

Diversity-sensitive social platforms and responsibility

Some ethical considerations

There is an ongoing debate on how algorithms and machine learning can and should deal with human diversity while avoiding the pitfalls of statistical stereotyping, the re-enforcement of clichés and the perpetuation of unjust discrimination. Computer scientists try to tackle these issues by developing algorithms and social-interaction protocols for mediating diversity-aware interactions between people, for instance on diversity-sensitive social platforms. At the same time, diversity-related data often comprise sensitive personal data, and their collection, storage and management increases the vulnerability of users to various misuse scenarios. Already this observation leads to the question, how do we need to conceptualize responsibility to do justice to the increased vulnerability? In this paper, I thus focus on the questions a diversity-sensitive social platform raises with regard to responsibility, and propose a tentative ethical framework of responsibility for these platforms.

Keywords: *algorithms, applied ethics, diversity, responsibility, social platforms*

Acknowledgement

The research for this paper was partially conducted as part of the European Union Horizon 2020 Project “WeNet - The Internet of us” (Grant no. 823783) and as part of the Baden–Württemberg Foundation Project “AITE – Artificial Intelligence, Trustworthiness and Explainability”. A draft version was presented at the Budapest Workshop on Philosophy of Technology in December 2019. I want to thank the participants of that workshop for the helpful discussions and comments that followed. I also want to thank Jessica Hesen, Moritz Hildt and an anonymous reviewer for their helpful suggestions.

Author Information

Karoline Reinhardt, Eberhard Karls Universität Tübingen

<https://uni-tuebingen.de/en/facilities/central-institutions/international-center-for-ethics-in-the-sciences-and-humanities/team/dr-karoline-reinhardt/>

How to cite this article:

Reinhardt, Karoline. “Diversity-sensitive social platforms and responsibility.”

Információs Társadalom XXI, no. 2 (2021): 43–62.

==== <https://dx.doi.org/10.22503/inftars.XXI.2021.2.4> ====

*All materials
published in this journal are licenced
as CC-by-nc-nd 4.0*

1. Introduction

Despite the apparent diversity of humans, technology in general and digital solutions in particular still struggle to take diversity into account. Thus, diversity poses not only an ethical but also a computational challenge. There is an ongoing debate on how computerized algorithms and machine learning algorithms can and should deal with human diversity while avoiding the pitfalls of statistical stereotyping, the re-enforcement of clichés and the perpetuation of discrimination against particular groups.

Computer scientists and developers are increasingly aware of these problems, and are trying to develop technical methods, algorithms and social-interaction protocols that enable machine-mediated diversity-aware interactions between people, for instance on diversity-sensitive social platforms (cf. Reinhardt 2020a, 272). To measure diversity and to make sure that diversity is instantiated in turn requires the collection of data. Diversity-related data, however, often comprise sensitive personal data. The collection of diversity-related data, thus, touches not only on questions of informational privacy, but also increases the risk of various misuse scenarios, online (e.g. hate speech, trolling, cyberbullying, misinformation) and offline (e.g. persecution).

The often, though not always, honourable goal, to make a social platform diversity-aware, thus, comes with an increased vulnerability for the users and data subjects of the platform – a vulnerability that is increasingly exploited by some service providers and other platform users. Hence, this vulnerability alone gives us sufficient reason to think about questions of responsibility with regard to diversity-sensitive social platforms. Additionally, technical innovations, especially those with an undeniable extent of societal impact, have consequences – both intended and unintended. It is part of responsible research, as well as of responsible action in general, to be mindful of the possible consequences in advance and to set up structures that do justice to this fact. This also includes a reflection on the appropriate concepts and structures of responsibility. Concerning social platforms, this has not yet been sufficiently done, though there are of course some notable exceptions¹. This article is, thus, intended also as a contribution to a more general debate on responsibility and social platforms.

In what follows, I outline some ethical questions that arise from the design process of diversity-aware social platforms. To do this, I first define what platforms and social platforms are and explain that the term diversity itself also raises some conceptual questions that we should bear in mind. Then, I will focus on the questions a diversity-sensitive social platform raises with regard to responsibility, and sketch a tentative ethical framework of responsibility that formulates some recommendations for relevant stakeholders.

2. Diversity-aware social platforms and their flip side

In computer science, the notion ‘platform’ refers to a unified basis on which application programs can be executed and developed. Platforms are digital structures

¹ With regard to responsibility and content moderation as well as to content generation, these include, among others, the works of Gillespie (2018a, 2018b) and Gorwa, Binns and Katzenbach (2020).

that store data and enable and mediate interactions on a large scale. There is a great variety of platforms on the market and their development and evolution are highly dynamic. Platforms include search engines, comparison and rating portals, marketplaces, trading platforms, media and content services, online games, social networks and communication services.

Herein, I refer to platforms as social platforms when their aim is to connect people, or to mediate interactions between individual persons – in contrast, for instance, to platforms that enable digital communication between industrial machines. This distinction is, of course, based on the implicit thesis that the interaction of humans – whether digitally mediated or not – raises ethical questions that are different from the ethical questions that are raised by the interaction of machines. Such interactions on social platforms include, but are not limited to, transactions, content production and consumption, or joint activities that remain within the virtual world or go beyond it. Social media sites are among the most prevalent examples of social platforms.

Increasingly, social platforms transcend the management and storage of user-created information and employ algorithmic decision-making procedures (cf. Gillespie 2018a; 2018b). This gives them considerable influence over their users – an influence that is often designed to be imperceptible. In this way, platforms have become important players in their respective fields with an often unknown and still insufficiently understood impact on people's behaviour and decisions.² However, although the algorithmic decision-making procedures deployed in social platforms are highly evolved, they often seem to struggle with the diversity of humans and tend to reproduce, perpetuate and amplify discrimination and stereotypes (cf. Heesen, Reinhardt and Schelenz 2021). As computer science has become more and more aware of this, some developers and designers have tried to make algorithms and interaction protocols more diversity-sensitive.

There are, of course, many different approaches to the topic of diversity. Some start from an instrumental understanding of diversity, in which diversity is understood as fostering other organizational goals; for instance, economic success (cf. Mensi-Klarbach 2012), while others take a normative approach to diversity; for instance, by emphasizing its intrinsic value (Parekh 2000). There are also representational and inclusive concepts of diversity (Young 1990), and concepts that emphasize the ambiguities of human diversity (Reinhardt 2020a). In any case, making algorithmic decision-procedures diversity-sensitive requires the collection of data that comprise – depending on the paradigm of diversity used – usually highly sensitive personal data.

A platform that collects, stores and works with data relevant to diversity, thus carries a considerable risk of increasing the vulnerability of individuals to all kinds of misuse scenarios such as harassment, stalking, political persecution and many others.³ If a platform is, furthermore, designed to not only connect people online but also to help

² In recent years, we have seen a number of publications that have taken a critical stance towards platforms and the underlying algorithms and mathematical models, among them O'Neil (2016), Gillespie (2018a; 2018b), van Dijk, Poell and de Waal (2018) and Zuboff (2019).

³ For an overview of the risks associated with social media platforms as one of the primary examples of social platforms, see Brake (2014).

them to organize and engage in joint activities in the offline world, these risks transcend the virtual lives we live and become real-world threats. That means creating a diversity-aware social platform, which *prima facie* seems to be a good thing, does in fact come with the risk of increasing the vulnerability of people to all kinds of risky and potentially horrible scenarios. This raises multiple questions, including: Who is responsible for the often unintended consequences of a diversity-aware social platform – and to what extent? How are we even to conceptualize responsibility with regard to platform-mediated interactions that by large part are powered by the aggregation of data?

3. Responsibility: Traditional concepts and evolving technologies

3.1. Responsibility in ethics and moral philosophy

There is an ongoing debate on (moral) responsibility in philosophy, which I will not attempt to summarize in its entirety here. Instead, in what follows, I start from a conception of responsibility that focuses on the responsibility for one's own actions, as formulated, for instance, by Nida-Rümelin (2011). This approach does not differ from common theories of responsibility in all aspects, but does so in one crucial respect, which I will highlight here. Since this text is, however, primarily concerned with conceptions of responsibility with regard to diversity-sensitive social platforms, I will only outline this action-centred approach here, but will not provide an independent justification of this perspective on responsibility.

Many takes on responsibility in ethics and moral philosophy share the following three elements as summarized by Noorman (2020): “A person is usually only held responsible if she had some control over the outcome of events”, and the person in question was “able to freely choose to act” in that way or another. However, we “tend to excuse someone from blame if they could not have known that their actions would lead to a harmful event” (Noorman 2020). Lenk argues that responsibility could be summarized in the following way: Somebody, the bearer of responsibility, is responsible for something – an action, consequences, a state of affairs, but also to fulfil certain tasks – to an addressee, possibly before a sanctioning authority with regard to a certain (prescriptive/normative) criterion in a given context (Lenk 2016, 7; see also Höffe 2008, 326).

As intuitive as it may seem at first glance, the list of what people are in general taken to be responsible for (actions, consequences, state of affairs, tasks) is not uncontroversial and it is exactly this point where Nida-Rümelin introduces an important specification. According to Nida-Rümelin, we are only responsible for our own actions – and not for all the consequences that follow causally from our actions.

Let me elaborate this point a little: That I am ultimately responsible for my actions does not mean that their consequences play no role: Each action has a context from which certain probability distributions for possible consequences of this action result. These probability distributions are part of a complete description of the action (Nida-Rümelin 2011, 113). To put it in Nida-Rümelin's words, this means: “I bear responsibility for the possible consequences of my action, weighted according to their respective probabilities” (Nida-Rümelin 2011, 111, my translation).

And I would add taking into account those measures that I take so that negative consequences do not occur for others. For example, the safety measures I take so that, when I brake at a red light, no-one is going to rear-end me, like driving with foresight, taking a look in the rear-view mirror before braking, and not braking suddenly but slowly if possible, help to prevent bad things from happening from my action (braking at a red light). If despite all these measures (provided they are executed properly) a rear-end collision happens, it is not my responsibility. The distractions of other drivers are not my responsibility as long as I give them enough time to react properly to my actions. The collision is, then, a causal consequence of my braking at the red light (it would not have occurred if I had not braked), but I am not responsible for it. The same holds true for “coincidences, which I cannot control, that lead to the realization of one sequence of actions and not the other” (Nida-Rümelin 2011, 111).

I bear, though, a greater responsibility to take countermeasures the more likely a negative consequence of my action is for others or when a certain probability is considered unacceptable for a particular consequence of action. To give an example: Even a 5% chance that my action will result in a person’s death is unacceptable, even though the number is relatively low.

So I am only responsible for those consequences that are part of the probability distribution of my action in a given context to the extent that those consequences are likely. Since I may seek to always do good to other persons, but not to harm them against their will (Nida-Rümelin 2011, 214), possible harm to others has to have priority in my deliberations on how to act.

This concept of responsibility, moreover, allows attributing responsibility even if there are no negative consequences, since it is concerned solely with the action and the probability distribution of its possible consequences. It is also irrelevant for the attribution of responsibility, according to this approach, whether the consequences actually occur. Nida-Rümelin uses the following example to illustrate this point: Someone throws paving stones from a bridge onto a busy highway. Even if, fortunately, no cars are hit and no-one is killed, the probability that this action will have fatal consequences is unacceptably high: “The favourable coincidence does not relieve the person of her responsibility. She bears responsibility for an action that involves unacceptably high risks to other people” (Nida-Rümelin 2011, 112).

On top of that, I would like to emphasize a point that is not stressed by Nida-Rümelin: Such an understanding of responsibility ultimately permits a high degree of proactivity. Since the approach does not focus on the consequences of action, it requires actors to take countermeasures against possible negative consequences before they even happen and, more importantly, to think with foresight. I will build on this action-centred understanding of responsibility in what follows, and apply it to the questions at hand and expand it where needed.

3.2. Responsibility in the age of digitalization

Traditional concepts of responsibility raise a number of fundamental philosophical questions: They work with assumptions about causality and accountability. They

also raise epistemological and psychological questions. They refer to the notion of freedom and choice and in general assume that we do have some kind of control over the outcome of our actions. All of these issues are highly controversial topics within philosophical debates and beyond. With regard to technological change and digitalization, even more problems arise, as described below.

Traditionally, philosophical ethics have focussed on human actions and interactions with respect to responsibility. Responsibility was about human actions, their intentions and consequences (Fisher 1999). However, as our world has become more and more defined by technological artefacts, it has become less and less appropriate to solely take human action into account when tackling questions of responsibility. Technological artefacts change how we view the world and ourselves (Weizenbaum 1976), but they also have an impact on our decisions and how we make them (Latour 1992). This seems in particular true with regard to computers and digital tools. They have become “active mediators” that “actively co-shape people’s being in the world” (Verbeek 2006, 364). A responsibility concept that solely focusses on human-to-human interaction, thus, seems in many ways inapt in a world that is increasingly shaped and defined by technologically mediated interactions.

Another prominent problem of a traditional concept of responsibility refers to the nexus of complexity, accountability and attributing responsibility. Since “computer technologies can obscure the causal connection between a person’s action and the eventual consequences” (Noorman 2020), attributing responsibility and defining who is accountable for a particular outcome of a specific technologically mediated interaction has become more difficult.

This observation is in some ways linked to a further issue, namely the complexity problem. Complex technologies can make it difficult to ascribe responsibility (Johnson and Powers 2005; cf. Noorman 2020), because humans increasingly do not understand the workings of the machines they use. Therefore, some argue: “The more complex computer technologies become and the less human beings can directly control or intervene in the behaviour of these technologies, the less we can reasonably hold human beings responsible for these technologies” (Noorman 2020; cf. Matthias 2004). Note that this point about complexity is even more true of social platforms where not only the technology is complex but also the interaction patterns on them.

A further complication when ascribing responsibility is that computerized human interactions possibly distance us (spatially and temporarily) from the outcome of our actions. With the help of a social platform, you can get in touch with people from different places, and the traces you leave on a social platform might, through their temporal persistence, have effects on others a long time after you have left them. This, in turn, leads to epistemic problems; whereby it might be quite difficult to fully comprehend what might happen with the content you produce now in the future, or what consequences it might have in other societal contexts that you might have little knowledge about.

Interactions on social platforms, furthermore, raise questions about collective responsibility, since in the case of a social platform, we not only deal with individuals that we interact with, but with a whole community of users whose actions affect

people online and offline. The notion of collective responsibility makes the group out as the bearer of responsibility – distinct from each individual member of that respective group. Cybermobbing and cyberbullying are important examples of how scale and publicity can have a particular destructive impact. Though mobbing and bullying are (sadly) not new phenomena, what is different in the online world is how persistent content is and how easily content can be replicated and traced (cf. Schmidt 2016, 290). Therefore, cyberbullying has also different psychological consequences: The knowledge that the defamations and insults used may be public and that access to them is not limited to the original perpetrators and bystanders, i.e. that they are stored and can in principle be found and shared by anyone, makes it much more difficult to deal with such attacks.

Collective responsibility is, however, a highly controversial topic in philosophy. One of the controversies is whether groups have what it takes to be attributed moral agency and thus moral responsibility. In particular, the question of whether and how we can make out something like intentionality for a collectives take centre stage. Another controversy about collective responsibility deals with how to distribute responsibility (or accountability and liability) among the members of the group in question. A third controversy revolves around the question about what effects a notion of collective responsibility would have with regard to individual responsibility. One prominent issue in this debate is whether we run the risk of replacing the notion of individual responsibility altogether.⁴ In this light, some argue that collective responsibility exists, if at all, only as a cooperative responsibility, that is, as a responsibility for joint action. Everything else, such as ascribing moral agency and intentionality to groups, would be simply ‘mystifications’ (Nida-Rümelin 2018, 393). Cooperative responsibility assumes that at the beginning of cooperative action there is an explicit or implicit consensus on the goals to be jointly achieved (Nida-Rümelin 2011, 119). Whoever then participates in a cooperative act is jointly responsible for its execution. The causal role of one’s own contribution is therefore irrelevant for the attribution of responsibility (Nida-Rümelin 2011, 125).

Coming back to the example of social platforms, as helpful as the distinction Nida-Rümelin made for acts in our offline lives, how helpful is it for highly complex, interconnected, technically mediated actions of various people who possibly have never met or not even communicated? Their actions, as in the case of cyberbullying, sometimes lead to hideous consequences, but we seem not to be always able to determine that there has been some sort of consensus on the nature of the action or its intended results. It is not the individual insult, although wrong in itself, that has disastrous consequences here, but, as mentioned above, the combination of scale, shareability and persistence. A strength of the notion of cooperative responsibility here is that that the quantitative contribution of the individual action is not decisive nor is the causal role of one’s contribution, but only the shared intention to act in a certain way. What it lacks, however, with regard to the interaction patterns on social

⁴ For an overview of the debate on collective responsibility, see Smiley (2017).

platforms, is that there is, in practice, not always a shared intention.⁵ In light of all this, it might be worthwhile to think about a notion of responsibility for ‘aggregated action’: This type of responsibility shares with the concept of cooperative responsibility the observation that the individual act and the blameworthiness attributed to it may be out of proportion regarding the damage caused by the sum of all the individual acts contributing to the outcome.

In any case, if one thinks that the notion of cooperative responsibility is sound, then it does not only apply to the community of users on a platform but also to the team that is designing the platform or the application. But also in this case the question arises as to whether it is really solely the agreement to commit a joint action for which responsibility can be ascribed, or whether, at least in some cases, responsibility for aggregated action is called for.

This leads us to the professional responsibility of researchers, developers and designers for the knowledge and products they produce and the question about whether they can be held responsible for the (often unintended) consequences of the inventions they have created (cf. Verbeek 2006, 379). One problem, among others, with attributing responsibility for unforeseen and, thus, unintended consequences is the ‘interpretative flexibility’ of users. Since “[p]eople often use technologies in ways unforeseen by their designers. This interpretative flexibility makes it difficult for designers to anticipate all the possible outcomes of the use of their technologies” (Noorman 2020).

Connected to the problem of attributing responsibility and professional responsibility is the “problem of many hands” (cf. Jonas 1984; Nissenbaum 1994 and 1996, 28-32; van de Poel, Royakkers and Zwart 2015): Since the development of a technology like a social platform involves many actors (designers, developers, computer scientists, sociologists, ethicists), it is often difficult to ascribe responsibility for a certain outcome created by that technology. In any case, an action-oriented (and not consequence-oriented) concept of responsibility holds that even if a technology in whose development I am involved has serious consequences that are even causally related to my actions, I may not ultimately be responsible for them, as shown above. Nevertheless, I am still responsible for ensuring I do not act negligently and that means also anticipating the possible consequences of my actions within human limits – and also, I would want to argue, anticipating the possible contribution of my actions to certain outcomes, if aggregated. In particular, when I am using a tool that works by aggregating information on actions like a social platform, my responsibility also entails including other perspectives in order to foresee as many side effects as possible and to adapt the product accordingly, as well as to regularly evaluate already existing products for their possibly

⁵ I am aware that a shared intention is often present even in mediated interaction. For example, if a group has the goal of harming somebody by insulting he or him online, then there is a shared intention to act. What I am getting at here is that somebody might also, for instance, ‘like’ and ‘share’ some of these insults out of different, though possibly still malicious, intentions. For instance, if the content is obviously insulting and somebody ‘likes’ and ‘shares’ it out of Schadenfreude, but with no intention of harming a real person – they might, for instance, assume that the person depicted is generic – this act is morally questionable but the person sharing and liking the content is not participating in a cooperative action with the original group that set up the insulting content.

unintended side effects and to improve them if necessary. In addition, the product should be tested on a small scale before use and all groups potentially affected by the use of the product, such as a platform, should be involved in the planning process.

Last but not least, it is worth noting that technologically mediated interactions do not take place in a vacuum, but in a societal and political framework. What about creating the legal framework that, for instance, defines mandatory certification systems that make it easier to understand which platform is trustworthy and which is not, prohibits problematic products as well as problematic ways of using products and ensures a certain level of consumer protection? In other words: Where does the political responsibility to set up a framework for the interactions begin, and where does it end?

In the face of all these problems and difficulties: Should we abandon the concept of responsibility in the age of digitalization? I do not think so. Although artefacts “may influence and shape human action, they do not determine it” (Noorman 2020). Although it might be hard to attribute responsibility, it is not impossible – in many cases, one might even argue, a digitalized society makes it all too easy to trace back who has done what and contributed to a given outcome in what way. The problem of distance through time and space also arises with regard to many other ethical issues and is not particular to responsibility under the conditions of an increasingly datafied and digitalized world. This issues is also under debate for instance with regard to global poverty relief and global distributive justice, and climate change mitigation.

The concept of responsibility is not rendered superfluous simply because it is sometimes difficult to determine who is responsible for what and in what way and to whom. If anything, the many layers to the concept show how relevant the concept is and that we cannot (and ought not) side-step it easily.

4. An ethical framework of responsibility for diversity-sensitive social platforms

In what follows, I do not propose a single unified concept of responsibility, but rather start from the observation that there are various dimensions of responsibility (cf. Höffe 2008, 326) and that we need a multidimensional concept of responsibility to capture all the relevant layers of the notion when dealing with a complex technological artefact that mediates complex human interaction patterns.

The sketch consists of seven main points: First, there is no such thing as “ethical neutrality” with regard to social platforms. Second, I stress the proactive side of responsibility (cf. Gotterbarn 2001), in particular with regard to social platforms. Third, I propose that in a computerized age we have to be not less, but rather more person-centred. This leads us, as the fourth point, to a positive conception of responsibility. Furthermore, a heightened sense of professional responsibility is warranted with regard to diversity-sensitive social platforms. Finally, the professional responsibility needs to be framed by appropriate standards and regulations.

4.1. *No ethical neutrality*

It is often argued that platforms or technologies in general are ‘ethically neutral’, that is, that they do not come with a perspective on what is right and wrong. Or, as Martin observed about the debate: “algorithms are implemented with the hope of being more neutral, thereby suggesting that the decisions are better than those performed solely by individuals. By removing individuals from decisions [...] algorithmic decisions are framed as less biased without the perceived irrationality, discrimination, or frailties of humans in the decision” (Martin 2019, 837).

By now quite a few authors have criticized the view that computing is an ethically neutral practice, or that there is such a thing as an ethically neutral technology at all (cf. Akrich 1992; Bijker 1995; Gotterbarn 2001; Friedman and Nissenbaum 1996; Latour 1992; Martin 2019; Winner 1980). Without wanting to decide the much wider debate on ethical neutrality here, I take the view that ethical neutrality cannot be assumed with regard to social platforms.⁶ At the same time, I hold the view that ethical non-neutrality does not depend solely on the possible or factual consequences of the deployment of a technology.

By inviting specific forms of action and prohibiting others, algorithms have an enormous impact on the behaviour of humans (cf. Verbeek 2006, 377; Martin 2019, 836). That they do have an impact, however, is not sufficient to make that impact ethically relevant. Let me elaborate this point a little: According to Martin, “algorithms are not neutral but value-laden in that they (1) create moral consequences, (2) re-enforce or under-cut ethical principles, or (3) enable or diminish stakeholder rights and dignity” (Martin 2019, 838). This definition, however, is too broad to define ethical non-neutrality, because it is also true for natural phenomena, like lightning. Lightning might create morally relevant consequences by destroying property and killing people, it can under-cut ethical principles like, for instance, the protection of property, and one might want to argue that it also diminishes possibly a stakeholder’s bodily integrity to be struck by lightning. Thus, clearly, algorithms potentially have an impact on our lives, but so does lightning.

Kramer, van Overheld and Peterson in my eyes capture the point about ethical non-neutrality more precisely than Martin, when they argue that some algorithms “cannot be designed without implicitly or explicitly taking a stand on ethical issues, some of which may be highly controversial” (Kramer, van Overheld and Peterson 2010, 251). What Kramer et al. stress is that it is humans, for instance the programmers and designers, that here take a stand on sometimes highly controversial matters – a stand that is very often not transparent to the users of that technology. As Mittelstadt et al. stress, algorithms are in fact intended to “privilege some values and interests over others” (Mittelstadt et al. 2016; cf. Martin 2019, 839), and since “[a]lgorithms are implicitly or explicitly designed within the framework of social customs” (Capurro 2019, 132), they are in fact in many ways (explicitly or implicitly) value laden. Algorithms are, thus, ethically non-neutral not because of the possibly morally relevant conse-

⁶ Also compare Gillespie’s illuminating discussion of platforms as constituents of public discourse, thus leaving their previously assumed role as mere intermediaries behind (Gillespie 2018a).

quences of their use, but because of the value-decisions that are already (necessarily) inscribed in their set up, regardless of whether they would ever generate negative consequences, even regardless of whether they would ever be put to use.

The non-neutrality thesis, by the way, also holds for supervised and unsupervised machine learning systems: If a machine is trained on biased data, it will ‘learn’ the biases from the data set and reproduce them in the decision-making process. To give an example from picture analysis and image search that I have discussed already in more detail elsewhere: In 2016, a tweet by Karbir Alli about the outcome of a search for “three black teenagers” in Google Images lead to a storm of protest (Beuth 2016), because it presented mugshots. A search for “three white kids” in contrast showed mostly pictures of happy white kids (cf. Zuiderveen Borgesius 2018, 16). Google’s response to that was that these search results merely reflected the portrayal of the respective subpopulation across the web and the frequency of that portrayal (York 2016). Since the market for images of “three happy white kids” is apparently much bigger than the market for images of “three happy black kids”, more stock pictures that are assumed to fit the first label are found on the web and come up as search results more frequently in Google Images (cf. York 2016).

In many ways, then, the search results reflect racism and the manifold injustices of our respective societies, such as the unequal distribution of resources and unequal market access, but also the unwillingness to acknowledge that reproducing the accompanying prejudices is not a neutral position: it is an ethically relevant decision.⁷

4.2. Not only remedial, but also proactive

Often when we talk about responsibility, we talk about it in hindsight. The question then is who is to blame for something that has gone wrong. Indeed, responsibility is closely linked to the idea of accountability and liability. But if responsibility is used synonymously with accountability, we tend to lose an important aspect; that is, the idea of foresight.⁸

Responsibility properly understood not only tells us who is to blame after something has gone wrong, it also demands us to act in a way that prevents bad things from happening.⁹ It furthermore demands that we collect the information necessary to make responsible decisions and to learn from any mistakes made. It requires a professional ethics that does not regard the problem to be solved in a given situation as isolated, but takes into account that it is part of a wider context in which the

⁷ For a more detailed interpretation and discussion of this example, see Reinhardt (2020a, 273f.).

⁸ Based on Ladd (1989), Gotterbarn makes a distinction between positive and negative responsibility that is similar to the idea of remedial and proactive responsibility. Note, that accounts that stress the incentives a strong culture of accountability would provide for a person to act responsibly under the threat of the sanctions that he or she would otherwise face, which tends to put less emphasis on the idea that a person should act responsibly no matter whether anybody could hold that person responsible for his or her actions (Gotterbarn 2001, 226–228).

⁹ Höffe (2008) distinguishes three levels of responsibility [Verantwortung]: 1) Aufgabenverantwortung [responsibility for a task] and Zuständigkeitsverantwortung [role responsibility], 2) Rechenschaftsverantwortung [accountability or answerability], 3) Haftungsverantwortung [liability].

proposed solution might have profound consequences – I will elaborate this point in the next section.

Platform developers should in any case implement algorithms that support and strengthen human agency (cf. Hartswood et al. 2016), thus making it easier for users to act responsibly in the above-mentioned person-centred way. Where possible, users should have a say in how the platform is regulated and its transparent working and easily accessible boards of appeal should be in place. Therefore, platforms should establish participatory structures for their users. Also, there have to be monitoring procedures in place that raise awareness of the signs of unintended algorithmic behaviour; for instance, discriminatory decisions. Generally, societies, governments, journalists but also platform owners and operators should invest in increasing data literacy and support users in becoming competent with regard to their online decisions. Establishing standing techno-ethical committees that have an eye on platform operations would also help to detect problems early on (ibid.). Political decision-makers can increase the plurality of platforms by, for example, supporting platform start-ups, thus, increasing the options for users to choose from.¹⁰The legal framework for social platforms has to be evaluated and updated regularly. Political decision-makers have to examine whether adjustments of civil and criminal law are necessary in certain areas and they have to ensure that the training of legal practitioners is adjusted accordingly. Monitoring and consultation procedures should be established to support the political decision-making process.

4.3. Not less, but more person-centred

The loss of control, autonomy and self-determination in a digitalized world is a common theme in fictional digital dystopias. These often depict a world in which humans lose or have given up everything that makes them capable of acting in an accountable and responsible way (cf. Reinhardt 2020b, 113). The current debate on responsibility in a digitalized world tends to lead away from persons as the ultimate bearer of responsibility:

Algorithms are surrounded by an “air of rationality or infallibility” (Zuiderveen Borgesius 2018, 8). They are commonly trusted because of the belief in their technical ‘superiority’ in terms of neutrality, objectivity, reliability and accuracy. This puts human decision-makers in a problematic situation: often they have an information deficit and little knowledge of how the system arrives at its results.¹¹ Prevalent time pressure in many work environments further encourages a pragmatic and approving attitude towards the recommendations of the algorithmic systems deployed (Heesen, Reinhardt and Schelenz 2021, 140).

¹⁰ *Plattform Lernende Systeme* formulated this idea for all systems that use machine learning (Plattform Lernende Systeme 2019).

¹¹ This fact is often associated with the keywords ‘black box’, lack of interpretability and explainability or lack of transparency (literature). However, the topic goes much further and also includes fundamental questions about the handling of technical expert systems, their use in decision-making processes and the time and resources we are willing to spend on important decisions, as well as the framework conditions of the work settings that involve algorithmic systems.

In the philosophical debate on responsibility and digitalization, there are now several ways to deal with these and other related observations. These, however, often re-enforce the impression of the supposed powerlessness of the persons involved: One strand in the debate is advancing responsibility from a new angle, namely asking the question whether computers and other computing entities ought to be regarded as moral agents that bear responsibility for their actions. Some authors in fact arrive at the conclusion that we should widen our picture of moral responsibility and expand our understanding of moral agency to computers and other machines (Misselhorn 2019). This is one way in which the current debate leads away from personal responsibility. Another strand is focused on enforcing responsibility, thus bringing about a “strong culture of accountability” (Nissenbaum 1996, 26), in which rule abiding behaviour is much more likely, simply because people would face sanctions. Note that this approach is based on an “instrumental” understanding of responsibility: Ultimately, it is about bringing about a desirable state of affairs, thus, also leading away from persons as bearers of responsibility as the focal point.

I think that we should instead stress the idea of personal responsibility – in particular, in a digitalized world. Responsibility should not only be seen as a means to an end, but rather as a feature of humaneness. The capacity to take on responsibility is an important feature of the human condition. We bring ourselves into this world as moral agents not by shying away from responsibility but by taking it on.¹²

4.4. A positive notion of responsibility

Stressing the concept of personal responsibility also stresses our nature as moral agents and represents an important counterpoint to a debate that focusses on the incomprehensibility and opacity of technological artefacts and the dependence of human decisions on them: Even if my actions are mediated by the interaction protocols on a social platform, I am still responsible for my actions and, as a consumer or user, I also bear a responsibility for my choices. So, for example, there is no context in which participation in cyberbullying, hate speech or any other harmful – and sometimes even illegal – activity is ever okay, simply because it is conducted online. From an ethical perspective, this is a trivial fact. Our current practices, however, seem to speak a different language.¹³ A positive notion does not reduce responsibility to accountability and liability for wrongs.

¹² Cf. Höffe’s notion of *sittliche Verantwortung* [moral/ethical responsibility]: “Accepting responsibility is moral [sittlich], as long as one does not take it on because of expected rewards and punishments, but because one recognizes oneself as responsible for fellow human beings, the world and oneself, and acts according to this responsibility as a person” (Höffe 2008, 327; my translation).

¹³ One example of legal frameworks going in the direction of stressing personal responsibility is the principle of human final decision of the GDPR: “The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her” (European General Data Protection Regulation §22(1)). This means that algorithmic decisions with the said effects must be professionally assessed by a person before they become effective. With regard to applied ethics, stressing personal responsibility also means that we are not to delegate decisions of moral relevance to machines. Or in other words: “An ethics of AI or algorithms can only be an ethics of those who deal with these techniques” (Lenzen 2020, 64; my translation), and not of ethical decisions by machines.

Taking on a positive concept of responsibility could with regard to social platforms, also entail that individuals or organizations set up collaborative online platforms themselves and develop techniques for fair machine learning via experimental data processing (cf. Veale and Binns 2017; Hagendorff 2019, 55). Watchdog institutions could be established and supported. Exchange forums could help to identify and counteract malpractices. Citizen journalism, blogging or professional data journalism could help with raising awareness about the possible effects of our conduct online and to the possible misconduct of and via platforms (cf. Heesen, Reinhardt and Schelenz 2021, 141).

4.5. A heightened sense of professional responsibility

There may be areas where computerized algorithmic decisions and AI applications have little or no effect on human interactions. The opposite is true for social platforms. They are aimed at people. Therefore, developers need to be aware that their actions will ultimately affect real people in the real world and they need to take this into account in their design and programming decisions.

What is more with regard to a diversity-sensitive social platform, as explained at the beginning, is that the collection of diversity-relevant data is associated with an increased degree of vulnerability for the users. In order to do justice to this fact, developers must exercise particular caution. These aspects call for a heightened awareness of the professional responsibility of developers.

It is important to note that the professional responsibility of developers cannot be side-stepped with reference to the choices of individuals: Though it is correct that people are (in general) responsible for their own choices and actions, we cannot impose our own duties on others by reference to their free choice. To give an example, securing the privacy of data and information is imperative with regard to the safety of the users of a social platform, in particular a diversity-sensitive one. This obligation should not to be transferred to the individual user via (possibly incomprehensible) general terms and conditions.

Overall, however, it is important that computer scientists and developers are not left alone with the high level of responsibility they bear, but are supported by guidelines and other instruments for implementing professional responsibility. These include legal standards and public control procedures, development standards and certifications, corporate codes, voluntary commitments to orient development work, appropriate protection of whistleblowers (Heesen, Reinhardt and Schelenz 2021, 138) and professional ethical training that raises awareness for ethical issues.

4.6. Framework responsibility

As has already been mentioned, social platforms do not exist in a vacuum but are integrated into social and legal structures. Political decision-makers as well as civil society are called upon to further develop this framework to cover the issues that

may arise from social platforms. Legal standards and binding certifications can set framework conditions for all providers on the platform market. They can provide criteria for trustworthy digital products and provide incentives to develop products accordingly. If necessary, it is advisable to supplement these standards with mandatory audits (Heesen, Reinhardt and Schelenz 2021, 139). Gillespie has enlisted a number of amendments to current law that might be worthwhile to implement with regard to platforms. To mention but a few: Legal standards could entail transparency obligations not only with regard to data being processed, but also on how interactions on the platforms are moderated (Gillespie 2018a, 213f.). There should be minimum standards for content moderation in place. “Platforms could each be required”, he adds, “to have a public ombudsman” and an expert advisory panel (ibid., 214). Major platforms could be required to invest in digital literacy programmes “to better address online harassment, hate speech and misinformation” (ibid.).

4.7. Responsibility for diversity on all levels

If a social platform wants to promote diversity and not only collect diversity-sensitive data for exploitative reasons, like for instance micro-targeting, then it should take on responsibility for diversity on all levels: While merely increasing diversity in technical development teams is no guarantee that the whole spectrum of the needs and requirements of different social groups will be taken into account, the composition of a development team can nevertheless help in reflecting the needs and requirements of different user groups (Heesen, Reinhardt and Schelenz 2021, 138). The teams developing and running the platform should consist, therefore, of developers with different perspectives and backgrounds. The platform community should, furthermore, be perceived, acknowledged, appreciated and welcomed as a diverse and heterogenous group. For the training of AI systems that are deployed, it is, moreover, important to ensure representation of the diverse societal groups in the training data (ibid.) and it should be closely monitored whether new grounds of discrimination arise from the specific nature of machine learning (Data Ethics Commission 2019). There should also be a platform constitution or charter in place that acknowledges and welcomes diversity as a core feature of the platform and that bans discriminatory behaviour.

Research and development in the field of algorithmic decision systems and, accordingly, in the field of artificial intelligence and machine learning, is increasingly concerned with diversity and non-discrimination. There are initiatives promoted by various professional associations, such as the Code of Ethics of the Association for Computing Machinery (ACM), or covered in conferences and research contexts, such as “Fairness, Accountability and Transparency in Machine Learning” (FAccTML) or “Discrimination-Aware Data Mining” (DADM). Such forms of diversity-sensitive research take on challenges such as data distortion and stereotyping (cf. Heesen, Reinhardt and Schelenz 2021, 137f.), thus showing an effort to act responsibly with regard to diversity. Some approaches explicitly focus on the concerns of marginalized groups, such as the Design Justice Network (cf. Constanza-Chock 2020).

5. Conclusion: A multidimensional notion of responsibility

Based on the observation that computer science is increasingly interested in issues of diversity and non-discrimination, while at the same time the representation of diversity in, for example, social platforms is linked to some ethical questions, I investigated what an ethical framework of responsibility for a diversity-sensitive social platform might look like. I started by outlining some of the dangers and risks that go hand in hand with making a social platform diversity-sensitive and argued that the increased vulnerability of the users of such a platform is a sufficient reason to take a closer look at the responsibility at stake in this particular context. Then, I went on to show how traditional concepts of responsibility face difficulties when dealing with algorithm-mediated computerized interactions between humans. I proposed that these difficulties make it all the more urgent to think through what responsibility might entail with regard to social platforms. My starting point was an action-centred account of responsibility as opposed to a consequences-centred account.

I then sketched a framework of responsibility for social platforms and proposed a multidimensional understanding of responsibility with regard to diversity-sensitive social platforms. The sketched framework stresses that platforms, in particular social platforms, are not ethically neutral. The responsibility concept we should employ should be person-centred, positive and deal with the remedial as well as the proactive side of responsibility. A social platform in addition calls for a heightened sense of professional responsibility on the side of the people developing, designing, operating and managing the platform. This professional responsibility, however, has to be framed by societal and legal formal and informal rules that give guidance to the individual decisions of the professionals involved. Beyond that, a diversity-sensitive platform has to establish procedures and work environments that are non-discriminatory and foster diversity on all levels – not only within the platform community. Finally, from the above, I arrived at a multidimensional concept of responsibility for diversity-sensitive social platforms that includes but is not limited to the following aspects:

- Professional responsibility: If I help creating technological items and products, I bear the responsibility to ensure, as far as humanly possible, that they will do no harm to people, the environment and democratic structures.
- Tool-mediated responsibility: If I create tools that could foreseeably lead to disastrous consequences, I am responsible for ensuring that the appropriate counter-measures are taken.
- Knowledge-mediated responsibility: I am also responsible for the foreseeable effects of the publication or non-publication of my research.
- Responsibility for my actions: Even if my actions are mediated by interaction protocols on a social platform, I am still responsible for them.

- Consumer/User responsibility: I am responsible for which platform I choose and for considering what consequences and side-effects my actions on a particular platform may have. I am responsible for deciding which tasks I delegate to the platform and for informing myself on how they are carried out — and for adjusting my decision accordingly.
- Framework responsibility: If I am in a position to work towards setting up mandatory and non-mandatory standards and guidelines, curricula for computer scientists, etc. that ensure the non-harmful use of technologies, I bear the responsibility to do so.

If we do not focus on one aspect of responsibility alone, we will get to see the different actors and types of responsibility that are at play with regard to a diversity-sensitive social platform. Through such a positive, proactive and multi-dimensional understanding of responsibility, we can avoid the impression that responsibility has only to do with the question of who can be called to account if something goes wrong. Instead, the empowering aspect of taking on responsibility should be stressed and a sense of foresight fostered while a dense net of overlapping dimensions of responsibility is woven to ensure that responsibility is taken seriously on all levels and all sides with regard to diversity-sensitive social platforms. The considerations here can, however, only be taken as a preliminary starting point for a much more thorough going debate on responsibility, social platforms and diversity-sensitivity.

In many cases, social platforms function as amplifiers where the single act does no, or little, harm, but the aggregation can do significant harm. Though everybody is responsible for their own wrongdoing, who is responsible for the amplified surplus effect of the aggregation? With regard to social platforms, but not only for social platforms, it might be worthwhile to work out a thorough concept of responsibility for aggregated actions different from the collective responsibility or cooperate responsibility. In future research, it would be, furthermore, useful to see how this framework could be applied to other technologies and what other dimensions of responsibility emerge from such an application, possibly leading to amendments and modifications of the framework.

References

- Akrich, Madleine. "The description of technological objects." In *Shaping technology/building society: Studies in sociotechnical change*, edited by Wiebe E. Bijker and John Law, 205–224. Cambridge, MA: MIT Press, 1992.
- Beuth, Patrick. "Nein, die Suchmaschine ist nicht rassistisch." Published 9 June 2016. Accessed March 30 2021.
<https://www.zeit.de/digital/internet/2016-06/google-three-black-teenagers-suchmaschine-rassismus>.

-
- Bijker, Wiebe. *Of bicycles, bakelite, and bulbs: Towards a theory of sociological change*. Boston MA: MIT Press, 1995.
- Brake, David R. *Sharing our Lives Online. Risks and Exposure in Social Media*. New York: Palgrave Macmillan, 2014.
- Capurro, Rafael. "Enculturating algorithms." *Nanoethics* 13 (2019): 131–137.
- Costanza-Chock, Sasha. *Design justice. Community-led practices to build the worlds we need*, Cambridge, MA: The MIT Press, 2020.
- Fisher, John Martin. "Recent work on moral responsibility." *Ethics* 110, no. 1, (1999): 93–139.
- Friedman, Batya and Helen Nissenbaum. "Bias in computer systems." *ACM Transactions on Information Systems*, 14, no. 3, (1996): 330–347.
- Gillespie, Tarleton. "Platforms are not Intermediaries." *Georgetown Law Technology Review*, 198, no. 2, (2018a): 198–216.
- Gillespie, Tarleton. *Custodians of the Internet. Platforms, Content Moderation and the Hidden Rules that Shape Social Media*. New Haven: Yale University Press, 2018b.
- Gotterbarn, Donald. "Informatics and professional responsibility." *Science and Engineering Ethics* 7 no. 2, (2001): 221–230.
- Gorwa, Robert, Reuben Binns and Christian Katzenbach. "Algorithmic Content Moderation: Technical and Political Challenges in Automation of Platform Governance." *Big Data & Society*, (2020): 1–15.
- Hagendorff, Thilo. "Maschinelles Lernen und Diskriminierung: Probleme und Lösungsansätze." *Österreichische Zeitschrift für Soziologie* 44, Supplement 1, (2019): 53–66.
- Hartswood, Mark et al. "A Social Charter for Smart Platforms". Published 2016. Accessed 30 March 2021.
https://eprints.soton.ac.uk/410307/1/SmartSocietySocialCharterforSmartPlatforms_final.pdf.
- Heesen, Jessica, Karoline Reinhardt, and Laura Schelenz. "Diskriminierung durch Algorithmen vermeiden: Analysen und Instrumente für eine demokratische digitale Gesellschaft." In *Diskriminierung und Antidiskriminierung. Beiträge aus Wissenschaft und Praxis*, edited by Gero Bauer, Maria Kechaja, Sebastian Engelmann, and Lean Haug, 129–147. Bielefeld: transcript, 2021.
- Höffe, Otfried. "Verantwortung." In *Lexikon der Ethik*, edited by Otfried Höffe, 326–327. München: C.H. Beck, 2008.
- Johnson, Deborah, Thomas Power. "Computer systems and responsibility: A normative look at technological complexity." *Ethics and Information Technology* 7 (2005): 99–107.
- Jonas, Hans. *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*, Frankfurt am Main: Suhrkamp, 1984.
- Kramer, Felicitas, Kees van Overheld, and Martin Peterson. Is there an ethics of algorithms? *Ethics and Information Technology* 13, 2010: 251–260.
- Ladd, John. "Computers and Moral Responsibility: A framework for an ethical analysis." In *The Information Web: Ethical and Social Implications of Computer Networking*, edited by Carol Gould, 207–228. Boulder CO: Westview Press, 1989.
- Latour, Bruno. "Where are the Missing Masses? The Sociology of a Few Mundane Artefacts." In *Shaping technology/building society: Studies in sociotechnical change*, edited by Wiebe E. Bijker and John Law, 225–258. Cambridge, MA: MIT Press, 1992.
- Lenk, Hans. "Verantwortlichkeit und Verantwortungstypen." In *Handbuch Verantwortung*, edited by Ludger Heidbrink, Claus Langbehn, Janina Sombetzki, 1–29. Wiesbaden: Springer VS, 2016.

- Martin, Kirsten. "Ethical Implications and Accountability of Algorithms." *Journal of Business Ethics* 160, (2019): 835–850.
- Matthias, Andreas. "The responsibility gap: Ascribing responsibility for the actions of learning automata." *Ethics and Information Technology*, 6 (2004): 175–183.
- Mensi-Klarbach, Heike. "Der Business Case für *Diversität und Diversitätsmanagement*". In *Diversität und Diversitätsmanagement*, edited by Regine Bendl, Edeltraud Hanappi-Egger, Roswitha Hofmann, 299–326. Wien: facultas, 2012.
- Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The ethics of algorithms: Mapping the debate." *Big Data & Society* 3, no. 2, (2016): 1–21.
- Misselhorn, Catrin. *Grundfragen der Maschinenethik*, Stuttgart: Reclam, 2019.
- Nida-Rümelin, Julian. *Verantwortung*. Stuttgart: Reclam, 2011.
- Nida-Rümelin, Julian. *Humanistische Reflexionen*. Berlin: Suhrkamp 2018.
- Nissenbaum, Helen. "Computing and Accountability." *Communications Association for Computing Machinery* 37, no.1, (1994): 72–80.
- Nissenbaum, Helen. "Accountability in a Computerized Society." *Science and Engineering Ethics* 2, no. 1, (1996): 25–42.
- Noorman, Merel. "Computing and Moral Responsibility." In: *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), edited by Edward N. Zalta. Accessed 30 March 2021. <https://plato.stanford.edu/archives/spr2020/entries/computing-responsibility/>.
- O'Neil, Cathy. *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. London: Penguin, 2016.
- Parekh, Bhikhu. *Rethinking Multiculturalism: Cultural Diversity and Political Theory*. London: Macmillan, 2000.
- Plattform Lernende Systeme. "Innovation nutzen, Werte schaffen. Neue Geschäftsmodelle mit Künstlicher Intelligenz Bericht der Arbeitsgruppe Geschäftsmodellinnovationen." Accessed March 30 2021: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG4_Bericht_231019.pdf.
- Reinhardt, Karoline. "Between Identity and Ambiguity. Some Conceptual Considerations on Diversity." *Symposion* 7, no. 2, (2020a): 261–283.
- Reinhardt, Karoline. "Digitaler Humanismus. Jenseits von Utopie und Dystopie." *Berliner Debatte Initial* 31, no. 1, (2020b): 111–123.
- Schmidt, Jan-Hinrik. "Ethik des Internets." *Handbuch Medien- und Informationsethik*, edited by Jessica Heesen, 284-292. Stuttgart: J.B. Metzler, 2016.
- Smiley, Marion. "Collective Responsibility." *The Stanford Encyclopedia of Philosophy* edited by Edward N. Zalta, (Summer 2017 Edition). Accessed March 30 2021. <https://plato.stanford.edu/archives/sum2017/entries/collective-responsibility/>.
- van de Poel, Ibo, Lambèr Royakkers, Sjoerd D. Zwart. *Moral Responsibility and the Problem of Many Hands*, London: Routledge, 2015.
- van Dijk, José, Thomas Poell and Martijn de Waal. *The Platform Society. Public Values in a Connective World*. New York: Oxford University Press, 2018.
- Veale, Michael and Reuben Binns. "Fairer machine learning in the real world. Mitigating discrimination without collecting sensitive data." *Big Data & Society* 4, no. 2, (2017): 1–17.

-
- Verbeek, Peter-Paul “Materializing Morality: Design Ethics and Technological Mediation.” *Science, Technology, and Human Values* 31, no. 3, (2006): 361–380.
- Weizenbaum, Joseph. *Computer Power and Human Reason. From Judgement to Calculation*. San Francisco: W.H. Freeman and Company, 1976.
- Winner, Langdon. “Do Artifacts Have Politics?” *Daedalus* 109, no. 1, (1980): 121–136.
- York, Chris. “Three black teenagers: Is Google Racist? It’s not them, it’s us.” Published 8 June 2016. Accessed March 30 2021.
https://www.huffingtonpost.co.uk/entry/three-black-teenagers-google-racism_uk_575811f5e4b014b4f2530bb5, published June 8 2016.
- Young, Iris. *Justice and the Politics of Difference*. Princeton: Princeton University Press, 1990.
- Zuboff, Shoshana. *The Age of Surveillance Capitalism. The Fight for a Human Future ar the New Frontier of Power*. London: Profile Books, 2019.
- Zuiderveen Borgesius, Frederik. “Discrimination, artificial intelligence, and algorithmic decision-making.” published by Directorate General of Democracy, Council of Europe, Strasbourg 2018.