involve young people in activities that centralise the importance of discussion and reflection, encouraging participants to share experiences, debate issues, form and change opinions, and put forward recommendations for improving or solving an issue or dilemma (Perez Vallejos et al., 2016). The juries were designed to enable the tackling of challenges related to the online world by engaging young people with real problems that they may encounter and enabling them to reflect on their own online behaviours.

Table 1
Summary of workshop participant demographics.

	Youth Juries n	Participants n	Females n (%)	Males n (%)	Mean age (SD)
Wave 1	14	144	67 (49.3)	77 (50.7)	15 (2.0)
Wave 2	11	116	47 (40.5)	69 (59.5)	14.5 (1.5)
Total	25	260	114 (43.8)	146 (56.2)	14.8 (1.7)

[2] https://5rightsfoundation.com/about-us/the-5-rights/, as of March 31, 2024

[3] Digital literacy educational resources: https://www.childnet.com/resources; https://www.lse.ac.uk/my-privacy-uk; https://saferinternet.org.uk/guide-and-resource/young-people as of March 31, 2024.

Both waves of juries started with a multiple-choice questionnaire, to assess the participants' level of knowledge and opinions about different aspects of the Internet. The juries ended with a second questionnaire, to establish any change in attitudes, to gauge learning and interest, and to consolidate some of the opinions discussed during the sessions. The questionnaires were modified for clarity in consultation with an advisory group of young people, following a pilot study in between the two waves of juries (Dowthwaite et al., 2020). This means the questionnaires differed across the two waves, but covered the same topics, and many of the questions remained the same or have close equivalents. In reporting, it is made clear which wave results are drawn from, or if it was a combination of both.

Wave 1 (W1)

These juries focussed on youth-led discussion and debate, with the facilitator using slides to introduce the concept of algorithms, and topics that encouraged young people to talk about their own experiences and voice their opinions. Participants were asked how they would like to see things change or issues to be tackled. The main body of the juries focussed on three main topics:

- Personalisation through algorithms: considered the use of data by online platforms and services, including filter bubbles and echo chambers.
- Autocomplete, search results, and fake news: considered how algorithms governing autocomplete and search results can become biased.
- Algorithm transparency and regulation: used a hypothetical legal case surrounding the results of an algorithm causing real harm to an individual (Carter, 2016) and asked participants to act as jurors.

The final discussion surrounded participant's queries and/or concerns about how algorithms work, what meaningful transparency is to them, and what such transparency would 'look like'. Finally, they were asked who they thought was responsible for ensuring that transparency is achieved. The sessions ended with a summary and participants provided recommendations and suggestions for increasing fairness and preventing bias in any or all of the scenarios presented.

Wave 2 (W2)

These juries focused on interactive activities to promote discussion and were co-created with the advisory group. The pre-session questionnaire was followed by introductory activities based on what the participants did online. This involved participants 'mapping' their online activity using stickers of common website logos on paper which provided context to the juries. The concept of algorithms was then introduced using a 'black box' (Fig. 1), which has an 'input' slot, but the contents of the box are obscured. This box was visible and referred to throughout the discussions of algorithms that followed. Further activities were based around 3 themes:

- Online personalisation: considered the use of data by online platforms and services, using a set of A6 laminated 'data cards' listing different types of data, and 'filter bubble' worksheets, both co-created with the advisory group and freely available online[4].
- The Regulation of Algorithms: used real-life and hypothetical examples of online bias, to discuss who was to blame, what should be done, and who was responsible for ensuring it is resolved.
- Algorithm Transparency: returned to the 'black box' to discuss what kind of transparency was meaningful and useful to participants.
- As in W1, the sessions ended with a summary, recommendations, and a final questionnaire.

[4] UnBias Youth Juries resources: https://uyj.wp.horizon.ac.uk/overview/ as of March 31, 2024.

Analysis

All of the Youth Juries were audio recorded with the permission of participants and a total of 50 hours recorded data was transcribed by a GDPR compliant external company, and then thematically analysed (Braun & Clarke, 2006). A single researcher used an inductive approach to code all the transcripts, before grouping them into themes using NVivo software. Three other members of the research team independently coded three random transcripts each. The results were validated collectively as a team, in order for any discrepancies to be discussed and reconciled. The coding was found to be consistent between all coders with only minor discrepancies being found and resolved. Key themes were then identified from this analysis. Written recommendations and notes were also added to the qualitative dataset and coded in the same way. Questionnaires were analysed using SPSS software. The main analysis is descriptive comparisons and summary data.

Findings

Knowledge and experience of algorithms

Most participants across both waves accessed search engines several times a day (80.4 %). However, nearly half felt they only knew only a little about how such services rank information (45.4 %), and a quarter knew not much or nothing at all (24.2 %). They did, however, feel that such knowledge was important (67.7 %), highlighting a common disparity between what users value, and what providers offer. Over the course of each jury, we learnt what the participants understood by the word 'algorithm' and about some of the functions and processes related to them, such as the collection and sale of data to third party companies.

Across all Youth Juries (YJ), participants were able to give a definition of an algorithm. A popular way to describe it was through a relatively technical understanding: *'a code with a purpose'* (W2, YJ6) and *'a series of steps you take to derive an output from an input'* (W1, YJ2). Definitions also focussed on the function of an algorithm, many of which were framed through the lens of their experiences of personalisation: *'Is it like how they choose what we see? Like, on Instagram…it's like what is popular'* (W2, YJ1). One participant framed their understanding of an algorithms' purpose through the lens of surveillance:

> *I feel like it's just basically like monitoring your life […] you're going on something and then it's recording and like monitoring what you're doing, and what you're clicking on and everything you're doing. (W2, YJ4).*

However, the level of knowledge related to how algorithms actually function was variable within and across juries. A few participants were open in declaring their lack of understanding of what an algorithm is, e.g., *'never heard of it'* (W1 YJ12), *'I think it's something to do with maths'* (W1, YJ13). Others often highlighted a lack of education and awareness around the subject: *'Possibly because not many people are educated on it'* (W1, YJ7) – an indicator of the need for the digital literacy and digital understanding discussed above. Another participant stressed their lack of knowledge of things such as targeted advertising: *'I get adverts for this 'Wish' site, and I don't know why'* (W2, YJ10).

Benefits and drawbacks of algorithmic tailoring

Participants were encouraged to discuss their personal experiences of algorithms. The key benefits surrounded automatic personalisation through algorithms: *'like maybe when you read something maybe you liked it and then they will pop up loads of like similar ones'* (W1, YJ1). This included removing undesirable content: *'I like social media and the filter because it filters out any stuff that I don't really want to see'* (W2, YJ2). Some also felt that creating a feed of shared interests and opinions was beneficial, and this was intertwined with the notion of convenience:

Fig. 1. The 'black box' algorithm, into which data is 'collected'. The box is fully covered throughout the session (left) to show that users do not know what happens to their data, whilst at the end of the session illustrations of providing 'transparency' to the algorithm are presented (right).

> *I mean isn't it kind of human nature to be selective towards the things you agree with? So, in a way it's kind of just cutting out the middle-man, not making you trawl through a load of articles to find stuff you agree with* (W1, YJ9).

According to the surveys, half of the participants across both waves (49.2 %) wanted their internet experiences to be more personalised; 43.9 % agreed that *'sites that recommend things make my life easier. I save time and effort'*. However, despite many participants recognising the benefits of personalisation, 45.0 % agreed that *'websites should not influence which information is given to users'*. Many participants pointed out their concerns about creating a filter bubble that only reflects the user's views and interests. Participants felt it was possible that *'sometimes there might be important stuff in the world going on that might be filtered out for me for some reason when actually I might be quite interested in it'* (W1, YJ1). Some participants felt that they were missing out; one participant expressed the irony that the Internet is marketed as something that can expose the user to endless possibilities of information, yet categorise users so they may only see a part of it:

> *They tell you to explore the world and everything, but then when they're, like, making stuff for you, it's […] limited everything* (W2, YJ10).

Others also noted that not seeing opposing views could be damaging:

> *If you […] don't get to see other people's, like, different opinions, so if someone disagrees with you, you are completely unaware about it* (W2, YJ11).

This demonstrates that some young people are able to make highly context-specific judgements about what type of algorithmically-driven recommendation might be appropriate, and when. The concern around seeing alternative views demonstrates an important understanding around the bias that can drive online interaction.

Another key drawback that was raised in all the juries is that attempts to tailor content towards the user's preferences are not always accurate, with participants often discussing advertisements they felt were completely irrelevant:

> *…well all my filter bubble is like Wix.com and pregnancy tests and then make up like L'Oreal and stuff and Netflix and I don't want that* (W2, YJ2).

Other participants believed that algorithms often missed the nuances in their searches, which they attributed to inaccuracy in profiling (despite the volumes of data collected), and some felt that the algorithms '*just kind of assume*' (W1, YJ10) the information that they would like to see. A few participants commented that this culminated in them feeling annoyed and frustrated by the algorithm persisting in recommending things that they do not like or wish to see.

Perceptions of algorithmic control and the need for transparency

Related to this last complaint around inaccuracy, some participants were concerned about the presumptions made by the algorithm-driven platforms related to what users want to see:

> *Why is it their decision to decide that? […] why should they be the ones to say, well, no, you can't see that because we don't think it's important for you? It's just like, well, it might be really important for you* (W1, YJ1).

Across many juries, participants had concerns about the amount of power and control held by companies due to the way that algorithms collect data, leading one participant to suggest that *'it seems like they're controlling you and they already know what you're going to do, so you follow what they show you'* (W1, YJ11), whilst another felt that *'they have control over what you see'* (W1, YJ12). According to the surveys, most of the participants in W2 agreed that the government should do more to make sure both the Internet and the digital world are safe and neutral (72.4 %); nearly half (45.7 %) thought that the big tech companies had a lot of power over their lives, and two thirds of those (67.9 %, 31.0 % of total) felt that it was too much. Just under half of the participants in W2 (46.6 %) felt the big tech companies had too much power overall.

All the juries agreed that some form of transparency in terms of **how algorithms were used** by companies and **how they worked** was a key element for a fairer and more unbiased online environment, providing a sense of control in making decisions: *'we should decide too*' (W1, YJ10). This was crucial for their feelings of safety online. One participant pointed out that increased algorithmic transparency was important to increase the accountability of companies, helping to ensure that discrimination does not take place:

> *there's discrimination laws in employment but when it comes as an algorithm that's a lot harder to realise, so like it's one thing firing someone because they're a woman etc. but it's another thing discriminating against her through your algorithm like that's not something that they can easily prove (*W1, YJ7).

Throughout the juries, participants were asked to consider how algorithmic transparency should be managed. Some believed that it was important for both parties (internet services/platforms and individuals) to take responsibility:

> *[It's] the responsibility of people to find out more. The responsibility of the Internet is to provide you with all the information available in an unbiased manner when you search for it. (W1, YJ9).*

Some felt that governments should exert more control, ensuring that companies provided documentation of their algorithms, but not necessarily controlling its function. Others raised concerns about the practicalities of implementing a system of transparency, particularly the compounding effect introduced by the speed of change:

It's how can you keep up with the technology and the changes that are happening all the time? (W1, YJ1)

Throughout the sessions it was evident that participants had a desire for a fairer online environment, and that accessible algorithmic transparency should be provided to users. However, one participant pointed out that the meaning of algorithmic transparency can vary according to many factors, including the motivations of different stakeholders, or the age group and preferences of individual users. They raised concerns that algorithmic transparency may be difficult to achieve if people are divided on what they mean by this term: *"they'll say well we've made our algorithms more transparent, and for a lot of people they don't know what that means"* (W1, YJ10). The YJ's conclusions point towards not simply transparency, but *meaningful* transparency.

Achieving meaningful transparency

In order to facilitate the discussion around what transparency 'meant' to them, participants were shown different versions of the way that the decision-making mechanisms that underpin an algorithm may be depicted. Participants were shown a flowchart, a picture of computer code, and the equation Facebook used to explain the algorithm (EdgeRank) that dictates what appears in the news feed. Participants gave feedback on each representation of the algorithm in terms of how transparent and useful it would be to them to aid their understanding. Opinions were mixed on using diagrams such as flowcharts to explain what was happening. A few responded positively but others felt that such a diagram could be used to avoid real transparency: *'I think it's trying to tell you as little information as possible'* (W1, YJ7). Similarly, when one participant was asked if Facebook has met their obligations for transparency through the visualisation of an equation, they explained *'they have, but they haven't shown enough of stuff'* (W2, YJ1). Another agreed, saying, *'it's really vague. It doesn't really answer all your questions'* (W2, YJ4). Others expressed frustration that the equation itself was hard to understand: *'you have to explain it, so it's not exactly simple'* (W2, YJ11), demonstrating the need for 'accessible' explanation.

Providing source code was met with a lack of enthusiasm, as a participant explained:

> … *realistically 95 % of people aren't going to even know what it means* (W2, YJ4).

They also recognised that companies wouldn't want to share their code and that if they did so, users would be *'overloaded with information that's irrelevant'* to them (W1, YJ9). One participant however, took a pragmatic approach in noting that if the platform is asked to reveal their algorithm, then they have met this demand and that to then *'put it in terms that everyone can understand [would be] almost impossible'* (W2, YJ4). Others believed that the code would not provide any form of meaningful transparency: *'Don't release the code, release what the code does'* (W2, YJ5). In fact, throughout the sessions, participants were clear-sighted about wanting to know *why* algorithms behave in a certain way:

> *it's useless being given the algorithm itself; you need to be told why the algorithm did that* (W1, YJ9).

At the end of the juries, participants were asked to consider how algorithmic transparency could be achieved in a useful manner. They pointed out that users would respond to different formats in different ways, and suggested that information about the algorithm should be presented in a variety of ways, providing choices to users:

> *So, you could have a video, you could have bullet points, so whatever, whichever suits a person you should have more ways of doing it. It will give a wider range for anyone that wants to view it can view it in the way they prefer to view it* (W2, YJ6).

Participants often highlighted the importance of knowing the purpose of the algorithm (i.e., functionality) and what data is being used for that purpose:

> *The …computing behind it isn't that important to us, it's more about like why it's happening and what information from us is being used* (W1, YJ7).

It also highlighted an additional need for the ability to assert some control over functionality:

> *Being more open about […] how it's working and why is it recommending this particular thing to you. For example, I think Amazon […] has an option to 'fix this recommendation' and then it will ask you why you don't want to be recommended that thing* (W1, YJ8).

Many of the discussions about achieving transparency were also intertwined with the ability to control their personal data more effectively: *'I think you should have the freedom to decide what you've been sharing'* (W1, YJ9). Some participants also expressed that they would like to be able to decide the type of content they see online: *'I'd feel like a bit more power over my own like recommendations I suppose. So, I could choose what ads I saw'* (W1, YJ4). It was generally felt that this control would be more achievable with access to more information. As such, a lot of the recommendations for transparency revolved around data handling and sharing practices:

> *…if they are going to sell it […]I'd maybe like to know where it went or who was going to use it* (W1, YJ7).

They would like procedures to be made clearer to the user, in particular by improving the terms and conditions of online user agreements:

> *The terms and conditions are so long that no one really wants to read them. There should be clearer, shorter conditions for people to read and a requirement for customers to read them* (W1, YJ10).

Throughout this study participants highlighted the need for more education on this topic to improve transparency, feeling that increased awareness of how algorithms work and what kinds of data might be collected about them would help with these efforts: *'Inform the public about algorithms and educate them'* (W1, YJ6). Some participants believed that schools played an important role:

There were also suggestions that this education should be extended beyond schools, to be accessible to others, including adults, who would benefit from this knowledge:

> *I think there should probably be some kind of schemes put up that people can choose to find out more about it* (W1, YJ11).

It is often presumed that children and young people as 'digital natives' (Davies et al., 2012), are so familiar with using the Internet that they do not care about their personal data, but research has shown that this is not the case (Coleman et al., 2017; Dowthwaite et al., 2020; Livingstone et al., 2019; Silva et al., 2017). Our study supported this, as participants showed a strong desire for transparency, mainly focused on learning how companies collect and manage their personal data, and understanding how algorithms function in decision-making processes, in a way that is accessible to them (Boyd & Marwick, 2011; Dowthwaite et al., 2020; Pangrazio & Selwyn, 2017). They want to know what data is collected and why, and who their personal information is shared with beyond the platform or service. Concern about third-party data sharing has also been seen in previous research (Hoofnagle et al., 2010; Fiesler & Hallinan, 2018).

Participants also felt they had a lack of choice and control over their data and the online content that is shown to them, which led to feelings of powerlessness and resignation, which have also been found in adults (doteveryone, 2018a, 2018b, 2020; Fiesler & Hallinan, 2018; Turow et al., 2015).

Participants' desire for understanding of high-level, or 'results'-based algorithmic function, as opposed to a release of source code or other proprietary technical information, agrees with recommendations

to the European Parliament on algorithmic transparency and accountability (Koene et al., 2019). Concrete explanations are the most helpful type when trying to understand how automated systems work (Kulesza et al., 2013). Building transparency through understanding of how algorithmic systems work should be a priority for technology designers, governments, and regulators worldwide.

Discussion

Algorithmically-generated content, timelines, recommendations, news etc, can be an enormously helpful way of dealing with the vast amount of content online (over 4.64 billon pages[5]) and allowing internet users to interact with a section of the online world that is relevant to them. In effect, algorithms seek to produce a curated experience for users. However, considering the large number of children and young people that are regularly engaged with online content, there has been too little research to date about their awareness of the implications of the algorithmically-mediated decisions they are exposed to daily, how they experience these online systems, and where they might have concerns. This paper has reported on a series of focus groups that examined young people's experiences of and attitudes towards algorithmically-mediated systems, and highlights their recommendations for increased and meaningful algorithmic transparency and accountability.

The importance of education

Participants' demand for transparency was intertwined with their requirement for education to improve digital understanding – a concern shared by many working in this area, discussed above. They felt that algorithmic behaviour and the consequences of personal data sharing with online companies should be taught in schools and through other resources available to all. The lack of knowledge and/or understanding of these issues puts young internet-users in a vulnerable position. Introducing these topics as part of digital literacy programmes in schools would help young people to comprehend the impact of algorithms in their lives and to make informed decisions when online. Digital literacy promotes digital citizenship (Carretero et al., 2018), and education fosters knowledge-inclusion, which also may empower citizens to contribute to processes of RI governance (Valkenburg et al., 2020).

In the UK, moves to include more data literacy in school curricula are underway (Department for Digital, Culture, Media and Sport, 2021; Hepburn & Christie, 2019). The UK government and online safety regulator (Ofcom) are working to improve national media literacy capabilities and help citizens managing their online safety (Ofcom, 2023; Year 2 Online Media Literacy Action Plan (2022/23), n.d.).

Outcomes from this study include sustainable educational materials co-created with young people and freely available online: the 'black box' algorithm and data cards (Dowthwaite et al., 2020) and the UnBias Fairness toolkit[6], developed for our research project with a designer in the UK, in collaboration with young people and researchers. A recent extension to the UnBias Fairness toolkit is the UnBias AI for Decision Makers toolkit[7] developed by Giles Lane, and commissioned by Ernst & Young Global Ltd, UK.

Transparency as a collective responsibility

To our knowledge this is the first study that incorporated specifically RI-based approaches, to include young people as active agents, co-creating, reflecting and providing their recommendations to influence the future design and governance of ICT.

Participants pointed out that algorithmic transparency may mean different things to different stakeholders. However, they agreed that transparency of some kind is important for helping users to obtain and maintain control over their online choices, and to hold companies accountable for their processes, as previous calls for transparency have emphasised (Koene et al., 2019; Lepri et al., 2018).

The majority of participants in W2 agreed that the government should do more to ensure that the Internet was safe for young people. A large part of creating a safe space online should be through making users aware of how their information is being used and how the content they see is created, in a transparent and accessible way, and to ensure that regulation for that is in place. Some participants also believed governments should ensure that tech companies provide meaningful algorithmic transparency. This matches the current public demand for greater accountability from technology companies and regulatory bodies (ACM, 2017; House of Commons, 2019; European Parliament, 2020) and is in line with the international guidelines aimed at creating awareness of the societal and ethical implications of AI (UNESCO, 2020).

Participants expressed their conviction that requirements for transparency have not yet gone far enough. They did however acknowledge that with the rapid development of online technologies, implementation of transparency can be challenging. This makes it more unlikely that young people are fully equipped to manage their digital life and choices. It would also be unfair to expect children and young people to be solely responsible for managing and protecting their online activities, and *'for handling the complex commercial environment'* (Livingstone et al., 2019, p. 35). Governance bodies must promote, and the tech industry must adopt, transparency of online systems by design and by default. Policy changes should be enforced to achieve not only meaningful but sustainable transparency, in agreement with Van Kleek et al. (2017).

This study demonstrated the value of two-way dialogue with the public by listening to young people's discussion of their needs online and responding to their concerns by creating educational materials. The study also enabled the authors to provide research evidence (Horizon Digital Economy Research Institute & UnBias, 2018) to the ICO when they were drafting the Age Appropriate Design Code, enforced in the UK (ICO, 2020). Young people's demands from this study are in line with the standards included in the Code. Compliance with the Code requires providers to offer transparent and accessible information to children about what to expect when they access that service, to minimise data collection, and to provide strong default privacy settings. Implementation of the Code in the UK is underway, and it represents an important milestone in children's online safety and data protection worldwide.

Value of a responsible innovation approach

Our RI approach enabled us to anticipate possible benefits and challenges that young people may encounter when interacting with algorithmic systems, which led to the identification of the topics and to development of the scenarios used in this study. Anticipation, stakeholder inclusion, reflexivity, and responsiveness were key to this work. Co-creation with young people of age-appropriate educational materials helped research participants to better understand how online algorithms operate, their impact in decision-making processes, and the importance of data in these processes, encouraging reflections about control of their personal data. Our user-centred commitment brings to the forefront young people's recommendations for meaningful algorithmic transparency and accountability to contribute towards shaping an inclusive, agile, and responsible digital economy.

In the spirit of RI, more interdisciplinary work is needed to better anticipate the challenges faced by young people when exposed to online technologies, to promote more formal avenues for engagement and action to address children and young people's needs, and to protect their human rights. In the UK, an example of work delivered by a

[5] https://www.worldwidewebsize.com/ , as of 31 March 2024.

[6] UnBias Fairness toolkit: https://unbias.wp.horizon.ac.uk/fairness-toolkit/ as of March 31, 2024.

[7] UnBias AI for Decision Makers toolkit: http://proboscis.org.uk/6346/ai-for-decision-makers/ as of March 31, 2024.

multidisciplinary group within the digital sector, is the Algorithmic Transparency Recording Standard (Algorithmic Transparency Recording Standard - Guidance for Public Sector Bodies, 2024), a guidance for public sector bodies promulgated by the Advisory Board[8] of the Centre for Data Ethics and Innovation, currently the Responsible Technology Adoption Unit.[9] Also, more action is also needed to include lay citizens on RI practices within this field, and to promote agile and sustainable changes for a fairer, more transparent, and responsible digital environment. The current state of the markets, the global challenges surrounding effective governance of digital tech (existing and emerging), and its overall impact on citizens, urges for a responsible approach to the design, deployment, decommissioning, and regulation of all digital technologies.

According to our study, we summarised below children and young people's Call for Action for meaningful algorithmic transparency and accountability, in three action areas:

1. Online companies should provide usefully transparent information on algorithmic function, and do this in an accessible way for their users. This should explain why conclusions have been reached or inferences drawn, in ways that allow users to engage with curating their own online experience. Tech companies should also detail how, why, and what personal data is being used, and inform users of the possible consequences of data sharing, including who users' information is shared with.
2. Users should have more agency in shaping their online lives, including more control over their personal data and over recommended content.
3. Governments must exert more control to make sure the Internet and digital world are safer -in particular- for young people, ensure tech companies comply with *meaningful* algorithmic transparency requirements, and ensure digital literacy and understanding is available to all.

These actions are not significantly different that those reported as desirable by older demographics. This itself is worthy of note, given the frequent characterisation of young people as 'digital natives' who are both relatively unconcerned with the amount of data being harvested and held by tech companies (Engels, 2019) and have a relatively high understanding of the ways content is surfaced to them. However, these requirements demonstrate that this narrative around unconcerned 'digital natives' is essentially flawed, and there is in fact an unanimity of voices calling for better and more meaningful transparency.

To make the above actions possible there is a need for a responsible, user-centred regulatory body that includes inter and multi-disciplinary stakeholders. Such a body would oversee the implementation of sustained algorithmic transparency by providers, including the adoption of meaningful transparency practices by default. Agile regulatory mechanisms should also be in place with associated measures to ensure compliance. For instance, the regulator should ensure online systems providers to apply a foresight approach throughout the life cycle of their technology, to anticipate the impact of their systems to young people, and to provide evidence of action taken to ensure young people would gain the most benefit of those systems (including meaningful transparency and more user agency by design and default), while minimising risks and preventing harms to end users. Promoting statutory education action is also key, to allow young people and citizens to understand what is happening 'behind the scenes' to help them making informed decisions when online, and to provide a safer Internet for all.

[8] List of Advisory Board Members of CDEI, UK, as of March 31, 2024

[9] Responsible Technology Adoption Unit (RTA), UK, as of March 31, 2024

Conclusions

This study shows that promoting useful and meaningful transparency is important to young people in many respects. In particular, they wish to understand the 'how' and 'why' of algorithmic function and the 'what' and 'who' of data management by online companies. This would help young people - and by extension all internet users - to gain more control over their personal data to make better decisions online. The requirement for relevant action to be taken to increase transparency empowers young people as possessing agency in the public conversation about the rights of citizens and the governance of online systems.

This work also showed that transparency is key but needs to be operationalised in tandem with education and increased choices and control for users, and it needs to be responsibly implemented and regulated. Technology designers, educators, governments, policymakers, regulators and researchers must consider what this underrepresented and often vulnerable group of online users needs and respond to those needs.

CRediT authorship contribution statement

Virginia Portillo: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Liz Dowthwaite:** Writing – review & editing, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Helen Creswick:** Writing – review & editing, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elvira Pérez Vallejos:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Carolyn Ten Holter:** Writing – review & editing, Validation. **Ansgar Koene:** Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization. **Marina Jirotka:** Writing – review & editing, Funding acquisition, Conceptualization. **Jun Zhao:** Writing – review & editing.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Acknowledgments

This project was supported by the UKRI under grants: EP/N02785X/1, EP/T022493/1 and EP/M02315X/1.

Data access statement: due to ethical concerns and the sensitive nature of the data - children's reserach data- it will not be available at repositories.

References

5Rights Foundation. (2019). *Towards an internet safety strategy*. https://5rightsfoundation.com/uploads/final-5rights-foundation-towards-an-internet-safety-strategy-january-2019.pdf.

ACM. (2017). *Statement on algorithmic transparency and accountability*. https://www.acm.org/articles/bulletins/2017/january/usacm-statement-algorithmic-accountability.

ALA. (n.d.). *Digital literacy*. Retrieved 13 May 2022, from https://literacy.ala.org/digital-literacy/.

Algorithmic Transparency Recording Standard—Guidance for Public Sector Bodies. (n.d.). GOV.UK. Retrieved 3 April 2024, from https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Angwin, J., Larson, J., Mattu, S., Kirchner, L., & ProPublica. (2016). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=lZ_nPrh6oVJEnMzcTH1Jr59Ibe3K8XZC.

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled

online for personalization. *MIS Quarterly, 30*(1), 13–28. https://doi.org/10.2307/25148715

Barnardo's. (2019). *Generation Digital. 30 years since the Children Act 1989 & the birth of the World Wide Web*. https://www.barnardos.org.uk/sites/default/files/uploads/Generation%20Digital%20-%20report.pdf.

Bentley, H., Burrows, A., Hafizi, M., Kumari, P., Mussen, N., O'Hagan, O., & Peppiate, J. (2019). *How safe are our children? 2019*. NSPCC. https://learning.nspcc.org.uk/media/1747/how-safe-are-our-children-2019.pdf.

Bodo, B., Helberger, N., Irion, K., Borgesius, K. Z., Moller, J., Velde, B.van de, Bol, N., Es, B.van, & Vreese, C.de (2018). Tackling the algorithmic control crisis -the technical, legal, and ethical challenges of research into algorithmic agents. *Yale Journal of Law and Technology, 19*(1). https://digitalcommons.law.yale.edu/yjolt/vol19/iss1/3.

Boyd, D., & Marwick, A. E. (2011). *Social privacy in networked publics: Teens' attitudes, practices, and strategies (SSRN Scholarly Paper ID 1925128)*. Social Science Research Network. https://papers.ssrn.com/abstract=1925128.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brunton, F., & Nissenbaum, H. (2015). *Obfuscation: A user's guide for privacy and protest*. MIT Press.

Bryce, J., & Klang, M. (2009). Young people, disclosure of personal information and online privacy: Control, choice and consequences. *Information Security Technical Report, 14*(3), 160–166. https://doi.org/10.1016/j.istr.2009.10.007

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), Article 205395171562251. https://doi.org/10.1177/2053951715622512

California Constitution. (2018). *The California consumer privacy act of 2018*. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375.

Carretero, S., Vuorikari, R., & Punie, Y. (2018). *DigComp 2.1: The digital competence framework for citizens with eight proficiency levels and examples of use. (EUR 28558)*. Publications Office of the European Union. https://doi.org/10.2760/38842

Carter, J. (2016). *Could a computer algorithm be put on trial?* TechRadar. https://www.techradar.com/news/world-of-tech/could-a-computer-algorithm-be-put-on-trial-1325901.

CDEI. (2020). *Centre for data ethics and innovation review of online targeting*. https://www.gov.uk/government/publications/cdei-review-of-online-targeting.

Children's Commissioner for England. (2018). *Who knows what about me? A Children's Commissioner report into the collection and sharing of children's data*. https://www.childrenscommissioner.gov.uk/wp-content/uploads/2018/11/cco-who-knows-what-about-me.pdf.

Coleman, S., Pothong, K., Perez Vallejos, E., & Koene, A. (2017). *'The internet on our own terms: How children and young people deliberated about their digital rights'*. https://5rightsfoundation.com/uploads/internet-on-our-own-terms-a4.pdf.

Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction, 18*(5), 455–496. https://doi.org/10.1007/s11257-008-9051-3

Creswick, H., Dowthwaite, L., Koene, A., Perez, V. E., Portillo, V., Cano, M., & Woodard, C. (2019). … They don't really listen to people": Young people's concerns and recommendations for improving online experiences. *Journal of Information, Communication and Ethics in Society, 17*(2), 167–182. https://doi.org/10.1108/JICES-11-2018-0090

Davies, H. C., Halford, S. J., & Gibbins, N. (2012). Digital natives?: Investigating young people's critical skills in evaluating web based information. In *Proceedings of the WebSci '12: 3rd annual ACM web science conference* (pp. 78–81). https://doi.org/10.1145/2380718.2380729

de Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology, 31*(4), 525–541. https://doi.org/10.1007/s13347-017-0293-z

Department for Digital, Culture, Media and Sport. (2021). *Online media literacy strategy*. July 14. GOV.UK https://www.gov.uk/government/publications/online-media-literacy-strategy.

DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). Algorithms ruin everything': #RIPTwitter, folk theories, and resistance to algorithmic change in social media. In *Proceedings of the CHI '17: 2017 CHI conference on human factors in computing systems* (pp. 3163–3174). https://doi.org/10.1145/3025453.3025659

Dogruel, L., Facciorusso, D., & Stark, B. (2020). 'I'm still the master of the machine.' Internet users' awareness of algorithmic decision-making and their perception of its effect on their autonomy. *Information, Communication & Society*, 1–22. https://doi.org/10.1080/1369118X.2020.1863999

doteveryone. (2018a). *People, power and technology: The 2018 digital attitudes report*.

doteveryone. (2018b). *People, power and technology: The 2018 digital understanding report*. https://doteveryone.org.uk/wp-content/uploads/2018/04/Doteveryone_PeoplePowerTechDigitalUnderstanding2018.pdf.

doteveryone. (2020). *People, power and technology: The 2020 digital attitudes report*. https://www.doteveryone.org.uk/wp-content/uploads/2020/05/PPT-2020_Soft-Copy.pdf.

Dowthwaite, L., Creswick, H., Portillo, V., Zhao, J., Patel, M., Perez, E., Koene, A., & Jirotka, M. (2020). It's your private information. It's your life." young people's views of personal data use by online technologies. In *Proceedings of the IDC '20: interaction design and children conference* (pp. 121–134). https://doi.org/10.1145/3392063.3394410

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society, 21*(5), 729–745. https://doi.org/10.1080/1369118X.2018.1428656

Eiband, M., Völkel, S. T., Buschek, D., Cook, S., & Hussmann, H. (2019). When people and algorithms meet: User-reported problems in intelligent everyday applications. In *Proceedings of the IUI '19: 24th international conference on intelligent user interfaces* (pp. 96–106). https://doi.org/10.1145/3301275.3302262

Engels, B. (2019). *Digital first, privacy second? Digital natives and privacy concerns*. 203–209. https://www.iadisportal.org/digital-library/digital-first-privacy-second-digital-natives-and-privacy-concerns.

EPIC. (n.d.). *AI Policy*. EPIC - Electronic Privacy Information Center. Retrieved 6 May 2022, from https://epic.org/issues/ai/ai-policy/.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). 'I always assumed that I wasn't really that close to [her]': Reasoning about invisible algorithms in news feeds. In *Proceedings of the CHI '15: 33rd annual ACM conference on human factors in computing systems* (pp. 153–162). https://doi.org/10.1145/2702123.2702556

Eslami, M., Vaccaro, K., Karahalios, K., & Hamilton, K. (2017). ' Be careful; things can be worse than they appear': Understanding biased algorithms and users' behavior around them in rating platforms. In *Proceedings of the ICWSM '17: eleventh international AAAI conference on web and social media* (pp. 62–71).

Eslami, M., Vaccaro, K., Lee, M. K., Elazari Bar On, A., Gilbert, E., & Karahalios, K. (2019). User attitudes towards algorithmic opacity and transparency in online reviewing platforms. In *Proceedings of the CHI '19: 2019 CHI conference on human factors in computing systems* (pp. 1–14). https://doi.org/10.1145/3290605.3300724

European Parliament. (2020). *Draft Report with recommendations to the Commission on a framework of ethical aspects of AI, robotics and related technologies*. https://www.europarl.europa.eu/doceo/document/JURI-PR-650508_EN.pdf.

European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. http://eur-lex.europa.eu/legal-con-tent/EN/TXT/?uri=OJ%3AL%3A2016%3A119%3ATOC.

Federal Trade Commission. (2016). *Big data: A tool for inclusion or exclusion? Understanding the issues* [FTC Report].

Fiesler, C., & Hallinan, B. (2018). We are the product': Public reactions to online data sharing and privacy controversies in the media. In *Proceedings of the CHI '18 2018 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3173574.3173627. Paper No. 53.

Fouquaert, T., & Mechant, P. (2021). Making curation algorithms apparent: A case study of 'Instawareness' as a means to heighten awareness and understanding of Instagram's algorithm. *Information, Communication & Society, 0*(0), 1–21. https://doi.org/10.1080/1369118X.2021.1883707

Garfinkel, S., Matthews, J., Shapiro, S. S., & Smith, J. M. (2017). Toward algorithmic transparency and accountability. *Communications of the ACM, 60*(9), 5. https://doi.org/10.1145/3125780

Gauch, S., Speretta, M., Chandramouli, A., & Micarelli, A. (2007). User profiles for personalized information access. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web: Methods and strategies of web personalization* (pp. 54–89). Springer. https://doi.org/10.1007/978-3-540-72079-9_2. Vol. 4321.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation. *AI Magazine, 38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Hepburn, D., & Christie, L. (2019). *Online safety education*. Houses of parliament. https://post.parliament.uk/research-briefings/post-pn-0608/.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the ACM conference on computer supported cooperative work* (pp. 241–250). https://doi.org/10.1145/358916.358995

Hoofnagle, C. J., King, J., Li, S., & Turow, J. (2010). *How different are young adults from older adults when it comes to information privacy attitudes & policies?* (p. 22). Annenberg School for Communication. http://repository.upenn.edu/asc_papers/399.

House of Commons. (2019). *Impact of social media and screen-use on young people's health*. [Fourteenth Report of Session 2017–19]. *House of Commons*. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/822.pdf.

Horizon Digital Economy Research Institute, UnBias and CaSMA. 2018. Response to Information Commissioner's Office "Age Appropriate Design Code" consultation.

ICO. (2022). *For organisations* (n.d.). ICO. Retrieved 6 Mayfrom https://ico.org.uk/for-organisations/.

ICO. (2020). Age appropriate design: A code of practice for online services. *Information Commissioner's Office*. https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services/.

Jirotka, 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

Koene, A., Clifton, C. W., Hatada, Y., Webb, H., Patel, M., Machado, C., LaViolette, J., Richardson, R., Reisman, D., European Parliament, European Parliamentary Research Service, & Scientific Foresight Unit. (2019). *A governance framework for algorithmic accountability and transparency: Study*. http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf.

Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I., & Wong, W.-K. (2013). Too much, too little, or just right? Ways explanations impact end users' mental models. In *Proceedings of the IEEE symposium on visual languages and human centric computing* (pp. 3–10). https://doi.org/10.1109/VLHCC.2013.6645235

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the CHI '15: 33rd annual ACM conference on human factors in computing systems* (pp. 1603–1612). https://doi.org/10.1145/2702123.2702548

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy & Technology, 31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

Livingstone, S., Cantwell, N., Özkul, D., Shekhawat, G., & Kidron, B. (2024). *The best interests of the child in the digital environment. Digital futures for children centre, LSE and 5Rights foundation. Digital Futures for Children Centre, LSE and 5Rights Foundation*. https://www.digital-futures-for-children.net/best-interests.

Livingstone, S., Carr, J., & Byrne, J. (2015). *One in three: Internet governance and children's rights (No. 22; Paper Series)*. Global Commission on Internet Governance. https://www.cigionline.org/sites/default/files/no22_2.pdf.

Livingstone, S., Davidson, J., Bryce, J., Batool, S., Haughton, C., & Nandi, A. (2017). *Children's online activities, risks and safety: A literature review by the UKCCIS Evidence Group*. https://www.lse.ac.uk/business-and-consultancy/consulting/assets/documents/childrens-online-activities-risks-and-safety.pdf.

Livingstone, S., Stoilova, M., & Nandagiri, R. (2019). *Growing up in a digital age: An evidence review*. London School of Economics and Political Science, Department of Media and Communications. http://eprints.lse.ac.uk/101283.

OECD. (2020). *Protecting children online: An overview of recent developments in legal frameworks and policies*. OECD Publishing. https://doi.org/10.1787/9e0e49a9-en

Ofcom. (2022). *Children and parents: Media use and attitudes report 2022*. https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2022.

Ofcom. (2023). *Making sense of media annual plan*. https://www.ofcom.org.uk/_data/assets/pdf_file/0027/256554/MSOM-Annual-Plan-2023-24.pdf.

Online Safety Act 2023. (2023). King's printer of acts of parliament. https://www.legislation.gov.uk/ukpga/2023/50/enacted.

Oulasvirta, A., & Blom, J. (2008). Motivations in personalisation behaviour. *Interacting with Computers, 20*(1), 1–16. https://doi.org/10.1016/j.intcom.2007.06.002. Interacting with Computers.

Owen, R. (2014). The UK engineering and physical sciences research council's commitment to a framework for responsible innovation. *Journal of Responsible Innovation, 1*(1), 113–117. https://doi.org/10.1080/23299460.2014.882065

Owen, R., & Pansera, M. (2019). Responsible innovation and responsible research and innovation. In D. Simon, S. Kuhlmann, J. Stamm, & W. Canzler (Eds.), *Handbook on Science and Public Policy* (pp. 26–48). Edward Elgar Publishing Ltd. https://doi.org/10.4337/9781784715946.

Pangrazio, L., & Selwyn, N. (2017). 'My data, my bad …': Young people's personal data understandings and (Counter)practices. *#SMSociety17: Proceedings of the 8th International Conference on Social Media & Society, 52*, 1–5. https://doi.org/10.1145/3097286.3097338

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin UK.

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Perez Vallejos, E., Koene, A., Carter, C. J., Statache, R., Rodden, T., McAuley, M., … Coleman, S. (2016). Juries: acting out digital dilemmas to promote digital reflections. *SIGCAS Comput. Soc., 45*(3), 84–90. https://doi.org/10.1145/2874239.2874252

Perez Vallejos, E., Koene, A., Portillo, V., Dowthwaite, L., & Cano, M. (2017). Young people's policy recommendations on algorithm fairness. In *Proceedings of the WebSci '17: 2017 ACM on web science conference* (pp. 247–251). https://doi.org/10.1145/3091478.3091512

Pew Research Center. (2018). *Public attitudes toward computer algorithms*. https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/.

Silva, C., Barbosa, G., Silva, I., Silva, T., Mourão, F., & Coutinho, F. (2017). Privacy for children and teenagers on social networks from a usability perspective: A case study on Facebook. In *Proceedings of the WebSci '17: 2017 ACM on web science conference* (pp. 63–71). https://doi.org/10.1145/3091478.3091479

Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). *EU Kids Online 2020: Survey results from 19 countries*. EU Kids Online. https://doi.org/10.21953/lse.47fdeqj01ofo

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

Stoilova, M., Nandagiri, R., & Livingstone, S. (2021). Children's understanding of personal data and privacy online – a systematic evidence mapping. *Information, Communication & Society, 24*(4), 557–575. https://doi.org/10.1080/1369118X.2019.1657164

Swart, J. (2021). Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media + Society, 7*(2), Article 20563051211008828. https://doi.org/10.1177/20563051211008828

Tindale, J., & Muirhead, O. (2019). Building data policies for the public good: Trust, transparency and tech. *Policy Connect*. https://www.policyconnect.org.uk/research/trust-transparency-and-technology-building-data-policies-public-good.

Turow, J., Hennessy, M., & Draper, N. (2015). *The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation* (SSRN Scholarly Paper ID 2820060). *Social Science Research Network*. https://doi.org/10.2139/ssrn.2820060

UNDESA. (n.d.). *Definition of youth*. Retrieved 5 May 2022, from https://www.un.org/esa/socdev/documents/youth/fact-sheets/youth-definition.pdf.

UNESCO. (2020). *First version of the draft recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000373434.

UNICEF. (2017). *Children in a digital world*. UNICEF. https://www.unicef.org/media/48581/file/SOWC_2017_ENG.pdf.

Valkenburg, G., Mamidipudi, A., Pandey, P., & Bijker, W. E. (2020). Responsible innovation as empowering ways of knowing. *Journal of Responsible Innovation, 7*(1), 6–25. https://doi.org/10.1080/23299460.2019.1647087

Van Kleek, M., Liccardi, I., Binns, R., Zhao, J., Weitzner, D. J., & Shadbolt, N. (2017). Better the devil you know: Exposing the data sharing practices of smartphone apps. In *Proceedings of the CHI '17: 2017 CHI conference on human factors in computing systems* (pp. 5208–5220). https://doi.org/10.1145/3025453.3025556

Wachter, S., & Mittelstadt, B. D. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review, 2*, 443–493. https://ora.ox.ac.uk/objects/uuid:d53f7b6a-981c-4f87-91bc-743067d10167.

Year 2 Online Media Literacy Action Plan (2022/23). (n.d.). GOV.UK. Retrieved 3 April 2024, from https://www.gov.uk/government/publications/year-2-online-media-literacy-action-plan-202223/year-2-online-media-literacy-action-plan-202223.