CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The dark side of digitalization and social media platform governance: a citizen engagement study
Authors	McCarthy, Stephen;Rowan, Wendy;Mahony, Carolanne;Vergne, Antoine
Publication date	2023-01-09
Original Citation	McCarthy, S., Rowan, W., Mahony, C. and Vergne, A. (2023) 'The dark side of digitalization and social media platform governance: a citizen engagement study', Internet Research, 33(6), pp. 2172-2204. https://doi.org/10.1108/INTR-03-2022-0142.
Type of publication	Article (peer-reviewed);journal-article
Link to publisher's version	10.1108/intr-03-2022-0142
Rights	© 2023, Stephen McCarthy, Wendy Rowan, Carolanne Mahony and Antoine Vergne. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/ legalcode - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-12 06:33:38
Item downloaded from	https://hdl.handle.net/10468/14399

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1066-2243.htm

Received 3 March 2022
Revised 21 July 2022
16 December 2022
Accepted 16 December 2022

The dark side of digitalization and social media platform governance: a citizen engagement study

Stephen McCarthy, Wendy Rowan and Carolanne Mahony
Department of Business Information Systems, Cork University Business School, University College Cork, Cork, Ireland, and
Antoine Vergne
Missions Publiques, Berlin, Germany

Abstract

Purpose – Social media platforms are a pervasive technology that continues to define the modern world. While social media has brought many benefits to society in terms of connection and content sharing, numerous concerns remain for the governance of social media platforms going forward, including (but not limited to) the spread of misinformation, hate speech and online surveillance. However, the voice of citizens and other non-experts is often missing from such conversations in information systems literature, which has led to an alleged gap between research and the everyday life of citizens.

Design/methodology/approach – The authors address this gap by presenting findings from 16 h of online dialog with 25 citizens on social media platform governance. The online dialog was undertaken as part of a worldwide consultation project called "We, the internet", which sought to provide citizens with a voice on a range of topics such as "Digitalization and Me," "My Data, Your Data, Our Data" and "A Strong Digital Public Sphere." Five phases of thematic analysis were undertaken by the authors to code the corpus of qualitative data.

Findings – Drawing on the Theory of Communicative Action, the authors discuss three dialogical processes critical to citizen discourse: lifeworld reasoning, rationalization and moral action. The findings point toward citizens' perspectives of current and future issues associated with social media platform governance, including concerns around the multiplicity of digital identities, consent for vulnerable groups and transparency in content moderation. The findings also reveal citizens' rationalization of the dilemmas faced in addressing these issues going forward, including tensions such as digital accountability vs data privacy, protection vs inclusion and algorithmic censorship vs free speech.

Originality/value – Based on outcomes from this dialogical process, moral actions in the form of policy recommendations are proposed by citizens and for citizens. The authors find that tackling these dark sides of digitalization is something too important to be left to "Big Tech" and equally requires an understanding of citizens' perspectives to ensure an informed and positive imprint for change.

Keywords Digitalization of the individual, Citizen participation, Dark sides, Qualitative study

Paper type Research paper

1. Introduction

Over the last 25 years, social media platforms have transformed human relationships and society as we know it. Recent statistics estimate that nearly 58.4% of the world's population is now connected to social media, with high growth rates projected in developing nations going forward (Statista, 2022). The proliferation of social networks has brought new opportunities

© Stephen McCarthy, Wendy Rowan, Carolanne Mahony and Antoine Vergne. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

The authors would like to acknowledge the helpful comments provided by the senior editor, associate editor and four anonymous reviewers. An earlier version of this paper was published at the Hawaii International Conference on Systems Sciences (McCarthy *et al.*, 2021). The authors are grateful for the feedback received from the chair, reviewers and attendees at the session.

Internet Research
Vol. 33 No. 6, 2023
pp. 2172-2204
Emerald Publishing Limited
1066-2243
DOI 10.1108/INTR-03-2022-0142

for individuals to connect almost instantaneously with friends, family, co-workers and other social groups across the world (Cheung *et al.*, 2015; Kapoor *et al.*, 2018; Richey *et al.*, 2018). Social media offers an open channel for user-generated content and communication with a global community, allowing users to engage both within and outside their existing networks (Richey *et al.*, 2018). This enables hobbyists to form online communities around their shared interests (Dwivedi *et al.*, 2018), politicians and celebrities to share content with citizens during election campaigns (Mallipeddi *et al.*, 2021) and entrepreneurs to collaborate with customers during new product development (Namisango *et al.*, 2021).

Research has primarily focused on the positive impacts of digital technology, with less attention directed toward its negative consequences (Rauf, 2021). However, despite recent advances, numerous ethical challenges remain for social media platform governance going forward (Aral, 2020). Firstly, significant concerns have been raised around the integrity of the information provided on social media, given the rise of fake news and propagation of misinformation and disinformation through social networks (Aral, 2020; Laato *et al.*, 2020; Torres *et al.*, 2018). This was notable during the coronavirus disease 2019 (COVID-19) pandemic when unverified claims about the virus spread through social media leading to "cyberchondria" among some user groups (Laato *et al.*, 2020). The World Health Organization (2022) has also highlighted risks associated with an online "infodemic" where false or misleading information on disease outbreaks quickly spreads across digital channels leading to confusion and mistrust among the public. Secondly, much of this digitalization of life has been brought about by large companies that seek to profile users' identities via social media for commercialization purposes. The profit motives of these companies have driven the emergence of 'surveillance capitalism' (Zuboff, 2015, 2019a), where social media is used to track citizens online to influence address this gap by presenting findings from 16 h of online dialog with 25 citizens' behavior, e.g. the Cambridge Analytica scandal. New developments in analytics bring an astonishing range of opportunities to "watch" citizens, raising new concerns around privacy and dignity in the digital age (Leidner and Tona, 2021).

Considering these issues, research is urgently needed to steer the future of digitalization and social media platform governance in a more responsible, ethical and inclusive direction (D'Arcy *et al.*, 2014; Tarafdar *et al.*, 2015; Turel *et al.*, 2019). Responsible governance aims to ensure an informed and positive imprint for change by listening to the voices of different actors affected by a system (McCarthy *et al.*, 2020; Stahl, 2012). The goal is to deliver socio-technical innovations that have beneficial consequences for all by highlighting the responsibilities of different stakeholders to make the "world a better place" (Walsham, 2012). In a similar vein, ethical governance deals with moral principles around how people should act and set rules for judging "right" from "wrong". This centers on ideals around how society should govern itself beyond formal institutions, rules and processes by empowering individuals to monitor and condemn immoral behaviors that violate principles such as privacy and fairness (Chen *et al.*, 2020; Molas-Gallart, 2012). Inclusive governance, meanwhile, aims to close the gap between scientists and citizens by promoting equitable, open and transparent collaboration through citizen engagement (Lukyanenko *et al.*, 2020). Inclusivity helps reduce the barriers to participation by engaging traditionally marginalized groups (Liu and Zheng, 2012). For instance, online platforms can support enhanced levels of inclusivity, allowing citizens with impairments and disabilities a chance to influence policymaking (Leible *et al.*, 2022, p. 2). Inclusive citizen engagement, therefore, incorporates a diversity of contributors who are each provided an equal opportunity to participate (Liu and Zheng, 2012). Interactions between citizens, researchers and other stakeholders can, in turn, foster knowledge creation and transform how we understand the world through meaningful collaboration (Kullenberg and Kasperowski, 2016).

Despite the importance of societal issues to Information Systems (IS) research, the voice of the citizen is often missing - as evident from the dearth of IS literature on citizen engagement

studies compared to other disciplines (Lukyanenko *et al.*, 2020; Weinhardt *et al.*, 2020). Our paper explores citizen engagement's role in supporting inclusive deliberation on the "dark sides" of digitalization (cf. Tarafdar *et al.*, 2015; Turel *et al.*, 2019) and social media platforms, more specifically. We investigate this objective through the following research question: *How do citizens perceive the current and future issues associated with social media platform governance?* Findings are presented from "We, the Internet", a citizen engagement project which sought citizens' thoughts and feelings around online governance using participation events and foresight methods. Discussions centered on how Internet technologies should be governed in the future, with support from strategic partners such as the United Nations, Internet Society, Google, The United Nations Educational, Scientific and Cultural Organization (UNESCO) and the Wikimedia Foundation. The primary aim was to explore different perspectives on social media governance and the digitalization of life more broadly. Questions included: (1) How should social media platforms be managed and governed? (2) What is the role of the different actors in this interdependent system? (3) What will be the impact of emerging technologies on governance? The dialog included not only researchers, practitioners and policymakers but also Irish citizens who were part of a larger participant group located across the world. The Internet enabled diverse citizen groups to gain equal access to policymaking roundtables through electronic citizen participation forums (Leible *et al.*, 2022).

Based on our findings, we present three primary contributions which will be of interest to academic and practitioner communities. Our first contribution is to reveal citizens' perceptions of issues associated with social media platform governance and their rationalization of the ethical dilemmas faced in addressing these issues. In doing so, we answer calls for increased citizen involvement in IS research on digitalization to move existing conversations beyond the political and commercial sphere (Kapoor *et al.*, 2018; Weinhardt *et al.*, 2020). Our findings reveal citizen concerns around the multiplicity of digital identities, sign-up consent for vulnerable groups and transparency in the moderation of content on social media timelines. Citizens also point towards ethical dilemmas in addressing each of these concerns, including tensions between the need for digital accountability vs data privacy, protection vs inclusion of user groups, as well as algorithmic censorship vs free speech. Drawing on the Theory of Communicative Action (Habermas, 1984, 1990), our second contribution is to explore how citizen engagement can harness the power of collectives to deliberate ethical dilemmas associated with Information Technology (IT). We discuss three dialogical processes critical to discourse ethics: *lifeworld reasoning*, *rationalization* and *moral action*. Lifeworld reasoning is a communicative act in which citizens share their pre-reflective knowledge and experiences, while rationalization involves argumentative processes where these speech acts are continuously exposed to criticism (Habermas, 1984; Ross and Chiasson, 2011). Moral actions then involve coordinating new rules for the external world (e.g. society) based on mutual understanding, legitimacy and consensus (Mingers and Walsham, 2010).

We further discuss the interplay between these three dialogical processes for inclusive deliberation on the ethical implications of IT. Thirdly, we discuss how citizen engagement can yield valuable insights into shaping a digital policy for the future. We assert that social media governance is not a trivial question to be left to "Big Tech' but a global issue demanding engagement from all members of society (Ransbotham *et al.*, 2016). This can enable citizens to become the new pioneers of evolved and ethical technology use, developing new meanings for our digital experiences online. We argue that this process is enhanced by the support of strategic partners who can help combat public skepticism in IS research and policymaking (Lukyanenko *et al.*, 2020; Weinhardt *et al.*, 2020), discussing how the United Nations" involvement in "We, the Internet" as part of their program on Digital Cooperation reinforced the message that governance cannot be left to individual organizations and remains a societal issue (Guterres, 2020).

The remainder of this paper is structured as follows: Section 2 provides the background to our study by reviewing literature on the dark sides of social media, citizen engagement and the Theory of Communicative Action. Section 3 provides an overview of our citizen engagement study, while findings are presented in Section 4. Section 5 discusses contributions from our study, while section 6 brings the paper to a close, emphasizing the importance of citizen input in IT governance.

2. Conceptual foundations

2.1 The dark sides of digitalization and social media

An emerging body of literature on the "dark sides" of digitalization points toward several negative consequences (both intended and unintended) from the use of ubiquitous technologies such as social media (D'Arcy *et al.*, 2014; Ransbotham *et al.*, 2016; Turel *et al.*, 2019). Our study focuses on three related areas of research: the emotional and social impacts of ubiquitous technology use, misinformation and the credibility of online content and data privacy loss.

A growing body of research suggests that the use of digital technologies, such as social media, is intimately linked to our social and psychological well-being (Agogo and Hess, 2018; Salo *et al.*, 2022). Psychological well-being is more than being free from depression, anxiety and stress; it is about being well-supported, empowered and satisfied with life (Winefield *et al.*, 2012). On the "bright side", social media can have a positive impact on users' sense of social connectedness (Brown and Kuss, 2020) by contributing to one's sense of identity and self-representation (Walther *et al.*, 2011). However, this can also lead to the development of "bad habits" when users engage in negative self-comparisons and habitually respond to social media notifications to avoid a fear of missing out (Walther *et al.*, 2011). Several studies have found a link between social media use and experiences of negative emotional states such as depression, attachment anxiety, personality distortion, mental exhaustion and attention deficiency, among others (Busch and McCarthy, 2021; Sriwilai and Charoensukmongkol, 2016). Literature also highlights that the relationship between psychological well-being and social media use is far from straightforward. Lin *et al.* (2016) found that while social media use was significantly associated with depression, depressed individuals were also more inclined to interact with social media.

Much of this negativity in reporting is related to the emergence of new digital vulnerabilities such as online harassment by cybermobs and "sock puppet" accounts, e.g. offensive speech and social shaming (Lowry *et al.*, 2016; Ransbotham *et al.*, 2016). Recent studies have shown a link between the level of anonymity offered by social media and the level of user-directed aggression displayed (Lowry *et al.*, 2016). For instance, McHugh *et al.* (2018) find that risk exposure through social media (e.g. cyberbullying, sexual solicitations and seeing explicit content) can lead to symptoms of post-traumatic stress disorder among teens. Johnson (2018) also finds that "mob rules" formed by online groups can sometimes make users unable to distinguish between just and unjust actions online. Trolling differs from other forms of online misbehavior as its central focus tends to be deception and mischief, with the troll seeking to get a reaction from the larger community (Cruz *et al.*, 2018; Dineva and Breitsohl, 2022). Trolling can escalate into cyberbullying (Cruz *et al.*, 2018), causing psychological distress for both the perpetrator and the victim (Dineva and Breitsohl, 2022).

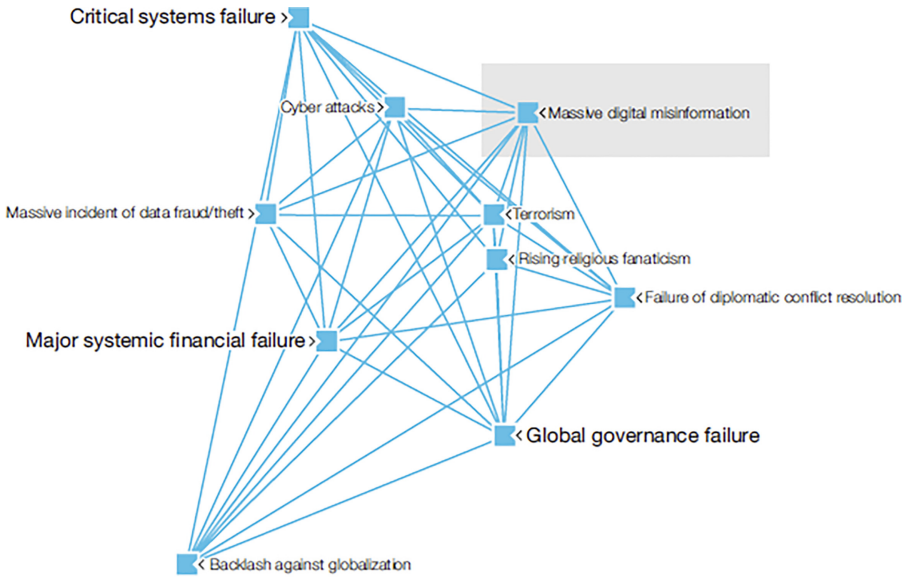
Secondly, digitalization can transform what we believe, how we think, feel and act: the most extreme case of this being when terrorist groups connect and radicalize citizens using social media to propagate hate speech and misinformation (Ransbotham *et al.*, 2016; Turel *et al.*, 2019; Thompson, 2011). For adolescents, their choice of online communities creates a custom "digital neighborhood", which can impact their behavior, attitudes and cultural norms (Brough *et al.*, 2020; Stevens *et al.*, 2016). Mihaylov *et al.* (2018) point toward the

importance of algorithmic classifiers to filter out comments by "opinion manipulation trolls" who seek to compromise the credibility of news forums. While censorship offers one means of tackling these issues, academics have also signaled that human dignity, liberty and autonomy can be compromised by online censorship (Leidner and Tona, 2021; Stoycheff *et al.*, 2020), aspects which are incompatible with debate, consensus and democracy (Berners-Lee, 2017; Robbins and Henschke, 2017). It is estimated that 71% of those with Internet access live in countries where they can be imprisoned for posting social media content on political, social, or religious issues (Stoycheff *et al.*, 2020). Emergency measures to combat misinformation and disinformation, such as automatic screening and the take down of content, can backfire by negatively impacting legitimate journalism, eroding trust in institutions and pushing users to other platforms (Radu, 2020). Algorithmic bias, caused by flawed data or unconscious assumptions, also poses a significant risk to users who may be falsely penalized (Ransbotham *et al.*, 2016).

Thirdly, digital surveillance on social media has also raised questions around data privacy loss. This creates an imbalance in power between "the watcher" and "the watched" as companies and governments increasingly use data analytics to profile social media users and promote desired behaviors for commercial or political gain (Ransbotham *et al.*, 2016; Zuboff, 2019a). Zuboff (2019b) explains that information represents a new form of power as personal data on users and their social contacts can be monetized and modified to create a new principle of social ordering. This centers on the surplus of behavioral data created in the information age. Like Zuboff (2019b), Sir Tim Berners-Lee expressed concern that the tech giants have become surveillance platforms. In an open letter on the 28th birthday of the World Wide Web, Sir Tim Berners-Lee (2017) highlighted the danger posed by companies and governments "*watching our every move online and passing extreme laws that trample on our rights to privacy*". However, Cheung *et al.* (2015) find that social influence has a more significant impact on self-disclosure through social media than perceived privacy risk, as users are often willing to accept privacy loss in exchange for desired benefits. Paradoxically, research suggests that users perceive others as suffering from the threat of Internet privacy risks rather than themselves, recommending privacy protections to others whilst displaying a decreased willingness to adopt these measures for themselves (Chen and Atkin, 2021). Questions, therefore, remain around the governance of social platforms in the ever-evolving landscape of privacy risks.

2.1.1 Social media platform governance. Concerns about the negative consequence of technology use have often been sidestepped *in lieu* of our fetish for innovation and commercial gain (Ransbotham *et al.*, 2016; Stahl, 2012; Zuboff, 2015, 2019a). After decades of relatively light intervention to enable social media platforms to flourish more freely in their infancy, policymakers and civil society have become increasingly aware of the imperative for dialog on how to harness the benefits while containing the drawbacks (Gorwa, 2019). A report by the World Economic Forum (2013) highlights the risk of "Digital Wildfires", which combine global governance failures with issues such as misinformation, fraud and cyber-attacks (see Figure 1). These wildfires can spread rapidly, spreading provocative content, such as misinformation or inflammatory messages, through open and easily accessible systems (World Economic Forum, 2013; Webb *et al.*, 2016).

Platform governance is a multifaceted concept that encapsulates areas such as content generation, technical infrastructure models, data subjects, policies and regulations (DeNardis and Hackl, 2015; Van Dijck, 2021). Social media platforms are public spaces where users can interact and engage in activities such as gaming, learning, consuming news and commerce. Considering the range of activities and variety of users, it can be difficult to predict future conflicts and governance issues (Almeida *et al.*, 2016). For Han and Guo (2022), governance can be understood according to two perspectives. In a broad sense, IT governance relates to issues such as information, network culture, domain name regulation, etc. In Floridi's



Source(s): World Economic Forum, 2013

Figure 1.
Digital wildfires in a hyperconnected world

philosophy of information and information ethics, to understand the impact of actions, we need to look beyond living creatures and their environment to include information itself. Floridi (2002) elaborates on the conceptual nature and basic principles of information and how they can be applied to philosophical problems. As information is a central component of social media platforms, Floridi's philosophy is worthy of consideration to help guide IT governance in a more ethical direction.

In the narrow sense, IT governance encapsulates various stakeholders working together to formulate common principles, rules, standards and cooperative decision-making procedures. The core proposition is that IT governance should follow a process of co-governance which is mirrored in the ideas of discourse ethics (Gorwa, 2019; Van Dijck, 2021; Han and Guo, 2022). Discourse ethics has its home in Habermas' (1984) Theory of Communicative Action and aims to study how different stakeholders select between technology frames, rejecting misaligned use or disengaging where there is an assessment of little value (Mingers and Walsham, 2010). This goes beyond an individual's own ethical decision-making process on technology use and considers the implications for wider society (Stahl, 2012; Walsham, 2012). Habermas (1993) recognized that debates need to go beyond the justification of norms to debate their application which can only occur when communities of actors are involved in discursive information ethics. We next discuss citizen engagement as an approach to discourse ethics.

2.2 Citizen engagement

Citizen engagement is an open movement that encourages voluntary participation by diverse citizen populations in research and policymaking (Lukyanenko *et al.*, 2020; Kullenberg and Kasperowski, 2016; Olphert and Damodaran, 2007). Citizen engagement ensures the *democratization* of research by allowing diverse communities to influence decision-making processes by contributing their thoughts, opinions and perspectives

(Levy and Germonprez, 2017). This aims to change the way research is conducted by involving the public in the research process, from generating ideas to conducting research and disseminating findings (Olphert and Damodaran, 2007; Weinhardt *et al.*, 2020).

While research has traditionally been the domain of academic experts who control the data-gathering process and analysis of findings, citizen engagement adopts a more *non-discriminatory* stance asserting that anyone impacted by a system should have a say in how it is designed, regardless of their educational background or subject matter expertise (Khan and Krishnan, 2021; Levy and Germonprez, 2017; Lukyanenko *et al.*, 2020; Weinhardt *et al.*, 2020). This requires ongoing collaboration between researchers and members of the public, as epitomized by the discourse ethics school of thought in IS research (Mingers and Walsham, 2010; Someh *et al.*, 2019). The aim is to make research easier to access and more inclusive by empowering citizens with the opportunity to participate. This supports freedom of expression, allowing individuals to realize their full potential through the development of ideas, exploration and self-discovery (Emerson, 1962).

Citizen engagement is also designed to encourage *consensus* building between a broad range of actors in research, innovation and public policy decisions (Ju *et al.*, 2019; McCarthy *et al.*, 2020; Olphert and Damodaran, 2007). This is achieved using various practices such as foresight processes and scenario planning (McCarthy *et al.*, 2020). Online communication tools can also be used to make IS research more collaborative by supporting knowledge creation and dialog across geographical locations (Khan and Krishnan, 2021). Lukyanenko *et al.* (2020) sees technology as an open door for practicing collaborative research, breaking down traditional barriers and encouraging diversity and inclusion by harnessing the "wisdom of the crowds". Indeed, social media can provide a platform for recruiting citizens in research and enable self-expression and the distribution of knowledge beyond the "soundbite" reporting of policy (Tacke, 2010).

The principle of *universality* aims to bring together volunteer citizen groups to collectively shape a better future for all (Mingers and Walsham, 2010; van der Velden and Mörtberg, 2021). For instance, citizen engagement can help promote new policy agendas through community self-organization (Hecker *et al.*, 2018). Participation efforts can, in turn, lead to increased interest and participation in the democratic process and contribute to a more scientifically literate society (Weinhardt *et al.*, 2020; Levy and Germonprez, 2017; Lukyanenko *et al.*, 2020). Table 1 outlines the participatory ideals of citizen engagement, moving away from centralized control and encouraging community involvement.

The next section draws on the Theory of Communicative Action as a lens for understanding how citizen engagement can support discourse ethics on the dark sides of digitalization.

Principle	Definition	References
Democratization	Citizens from diverse backgrounds (e.g. education) are invited and actively encouraged to participate in the research design and development process	Kullenberg and Kasperowski (2016), Lukyanenko *et al.* (2020)
Non-discrimination	Participation is not limited to experts, and all contributions are treated equally. This implies the facilitation of dialog, free from censorship	Khan and Krishnan (2021), Lukyanenko *et al.* (2020), Weinhardt *et al.* (2020)
Consensus	Shared understanding is sought through negotiation and transparency during the participatory process	Ju *et al.* (2019), Olphert and Damodaran (2007)
Universality	The aim is to formulate recommendations that go beyond a single context or the interests of a single community	Mingers and Walsham (2010), van der Velden and Mörtberg (2021)

Table 1.
Principles of citizen engagement

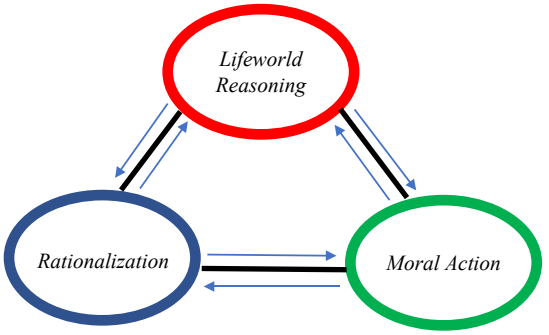
2.3 Theory of communicative action

The Theory of Communicative Action proposes that collective deliberation between diverse groups can deliver informed and argumentative perspectives on issues of public concern (Habermas, 1984, 1990). Discourse ethics defines the theories on communicative rationality and consensus decision-making of the philosopher Jürgen Habermas which aims to promote involvement by diverse actors to expose decision-makers to a wider variety of perspectives, needs and potential solutions (Mingers and Walsham, 2010). Discourse ethics differs from moral theories, such as principles of human rights which have a catalog of clear morality (Beschorner, 2006). Instead, discourse ethics is a procedural moral theory where norms are created during open discourse and agreed upon based on mutual understanding and consensus (Beschorner, 2006; Lyytinen and Klein, 1985; Yuthas *et al.*, 2002). Discourse ethics proposes that morality can only be achieved through "an ideal-speech situation" where stakeholders can equitably debate the ethical concerns of different proposals (Someh *et al.*, 2019).

While it may seem unattainable and utopian at face value, Habermas (1990) later offered clarifications on how an ideal-speech situation can be approximated in practice. He acknowledges that speech acts are socially constructed and subject to different interests, which may not always align with the pursuit of mutual understanding. Habermas (1990) nevertheless asserts that ideal-speech situations are a critical standard by which discourse should be evaluated. Therefore, researchers should aim to approximate these critical standards by guiding discussion and encouraging inclusivity (Ross and Chiasson, 2011).

The aim is to bring together different participant groups to collectively shape a better future through the interplay of different dialogical practices (Habermas, 1984, 1990). The Theory of Communicative Action presents three practices central to social cooperation: *lifeworld reasoning*, *rationalization* and *moral action* (see Figure 2).

Lifeworld Reasoning is pre-reflective and centers on participants' own perspectives, assumptions and ideals for a given situation. Participants are invited to express their inner world by freely sharing honest and personal thoughts on the topic of discussion (known as *personal sincerity*). Each participant should be given an equal voice, with legitimacy for everyone's contribution. Lifeworld reasoning seeks contributions from those directly affected by proposals and embeds their input into decisions (Mingers and Walsham, 2010). It recognizes that systems impact the lifeworld of participants and vice versa that the lifeworld impacts systems. Participants need to feel comfortable not only giving truthful information but also projecting a genuine picture of themselves to the group; this is vital for developing mutual understanding (Lyytinen and Klein, 1985; Yuthas *et al.*, 2002).



Source(s): Adapted from Habermas, 1984, 1990

Figure 2.
Theory of communicative action

Rationalization then refers to an argumentative process through which speech acts are increasingly exposed to criticism. During rationalization, the lifeworld reasoning of participants is challenged and negotiated through questioning (Ross and Chiasson, 2011). The aim is for participants to justify their position and deliver better arguments by testing them with other participants. Habermas (1984, 1990) argues that decisions subjected to rationalization suffer less polarization and have a much lower level of volatility than those that have not. Rationalization is, therefore, an "emancipatory communicative act" in which participants build a shared understanding of situations through reasoned argument, negotiation and consensus. The discourse can lead to changes in position due to a better understanding of the issue or impact on others (c.f. Beschorner, 2006). A speech act is successful when judged as merit-worthy, authentic and justified by the hearer, who then adopts "*an affirmative position*". In contrast, a speech act is unsuccessful if it fails to gain uptake from others and must then shift toward new speech acts (Ngwenyama and Lee, 1997).

Lastly, *Moral Action* refers to the creation and coordination of new rules for the external world (e.g. society) based on consensus and collective interests. Habermas (1984, 1990) asserts that rules can be legitimate only if they arise from the will of citizens and represent the will of all. Moral action centers on rules which are said to be equally good for every citizen, beyond the interests of any one community or context. Mingers and Walsham (2010) recognize that while this ambitious ideal may never be realized, all debates should aspire toward it. The ideal is established through an open and fair debate process (or *ideal speech situation*) and cannot pre-exist dialog or be imposed by more powerful stakeholders (Someh *et al.*, 2019). Moral actions seek to reduce the ethical impacts of systems on the lifeworld of participants by coordinating actions and developing restraining barriers. Moral actions focus on questions of the good life ("*ethical goodness"*) for individuals ("*ethical-existential*" discourse) or collectives ("*ethical-political*" discourse).

Three conditions are presented for evaluating moral actions and the validity of claims made during dialogical processes: sincerity, truth and rightness. The condition of sincerity relates to the speaker's inner world and evaluates whether claims of truthfulness are based on subjective perceptions and lifeworld reasoning (e.g. personal experiences). Truth shifts our focus to the external world and whether the speaker's claims are based on statements with fair assumptions of the objective world (e.g. the performance of an information system). Rightness then directs attention towards the wider context of society to assess the speaker's claims of legitimacy in relation to "our" social world. All speakers must defend their claims to validity according to these three conditions (Lyytinen and Klein, 1985).

Another key aspect of Habermas' work is "deliberative democracy", which proposes that citizens are invited to participate in developing problem-solution pairings to complex challenges. Encouraging citizen participation in this way ensures legitimacy as diverse voices are engaged in creating new norms. Inclusive deliberation can similarly support democratic IS processes, allowing citizens to share their opinions on new systems before they are implemented in an organization or wider society (Lyytinen and Klein, 1985; Ross and Chiasson, 2011). We next present the research design behind our citizen engagement study, which follows a Habermasian perspective.

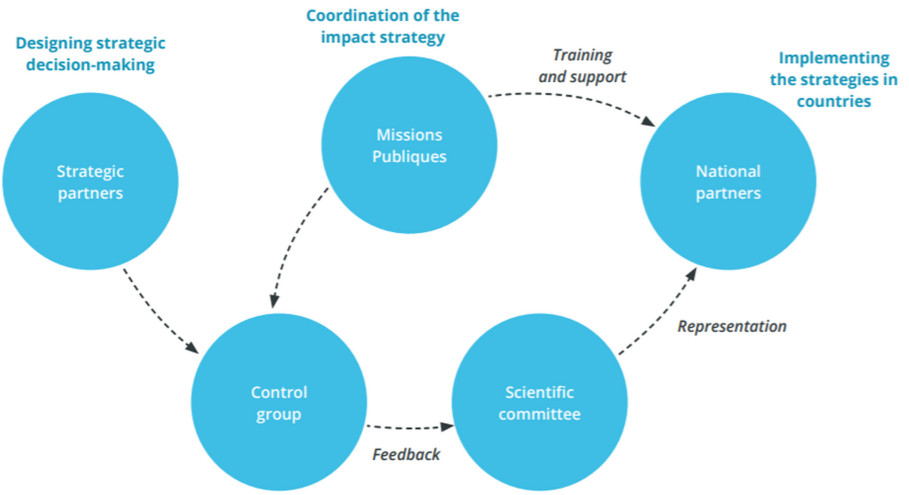
3. Research design

Citizen engagement was selected as an appropriate research approach as it supports the investigation of contested, fragmented and multi-dimensional phenomena through discourse ethics (cf. Mingers and Walsham, 2010). Critical social theory was then chosen as the foundational philosophy of this study to develop explanations and understandings of a social situation and critique inequitable conditions (Ngwenyama and Lee, 1997). This methodology is in line with Habermas' view that citizens, given the right environment, "*may become the*

supreme judges of their own best interests . . . [and] the idea is thus to open up public, democratic processes, based on dialogue between citizens" (Alvesson, 1996, p. 139). The critical standard of an "ideal-speech situation" was approximated by a trained group of facilitators who followed a set of guidelines that aimed at supporting intersubjectivity (see Appendix 1). During the discourse process, efforts were made to ensure underlying power dynamics in the group were addressed, and any marginalized participants were invited into the discussion to overcome any barriers to consensus.

Our study centered on the global consultation project "We, the Internet" (https://wetheinternet.org/), which explored citizens' attitudes toward the opportunities and challenges provided by Internet platforms and future developments in this technology (McCarthy *et al.*, 2021). The project was coordinated by Missions Publiques (France) in collaboration with national organizers in over 80 countries worldwide. These national partners recruited citizens in their respective countries and were part of the facilitation team during the online dialog. In addition, the project had support from public and private strategic partners such as the German Federal Foreign Office, the United Nations, European Commission, World Economic Forum, Wikimedia Foundation, Internet Society and Google. The strategic partners constituted the advisory board and scientific committee to provide conceptual and scientific guidance. This network of partners (see Figure 3) was essential in ensuring global outreach with a diversity of participants, as well as enhancing the impact on decision-making processes.

On October 10 and 11, 2020, citizen assemblies were held in over 80 countries simultaneously, representing about 25–100 citizens per country. In this paper, we refine our scope to concentrate on results from Ireland, where the co-authors facilitated over 16 h of dialog between 25 citizens. An open registration process was utilized to ensure a diverse sample of citizen participants and encourage discussion and reflection based on different experiences (Chapple *et al.*, 2020). An online platform allowed volunteers to join two synchronous four-hour sessions across two days. Volunteers did not require expertise on the topic to register and the final sample had representation from different demographics (see Figure 4). Following Patton (2002), we adopted a snowball sampling strategy using various recruitment channels to identify people with interest in the topic of digitalization. Public



Source(s): https://wetheinternet.org/

Figure 3.
We, the Internet consortium

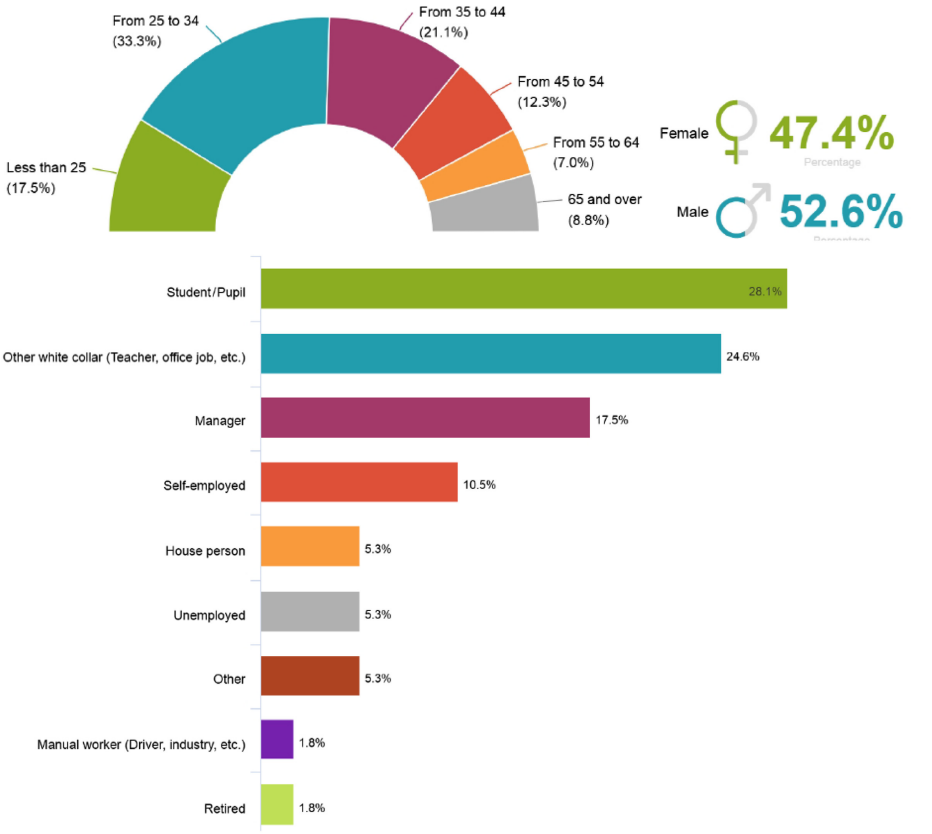


Figure 4.
Overview of participating citizens in Ireland

adverts were disseminated across several digital and printed outlets, including the university website, social media and local and national news sites. While the demographics included in the final sample were not directly representative of the Irish population, we nevertheless ensured that different age ranges, genders and professions were represented, which supported our aim of inclusivity.

The study received ethical approval from the Social Research Ethics Committee in [*university name withheld for review*], Ireland (Log number 2020-115). All recruited citizens received an information leaflet about the study and were asked to provide their informed consent prior to participating.

3.1 Data collection

Data was collected through 16 h of dialog with citizen participants in Ireland. This translated into an 87-page transcript with 80,662 words (around 20,000 words per group interaction) and 72 pages of field notes. The views of volunteer citizens (see Appendix 2) were collected through a set of structured steps during roundtable discussions in virtual breakout rooms. A World Café approach (Fouché and Light, 2011) was used to divide participants into sub-groups with a facilitator present to help guide the emerging discourse on the dark sides of digitalization. The facilitator was responsible for providing an inclusive space for different

debates and viewpoints to emerge, moderating potential conflicts between participants and supporting them in working toward a resolution.

The sub-groups encompassed about 5 participants each to provide a suitable atmosphere for everyone to engage in meaningful discourse. High levels of rationalization were vital to ensure that participants had opportunities to declare their position and justify or update it where required based on the evolving group discourse (Alvesson, 1996). It also allowed group members to evaluate statements according to the conditions of sincerity, truth and rightness (Habermas, 1984). Google Docs was used as a collaborative note-taking platform, with all participants invited to record key points from the dialog using designated worksheets. The participants could also include comments and share hyperlinks using the chat function available through Zoom Video Conferencing.

Discourse was organized around seven sessions that covered critical topics on digitalization, social media governance and the future of the Internet (see Table 2). Each session followed structured templates which aimed at supporting participants in their discussion (see sample templates in Appendix 3). The template designs were informed by the High-Level Panel of Digital Cooperation launched by Antonio Guterres, Secretary General of the United Nations and aimed to directly relate to current issues being discussed by policymakers and strategic partners. This encouraged reflection and discussion around a set of prompting questions derived from the report. Templates were uploaded to a virtual collaboration platform and acted as a visual canvas for participants to record ideas using post-it notes. After each session (typically lasting 45-60 min), participants were asked to record their contributions. Following Habermas' (1984) fundamental principle of deliberation and open dialog, a semi-structured approach to facilitation was adopted to ensure that discussions were not constrained by the templates. Nevertheless, the facilitators also made certain that all prompted questions were answered and that the templates were used consistently across sessions. Participants in the deliberation were not required to read the material beforehand, and their contributions were not differentiated based on prior level of knowledge or expertise. Instead, the citizen engagement study aimed to support open debate on complex issues. For the purposes of this paper, our findings will center on the first four sessions.

Session	Description	Objective
Introduction	Facilitators presented the study's objectives and program for dialog	To welcome participants and ensure they are aware of rules for good dialog, e.g. active listening and respect
Digitalization and me	Participants explored personal and collective experiences with the Internet and, more specifically, social media	To understand basic concepts of digital governance from the perspective of citizens
My data your data our data	Participants explored their understanding of digital identities on social media and the data footprint left through interactions on social media	To reflect on citizens' understanding of the data that is gathered online
A strong digital public sphere	Participants explored their wishes for a strong digital public sphere	To explore the impacts of disinformation and approaches to tackle it online
Exploring artificial intelligence	Participants exchanged their thoughts and feelings about AI through targeted advertising and personalization of content	To explore citizens' basic understanding of the data analytic practices adopted by social media platforms
Internet for and with citizens	Participants proposed the policy actions for governance going forward	To discuss and prioritize policy actions for addressing ethical issues with social media
Conclusion	Facilitators closed the session and summarized the results	To outline the next steps of the project and thank participants

Table 2.
Overview of the citizen engagement sessions

3.2 Data analysis

The Gioia Methodology (Gioia *et al.*, 2013) was followed to analyze transcribed material from our study and cluster interesting findings. The first and second authors coded the qualitative data across five phases to develop first-order concepts, second-order themes and aggregate dimensions.

In phase one, the first and second authors began by continuously reading and rereading 16 h of transcribed content from the citizen engagement study as well as field notes to generate a set of initial codes which they judged as meaningful and important to the study in question. Based on this process of familiarization, the authors noted initial ideas on the transcribed data. In phase two, the authors systematically coded interesting features of the entire data set and collated data relevant to each code. This process allowed strongly expressed ideas to emerge based on citizens' evaluations of the issues associated with social media governance.

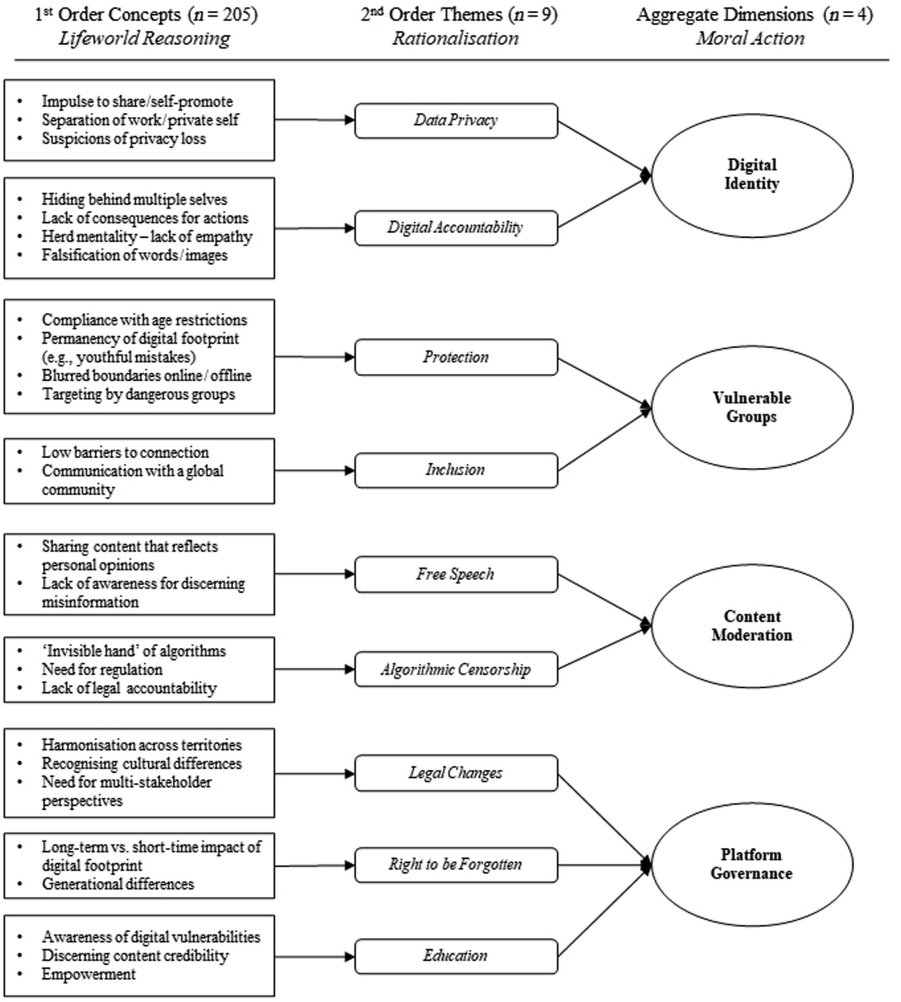
Phase three involved grouping initial codes together to form overarching categories of codes which helped organize the content. The collated codes were sorted into potential themes, with all data on these themes included. In phase four, themes were reviewed and critically appraised by each co-author to ensure it was representative of coded extracts and the entire data set. The authors refined each theme and formed an overall story of the analysis based on clear definitions and names. Finally, themes of interest to the citizen participants were collated, aggregated and reported by the co-authors. The authors selected vivid and compelling extracts guided by the research question and literature review.

Figure 5 illustrates our data structure and the full list of codes. In total, 205 first-order concepts were created during phases one and two of data analysis. This subsequently led to the definition of nine second-order themes during phase three and four aggregate dimensions during phase four. In the final phase, abductive reasoning was employed to draw on plausible theories that might give the authors deeper insights into the coded observations. The authors sought to better understand the dialogical process through which citizens engaged in discussion around the dark sides of digitalization. To do this, we adapted three constructs from the Theory of Communicative Action (Habermas, 1984), *Lifeworld Reasoning, Rationalization* and *Moral Actions*. Working with the data, we realized that open codes primarily related to citizens' own *lifeworld reasoning*, which was aggregated as personal experiences (feelings and intentions) as well as thoughts on states of affairs/norms in the social world (Mingers and Walsham, 2010). The process of *rationalization* was then coded for citizens' argumentative discussion around the ethical dilemmas faced in addressing concerns from their lifeworld reasoning and was coded as to where the validity of claims is challenged (Mingers and Walsham, 2010). Finally, platform governance recommendations were aggregated as *moral actions* (summarized in Table 3 of the discussion section) and centered on suggestions to resolve issues that transcend the interests of a particular group (Mingers and Walsham, 2010). While some coded themes appear to overlap (e.g. data privacy and free speech), the reported findings are distinctive and cover different aspects of the discussion topic. Overlap is also indicative of the complex and interconnected nature of digitalization and social media governance.

The next section provides an overview of the findings from our discourse ethics study.

4. Findings

Findings center on the issues associated with social media platform governance from the citizens' perspective and point towards the ethical dilemmas they identified in addressing governance issues. Drawing on the Theory of Communicative Action (Habermas, 1984, 1990), we investigate citizens' contributions through the lens of *lifeworld reasoning* (citizens' personal reflections on social media governance issues) and *rationalization* (critical debate



Source(s): cf. Gioia *et al*., 2013

Figure 5.
Data structure

around the ethical dilemmas associated with social media governance). In the discussion section of our paper, we then explore *moral actions* aimed at protecting the dignity, autonomy and liberty of citizens on social media. All names are anonymized for confidentiality reasons.

4.1 The multiplicity of digital identities (lifeworld reasoning)

Discussions during Session two, "*Digitalization and Me",* centered on the current loopholes in social media governance that enable citizens to create multiple "*fake identities*" online, often with the purpose of malicious intent. Several participants shared personal accounts of abuse that they experienced either directly or indirectly through social media platforms and the relative freedoms that anonymity affords "sock puppet" accounts due to an overall lack of accountability. One story centered on a difficult experience involving a "troll" with multiple

fake profiles who sought to inflict abuse on innocent users. "Amanda", a 40-year-old public figure and advocate for employee well-being in the IT sector, noted how "*an individual with numerous fake identities targeted a few people, including myself. A stranger (I) don't know this person, and they literally put out so much stuff, all our profiles, and photographs. It was huge online bullying harassment*". "Amanda" also discussed the impact this experience had on them emotionally, in no small part due to the lack of mitigation mechanisms available through the social media platform in question. She described efforts in seeking litigation action for this attack on her privacy; however, the case was eventually dismissed by the courts: "*it led to legal threats at one particular point made against me by a completely mischievous person. Eventually, that case was dropped*".

Participants also spoke about the potential for collaborative cyberbullying on social media, where users develop a "*herd mentality*" and work together to attack a fellow user of the platform. Reflecting again on the affordances of anonymity, citizens noted how "*hiding behind their digital identity, they can be as cruel and as nasty as they like*" and the lack of real-world implications can empower groups to rapidly spread rumors on social media. The posting of altered photographs ("*Instagram Fakery*") to a broad audience was also discussed, leading to negative consequences such as body dysmorphia among young users. "Juliet", a 35-year-old public school teacher working in the local area, reflected on their conversations with a child around Instagram Fakery and how it affects their perceptions of identities in real life: "*[the child said] "yeah everything is fake, everyone's edited it' but then on the flip side of that she said yeah 'I also wouldn't talk to someone if I don't like the look of them [on social media] I just wouldn't even bother trying to get to know them"*. This prompted others to share childhood experiences of bullying before the advent of social media and the difficulties that younger generations experience today. "John", a 29-year-old law graduate now working in the field of digital rights, noted how prior to social media, victims of bullying would only have to deal with abuse during school hours, as they had reprieve once in the comforts of their family home. However, with social media, cyberbullying can now permeate the safe space of a young person's home: "*originally, if you were a boy in school [bullying] was only during the day. Nowadays, because everyone has online access, you see that people are bullied so badly because, like, basically the bullying can go on 24/7*".

4.2 Digital accountability vs data privacy (rationalization)

Toward the end of session two, participants began to challenge the notion of digital accountability and the multiplicity of digital identity. Some argued the need for "*authenticity*" on social media platforms through new governance mechanisms which ensure real-world repercussions for instigators of abuse, i.e. *"naming and shaming*" users in real life. Others argued that accountability was essential to protect victims and respect basic human rights to dignity on social media platforms: "*people seem to lose their normal human inhibitions when they go online. When I'm interacting with people face-to-face, there are certain ways [of] civilized behavior and empathy for other people. An ability to understand that even if you disagree with what he or she is saying, there are ways in which you can do that which are sensitive to their personhood*." Similarly, the lack of governance around digital identities on social media was argued to create more extreme views "*we feel free to make the most bombastic and black and white judgments about other people … taking ridiculously over the top positions about everything and forgetting about the fact they're talking to other human beings and what you're saying has an emotional impact*."

However, some citizens countered by highlighting the need for "fake" digital identities today to protect real-world privacy. They argued that "different personas" are required for separating professional and private contexts, asserting these two worlds should not necessarily mix: "*you might [use] a different persona with government organizations,*

companies, or different social media outlets. [Some are] a lot closer to what your real-world identity would be because you're logging on to use essential services". "Jane", a senior manager with high visibility in her field, made a similar assertion that "*I have my public profile, which is my professional profile, but I refuse to have my private profile at all in most social media … I think it's very dangerous to mix these. Yet, most people [do]*". The facilitator "Johnathon" however, welcomed contrasting approaches to data privacy which were less rational in nature: "*So anybody else got a different perspective on the way that they do it?*". In response, one citizen noted that younger generations today wish to be transparent online and represent their authentic self, regardless of the potential impacts on their future employability: "*if an employer didn't find their drunken photos funny on their publicly accessible Facebook page, which they had no intention of privatizing, then they didn't want to work for that company*". Younger generations were generally perceived as less concerned about mixing the professional and private worlds. "Thomas", a 33-year-old male, also explained that this is driven by peer pressure and the demands for continuous self-promotion on social media to present a more socially attractive persona (e.g. opinionated, socially active and adventurous): "*[they're] chasing social gratification … people used to call TV the opiate of the masses … I don't even think kids are interested in drugs and cigarettes anymore because they're on Instagram, and they get their dopamine from that*". Education was highlighted as a critical intervention to ensure that younger generations are aware of the risks posed by privacy loss and the dangers of mixing personal and professional identities online.

4.3 Consent for vulnerable groups (lifeworld reasoning)

In session three, "*My data, your data, our data*", discussions centered on the issue of consent online and the minimum age requirement for sign-up (typically 13 and above). "Juliet" again shared personal experiences of when proposed consent policies were properly governed in practice and the real ethical risks that this carries for underage users: "*I'd be teaching classes maybe where there are children at eight, nine, ten years old and they all have smart devices, and all have accounts for TikTok or Instagram or various social media sites despite the being much younger than the minimum [age] to actually sign up"*. She then reflected on a conversation with parents who allowed their child to bypass the minimum age of consent for social media, conceding that the decision was made due to peer pressure from other children in the class: *"I said, how old is your child? (She said) "my child is seven". Okay, so you obviously put in a fake date of birth to set up this account [. . .] there's a lot of kids hammering their parents to get a TikTok account.*" Others agreed that peer pressure was a primary driver of social media adoption by vulnerable groups: "*10- or 11-year-olds … all have Snapchat and TikTok and Facebook and Twitter despite there being age restrictions. Think this age restriction is ridiculous anyway … [13-year-olds] are not informed about the dangers of the apps at that stage in your life"*.

Participants then discussed the risk of underage social media users being targeted by sex offenders and extremist groups. One shared a harrowing story of an underage user who was stalked by a known pedophile in their local area using location data on their social media posts to determine the child's location: "*if you go to Snapchat and you turn on the map, you can see everybody in Snapchat that you're connected to, where they are and what they're doing … one of my one of my students got contacted by the guards (police) because this pedophile was tracking his daughter and other pre-teens"*. The maturity of underage users to assess their exposure to such risks on social media was discussed ("*too young to understand*") and some felt they were often more willing to share information freely online, without concern for any negative consequences: "*they're sitting inside the living room on apps so that's their real-world [and] there's no difference between the real world and the online world for a younger generation. They're living through their phones*". "Anthony", a parent with young children, also shared his

struggles when trying to communicate the risks of social media engagement, noting that warnings continuously fell on deaf ears: "*if it's a parent telling them (about the risks online), it's just like "another lifelong lesson" and they don't tend to listen. Trying to explain that to young people, especially as a parent, they shut down a lot of the time*".

4.4 Protection vs inclusion (rationalization)

Building on the dialog presented above, participants began to challenge whether protection is the best way to mitigate social media adoption by underage users. One rationale in support of protection was to give young users the freedom to make mistakes without repercussions later in life: "*you know, there are pictures of me probably smoking weed when I was 15 or something up online somewhere, but there's nothing I can do about that. But if you're even younger, you know that's crazy . . . I suppose reality in the physical world and the online world is becoming blurred*". Young users' discernment of appropriate content to share online was discussed, as well as the impact this might have on their reputation in adulthood. "Thomas", whose sibling is on the autistic spectrum, shared a personal story of a young relative with special needs who continuously share large volumes of information on social media without appreciating the risks it poses to their safety online: "*I know a person that has special needs and really didn't understand the concept of private information and what you put out into the public. You kind of end up then opening yourself up to vulnerabilities the more you share online; the more you think it's acceptable to share online*".

However, others later challenged the recommendation of universal protection, countering the equal need for inclusion across diverse groups. Inclusion was argued to be an inherent advantage of social media: "*I think one of the advantages obviously is that you can send out multiple messages and reach so many more people than you would if you were just trying to do it and physically. You'd never be able to do that, and so you are able to share good news stories, you're able to share positive photos and messages"*. Some argued that the current low barriers to sign-up helped promote diversity on social media, allowing users to freely contribute content and communicate with others across the world: "*I think it [social media] gives people the opportunity to promote themselves in good ways. Like you can tell people: look at this interesting stuff we are doing; come and get involved or . . . spread the word.*" Inclusion was also highlighted as a way of empowering citizens who may have had limited access to content and resources in the past: "*We take for granted how easy it is to access all of those things now compared to not even ten years ago. Just the access and how quick everything has become at the end of our fingertips. We have access to information but also access to resources that ordinary citizens might not necessarily be involved in*".

4.5 Transparency in content moderation (lifeworld reasoning)

During session five, "*A strong digital public sphere",* participants shared their personal concerns about the moderation of content on social media timelines. The increase of fake news during the COVID-19 pandemic was discussed as well as its transcendence from the digital world, impacting how people act in the real world, e.g. anti-vax groups. The "*invisible hand*" of algorithms that shape what is displayed on social media timelines was also criticized for lacking transparency: *"how can we discern what data is coming from reliable sources and what is not? And how do we deal with the stuff that's not without simply saying, oh yeah, just ban all that? It's a question really of judicial discernment and digital awareness."* Many noted that they "*didn't understand the question*" as they felt AI was a topic beyond their understanding. However, the facilitator "Gary" assured them that "*there are no wrong answers, just reflect on how [social media] is now and what you foresee in the future. It could be kind of a utopian or dystopian view"*. In response, one individual described social media timelines as a "*kind of free-floating world which is not grounded in real-life experience, whatever that experience may*

be . . . it creates a kind of a sense in which people are living at least partially in a world of fantasy." Reflecting on earlier discussions around digital identities, participants agreed that social media had become an increasingly unruly environment of public discourse and the impact of "*catfishing*" was noted as another driver of fake news dissemination: *"There are a lot of fake identities online and we have heard a lot about people doing catfishing. There is a difference . . . online; you can be anybody."*

The power that social media companies possess in controlling content, banning users and posting disclaimers on content flagged as fake news was also observed: *"Facebook banned a marketing agent's PR agency in the US who were using teenagers who were using (sock puppet accounts) on Facebook to support Mr. Trump to amplify and comment. So, they were able to shut down that kind of a micro-network [of] groups trying to influence the elections in the US"* However, questions remained around whether it is wise to empower private organizations with absolute control over the censorship of information online, with the lack of transparency around content moderation again noted. Participants felt it was important to continue dialog on the design of moderation policies, how these are enacted and by whom. This was seen as essential to avoid "blind trust" in the algorithms of private companies that moderate content: "*as things become more automated, we begin to trust more that 'oh look, it's all done in the background, I don't need to worry', but that's exactly the [concern] . . . if we had accurate information that's properly assessed using these really smart technologies then the world would benefit more than having all the information at our fingertips, which none of us have the expertise to analyze"*. In the spirit of mutual understanding, the facilitator "Brenda" summarized that "So *if we were to come to some sort of consensus [. . .] it all relates to a strong digital public sphere [which] organizations, government bodies, and non-government agencies [should] offer*".

4.6 Algorithmic censorship vs free speech (rationalization)

Through ongoing discussion, participants next explored the balance between algorithmic censorship and freedom of speech. The European Convention on Human Rights' (ECHR) position on Internet content as distinct from printed media (subject to different regulations and controls) was noted by several participants, with its underlying rationality questioned: "*[social media] is very much a free for all [on] how this information is disseminated . . . These platforms defend themselves saying they are just platforms and not publishers. I think they should be publishers and should be held legally accountable for what they do. That would be a good start and something which would be a major departure from how things are now.*" As a result of ongoing dialog, some felt that regulating social media platforms as "*publishers*", like traditional media (e.g. print, television, radio), could help address the concerns around censorship: *"I do think there needs to be more accountability to people because of the amount of money and power that [social media companies] have . . . I don't know who should regulate [social media companies], but I don't want the companies self-regulating themselves. There is an international convention, but that doesn't mean that all countries in the world are following those"*.

There was general agreement that the balance of algorithmic censorship and free speech would be a complex goal and trade-offs may come at a price. Some argued that governance should be driven at an international level and *"there needs to be an international body set up for the governance of the internet"*. In contrast, others felt that judicial pressure from regional legislators is the only way to drive change: *"it's normally judicial pressure to actually get something made. The only reason that this Digital Media Bill is coming in is because the EU is making us."* "Jonas", a 50-year-old digital inclusivity champion who works with engagement groups across the world, reaffirmed this by arguing that cultural differences around the expectations of censorship and freedom of speech must be accommodated at a regional level:

"the issues that we have with Facebook and LinkedIn result from the fact that they see from the perspective of the United States only. They don't consider the needs of other people. I believe there's cultural bias in that model of scaling".

5. Discussion

With this study, we respond to calls by Kapoor *et al.* (2018) and Weinhardt *et al.* (2020) for engaged IS research that offers citizens a voice in discussions around ubiquitous platforms such as social media, moving existing conversations beyond the political and commercial sphere. Despite their representation as both users and the "product" of such platforms (cf. Zuboff, 2015, 2019a), citizens and other "non-experts" are often missing from conversations on topics such as digitalization, raising questions of legitimacy and representation (Lukyanenko *et al.*, 2020; Weinhardt *et al.*, 2020). Building on findings from 16 h of online dialog with citizens, we take steps towards addressing the current dearth of citizen engagement studies in IS by investigating their perceptions of issues such as the digitalization of the individual and the digital public sphere (D'Arcy *et al.*, 2014; Tarafdar *et al.*, 2015; Turel *et al.*, 2019).

Our first research contribution centers on the exploration of citizens' perceptions of the current issues associated with social media platform governance, as well as their rationalization of the ethical dilemmas faced in addressing these issues going forward (Kapoor *et al.*, 2018). This includes the multiplicity of digital identities, sign-up consent for vulnerable groups and transparency in the moderation of content, among others. While existing research is mostly focused on the self-disclosure of singular "true" identities online and the impact of privacy risks (Cheung *et al.*, 2015), our research points towards the complex nature of digital identity and its impact on accountability online. Some citizens discussed the creation of fake personas to represent their ideal selves to certain audiences. By dichotomizing social connections this way, the user can control the type of persona they portray, returning some element of control to the user (Kang and Wei, 2020). However, despite the potential for employers to assess candidates using social media, our findings also suggest that younger users may underestimate the risk of privacy loss (Cheung *et al.*, 2015) by failing to separate their private-professional identities using different accounts. We also extend previous discussions by McHugh *et al.* (2018) around risk exposure on social media by providing real-life accounts of how vulnerable groups can be stalked by predatory groups online. Anonymity ("digital cloaking") was identified as a key enabler of online harassment and the omnipresent nature of cybermobs (Lowry *et al.*, 2016; Johnson, 2018; Ransbotham *et al.*, 2016). We also find that although parents can take steps toward protecting their children, further protections are needed at a platform level. This corresponds to accounts of 21% of children in the UK who are between 811 years of age owning social media profiles (Ofcom, 2020). Our findings show citizens' support for Mihaylov *et al.*'s (2018) assertion that algorithmic censorship can provide a better digital public sphere by filtering out malicious comments and fake news by trolls; however, they also highlight the need for algorithmic transparency to avoid false positives (algorithmic bias) and protect freedom of speech (Emerson, 1962). As highlighted by Radu (2020), the use of algorithmic censorship to fight misinformation and disinformation during the COVID-19 pandemic was not always successful and sometimes had unintended consequences. This further emphasizes the need for new and improved governance structures in the digital ecosystem.

Our second contribution is to adapt the Theory of Communicative Action (Habermas, 1984, 1990) to our discussions on how citizen engagement can harness the power of collective deliberation and rationalize potential ethical dilemmas around the governance of digital platforms (cf. Mingers and Walsham, 2010; Someh *et al.*, 2019). We highlight the importance of providing citizens with a "stage" to share their concerns and hopes for social media in the future by inductively revealing their perspectives based on a thematic analysis of transcribed

dialog. We further uncover the interplay between key dialogical practices of discourse ethics and explore their role in the collective deliberation of governance issues associated with digitalization. Our findings showcase how ordinary citizens, regardless of their educational background or subject matter expertise, can contribute to discussions through *lifeworld reasoning* and sharing their personal experiences (Khan and Krishnan, 2021). This draws on citizens' interest in the public good to bring together volunteers who wish to have their say on research and policy (Lukyanenko *et al.*, 2020). Findings also showcase how citizens engaged in the *rationalization* of ethical dilemmas faced in addressing governance issues going forward, exploring tensions such as digital accountability vs data privacy, protection vs inclusion and algorithmic censorship vs free speech. Consistent with previous research, we find that the diversity of citizens' backgrounds can stimulate *rationalization* by leveraging the plurality of views they hold (Lukyanenko *et al.*, 2020; Mingers and Walsham, 2010). In doing so, we take steps towards answering the call for more discourse ethics studies on IS research topics of societal concern, aspiring towards "an ideal-speech situation" where citizens can equitably debate different proposals (Mingers and Walsham, 2010; Someh *et al.*, 2019).

Discourse ethics guide us to the principles of engagement over norms, through consideration of pragmatics and ethical values, but also by negotiation – being willing to take the other's perspective, modify one's perspective and by ensuring that such agreements are based on the "force of argument" not the force of power (Mingers and Walsham, 2010). The inequalities of society are often reflected in how information systems are designed, which calls into question any claims that technology is value-neutral and free from bias. For instance, algorithms can funnel vulnerable users into "echo chambers", which proliferate disinformation, while trolls and bots can produce automated messaging content that triggers intergroup conflict (Jamison *et al.*, 2019). Discourse ethics centers on moral, pragmatic and ethical questions such as: How do value systems affect our actions? What are the consequences of our actions? How do we ensure justice for all (Mingers and Walsham, 2010). We invite IS scholars to consider the importance of citizen engagement for supporting discourse ethics across three frames. This includes collective thought and discourse work whereby ethics is an open-ended process of joint deliberation, the balancing a continuum of values and the appropriation of a technical fix which is solvable through technical or instrumental means (cf. Wehrens *et al.*, 2021).

For our third research contribution, we present a set of policy recommendations (*moral actions*) proposed by citizens and for citizens. Building on our findings and theoretical model, Table 3 summarizes citizens' lifeworld reasoning of governance issues in social media and their rationalization of ethical dilemmas. In any moral reasoning, the critical issues include "*Do the means justify the ends? Or do the ends justify the means"* Habermas (1984) argues that citizen engagement on such important questions can provide insights into what "the people" want from technology going forward. Habermas (1984, 1990) asserts that morality and law are not separate and instead constitute an incremental and developing system of interplay. Moral consciousness requires a transition from "*pre-conventional*" expectations of behavior to "*post-conventional*" principles - where the ethics of conviction and responsibility are underpinned by formal laws and social norms (Habermas, 1984, 1990).

Table 3 outlines "*post-conventional"* principles and calls for future research and policies on new governance models which are both fair and responsible. Although our study focuses on Habermas' conceptualization of moral action, it should be noted that other versions exist. For example, Rest's (1994) model of ethical decision-making focuses on the individual over Habermas' collective interest. In his model, Rest proposes a four-stage process that occurs after the individual has been presented with an ethical dilemma, (1) recognizing the situation (*moral awareness*), (2) evaluating choices and impacts (*moral judgment*), (3) choosing how to act (*moral intention*) and (4) the actual behavior (*moral action*) (Agag *et al.*, 2016). Technology

Digital accountability vs data privacy	
Lifeworld reasoning	Citizens shared personal experiences where the guise of multiple digital identities allowed trolls to engage in acts of hate speech without consequence. Collective digital identities were also discussed, where users engage in collaborative acts of hate speech driven by a herd mentality
Rationalization	Citizens felt that the right to human dignity is currently set against a backdrop of "anything goes" with little to no accountability for harmful acts on social media. Herein lies the dilemma between social media as an open platform for connection and contribution and the need for content moderation
Moral action	Citizens believe education on what constitutes digital accountability and how this should be addressed in the future is crucial. This should be based on an assessment of human dignity for people from all levels of society. They believe moral caveats of decency and respect for self and others also require further consideration
Protection vs inclusion	
Lifeworld reasoning	Citizens were concerned for young (often underage) users who feel increased pressure for peer validation on social media. The expectations of social conformity are pitched against the risks of vulnerable users being targeted by dangerous groups through their publicly visible profiles
Rationalization	A dilemma is perceived by citizens between the desire for inclusion on social media through equitable access to services and the need to protect vulnerable groups against risks online
Moral action	Citizens called on policymakers to increase protection by monitoring and enforcing the regulations around age requirements. Such consent policies should be inclusive but protectionist. They believe more clarity and openness are needed on 'the right to be forgotten' as per the General Data Protection Regulation (GDPR)
Algorithmic censorship vs free speech	
Lifeworld reasoning	Citizens discussed how actions taken in the digital public sphere can impact different social groups and society more broadly. Human rights laws are seen as set against ambiguity around social media platforms as "publishers"
Rationalization	Dilemmas are perceived between a duty to respect freedom of speech and the challenge of making social media a trustworthy and safe place for all. Citizens questioned who (e.g. regulators) or what (e.g. algorithms) should moderate the content
Moral action	Citizens called for legal changes at both an international and regional level. This requires consensus and cooperation among international organizations, nation-states and social media platforms. Actions must consider individual and cultural needs

Table 3.
Summary of findings from the citizen engagement study

moral action is a complex concept that examines the system, user and designer across three types of responsibility (causal, moral and role), which sometimes overlap (Johnson and Powers, 2005). The doctrine of double effect (DDE), meanwhile, attributed to St. Thomas Aquinas, focuses on the intention of the actor (McIntyre, 2019). DDE postulates that it is acceptable to cause harm when it is an inescapable consequence of achieving a greater good (Cushman, 2016). The fact that the action will cause harm can be foreseen, however, the harm must not be the goal. Foot (1967) offered the "trolley problem" as a method to examine DDE. According to the DDE, it is acceptable to redirect a runaway trolley away from five people to a side track where it will kill one person. This is because the predicted side-effect of killing one person is outweighed by the benefit of saving five lives. However, throwing a person in front of the train is not permissible to slow it down, thereby preventing it from hitting five because, in this version, you intend death (Cushman, 2016). Studies suggest that moral judgments by ordinary people are often consistent with DDE, although this process is unconscious and automatic (Cushman *et al.*, 2006). By examining the trolley problem in different scenarios, Foot (1967) highlights the complexity of moral judgment and moral actions, opining that other factors must be considered, such as avoiding injury and bringing aid.

Social media was originally intended as a place for sharing content and communicating with fellow users across the globe (Dwivedi *et al.*, 2018; Mallipeddi *et al.*, 2021; Namisango *et al.*, 2021). More recently, however, the role of such digital platforms in society has started to change, leading to numerous unintended consequences in what is now a global communication channel (Aral, 2020; Laato *et al.*, 2020; Torres *et al.*, 2018). In our "always online world", ideals about morality, human dignity and law are continuously tested by ubiquitous technologies. Moreover, many of the conventional norms of society are under attack from phenomena such as fake news and the rise of hate speech, showing existing governance models to be inadequate. Our findings suggest the need for a re-evaluation of governance models to reduce users' exposure to potential negative consequences of social media platforms. This requires a coordinated approach across different stakeholder groups.

In terms of practical contributions, our study provides insights into how citizen engagement practices can be used to engage diverse groups in IS research and policymaking processes (Lukyanenko *et al.*, 2020; Weinhardt *et al.*, 2020). Our findings contribute insights into engaged scholarship by building on the ethos that collective intelligence emerges from constructive, non-partisan forums. The primary objective of citizen engagement is to encourage transparent collaboration and transform how we understand the world (Lukyanenko *et al.*, 2020). It is, therefore, not only about deliberating for the sake of deliberating; the aim is to further improve decision-making and governance by delivering recommendations to political authorities. Our study displays the potential of citizen engagement to influence future-oriented policymaking (Lukyanenko *et al.*, 2020; Levy and Germonprez, 2017; Weinhardt *et al.*, 2020) once intertwined with the decision-making processes of strategic partners. In the "We, the Internet" project, results from citizen assemblies were later recognized in the Roadmap on Digital Cooperation issued by the UN Secretary General's (UNSG) Office and are well aligned with the roadmap presented in a UNSG options paper on the Future of Digital Cooperation. Outputs from the project were also endorsed by the German Government and the German Federal Foreign Office as part of the High-Level Panel's follow-up process, with global findings incorporated into relevant publications.

To enable new forms of informed action, we suggest the need for a "deliberative imperative" in the IS field (Habermas, 1984, 1990; Lukyanenko *et al.*, 2020). IS researchers can seek to go outside the "organizational walls" to engage diverse groups in formulating problems and developing concrete solutions that are in the best interests of civil society. Our study shows how citizen engagement can offer a mechanism for different stakeholder groups to have an active role in the definition of public policies (Mingers and Walsham, 2010; Someh *et al.*, 2019). In this model, divergent mindsets are put aside, and everyone is given a chance to speak and form enlightened viewpoints that can inspire policymakers. Most importantly, this enables citizens, as a previously underrepresented group, to discuss issues of public concern, debating proposals related to both their daily lives and those of future generations.

5.1 Limitations and future research

There are, nevertheless, limitations inherent in our study that future research can seek to address. Firstly, we recognize limitations associated with some of the primary assumptions of Habermas' Theory of Communicative Action, which can seem overly idealistic and utopian when taken at face value (Ross and Chiasson, 2011). For instance, definitions of an "*ideal speech situation*", "*consensus*" and "*moral action*" may seem to neglect some of the realities of social interactions, such as power, conflict and context. We, therefore, suggest that these theoretical assumptions must be adapted to ensure that they are practically useful for the context under investigation. Secondly, the study was primarily focused on the initial stages of engaging citizens in dialog around current challenges. As a result, an in-depth study of potential solutions and the impact derived from the project outcomes on the future

development of social media governance was beyond our paper's scope. Future studies can seek to longitudinally explore the practical impact of citizen engagement studies on mitigating the dark sides of digitalization. Findings from our study highlight urgent questions for researchers going forward, including: How can we ensure accountability with digital identities? Who decides what is to be done and who has legitimate authority to govern? Further questions abound on the introduction of sanctions, with punishment for offenders of regulated systems. This calls for action by both industry and policymakers.

Our understanding of citizen engagement in IS research and policy are only emerging, with several factors yet to be considered. Firstly, limitations associated with the information quality and the self-selection biases of participants involved in citizen engagement studies should be recognized (Lukyanenko *et al.*, 2020). We suggest that future research also explore the fit between research purpose and citizen engagement design, as well as the policymakers' assessment of whether the resulting findings stand up to political scrutiny (Ju *et al.*, 2019). Future research can also contribute insights into the potential of discourse ethics to foster change around technology use - beginning at the individual level and moving through groups to eventually support societal-level acceptance. Our findings on the need to protect underage users on social media suggest that future research should include younger users in discussions around the design and governance of online platforms. Lastly, future studies can seek to explore other "dark sides" of digitalization through engagement with the citizens directly affected by technologies as well as the industry practitioners and policymakers responsible for their governance. Our research provides initial insights into how IS researchers might support openness by engaging diverse citizen groups on socio-political issues.

6. Conclusion

In this paper, we explored concerns and ethical dilemmas surrounding the dark sides of digitalization based on findings from a citizen engagement study. Drawing on the Theory of Communicative Action (Habermas, 1984, 1990, 1993), we reveal that the interplay between different dialogical practices (*lifeworld reasoning*, *rationalization* and *moral action*) can promote critical input for the future of social media governance. We further discuss the study's practical implications and highlight the need for consultation processes to be intertwined with the policymaking processes of high-level strategic partners to deliver concrete actions that address ethical dilemmas associated with social media platforms. Research on this citizen-led process can provide valuable insights and the impulse for transformative change, beginning with an inclusive, deliberative process that explores the attitudes of individuals towards the dark sides of digitalization. Combined with targeted policy recommendations, citizen discourse can, in turn, act as input for political decision-making. Future research efforts are vital for building a deeper understanding of complex issues, such as the interaction between open dialogs and the governance of digital technologies going forward.

References

Agag, G., El-masry, A., Alharbi, N.S. and Ahmed Almamy, A. (2016), "Development and validation of an instrument to measure online retailing ethics: consumers' perspective", *Internet Research*, Vol. 26 No. 5, pp. 1158-1180, doi: 10.1108/IntR-09-2015-0272.

Agogo, D. and Hess, T.J. (2018), "How does tech make you feel? A review and examination of negative affective responses to technology use", *European Journal of Information Systems*, Vol. 27 No. 5, pp. 570-599, doi: 10.1080/0960085X.2018.1435230.

Almeida, V.A.F., Doneda, D. and Córdova, Y. (2016), "Whither social media governance?", *IEEE Internet Computing*, Vol. 20 No. 2, pp. 82-84, doi: 10.1109/MIC.2016.32.

Alvesson, M. (1996), "Communication, power, and organization", Kieser, A. (Ed.), *De Gruyter Studies in Organization: Organizational Theory and Research*, Vol. 72. Walter de Gruyter, Berlin, NY, doi: 10.1515/9783110900545.

Aral, S. (2020), *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy and Our Health – and How We Must Adapt*, Harper Collins, New York, NY.

Berners-Lee, T. (2017), *Three Challenges for the Web, According to its Inventor*, World Wide Web Foundation, available at: https://webfoundation.org/2017/03/web-turns-28-letter/ (accessed 16 March 2022).

Beschorner, T. (2006), "Ethical theory and business practices: the case of discourse ethics", *Journal of Business Ethics*, Vol. 66 No. 1, pp. 127-139, doi: 10.1007/s10551-006-9049-x.

Brough, M., Literat, I. and Ikin, A. (2020), *Good Social Media? Underrepresented Youth Perspectives on the Ethical and Equitable Design of Social Media Platforms*, Vol. 6 No. 2, Social Media + Society, pp. 1-11, doi: 10.1177/2056305120928488.

Brown, L. and Kuss, D.J. (2020), "Fear of missing out, mental wellbeing, and social connectedness: a seven-day social media abstinence trial", *International Journal of Environmental Research and Public Health*, Vol. 17 No. 12, p. 4566, doi: 10.3390/ijerph17124566.

Busch, P.A. and McCarthy, S. (2021), "Antecedents and consequences of problematic smartphone use: a systematic literature review of an emerging research area", *Computers in Human Behavior*, Vol. 114 No. 106414, pp. 1-47, doi: 10.1016/j.chb.2020.106414.

Chapple, W., Molthan-Hill, P., Welton, R. and Hewitt, M. (2020), "Lights off, spot on: carbon literacy training crossing boundaries in the television industry", *Journal of Business Ethics*, Vol. 162 No. 4, pp. 813-834, doi: 10.1007/s10551-019-04363-w.

Chen, H. and Atkin, D. (2021), "Understanding third-person perception about Internet privacy risks", *New Media and Society*, Vol. 23 No. 3, pp. 419-437, doi: 10.1177/1461444820902103.

Chen, X., Huang, C. and Cheng, Y. (2020), "Identifiability, risk, and information credibility in discussions on moral/ethical violation topics on Chinese social networking sites", *Frontiers in Psychology*, Vol. 11, doi: 10.3389/fpsyg.2020.535605.

Cheung, C., Lee, Z.W. and Chan, T.K. (2015), "Self-disclosure in social networking sites: the role of perceived cost, perceived benefits and social influence", *Internet Research*, Vol. 25 No. 2, pp. 279-299, doi: 10.1108/IntR-09-2013-0192.

Cruz, A.G.B., Seo, Y. and Rex, M. (2018), "Trolling in online communities: a practice-based theoretical perspective", *The Information Society*, Vol. 34 No. 1, pp. 15-26, doi: 10.1080/01972243.2017.1391909.

Cushman, F. (2016), "The psychological origins of the doctrine of double effect", *Criminal Law, Philosophy*, Vol. 10 No. 4, pp. 763-776, doi: 10.1007/s11572-014-9334-1.

Cushman, F.A., Young, L. and Hauser, M.D. (2006), "The role of conscious reasoning and intuition in moral judgment: testing three principles of harm", *Psychological Science*, Vol. 17 No. 12, pp. 1082-1089, doi: 10.1177/1461444820902103.

DeNardis, L. and Hackl, A.M. (2015), "Internet governance by social media platforms", *Telecommunications Policy*, Vol. 39 No. 9, pp. 761-770, doi: 10.1016/j.telpol.2015.04.003.

Dineva, D. and Breitsohl, J. (2022), "Managing trolling in online communities: an organizational perspective", *Internet Research*, Vol. 32 No. 1, pp. 292-311, doi: 10.1108/INTR-08-2020-0462.

Dwivedi, Y.K., Kelly, G., Janssen, M., Rana, N.P., Slade, E.L. and Clement, M. (2018), "Social Media: the good, the bad, and the ugly", *Information Systems Frontiers*, Vol. 20 No. 3, pp. 419-423, doi: 10.1007/s10796-018-9848-5.

D'Arcy, J., Gupta, A., Tarafdar, M. and Turel, O. (2014), "Reflecting on the 'dark side' of information technology use", *Communications of the Association for Information Systems*, Vol. 35 No. 1, pp. 109-118, doi: 10.17705/1CAIS.03505.

Emerson, T.I. (1962), "Toward a general theory of the First Amendment", *The Yale Law Journal*, Vol. 72, pp. 877-956.

Floridi, L. (2002), "What is the philosophy of Information?", *Metaphilosophy*, Vol. 33 Nos 1-2, pp. 123-145, doi: 10.1111/1467-9973.00221.

Foot, P. (1967), "The problem of abortion and the doctrine of the double effect", *Oxford Review*, Vol. 5, pp. 1-6.

Fouché, C. and Light, G. (2011), "An invitation to dialogue: 'the world Café' in social work research", *Qualitative Social Work*, Vol. 10 No. 1, pp. 28-48, doi: 10.1177/1473325010376016.

Gioia, D.A., Corley, K.G. and Hamilton, A.L. (2013), "Seeking qualitative rigor in inductive research: notes on the Gioia methodology", *Organizational Research Methods*, Vol. 16 No. 1, pp. 15-31, doi: 10.1177/1094428112452151.

Gorwa, R. (2019), "What is platform governance?", *Information, Communication and Society*, Vol. 22 No. 6, pp. 854-871, doi: 10.1080/1369118X.2019.1573914.

Guterres, A. (2020), "Report of the secretary-general roadmap for digital cooperation", United Nations, available at: https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf (accessed 28 March 2022).

Habermas, J. (1984), in McCarthy, T. (Ed.), *The Theory of Communicative Action: Reason and the Rationalization of Society*, Beacon Press, Boston, MA.

Habermas, J. (1990), "Discourse ethics: notes on a program of philosophical justification", in Habermas, J. (Ed.), *Moral Consciousness and Communicative Action*, Polity Press, Cambridge, MA.

Habermas, J. (1993), *Justification and Application*, Polity Press, Cambridge.

Han, Y., Guo, Y., E-Business, E-Management and E-Learning (2022), "Literature review of the concept of 'internet governance' based on the background of E-society", *13th International Conference on E-Education, E-Business, E-Management, and E-Learning,(IC4E 2022)*, January 14-17, 2022 Tokyo, pp. 611-618, doi: 10.1145/3514262.3514263.

Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J. and Bonn, A. (2018), "Innovation in open science, society and policy–setting the agenda for citizen engagement", *Citizen Science: Innovation in Open Science, Society and Policy*, UCL Press, London, pp. 1-23, doi: 10.14324/111.9781787352339.

Jamison, A.M., Broniatowski, D.A. and Quinn, S.C. (2019), "Malicious actors on Twitter: a guide for public health researchers", *American Journal of Public Health*, Vol. 109 No. 5, pp. 688-692, doi: 10.2105/AJPH.2019.304969.

Johnson, B.G. (2018), "Tolerating and managing extreme speech on social media", *Internet Research*, Vol. 28 No. 5, pp. 1275-1291, doi: 10.1108/IntR-03-2017-0100.

Johnson, D.G. and Powers, T.M. (2005), "Computer systems and responsibility: a normative look at technological complexity", *Ethics Information Technology*, Vol. 7 No. 2, pp. 99-107, doi: 10.1007/s10676-005-4585-0.

Ju, J., Liu, L. and Feng, Y. (2019), "Design of an O2O citizen participation ecosystem for sustainable governance", *Information Systems Frontiers*, Vol. 21 No. 3, pp. 605-620, doi: 10.1007/s10796-019-09910-4.

Kang, J. and Wei, L. (2020), "Let me be at my funniest: instagram users' motivations for using Finsta (aka, fake Instagram)", *The Social Science Journal*, Vol. 57 No. 1, pp. 58-71, doi: 10.1016/j.soscij.2018.12.005.

Kapoor, K.K., Tamilmani, K., Rana, N.P., Patil, P., Dwivedi, Y.K. and Nerur, S. (2018), "Advances in social media research: past, present and future", *Information Systems Frontiers*, Vol. 20 No. 3, pp. 531-558, doi: 10.1007/s10796-017-9810-y.

Khan, A. and Krishnan, S. (2021), "Citizen engagement in co-creation of e-government services: a process theory view from a meta-synthesis approach", *Internet Research*, Vol. 31 No. 4, pp. 1318-1375, doi: 10.1108/INTR-03-2020-0116.

Kullenberg, C. and Kasperowski, D. (2016), "What is citizen science? – a scientometric meta-analysis", *PLoS ONE*, Vol. 11 No. 1, pp. 1-16, e0147152, doi: 10.1371/journal.pone.0147152.

Laato, S., Islam, A.N., Islam, M.N. and Whelan, E. (2020), "What drives unverified information sharing and cyberchondria during the COVID-19 pandemic?", *European Journal of Information Systems*, Vol. 29 No. 3, pp. 288-305, doi: 10.1080/0960085X.2020.1770632.

Leible, S., Ludzay, M., Götz, S., Kaufmann, T., Meyer-Lüters, K. and Tran, M.N. (2022), "ICT application types and equality of E-participation - a systematic literature review", *Pacific Asia Conference on Information Systems, 2022 Proceedings*, Vol. 30, available at: https://aisel.aisnet.org/pacis2022/30 (accessed 5 July 2022).

Leidner, D. and Tona, O. (2021), "The CARE theory of dignity amid personal data digitalization", *MIS Quarterly*, Vol. 45 No. 1, pp. 343-370, doi: 10.25300/MISQ/2021/15941.

Levy, M. and Germonprez, M. (2017), "The potential for citizen science in information systems research", *Communications of the Association for Information Systems*, Vol. 40 No. 1, pp. 22-39, doi: 10.17705/1CAIS.04002.

Lin, L.Y., Sidani, J.E., Shensa, A., Radovic, A., Miller, E., Colditz, J.B., Hoffman, B.L., Giles, L.M. and Primack, B.A. (2016), "Association between social media use and depression among U.S. young adults", *Depression and Anxiety*, Vol. 33 No. 4, pp. 323-331, doi: 10.1002/da.22466.

Liu, X. and Zheng, L. (2012), "Government official microblogs: an effective platform for facilitating inclusive governance", *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, pp. 258-261, available at: https://dl.acm.org/doi/pdf/10.1145/2463728.2463778?casa_token=qg85vmOL-CMAAAAA:QhCq9a1lSCVDlDwItMHKc4nZz7twQCTS3-othb_1jOJJu3DPjA89lgSTUkVmZFzWraptI-jWMaqSaw (accessed 11 March 2022).

Lowry, P.B., Zhang, J., Wang, C. and Siponen, M. (2016), "Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model", *Information Systems Research*, Vol. 27 No. 4, pp. 962-986, doi: 10.1287/isre.2016.0671.

Lukyanenko, R., Wiggins, A. and Rosser, H.K. (2020), "Citizen science: an information quality research Frontier", *Information Systems Frontiers*, Vol. 22 No. 4, pp. 961-983, doi: 10.1007/s10796-019-09915-z.

Lyytinen, K.J. and Klein, H.K. (1985), "The critical theory of jurgen Habermas as a basis for a theory of information systems", in Mumford, E., Hirschheim, R., Fitzgerald, G. and Wood-Harper, T. (Eds), *Research Methods in Information Systems*, Elsevier Science Publishers B.V., Amsterdam, North-Holland, pp. 207-226.

Mallipeddi, R.R., Janakiraman, R., Kumar, S. and Gupta, S. (2021), "The effects of social media content created by human brands on engagement: evidence from Indian general election 2014", *Information Systems Research*, Vol. 32 No. 1, pp. 212-237, doi: 10.1287/isre.2020.0961.

McCarthy, S., Rowan, W., Lynch, L. and Fitzgerald, C. (2020), Blended stakeholder participation for responsible Information Systems research, *Communications of the Association for Information Systems*, pp. 124-149, doi: 10.17705/1CAIS.04733.

McCarthy, S., Mahony, C., Rowan, W., Tran-Karcher, H. and Potet, M. (2021), The pragmatic school of thought in open science practsice: a case study of multi-stakeholder participation in shaping the future of internet governance, *54th Hawaii International Conference on System Sciences, Kauai, Hawaii, University of Hawai'i at Manoa, USA*, 4-8 January 2021 pp. 670-679, available at: http://hdl.handle.net/10125/70693 (accessed 1 February 2022).

McHugh, B.C., Wisniewski, P., Rosson, M.B. and Carroll, J.M. (2018), "When social media traumatizes teens: the roles of online risk exposure, coping, and post-traumatic stress", *Internet Research*, Vol. 28 No. 5, pp. 1169-1188, doi: 10.1108/IntR-02-2017-0077.

McIntyre, A. (2019), "Doctrine of double effect", in Zalta, E.N. (Ed.), *The Stanford Encyclopaedia of Philosophy*, Spring 2019 Edition, available at: https://plato.stanford.edu/archives/spr2019/entries/double-effect (accessed 10 March 2022).

Mihaylov, T., Mihaylova, T., Nakov, P., Màrquez, L., Georgiev, G.D. and Koychev, I.K. (2018), "The dark side of news community forums: opinion manipulation trolls", *Internet Research*, Vol. 28 No. 5, pp. 1066-2243, doi: 10.1108/IntR-03-2017-0118.

Mingers, J. and Walsham, G. (2010), "Toward ethical information systems: the contribution of discourse ethics", *MIS Quarterly*, Vol. 34 No. 4, pp. 833-854, doi: 10.2307/25750707.

Molas-Gallart, J. (2012), "Research governance and the role of evaluation: a comparative study", *American Journal of Evaluation*, Vol. 33 No. 4, pp. 583-598, doi: 10.1177/1098214012450938.

Namisango, F., Kang, K. and Beydoun, G. (2021), "How the structures provided by social media enable collaborative outcomes: a study of service Co-creation in non-profits", *Information Systems Frontiers*, Vol. 24 No. 2, pp. 517-535, doi: 10.1007/s10796-020-10090-9.

Ngwenyama, O. and Lee, A. (1997), "Communication richness in electronic mail: critical theory and the contextuality of meaning", *MIS Quarterly*, Vol. 21 No. 2, pp. 145-167, doi: 10.2307/249417.

Ofcom (2020), *Children and Parents: Media Use and Attitudes Report 2019*, Ofcom, London, available at: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2019 (accessed 2 February 2022).

Olphert, W. and Damodaran, L. (2007), "Citizen participation and engagement in the design of e-government services: the missing link in effective ICT design and delivery", *Journal of the Association for Information Systems*, Vol. 8 No. 9, pp. 491-507, doi: 10.17705/1jais.00140.

Patton, M.Q. (2002), "Qualitative research and evaluation methods", *Qualitative Research and Evaluation Methods*, 3rd ed., Sage, Thousand Oaks, CA.

Radu, R. (2020), "Fighting the 'infodemic': legal responses to COVID-19 disinformation", *Social Media + Society*, Vol. 6 No. 3, pp. 1-4, doi: 10.1177/2056305120948190.

Ransbotham, S., Fichman, R.G., Gopal, R. and Gupta, A. (2016), "Special section introduction—ubiquitous IT and digital vulnerabilities", *Information Systems Research*, Vol. 27 No. 4, pp. 834-847, doi: 10.1287/isre.2016.0683.

Rauf, A.A. (2021), "New moralities for new media? Assessing the role of social media in acts of terror and providing points of deliberation for business ethics", *Journal of Business Ethics*, Vol. 170 No. 2, pp. 229-251, doi: 10.1007/s10551-020-04635-w.

Rest, J. (1994), "Background: theory and research", in Rest, J. and Narvaez, D. (Eds), *Moral Development in the Professions: Psychology and Applied Ethics*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 1-26.

Richey, M., Gonibeed, A. and Ravishankar, M.N. (2018), "The perils and promises of self-disclosure on social media", *Information Systems Frontiers*, Vol. 20 No. 3, pp. 425-437, doi: 10.1007/s10796-017-9806-7.

Robbins, S. and Henschke, A. (2017), "The value of transparency: bulk data and authoritarianism", *Surveillance and Society*, Vol. 15 Nos 3/4, pp. 582-589.

Ross, A. and Chiasson, M. (2011), "Habermas and information systems research: new directions", *Information and Organization*, Vol. 21 No. 3, pp. 123-141, doi: 10.1016/j.infoandorg.2011.06.001.

Salo, M., Pirkkalainen, H., Chua, C. and Koskelainen, T. (2022), "Formation and mitigation of technostress in the personal use of IT", *MIS Quarterly*, Vol. 46 No. 2, pp. 1073-1108, doi: 10.25300/MISQ/2022/14950.

Someh, I., Davern, M., Breidbach, C.F. and Shanks, G. (2019), "Ethical issues in big data analytics: a stakeholder perspective", *Communications of the Association for Information Systems*, Vol. 44, doi: 10.17705/1CAIS.04434.

Sriwilai, K. and Charoensukmongkol, P. (2016), "Face it, don't Facebook it: impacts of social media addiction on mindfulness, coping strategies and the consequence on emotional exhaustion", *Stress and Health*, Vol. 32 No. 4, pp. 427-434, doi: 10.1002/smi.2637.

Stahl, B.C. (2012), "Responsible research and innovation in information systems", *European Journal of Information Systems*, Vol. 21 No. 3, pp. 207-221, doi: 10.1057/ejis.2012.19.

Statista (2022), Social media – statistics and facts, available at: https://www.statista.com/topics/1164/social-networks/#dossierKeyfigures (accessed 3 March 2022).

Stevens, R., Dunaev, J., Malven, E., Bleakley, A. and Hull, S. (2016), "Social Media in the sexual lives of African American and Latino youth: challenges and opportunities in the digital neighborhood", *Media and Communication*, Vol. 4 No. 3, pp. 60-70, doi: 10.17645/mac.v4i3.524.

Stoycheff, E., Burgess, G.S. and Martucci, M.C. (2020), "Online censorship and digital surveillance: the relationship between suppression technologies and democratization across countries", *Information, Communication and Society*, Vol. 23 No. 4, pp. 474-490, doi: 10.1080/1369118X.2018.1518472.

Tacke, O. (2010), "Open science 2.0: how research and education can benefit from open innovation and Web 2.0", in Bastiaens, T.J., Baumöl, U. and Krämer, B.J. (Eds), *On Collective Intelligence. Advances in Intelligent and Soft Computing*, Vol. 76, Springer, Berlin, Heidelberg, pp. 37-48, doi: 10.1007/978-3-642-14481-3_4.

Tarafdar, M., Gupta, A. and Turel, O. (2015), "Special issue on 'dark side of information technology use': an introduction and a framework for research", *Information Systems Journal*, Vol. 25 No. 3, pp. 161-170, doi: 10.1111/isj.12070.

Thompson, R. (2011), "Radicalization and the use of social media", *Journal of Strategic Security*, Vol. 4 No. 4, pp. 167-190, doi: 10.5038/1944-0472.4.4.8.

Torres, R., Gerhart, N. and Negahban, A. (2018), "Epistemology in the era of fake news: an exploration of information verification behaviors among social networking site users", *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, Vol. 49 No. 3, pp. 78-97, doi: 10.1145/3242734.3242740.

Turel, O., Matt, C., Trenz, M., Cheung, C.M.K., D'Arcy, J., Qahri-Saremi, H. and Tarafdar, M. (2019), "Panel report: the dark side of the digitization of the individual", *Internet Research*, Vol. 29 No. 2, pp. 274-288, doi: 10.1108/INTR-04-2019-541.

van der Velden, M. and Mörtberg, C. (2021), "Citizen engagement and design for values". In van den Hoven, J., Vermaas, P.E. and van de Poel, I. (Eds), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, Dordrecht, Springer Netherlands, pp. 1-22.

Van Dijck, J. (2021), "Seeing the forest for the trees: visualizing platformization and its governance", *New Media and Society*, Vol. 23 No. 9, pp. 2801-2819, doi: 10.1177/1461444820940293.

Walsham, G. (2012), "Are we making a better world with ICTs? Reflections on a future agenda for the IS field", *Journal of Information Technology*, Vol. 27 No. 2, pp. 87-93, doi: 10.1057/jit.2012.4.

Walther, J.B., Liang, Y.J., De Andrea, D.C., Tong, S.T., Carr, C.T., Spottswood, E.L. and Amichai-Hamburger, Y. (2011), *The Effect of Feedback on Identity Shift in Computer-Mediated Communication*, Vol. 14, Media Psychology, pp. 1-26, doi: 10.1080/15213269.2010.547832.

Webb, H., Jirotka, M., Stahl, B.C., Housley, W., Edwards, A., Williams, M., Procter, R., Rana, O. and Burnap, P. (2016), "Digital wildfires: hyper-connectivity, havoc and a global ethos to govern social media", *ACM SIGCAS Computers and Society*, Vol. 45 No. 3, pp. 193-201, doi: 10.1145/2874239.2874267.

Wehrens, R., Stevens, M., Kostenzer, J., Weggelaar, A.M. and de Bont, A. (2021), "Ethics as discursive work: the role of ethical framing in the promissory future of data-driven healthcare technologies", *Science, Technology, and Human Values*, pp. 1-29, doi: 10.1177/01622439211053661.

Weinhardt, C., Kloker, S., Hinz, O. and van der Aalst, W.M.P. (2020), "Citizen science in information systems research", *Business Information Systems Engineering*, Vol. 62, pp. 273-277, doi: 10.1007/s12599-020-00663-y.

Winefield, H.R., Gill, T.K., Taylor, A.W. and Pilkington, R.M. (2012), "Psychological well-being and psychological distress: is it necessary to measure both?", *Psychology of Well-Being*, Vol. 2 No. 1, pp. 1-14, doi: 10.1186/2211-1522-2-3.

World Economic Forum (2013), "Digital wildfires in a hyperconnected world. Global risks report", World Economic Forum, available at: http://reports.weforum.org/global-risks-2013/risk-case1/digital-wildfires-in-a-hyperconnected-world/ (accessed 30 March 2022).

World Health Organization (2022), "Overview of the infodemic", WHO Health Topics, available at: https://www.who.int/health-topics/infodemic#tab=tab_1 (accessed 7 February 2022).

Yuthas, K., Rogers, R. and Dillard, J.F. (2002), "Communicative action and corporate annual reports", *Journal of Business Ethics*, Vol. 41 No. 1, pp. 141-157, doi: 10.1023/A:1021314626311.

Zuboff, S. (2015), "Big other: surveillance capitalism and the prospects of an information civilization", *Journal of Information Technology*, Vol. 30 No. 1, pp. 75-89, doi: 10.1057/jit.2015.5.

Zuboff, S. (2019a), *The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power*, Profile Books, London.

Zuboff, S. (2019b), "Surveillance capitalism and the challenge of collective action", *New Labor Forum*, Vol. 28 No. 1, pp. 10-29, doi: 10.1177/1095796018819461.

Appendix 1
Guidelines for discourse facilitation

As a facilitator, you have one of the most important roles during the day. You are the guarantor of the quality of the deliberation at your table, and hence at all tables and for all the dialogs. The quality of the results often depends on the quality of the facilitation.

In a nutshell, a good facilitator must:

(1) *Be neutral:* be aware of your "power" on participants and not misuse it. Your role is not to influence deliberations, but to facilitate them by reformulating, and helping participants to express their thoughts.

(2) *Be aware of the group's dynamics*: pay attention to the participants and be able to identify if someone at the table is uncomfortable.

(3) *Be clear*: reformulate and make sure participants understand the purpose of the discussions.

(4) *Be inclusive*: identify participants that are participating less and try to make them comfortable enough to speak out.

(5) *Be polite*: always be friendly, even with participants that are less polite.

The dialog must above all respect one value: inclusiveness. We are committed to it.

How do we do this?

(1) By welcoming each participant individually at your table (virtual or real), and by making sure that they feel comfortable and have a good day.

(2) By introducing yourself at the beginning of the day and when participants change tables (if the participants change table/virtual room). That way, citizens at your table can identify you and their fellow citizens.

(3) By explaining the principle of inclusiveness to citizens from the very beginning of the day: *"No one here is an expert on the Internet and its issues. We are here to listen to you and to give you the keys you need to deepen your opinions. You are here to exchange with another"*.

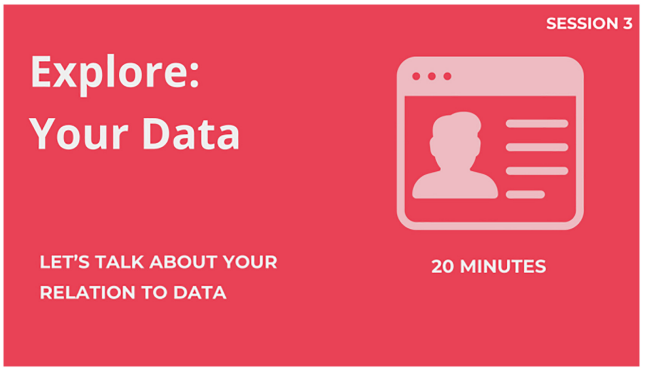
As a facilitator, you must regulate speaking: some people are more comfortable speakers than others. In a group there may be some people who monopolize the floor and others who are naturally withdrawn. In this case, do not hesitate to ask the latter if they have anything to add or supplement to what has been said, without pushing them to speak if they do not want to. On the contrary, you will have to regulate participants that speak a bit too much and do not leave space for others.

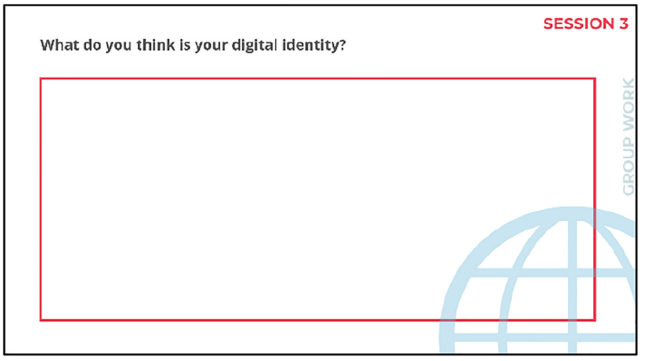
Appendix 2
Anonymized list of volunteer participants

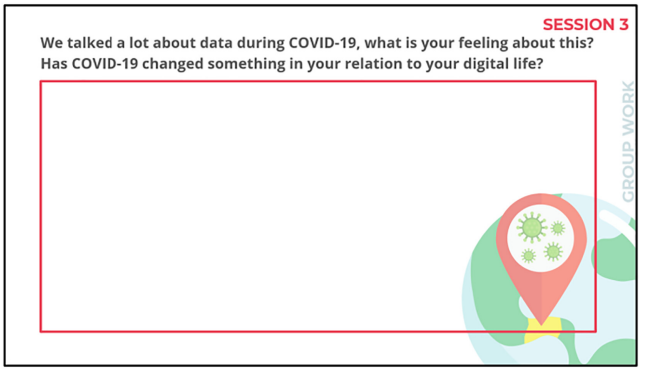
ID	Background	Subgroup
Participant 1	Retired senior citizen	Subgroup 1
Participant 2	Small business owner	Subgroup 1
Participant 3	Head of social media in an advertising agency	Subgroup 1
Participant 4	Teacher in a primary school	Subgroup 1
Participant 5	Unemployed	Subgroup 1
Participant 6	IT Director	Subgroup 2
Participant 7	Coach and Mentor	Subgroup 2
Participant 8	Undergraduate student	Subgroup 2
Participant 9	Social media business evangelist	Subgroup 2
Participant 10	University lecturer	Subgroup 2
Participant 11	Marketing and communications professional	Subgroup 3
Participant 12	PhD Student	Subgroup 3
Participant 13	Postgraduate student (Law)	Subgroup 3
Participant 14	Independent non-executive Chairman and Director	Subgroup 3
Participant 15	Retired senior citizen	Subgroup 3
Participant 16	Information scientist	Subgroup 4
Participant 17	University lecturer	Subgroup 4
Participant 18	Researcher/legal practitioner	Subgroup 4
Participant 19	Chief Executive Officer	Subgroup 4
Participant 20	Student (Applied Psychology)	Subgroup 4
Participant 21	Undergraduate student	Subgroup 5
Participant 22	Marketing manager	Subgroup 5
Participant 23	Undergraduate student	Subgroup 5
Participant 24	Administration professional	Subgroup 5
Participant 25	Owner of HR firm	Subgroup 5

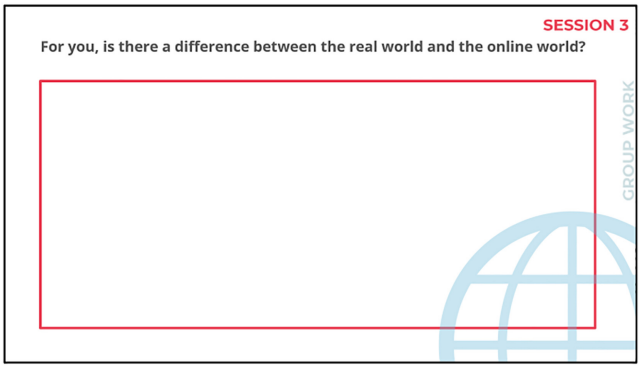
Appendix 3

Sample templates from the citizen engagement study

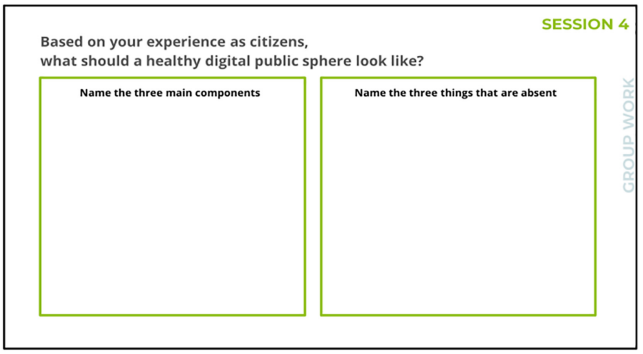


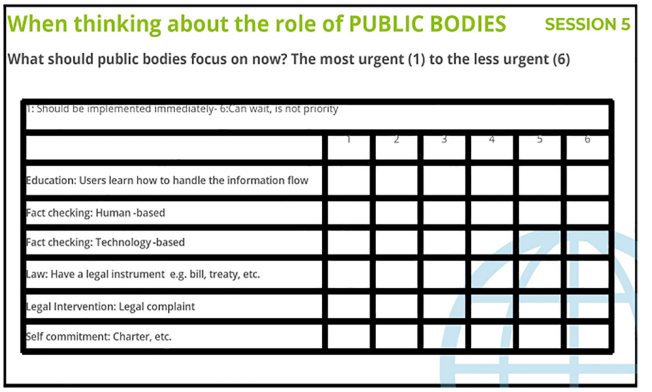


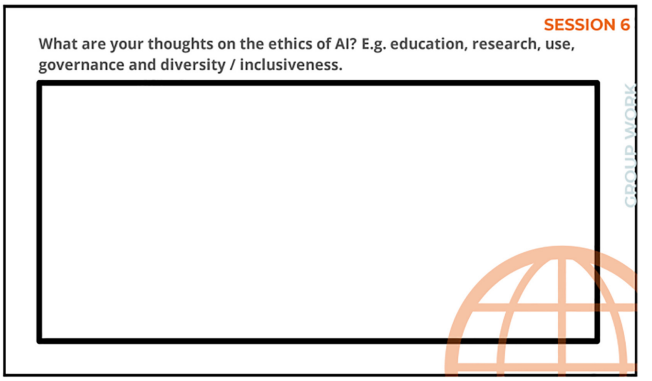



SESSION 3
For you, is there a difference between the real world and the online world?
GROUP WORK

SESSION 5
Choose:
Tools to fight disinformation
HOW DO YOU SEE THE TOOLS TO FIGHT DISINFORMATION?


SESSION 4
Based on your experience as citizens, what should a healthy digital public sphere look like?
Name the three main components
Name the three things that are absent
GROUP WORK





Corresponding author
Stephen McCarthy can be contacted at: stephen.mccarthy@ucc.ie

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**