Algorithmic Pollution: Making the Invisible Visible

Professor Olivera Marjanovic (contact author)
School of Professional Practice and Leadership
Building 11, Office 11.10.213
Faculty of Engineering and IT
University of Technology Sydney
Sydney, NSW 2007
T. +61 (02) 9514 3175
PO Box 123 Broadway NSW 2007 Australia
Email: olivera.marjanovic@uts.edu.au

Professor Dubravka Cecez-Kecmanovic
UNSW Business School
The University of New South Wales
Room 54640, Quadrangle building
Sydney, NSW, 2052, Australia.

Professor Richard Vidgen
UNSW Business School
The University of New South Wales
Room 2099, Quadrangle building
Sydney, NSW, 2052, Australia.

> *"We have become accustomed to thinking of pollution exclusively in terms of environmental degradation. This approach so pervades the societal mindset that people often dismiss references to cultural pollution, light pollution, spiritual pollution, and other nonenvironmental pollution as a mere rhetorical device … Pollution has always had dual meanings: a broad reference to all sorts of effects upon human environments, and a narrow focus upon natural environments. In fact, until less than a century ago society applied the term to human environment more often than natural environment" (Nagle, 2009, p.6).*

Abstract

In this paper, we focus on the growing evidence of unintended harmful societal effects of automated algorithmic decision-making (AADM) in transformative services (e.g., social welfare, healthcare, education, policing and criminal justice), for individuals, communities and society at large. Drawing from the long-established research on social pollution, in particular its contemporary 'pollution-as-harm' notion, we put forward a claim - and provide evidence - that these harmful effects constitute a new type of digital social pollution, which we name 'algorithmic pollution'. Words do matter, and by using the term 'pollution', not as a metaphor or an analogy, but as a transformative redefinition of the digital harm performed by AADM, we seek to make it visible and recognized. By adopting a critical performative perspective, we explain how the execution of AADM produces harm and thus performs algorithmic pollution. Recognition of the potential for unintended harmful effects of algorithmic pollution, and their examination as such, leads us to articulate the need for transformative actions to prevent, detect, redress, mitigate, and educate about algorithmic harm. These actions, in turn, open up new research challenges for the information systems community.

DOI: 10.1177/02683962211010356

1. Introduction

The widespread adoption of automated decision-making algorithms (AADM[1]), powered by AI, analytics, and big data, in human services in sectors vital for any society – such as social welfare, education, healthcare, employment, public housing, policing and criminal justice – is motivated and justified by intended and expected positive effects. These include, for example, increased efficiency and speed of service delivery, better compliance with government policies, greater transparency and accountability, reduction of costs and, most importantly for service beneficiaries, improved overall service quality (Redden, 2018; Caplan et al. 2018; Alston, 2019a, 2019b; Park and Humphry, 2019). However, in spite of positive intentions, there is growing and disturbing evidence of the harmful societal effects of AADM (see for example O'Neil, 2016a; Eubanks, 2018; Caplan, et al. 2018; Park and Humphry, 2019; Benjamin, 2019; Alston, 2019a, 2020; UN, 2020). Moreover, unintended harmful effects on citizens, students, patients, employees, and other intended beneficiaries of these services, continue to spread through systems of algorithms (Pasquale, 2015) and be amplified in unknowable ways through ongoing datafication practices (Galliers, et al., 2017; Marjanovic and Cecez-Kecmanovic, 2017). These effects, however, remain largely invisible (Eubanks, 2018; Caplan, et al. 2018; Benjamin, 2019). Worryingly, they are often dismissed in the literature as unintended, and sometimes even inevitable 'transactional side effects' (Markus, 2015). As a result, they have been neither thoroughly examined nor effectively addressed.

This paper is motivated by our critical concerns for the wellbeing of individuals and communities that continue to experience the harmful effects of fully automated algorithmic decision-making. As researchers, we feel responsible for making these harmful effects visible and recognized and to act as a catalyst for transformation (Alvesson and Deetz, 2000). Taking a critical research perspective (Alvesson and Deetz, 2000; Myers and Klein, 2011), we seek to provide insights into these harmful effects and propose a transformative redefinition that has the potential to address the social and technological arrangements that produce them. The objectives of this paper therefore are to:

i) offer *critical insights* into the new digital forms of harmful social effects performed by automated algorithmic decision-making (AADM), based on the literature;

ii) articulate and theorize the new phenomenon of *algorithmic pollution* as a digital type of social pollution, which we put forward as a transformative redefinition of these unintended harmful effects;

iii) propose *transformative actions* to prevent, detect, redress, mitigate, and educate about algorithmic harm, based on our sociomaterial theorisation of algorithmic pollution, along with the identification of associated IS research challenges.

To achieve our objectives, we first draw on a broad IS and social science literature as well as various government reports and algorithmic primers, to provide critical insights (Alvesson and Deetz, 2000) into the reported harmful effects of automated algorithmic decision-making and its associated datafication practices. We focus on the domain of human services (also called transformative services), such as those offered in the areas of social welfare, education, healthcare, criminal justice, housing, and employment. These services are considered to be vital for any society as they have a direct transformative impact on the wellbeing of individuals, their families, communities, and society at large (Anderson and Ostrom, 2015; Anderson et al., 2018).

[1] In this paper we use the following definition of AADM from the information commissioner in the UK: "Automated decision-making is the process of making a decision by automated means without any human involvement. These decisions can be based on factual data, as well as on digitally created profiles or inferred data. Examples of this include: an online decision to award a loan; and an aptitude test used for recruitment which uses pre-programmed algorithms and criteria. Automated decision-making often involves profiling, but it does not have to." (ICO, 2018, p.7).

DOI: 10.1177/02683962211010356

Based on these critical insights we demonstrate that the harmful effects on individual and collective wellbeing constitute a new form of digital harm. We name this harm *algorithmic pollution*.

To articulate and justify the notion of algorithmic pollution, we draw from the long-established research on *social pollution*, see for example Norman (2004), Nagle (2009), Sarine (2012) and others, which is based on an anthropological notion of 'pollution' (Douglas, 1966, 1969, 1975). As Nagle (2009) explains, the concepts of social and moral pollution appeared in the research literature and practice of law long before environmental pollution. Social pollution focuses on pollution as *harm* to individuals, communities and societies, rather than environmental pollution and the discharge of harmful particles.

Thus, taking an outside-in perspective (Deville and van der Velden, 2016), we define algorithmic pollution as a new kind of social pollution manifested as digital harm *performed* by automated algorithmic decision making (AADM). We frame algorithmic pollution as a phenomenon in the sociomaterial environment, defined as the entanglement of humans and technologies (and other non-humans), discourses and materialities, which perform the social and how the social is done (Gherardi, 2019). Enacted in the sociomaterial environment, AADM performs algorithmic pollution through material-discursive practices in which the human/social are co-constituted through technology (Law, 2004; Orlikowski, 2007; Orlikowski and Scott, 2008) often in invisible ways.

Our critical research perspective is, therefore, informed by a performative view of sociomateriality (Cecez-Kecmanovic et al., 2014; Introna, 2019; Orlikowski, 2007). More precisely, we take a *critical performative perspective*. Such a perspective leads us to theorize algorithmic pollution by questioning first the assumption that AADM technologies exist separately from human beings – subjects of decision-making – and then that they 'affect' them as intended, in a pre-determined, desirable manner. Instead, we argue that to understand algorithmic pollution, that is the harmful social effects of AADM technologies, it is necessary to understand what their execution actually does "not just empirically, but also *ontologically*" (Introna, 2019, 316, emphasis in the original). In other words, we need to understand how these technologies *perform* the subjects about which they acquire data and use 'datafied individuals' to make decisions that determine their possibilities to *be*. Our critical performative perspective is therefore grounded in an *ontology of becoming* that assumes inseparability of the human/social and the material/technological in an ongoing process or flow that is never complete (Barad, 2007; Cecez-Kecmanovic, 2016; Introna, 2019). Such a critical performative perspective allows us to expose an ongoing flow of performing the subjects in the image inscribed in the AADM technologies. As decisions are made, the subjects *become* what the AADM technologies assume they are (Introna, 2019). This is how harm is done. In other words, by revealing the ontological assumptions underlying AADM we demonstrate how their execution is ontological, performing the subjects and thus producing, propagating and amplifying harmful effects, which often remain invisible.

Based on such a critical performative theorizing concerning how algorithmic pollution is performed, we then discuss possible transformative actions (Alvesson and Deetz, 2000; Myers and Klein, 2011) of prevention, detection, redress and mitigation of, and education about algorithmic pollution. As critical IS researchers, we see these actions, and the research challenges they entail, not only as an interesting direction for IS research, but also as an integral part of our shared societal responsibility for urgent research-informed collective action to address algorithmic pollution. Further, words do matter, especially in the public discourse. By giving the concept of algorithmic pollution a name, not as a metaphor or an analogy, but as a transformative redefinition (Alvesson and Deetz 2000) of the harmful effects of AADM, we seek to make these harms visible and recognised. In doing this we also issue a call for research-informed action that we, the IS research community, are well positioned to inform and lead.

The main theoretical contributions of our research are: 1) articulation of the new concept of algorithmic pollution as a new form of social pollution that presents a transformative redefinition of

DOI: 10.1177/02683962211010356

unintended harmful effects of automated algorithmic decision-making; 2) a theoretical elaboration of algorithmic pollution that explains how automated decision-making performs harm in the sociomaterial environment; and 3) setting the research foundations for an agenda of transformative actions.

The main practical contributions come from naming and making algorithmic pollution visible not only for researchers, but also for developers, users, regulators, and the general public, in order to inform, inspire and mobilise broader changes. By using the term 'pollution' – something we can all relate to – we also signal it as harmful and in need of urgent attention and regulation.

The paper is organised as follows. In Section 2, we introduce the domain of transformative services and review relevant literature on the harmful effects of AADM and the underlying datafication practices. In Section 3 we introduce and define the concept of algorithmic pollution, building upon the multidisciplinary literature of social pollution (which was long established before the contemporary and familiar notion of environmental pollution). In Section 4 we elaborate further on the concept of algorithmic pollution through a critical performative perspective and explain how algorithmic harm is performed in the sociomaterial environment. This forms the basis for proposing transformative actions needed to prevent, detect, redress and mitigate algorithmic harm, together with identifying the associated IS research challenges, in Section 5. Finally, in Section 6, we offer some concluding remarks, reflect on the limitations of our research and discuss possible ways forward.

2. Foundation Concepts and Related work

2.1. Transformative services

The emerging multidisciplinary field of Transformative Services Research (TSR) focuses on services that transform human lives by having a *direct* impact on the wellbeing of individuals, communities and the wider society (Anderson, et al., 2013; Ostrom et al., 2015; Anderson and Ostrom, 2015). Well-known examples of transformative services include various social and other human care services in contexts such as social welfare, housing, healthcare, aged-care, employment, and education (Anderson et al., 2013; Anderson and Ostrom, 2015; Danaher and Gallan, 2016; Anderson, et al., 2018). Importantly, the very nature of these services is directly related to the wellbeing of the intended service beneficiaries (e.g., patients, students, citizens), because of their direct and profound impact on human wellbeing outcomes, such as health, livelihood, access, inclusion and, ultimately, human dignity and human rights – both at individual and collective levels (Anderson et al., 2013). Indeed, calling upon the United Nations Universal Declaration of Human Rights, Anderson at al. (2013) describe the "moral imperative" of transformative services "founded on the concept of human dignity, which advances the development of rights and responsibilities" (p.1203).

Research shows that transformative services are increasingly supported by AADM, powered by AI, big data and analytics. Various examples found in a multidisciplinary literature show that this type of technology is being adopted all over the world at "breath-taking speed", as Eubanks (2018, p,11) warns. Important decisions are now made by fully automated algorithms in vital societal services in social welfare, housing, healthcare, education, criminal justice, and employment (Caplan et al. 2018). These algorithms are used, for example, to predict "children at risk" in children and family services in New Zealand and the USA, to calculate citizen scores used for predictive policing and criminal sentencing in the criminal justice services, to select and score suitable and promising candidates in employment services, to determine eligibility and access to social welfare payments in digital welfare systems, to manage priority lists in housing services, and in many other domains (Pasquale, 2015; O'Neil 2016; Wachter-Boettcher, 2017; Noble, 2018; Caplan et al. 2018; Eubanks, 2018; Alston, 2019a, 2019b; Marda, 2019, Zalnieriute et al, 2019). The intended beneficiaries of AADM are often the most vulnerable groups in society – groups who depend on

DOI: 10.1177/02683962211010356

provision of these services for their health, wellbeing, livelihood, access and, ultimately, human dignity and human rights.

Even without AADM, decision-making in transformative services is extremely complex and with high stake outcomes, due to highly contested aims, political issues, institutional mechanisms, ideological underpinnings, and even institutional and societal value systems (Eubanks, 2018; Keddell, 2019; Zavrsnik, 2019). For example, Keddell (2019) illustrates the complexities of the institutional system of child protection: "Whether a child protection system is based on a child protection-, child welfare-, or child-focussed policy orientation, for example, will shape its philosophical basis, broad institutional structures, preferred priorities and methods of social work practice" (p.2). In other words, the institutional system of child protection itself could be punitive or, instead, focused on providing support. This in turn, influences both decision-making and its outcomes. For example, "[t]o be offered a voluntary, in-home support service or better housing has different ramifications than being investigated for child abuse" (Keddell, 2019, p.5). As this example illustrates, decision making in transformative services includes unique ethical and moral challenges, which are made even more complex through the use of AADM (Zavrsnik, 2019; Eubanks, 2018, Keddell, 2019; Sun and Gerchick, 2019).

Research in IS has only recently recognised the unintended harmful consequences of AI and AADM (Galliers et al., 2017; Markus, 2017; Schultze et al., 2018; Bohme, 2019; Cecez-Kecmanovic, 2019; Clarke, 2019a; 2019b; Leidner, 2019; March, 2019; Gal et al., 2020), but it is yet to consider the context of transformative services. The transformative services literature, on the other hand, has identified the use of AI and automated algorithmic decision-making as "gnarly issues" in wellbeing and service research (Blocker and Davis, 2019). Both IS and transformative services researchers agree that the unintended negative and harmful consequences of AADM (Blocker and Davis, 2019; Mikalef et al. 2019) are of critical social concern and in urgent need of our attention.

2.2. Harmful effects of automated algorithmic decision-making

Without any doubt the intentions behind AADM are positive. As the information commissioner in the UK explains: "Profiling and automated decision making can be very useful for organisations and also benefit individuals in many sectors, including healthcare, education, financial services and marketing. They can lead to quicker and more consistent decisions, particularly in cases where a very large volume of data needs to be analysed and decisions made very quickly" (ICO, 2018, p.7). Despite positive intentions behind the adoption of algorithmic decision-making in transformative services, AADM comes with significant risks and unintended harmful effects for individuals, their families, and wider communities. For example, the ICO (2018) identifies the following risks of AADM: "i) Profiling is often invisible to individuals; ii) People might not expect their personal information to be used in this way; iii) People might not understand how the process works or how it can affect them; and iv) the decisions taken may lead to significant adverse effects for some people" (p.8).

The related literature reports growing evidence of unintended harmful effects. They include restricted access to services, digital exclusion and other forms of 'digital discrimination' and 'technological redlining' (Caplan et al., 2018:3; Eubanks, 2018; Noble 2018). AADM are also found to further amplify en masse existing inequalities and other systemic issues, such as poverty and discrimination (Caplan et al., 2018; Noble, 2018; Alston, 2019a, 2019b; Marda, 2019). Seeing algorithmic harm as a human-rights issue, Marda (2019) explains: "It is not simply a matter of ensuring accuracy and perfection in a technical system, but rather a reckoning with the fundamentally imperfect, discriminatory and unfair world from which these systems arise, and the underlying structural and historical legacy in which these systems are applied" (p.8).

DOI: 10.1177/02683962211010356

Unregulated and often hidden and invisible, AADM systems continue to create these harmful effects without anyone taking responsibility or being identified as responsible (O'Neil, 2016; Eubanks, 2018; Caplan et al., 2018; Alston, 2019a, 2019b). With the intended beneficiaries of AADM often being the most vulnerable members of society, these algorithms, Eubanks (2018:11) warns, are fast becoming "the new tools for digital poverty management", sometimes with life-threatening consequences (Eubanks, 2018; Carney, 2018; Zalnieriute et al., 2019).

There are a few notable IS studies that focus on the harmful effects of AI and automated algorithmic decision-making on individuals, organisations and society – see for example Loebbecke and Picot (2015), Newell and Mirabelli, (2015), Zuboff (2015, Galliers et al. (2017), Markus (2017), Schultze et al. (2018), Clarke (2019a; 2019b), Cecez-Kecmanovic (2019), Bohme (2019), Leidner (2019), March (2019) and Gal et al. (2020), However, the IS research community is yet to focus on this important topic on a larger scale and explain how unintended harmful effects of AADM are created. Moreover, as Mikalef et al. (2019) observe, "[m]ainstream information systems research generally celebrates the proliferation of analytics and AI", and is yet to focus on the "dark side of AI and big data" (pg. 1). In this context, it is of note that critical IS research is conspicuously missing.

While it remains unclear how algorithms create harmful effects, the related literature points to the role of underlying datafication practices (Galliers et al., 2017, Markus, 2017). The role of datafication in the context of AADM is reviewed next.

2.3. Datafication and automated algorithmic decision-making

At the very core of algorithmic decision-making are several - often hidden - mechanisms of *datafication*, which are enacted as algorithms use and produce data. Datafication (also known as datification) is a process of representing various phenomena (including people) by data (Lycett, 2013; Newell and Mirabelli, 2015; Galliers et al., 2017; Markus, 2017). While we recognise that datafication may have positive effects (Lycett, 2013), in this section we focus on datafication practices in transformative services and how those practices, in the context of AADM, can cause harm.

The algorithmic primers, published by Data & Society (n.d) in collaboration with practitioners offer illustrative examples of datafication used in AADM and the resulting algorithmic harm in various transformative service sectors, such as healthcare (Rosenblat et al., 2014a), public housing (Rosenblat et al., 2014b), employment (Rosenblat et al., 2014c), criminal justice (Rosenblat et al., 2014d), and education (Alarconn, et al., 2014). Further evidence of algorithmic harm is provided by influential government and other public reports (Caplan et al., 2018; Redden, 2018; Alston et al., 2019a, 2019b; Marda et al. 2019; UN 2020). By combining this evidence with related research literature from social sciences and IS, we identify the following datafication practices and their harmful effects.

- *Use of proxy, inferred and unrelated data to describe an individual*

Algorithms are applied to 'datafied individuals' that is, individuals are represented by a limited number of attributes that are chosen as relevant in the context of a particular transformative service. However, what constitutes "relevant" data remains highly problematic (Caplan et al., 2018; Alston, et al., 2019a, 2019b) as it often includes proxy, inferred and even unrelated data. For example, in criminal justice a person's postcode may be augmented with additional "proxy data" such as "crime hot spots" which are then used to determine the length of their sentence (Rosenblat et al., 2014d; Caplan et al., 2018). In another reported example, the hiring algorithm Evolv uses the distance between an employee's workplace and residential postcode to infer their intention of staying in the job (Rosenblat et al., 2014c). Those who live 0-5 miles from their workplace are predicted to stay in their jobs 20% longer, regardless of many other more compelling reasons why people may stay or leave their jobs. When not available, personal data may be inferred from other data. For example, a

DOI: 10.1177/0268396221101035

person's health or criminal status may be inferred in part from a relative's medical status (Rosenblat et al., 2014a:1) or a relative's past criminal record (Rosenblat et al., 2014d; Ferguson, 2017).

Inferred and unrelated data have also been widely used in business by marketers when selling products or services to targeted customers. Although the consequences of such practices may appear low-stake (for instance, products being recommended to a particular customer, based on their datafied profile, are not sold or advertisements are not shown to the most profitable customers) the result is the yet further datafication of individuals. However, in the case of transformative services, such datafication practices of using proxy, inferred, and unrelated data result in so-called "representational harm", which can have long-lasting and unknowable future consequences (Reisman et al., 2018). Inaccurate inferences may have serious, even tragic consequences, for example, for people needing access to healthcare (Rosenblat et al., 2014a), including life-saving medical services (Eubanks, 2018).

- *Use of poor-quality historic data to train algorithms*

Algorithms are trained using past data about (datafied) individuals to predict their future behaviour and make related decisions. Such datafication practices are highly problematic for several reasons. First, past data may be outdated. For example, data collected on historical "gang districts" are now used by police for predictive policing even though these districts may no longer be representative (Ferguson, 2017). Similarly, past assumptions about traits (data values) that correlate with crime, which are no longer considered valid, may continue to be reflected in discriminatory outcomes of the algorithms trained on such data (Rosenblat et al., 2014d).

Second, various biases contained in past data will lead to more biases in the algorithmic outcomes. For example, "[w]hen algorithms rely on the characteristics of convicted or arrested populations to predict the persons who are likely to commit crime, they solidify a history of bias against those already disproportionately targeted by the criminal justice system" (Rosenblat et al., 2014d:2-3). As O'Neil (2016:1) observes: "if we allowed a model to be used for college admissions in 1870, we'd still have 0.7% of women going to college. Thank goodness we didn't have big data back then".

These datafication practices of using past data can create harmful effects, whereby past disadvantages are reinforced and a history of bias is further solidified. Over time, they create harmful effects known as "cumulative disadvantage" (Rosenblat et al., 2014d).

- *Further datafication through scoring and ranking*

Individual scores and ranks are very common algorithmic outputs found across all types of transformative services. These scores/ranks are used to determine outcomes, to offer (or not offer) services, or to predict future behavior. For example, citizen scores can be used to unlawfully arrest predicted 'future criminals', before they even commit any crime (Rosenblat et al., 2014d; Ferguson, 2017; Caplan et al., 2018). The resulting harmful effects are long lasting and affect individuals, families, neighbourhoods and whole communities.

Yet, the ways these scores and ranks are calculated and used is invisible and as such very hard to challenge and change. For example, almost 400,000 Chicago citizens have an official police risk score of which they are not aware (Ferguson, 2017). Moreover, "[t]his algorithm – still secret and publicly unaccountable – shapes policing strategy, the use of force, and threatens to alter suspicion on the streets. It is also the future of big data policing in America – and depending on how you see it, either an innovative approach to violence reduction or a terrifying example of data-driven social control". (Ferguson, 2017:1)

Algorithmic outcomes of scoring and ranking result in further datafication of already datafied individuals. Thus, patients become high-risk patients, students are identified as future failures, employees become bad employees, and so on. Being more than labels, these scores and ranks start to perform new worlds for future service encounters with the same providers and beyond.

DOI: 10.1177/02683962211010356

- *Networked harm resulting from cumulative datafication practices across different contexts*

Algorithmic outputs are further propagated through systems of algorithms and consequently reused in new contexts for different (unknowable) purposes, both across transformative services and beyond. These society-wide datafication practices result in data about individuals being perpetually reconstructed (Cheney-Lippold, 2018) in unknowable ways. The harmful effects experienced in one context, are propagated and further amplified throughout subsequent services, with long lasting and unknowable cumulative effects. For example, educational data are already used, or have a potential to be reused, in other contexts. As stated in the Education Primer (Alarconn, et al., 2014:5) "…there is some worry that information such as attendance records will affect financial decisions in other domains". Credit scores (created elsewhere) are commonly used as an input for hiring algorithms, "even though connections between credit history and work capability are dubious at best" (Caplan et al., 2018:7). The credit agency, Experian, is making its demographic segmentation software called Mosaic (2019) and data available to local government agencies in the UK where it is used in a variety of ways, such as for risk assessment of defendants in court cases (Dencik, Hintz and Cable, 2019). As a consequence, the crude labelling of people that might be appropriate for market segmentation in commerce is propagated to local government agencies that make potentially life-changing decisions about individual citizens.

These practices result in 'networked harm'. As explained in the employment algorithmic primer (Rosenblat et al., 2014c), "[h]arms from networked information stem from the sudden availability of large amounts of data on individuals that is gathered and shared beyond their control. Legal remedies for individual harm are not structured in a way that accounts for networked harms" (pg. 5).

2.4. Summary

In summary, our critical insights into the growing literature on automated algorithmic decision-making in transformative services reveal wide-ranging discriminatory and other harmful societal effects. Moreover, these insights raise "novel questions about objectivity, legitimacy, matters of inclusion, the black-boxing of accountability, and the systemic effects and unintended consequences of algorithmic decision-making" (Schultze, et al., 2018:7). What makes the raising of these questions even more challenging, we argue, is the invisible nature of these harmful effects. Through datafication practices, algorithms are starting to perform new worlds[2], by creating and enacting "new behaviours, new expressions, new actors and new realities" (Muller, 2015:29) and exerting power over us (Beer, 2017) through these effects (Diakopoulos, 2013). While the harmful effects of datafication practices and 'algorithmic doing' (Introna, 2016) continue to be reported in the literature by more and more studies, there is a paucity of research that engages with a theoretical explanation of *how* algorithms *perform* these effects. An evident lack of IS studies in this area makes it urgent that the IS community engage with these questions, taking a critical approach, rather than the currently prevailing celebratory approach. With this objective in mind, in the next section we propose a transformative redefinition (Alvesson and Deetz 2000) of the harmful social effects of automated algorithmic decision-making as *algorithmic pollution*.

3. Algorithmic Pollution

To articulate and justify the notion of algorithmic pollution, we ground our elaboration in the long-established concept of *social pollution*, which has been used in different fields, including social

[2] A vivid example of algorithms, or more precisely datafication practices, performing new worlds is the story of My Shed (Butler, 2017). A fake restaurant, made into a top restaurant on Trip Advisor through a deliberate datafication experiment (scoring), led to the opening of a physical restaurant in the author's shed.

DOI: 10.1177/02683962211010356

sciences, law (both in research and in practice), education, political and cultural studies. The concept of social pollution originates from the concept of moral pollution in theological and cultural studies (Douglas, 1966, 1969, 1975). Both concepts – social and moral pollution – appeared in the research literature and the practice of law long before our "contemporary understanding of pollution as a uniquely environmental phenomenon" (Nagle, 2009, p.39).

Drawing upon Douglas' anthropological understanding of pollution, Sarine (2012) shifts the locus of pollution to *harm*. Focusing on discrimination as a specific form of harm, Sarine (2012) defines social pollution "as encompassing systematic discrimination created by implicit bias" (p.1359). Sarine's (2012) interpretation of social pollution-as-harm is not unique. Our multidisciplinary literature review reveals numerous other examples of social pollution interpreted as harm. For example, forms of social pollution include racial discrimination and racism (Vesely-Flad 2017; Norman, 2004; Bhattacharyya et al. 2002; Sherman and Clore, 2009), political judgment (Inbar and Pizzaro, 2014), various forms of workplace maltreatment such as bullying, harassment and gender bias (Fedorova and Menshikova, 2014; Paradis et al. 2014; Pietrulewicz, 2016; Dunham, 2017), pollution of privacy by mass surveillance (Froomkin, 2015), as well as forms of pollution in education, caused by the internet (Hope, 2008) or test scores (Haladyna and Nolen, 1991). Recently, the ideas from social pollution, in particular the notion of pollution as a harm to people rather than pollution as discharge, have been used to argue the case for visual pollution, which is now formally recognized as a new type of environmental pollution (Nagle, 2009, Wakil, et al. 2019)

Following Sarine (2012), Vesely-Flad (2017), Nagle (2009) and other contemporary scholars of social pollution, our notion of algorithmic pollution is based on the idea of social pollution-as-harm. As mentioned earlier, we use the term algorithmic pollution, not as an analogy or a metaphor, but to signify a new kind of social pollution of the human/digital environment. In doing so, we also draw inspiration and encouragement from Nagle's (2009) argument about the need to recognise new types of pollution by looking beyond the narrow interpretation of environmental pollution. As Nagle (2009) explains: "pollution has always had dual meaning: a broad reference to all sorts of effects upon human environments, and a narrow focus upon natural environments" (p.6). In this research we take the former broad meaning of pollution as a reference to harmful effects and recognize algorithmic pollution as a social pollution caused by AADM and datafication. In particular, we argue that algorithmic pollution is a new form of digital social pollution, distinct from other forms of social pollutions previously studied in social sciences.

When AADM, underpinned by datafication, leads to harmful effects we call this phenomenon **algorithmic pollution** and define it as follows:

> *Algorithmic pollution is a new form of social pollution which denotes the unjustified, unfair, discriminatory, and other harmful effects of automated algorithmic decision-making for individuals, their families, groups of people, communities, organizations, sections of the population, and society at large.*

Algorithmic pollution is an appropriate and important framing of the negative effects of AI and algorithms for IS research and practice for several reasons. First, while it assumes a baseline position that AI and algorithms can be a force for good with a significant potential for positive business and social impacts, it recognizes that there are both intended and unintended, yet harmful consequences for individuals, communities and society. Second, it gives visibility to and promotes understandings of these consequences which have not been recognized or expressed in public debates. Third, it creates awareness and provides a reminder for adopters and advocates of automated algorithmic decision-making to consider not only the potential benefits but also the risks of serious harmful effects that may arise as unintended side-effects of AADM initiatives that "often

DOI: 10.1177/02683962211010356

serve legitimate social purposes" (Sarine, 2022, p.1383). Fourth, the notion of algorithmic pollution signals the need for society-wide monitoring of the effects of algorithmic decision-making and identifying and addressing harm to individuals and communities. Finally, the increased awareness and the attention to evidence of algorithmic pollution may help instigate public debates about the future of decision-making in transformative services.

Learning the lessons from the history of pollution of all kinds, we understand why naming a new pollution phenomenon matters for the recognition and examination of its consequences (Nagle, 2009, Wakil, et al. 2019). We further emphasize that our proposal and articulation of the concept of algorithmic pollution as *harms* to individuals, communities and ultimately society, performed by automated algorithmic decision-making, represents a transformative redefinition (Alvesson and Deetz, 2000) as it undermines the dominant discourses that neglect or discount these harms and also encourages alternative ways of seeing and understanding reality enacted by algorithms and datafication. This is an important contribution of our critical research: it encourages alternative ways of seeing a phenomenon, in our case, the overall effects of automated algorithmic decision-making, and promotes the setting of new agendas in both research and public debates. In the following section, we examine in greater depth how algorithmic pollution is performed and how it is spreading.

4. A critical performative view of algorithmic pollution

To understand and explain how algorithmic pollution arises and how it is performed, we adopt a critical performative perspective (Barad, 2007; Introna, 2019). This perspective allows us to expose the ontological assumptions underlying AADM and the ways in which the execution of algorithms enacts the subjects of decisions and reconstructs the sociomaterial environment.

The ontological assumptions underlying AADM, and specifically those about the subjects of decision-making (clients, citizens, children at risk, offenders, welfare recipients, students, and others) – what they are, how they are represented by data sets and how these data sets are used to compute decisions about them – are often taken for granted and as such, not discussed by organizations that adopt AADM. Revealing and exploring these assumptions is not only a matter for academic debate - it is, we argue, fundamental to understanding how the execution of AADM systems interfere in and perform reality and how the harms done to individuals and communities remain a non-issue, rejected or tacitly accepted as inevitable and justified in the business or public sector organizations that deploy them. Questioning ontological assumptions and the ways AADM technologies are designed, deployed, and executed in *specific sociomaterial environment* is also highly important for IS and all other researchers who are concerned with and seek to explain their harmful human and social implications (for which any responsibility is yet-to-be taken).

AADM assumes that individuals that are subjects of decision-making exist as externally bounded and self-contained entities with given properties. To acquire data about relevant properties of the targeted individuals ('entities'), algorithms draw, as we discussed above, from various available data sources that are consolidated per individual (Bucher 2018; Clarke, 2019a; Marsh, 2019) as well as proxy, inferred and unrelated data, often produced by other algorithms in other contexts and for unknown purposes. In addition, targeted individuals are not aware of such data collection, nor do they give permission for the use of such data. Nevertheless, when adopting AADM, organizations take for granted that the collected data sets about individuals 'represent' them *sufficiently accurately and fairly enough* so that correct and fair decisions are made (Gitelman 2013; Bucher, 2018; Dencik, Redden et al. 2019; Dencik, Hintz and Cable, 2019; Marsh, 2019; Cheney-Lippold, 2018). Consequently, organizations are confident that they can use such data for automating their decision-making (Seaver, 2013; Kitchin, 2017): for instance, to calculate scores (risk scores; credit scores); to make predictions (the likelihood of failing at a university or of reoffending); to determine sentencing in court proceedings; to shortlist job applicants; and, to decide on loan approvals.

DOI: 10.1177/02683962211010356

Further, to automate decision-making processes in transformative services it is assumed that the required knowledge possessed by human decision-makers in a given domain, can be 'acquired', 'inferred', and 'contained' by algorithms. In other words, through training based on past data sets in specific domains of decision-making, AADM is assumed to acquire relevant knowledge to predict the outcomes for new cases (such as those committing offences, job applicants, or social security claimants) and thus make appropriate decisions. Irrespective of the complexity and equivocality of knowledge and the decision-making process and the ways in which human beings (such as judges, recruiters or social security case workers) come to their decision in any concrete case, AADM is assumed to be able to 'replicate' the decision-making based on learning from past data sets (i.e., decisions in past cases) and even outperform human decision-makers (Baer and Kamalnath 2017; Kitchin 2017).

This suggests that AADM in transformative services is based on two important assumptions. First, that decision-making practices, involving situated knowledges, professional discourses, cultural-historic experiences and moral and ethical reasoning can be abstracted and generalized (patterned) based on a sufficiently large number of past cases. Consequently, achieving the desired quality (accuracy, fairness) of decision-making by AADM becomes a question of the size and quality of data sets of past decisions. Additionally, quality can be improved by advancing the learning algorithms themselves. AADM thus assumes that complex and value-laden practices of transformative services are, as David (2019) argues, reducible to opaque, incomprehensible correlations derived from masses of past data, a learning that seems unlimited and unattainable by human beings.

The second assumption underpinning AADM is that future decisions will largely resemble past decisions. This ignores the novelty that shows up in any new case, making it unique, and uniquely challenging for decision-makers. In any domain of transformative services, practices also change and evolve over time, responding for example to changes in society, advancement of professional knowledge and regulatory changes. Recalling Bergson (1911), Maria David observes that the "future is not a permanently recomposed past" (2019: 892).

To summarize, the ontology underlying AADM in practice is an entitative, substantialist ontology that assumes the separate existence of human beings and decision-making technologies, with each considered bounded, self-contained entities (Cecez-Kecmanovic, 2016). Even more radically, it is a perfect mechanistic, reductionist ontology that exiles decision-makers and their sociomaterial practices. The sociomaterial practices of decision-making as collective, knowledgeable doings (Gherardi, 2019) are black-boxed, reduced to algorithms that only 'know' and deal with datafied individuals (data sets) as the objects of decision-making. The complex, dynamic, and uncertain reality of decision-making practices in transformative services is thus reduced to a mechanical, rational, clock-like working reality. Such an ontology provides grounds for the belief that AADM is wholly independent of and exterior to knowledge of the actors in sociomaterial environments in which algorithms are deployed and executed. AADM is therefore assumed as an independent, external factor in this environment - one that improves efficiency, correctness and fairness of decision-making while reducing its costs. If this is so, why should we be worried about such an ontology and why is it relevant for understanding algorithmic pollution?

While such an ontology might seem common-sensical, it is, we agree with Introna (2019), in many ways misleading. It prevents us from understanding the "radical openness of sociomaterial becoming" (Introna, 2019:317) of that which is assumed as pre-given and fixed – the human/social, the technological, and their entanglements in sociomaterial practices. Moreover, such an ontology underlies the claim that automated algorithmic decision-making is objective, fair, ethical and moral which thus justifies any 'impacts' on subjects made by the AADM (technologies) as objective and fair. In other words, AADM that is designed (and continually improved) as objective, fair, ethical and moral is a guarantor that its execution performs (equally) objective, fair, ethical and moral decisions. When evidence shows that this actually is not the case and that biases, discrimination and unfair decisions are made, there is still an assumption, held by many as Gangadharan & Niklas

DOI: 10.1177/0268396220915600

(2019) observe, that this is a technical problem that could be fixed by improving algorithms and de-biasing data by technical means. For instance, the long-established traditional stream of research on Algorithmic Fairness, Accountability and Transparency in Machine Learning (FAT-ML) is seeking to develop technical solutions that would ensure the desired qualities of algorithmic decision-making and its outcomes are achieved (see for example Zemel et al. 2013; Celis et al. 2018; Bellamy et al. 2018). Calling it highly influential, Gangadharan & Niklas (2019) argue that the mainstream FAT-ML field[3] focuses on identifying criteria to assess if machine learning is fair, while failing to articulate their underlying assumptions about antidiscrimination or fairness (assumptions which are, in themselves, value-based). This suggests that the ontology underpinning AADM prevents recognition of harms done to individuals (i.e., the subjects of decision-making) and thus limits and potentially disables our understanding of algorithmic pollution.

By adopting a critical performative perspective grounded in the ontology of becoming (Barad, 2007; Cecez-Kecmanovic, 2016; Introna, 2019), we are able to expose how the execution of AADM *performs* the subjects (i.e., datafied individuals) in the ongoing flow of their sociomaterial becoming. To do that we draw attention to the reconfiguration of sociomaterial practices of transformative services through 'intra-acting' triggered by the execution of AADM. We use Barad's (2007) concept of intra-acting to describe how actors (subjects and objects) emerge from, rather than precede, the relations that produce them (see e.g., Orlikowski, 2007; Orlikowski and Scott, 2008; Cecez-Kecmanovic et al., 2014). When algorithms are executed and decisions implemented – social security payments determined and administered; court sentences issued; loans granted/declined – the intra-actions are triggered in targeted sociomaterial environments. Through such intra-acting, the individuals who are the subjects of these algorithmic decisions are performed in ways assumed by the algorithm and made real as part of a reconfiguration of the sociomaterial environment. By focusing on the performing and reconfiguration we can now explain how algorithmic pollution is generated in sociomaterial environments.

First, the intra-acting triggered by algorithmic decisions involves the clash of the entitative, mechanistic ontology (assumptions) underlying AADM and the real-life, complex and dynamic sociomaterial practices in which it is deployed and executed. While this ontological clash is not observable it is experienced by the subjects of AADM and also becomes revealed in practice when, for instance, citizens or neighbourhoods are wrongly identified as high-risk based on past police data and proxy and inferred data (Ferguson 2017; Cino 2018); or when court sentencing is evidently discriminatory (Caplan et al. 2018; Denick, Redden et al., 2019). We suggest that the clash of ontology underlying AADM and the real-life practices (purportedly reflected in the algorithm) enacted through intra-acting is central to understanding and explaining the unfolding of algorithmic pollution.

Second, such a clash of ontologies produces performative effects. Through intra-acting in particular sociomaterial practices, AADM *performs* individuals: concrete individuals become what the algorithm claims they are – unsuccessful job applicants, 'failing' students, high-risk citizens, or security suspects. The repeated execution of AADM thus continually reconstructs actual people – employees, clients, citizens – in the image of datafied individuals. When such performing creates harmful material implications for such individuals (and communities) algorithmic pollution is generated. Moreover, pollution continues to spread through systems of algorithms, resulting in

[3] While we focus on the mainstream FAT-ML research, it is important to acknowledge its emerging streams. For example, the 2020 ACM conference on Fairness, Accountability and Transparency (now known as FAccT), has broadened its scope to include "ethics and policy", with papers investigating the social good (Washington and Kuo, 2020), collective freedom (Terzis, 2020), and algorithmic targeting of social policies (Noriega-Campera, et al., 2020). While the underlying ontology may still be one that favours technology-based solutions, it is encouraging to see a wider concern for society. It allows for recognition of harms done to individuals (the subjects of decision-making) and for society at large, and may thus assist in addressing algorithmic pollution.

DOI: 10.1177/02683962211010356

ongoing networked harm that is perpetually amplified, such that, for example, 'failing students' become 'unsuccessful job applicants' or 'high-risk' citizens. Harm is spread across different contexts in a hard-to-trace network of interconnected transformative services – all designed to help the same individual. Thus, networked algorithmic pollution ends up amplifying harm, and performing new forms of algorithmic pollution.

Third, algorithms not only *perform* datafied individuals (i.e., digital versions of individuals), they also reconfigure their relations with institutions – companies, governments, police, courts, social security departments, and the like. The intra-acting triggered by repeated execution of AADM in sociomaterial environments is materially constrained by mechanical, one-way transactions, that often disable and exclude individuals' ability to reply, complain or give feedback. In the case of court sentencing, or police profiling and targeting of citizens that have not committed any crime (Ferguson 2017; Caplan et al. 2018), individuals are given little or no opportunity to object to and demonstrate that an algorithmic decision isn't right or isn't legal (O'Neil 2016; Eubanks 2018; Benjamin, 2019). In such cases AADM reconfigures relations between citizens and institutions, often transforming them into coercive power relations and strict control mechanisms. In the words of Benjamin "[w]e should acknowledge that most people are forced to live inside someone else's imagination, and one of the things we have to come to grips with is how the nightmares that many people are forced to endure are really the underside of elite fantasies about efficiency, profit, safety and social control" (Johnson, 2020 p.1.)

As algorithmic decision-making permeates transformative services, sociomaterial practices get reconfigured: targeted subjects become performed as particular individuals (risky, suspect, or guilty) subordinated to institutions that efficiently exercise power over them through algorithmic acting (Diakopoulos 2013; Eubanks, 2018; Keddell, 2019). In the case of network harm, this algorithmic acting propagates and amplifies harm across different contexts. By deploying AADM and replacing human decision-making practices, institutions reconstruct their sociomaterial environments in the image of an embedded economic-rational logic concerned solely with efficiency and cost cutting (Bucher, 2018). Algorithmic harm thus becomes both generated and generative, causing harm now and harm in the future.

5. Addressing algorithmic pollution through transformative actions

As the use of AADM proliferates and algorithmic pollution rapidly advances there is a sense of urgency to act promptly to address the damage done and to prevent or mitigate further polluting. As the first step, we recognize the need to start from a fundamental question: *What kind of problem is algorithmic pollution*? Our theorisation has surfaced the existing ontological clashes, which offer different answers to this question. For example, some researchers including those from the traditional stream of FAT-ML, as Gangadharan & Niklas (2019) observe, see the problem of algorithmic harm primarily as a technical issue – one that can be addressed by better quality data and more accurate and transparent algorithms. Consequently, the solutions suggested in the related literature (see for example Zemel et al. 2013; Celis et al. 2018; Bellamy et al. 2018) are also grounded in a technical rationality. Contrary to this view, we argue that algorithmic pollution should be treated similarly to other types of social pollution and as a matter of social justice, as discussed earlier[4]. Following our critical performative perspective, we propose here possible

[4] This argument is also inspired by pioneering work of an emerging group of multidisciplinary researchers (such as (Keddell, 2019; Marda, 2019; Dencik, Hintz, Redden and Trere, 2019; Gillingham, 2019; Sloan and Warner, 2020; Zavrsnik, 2020; Mann, 2020).

DOI: 10.1177/02683962211010356

transformative actions[5] to address algorithmic pollution. While we discuss these transformative actions individually it should be noted that they are interrelated and overlapping.

- ***Prevention of algorithmic pollution***

This transformative action focuses on the key question: *What can we do to stop algorithmic pollution from occurring in the first place*?

While answering this complex question requires further research, we argue that the very concept of AADM that excludes human involvement, oversight and responsibility should be questioned. Especially in the highly sensitive context of transformative services, which are, as we discussed, critical for the well-being of citizens and communities. The ontological clash among AADM systems and concrete decision-making practices, that ultimately leads to social pollution of sociomaterial environments, cannot be remedied by technological and data improvements alone. There is no reason to question the best intentions in designing the technologies and in using the best available data. However, no matter how advanced and sophisticated the technologies become and how much data sources improve, the automation of decision-making processes that are complex, uncertain, and equivocal remains an elusive goal (as Jarrahi, 2018; Davenport and Kirby, 2016; among many others, show). There are already calls to abandon the idea of automating decision-making and instead rely on human-machine collaboration in decision making processes. Thus, instead of using AI to replace humans (knowledge workers, managers) and automate decision making processes, Jarrahi (2018), for example, argues that 'human-machine symbiosis'[6] and collaborative decision making are more promising. To this, we add the importance of considering a particular *context* of transformative services.

As previously discussed, decision-making processes in transformative services are characterized by complexity, uncertainty and equivocality, often involving ethical and moral judgements. Uniquely human faculties such as intuitive and creative thinking, holistic vision, ethical and moral reasoning, emotional intelligence, compassion and empathy, and the ability to get deep insights into and assess intangible social aspects, are indispensable in this context. On the other hand, computational information processing, mathematical modelling, AI and analytics are far superior in dealing with large data sets and complexity of decision-making, compared to humans. These are the arguments for proposing human-machine collaboration in which humans and AI technologies would have complementary roles, drawing on their comparative strengths (Davenport, 2016; Jarrahi, 2018). Instead of automating human decision-making, the role of AI technologies would be to augment and enhance human intelligence and advance decision-making processes above and beyond what is possible by either humans or machines on their own. In the words of Ginni Rometty, the president of IBM, "this is about man and machine, not man vs. machine. This is an era—really, an era that will play out for decades in front of us." (Murphy 2017).

Indeed, this is a long-term prospect for imagining, exploring, developing and testing in practice human-machine symbiotic working and cooperative decision-making. It would require a fundamental rethinking of decision-making problems in transformative services and an exploration of complementary roles that both humans and technologies could, and should, play in the *context* of transformative services. For any type of decision making the forms of human-algorithm cooperative working and acting have to be examined together with possible configurations of agency distribution while preserving human responsibility for the outcomes in sociomaterial practices. Possible configurations have to be tested and monitored in practice with particular sensitivity to the fairness, ethicality and morality of outcomes. Further, we expect that any cooperative form of human-machine decision-making would evolve in time. Through collaborative work with AI technologies, human actors (knowledge workers, managers, citizens) will learn about

[5] Although they are both using the word "transformative", the notion of "transformative actions" from critical research (Avesson and Deetz, 2000) is unrelated to its use in "transformative services".

[6] This is inspired by the original idea of Licklider (1960) from MIT Labs.

DOI: 10.1177/02683962211010356

these technologies, what they can (and cannot) do, and also about the relevant data sets, their quality and their ethical use. Given rapid developments in AI capabilities, decision-makers would also need to become more knowledgeable about new analytic techniques in order to be able to explore new opportunities for advancing decision-making processes (e.g., new configurations of task allocation and agency distribution in a particular context).

While adopting human-machine partnerships and cooperative decision-making, organizations would need to take full responsibility for the outcomes and their social implications. Any organization providing transformative services to citizens, including government social security and other agencies, hospitals, schools, police departments, or courts, would continue to act in accordance with the norms, rules, and regulations established in society. Holding organizations accountable for their actions, which translates to individual decision-maker's responsibility in their specific domains across an organization, is critical for preventing algorithmic pollution.

The adoption of human-machine symbiotic working and cooperative decision-making that we propose as a key transformative action to prevent algorithmic pollution would also, as the above discussion shows, advance the decision-making processes beyond what would be possible by AADM or human decision-making alone. For such a transformative action, however, there are no ready-made simple solutions as to how decision-making tasks would be shared between human and algorithm and how they might work cooperatively and make decisions in a particular context. This opens a new domain of research into configurations of human-machine cooperative decision-making in the context of transformative services and their evolution over time. Emerging research questions, among many, include: What are the distinctly human roles and responsibilities in transformative services? How can AI and analytic processing be employed to augment and enhance human capacities: How can they together make decisions not only more efficiently, but also in a socially responsible, ethical and moral way?

Informed by the social pollution literature, we see important steps in this direction. The anthropological notion of pollution also draws attention to a system of values (Douglas, 1966). For instance, in designing and practicing human-machine cooperative decision-making there needs to be recognition and articulation of competing values (e.g., efficiency and cost reduction versus care for people) and understanding how these values are guiding the decision-making process. Different scenarios can be experimented with (using algorithmic calculations and predictions) to assess impacts on these values, that would ultimately inform decisions.

Finally, in line with a number of social pollution scholars who advocate elimination of social pollution (Sarine, 2012), we would like to emphasize that the goal of prevention of algorithmic pollution should be *elimination*, rather than living with an 'acceptable level of harm'. Adopting the social justice perspective, we propose that any level of harm is still harm, and as such should not be tolerated in a civil society.

- ***Detection of algorithmic pollution***

The transformative action of detection focuses on the key question: *How do we know algorithmic pollution has occurred?*

Based on our research, we suggest that detection of algorithmic pollution needs to consider society-wide datafication practices. Moreover, detection should not be implemented by or left to any single authority, including government legislators. This is due to the complex and unknowable nature of society-wide datafication, with harmful effects being propagated, amalgamated and amplified on an ongoing basis and in unknowable ways. Instead, we argue, detection of algorithmic pollution needs to be an ongoing, society-wide initiative, enacted through systematic means, and made visible through appropriate channels.

Therefore, further IS research is needed to understand what these detection mechanisms entail and how they might be implemented. Here we see two IS research challenges: (i) society-wide tracing of algorithmic pollution through systems of algorithms, following the trails of datafication

DOI: 10.1177/02683962211010356

practices, and (ii) society-wide detection and reporting of algorithmic pollution, which needs to be ongoing and systematic. Although more research is required, the existing literature offers some starting points. For example, the emerging research on data activism and civil society actions (ACLU, 2016; Gutierrez, 2018; Datactive n/d, NotMyDebt, n/d) point to the need for grass-root reporting of harm. We also observe that the current initiatives of various data activist groups are isolated and focused on one-directional change through activism, including collective pressure for change and/or legal actions.

Inspired by these insights, we see the need for a large-scale society-wide information system that could meet the previously identified IS research challenges by enabling a coordinated grass-root reporting of algorithmic pollution by the affected stakeholders and/or those who have the power and resources to act on their behalf. Further characteristics of this type of systems are discussed below in relation to mitigation of algorithmic pollution.

- ***Redress of algorithmic pollution***

The transformative action of redress focuses on the key question: *What can we do to redress harm suffered by individuals exposed to algorithmic pollution on a case-by-case basis?*

Our research reveals that algorithmic pollution involves different types of harm. When harms are detected they need to be addressed. We recognize the redress of algorithmic harm to be a multidisciplinary challenge that is currently discussed by legal scholars and practitioners, such as Zavrsnik (2019), Zalnieriute et al. (2019), social scientists (Keddell, 2019), social justice researchers (Marda, 2019), as well as multidisciplinary researchers (Metcalf et al., 2021). The questions about how to determine the level of harm and who is responsible for the assessment and redress of harm, as Metcalf et al. (2021) explain, are domain specific and regulated by different norms about what constitutes harm. We consider these important questions to be outside of our collective IS expertise. Instead, in this paper we focus on the IS perspective that we observe is currently missing from this multidisciplinary discourse about the redress of harm.

Here, we see the important IS research challenges that are again focused on datafication practices, in relation to individuals. They could be captured by the following research questions: How can we trace and disentangle 'datafied individuals', back to the sources of data and datafication practices used to construct the individual's datafied representation? Which of these datafication steps caused and/or contributed to algorithmic harm experienced by an individual? How can we disentangle networked harm, in order to trace and determine responsibilities when harm is the result of a system of algorithms? Who is responsible for networked harms that are created as more than the sum of individual services?

Broad (2018) describes a possible first step toward redressing algorithmic harm: "Perhaps at the minimum, any organisation deploying AI systems in decision-making contexts should be required to provide documentation publicly, and to purchasers of their system, about the data they've used to train their system: when it was collected, for what purpose, the characteristics it includes, its limitations and omissions." (p.52). The same practice could be also used to the tracing and detection of algorithmic pollution.

- ***Mitigation of algorithmic pollution***

The transformative action of mitigation focuses on the key question: *What can we do to address harm involved in algorithmic pollution at the societal level?*

In seeking possible answers to this question, we observe a growing number of various AI ethics frameworks and guiding principles. They are being generated by commercial enterprises, activist bodies, international organisations, and government agencies - see for example those created by the Australian Human Rights Commission (2019), the Toronto Declaration (Brandom, 2018), the European Commission (2019) and the United Nations (2020).

DOI: 10.1177/02683962211010356

While these initiatives are of tremendous importance, we also observe a serious limitation, i.e., they are primarily focused on the developers of new AI-enabled algorithms. As such, they do not address ongoing algorithmic pollution and related issues of society-wide datafication, which may still occur in spite of best intentions and actions of a single organisation.

Drawing from the work by social justice theorist Nancy Fraser (2008), we propose that mitigation of algorithmic pollution requires an ongoing dialogical process, between a civil society track with grass-root insights into harm experienced by individuals and an institutional track (e.g., government regulators), with legislative power and the capacity to make decisions.

In the case of algorithmic harm, we also observe the ongoing tension created by mutually competing goals of different stakeholders, such as efficiency and cost reduction en masse, versus the need to prevent and mitigate harmful effects of algorithmic decision-making. Consequently, we argue the need for independent regulatory oversight of transformative services with the authority to act in cases of algorithmic harm as well as having the power to influence the formation and content of policies and regulations. This idea is further supported by recent research by Sun and Gershik (2019) and recommendations made by the US Government Accountability Office (2019). They both argue for the establishment of an oversight agency with relevant expertise to deal with society-wide issues of algorithmic harm.

The need for an ongoing, society-wide, dialogical process of mitigation, overseen by an independent regulator brings us back to an IS research challenge concerning the design and implementation a society-wide information system to support such a process. This in turn leads to a number of research questions: What kind of IS is it? How might we design such a system and who should be involved? Will it require new IS design methodologies (for example, community-based approaches)? What is the most effective way to implement such an IS to support an ongoing society-wide dialogue? How will it support the algorithm mitigation work of an independent regulator?

- ***Education about algorithmic pollution***

Fundamental to prevention, detection, redress and mitigation of algorithmic pollution is a transformative action of education. First, there is a clear need for education of managers, developers and, in particular policy makers and regulators (Caplan et al. 2018) about algorithms and their use in AADM. We argue that these stakeholders need to be educated about the myths and limitations of 'data objectivity' and 'accurate representation' of 'entities' (i.e., datafied individuals) that are used and produced by algorithms in the area of human services. As Broad (2018) advises, we can learn from anthropologists, sociologists, historians, librarians, social workers, health care administrators and others who have been collecting and analyzing data about humans for some time.

Second, it is also important to educate various stakeholders about the notion of fairness and 'correctness' of algorithms. In particular, they need to be made aware of the notion of fairness in social justice, which is much broader than fairness in statistics and computer science (Keddell, 2019). Moreover, as social justice scholars now argue, the widely-used ethical frameworks of fairness, accountability, and transparency (FAT-ML) of algorithms do not go far enough, and need to be augmented with the concepts of justice and human rights when considering AADM in transformative services (Keddell, 2019; Gurses et al., 2019; Zavrsnik, 2019, 2020; Marda, 2019; Chouldechova, 2017).

These insights from the social justice literature, combined with an awareness of a growing influence of techno-solutionism, (Morozov, 2013; Zavrsnik, 2019), may enable managers, developers and other stakeholders to better understand the important limitations of the widespread claims about the superiority of algorithms over human decision makers, in their own contexts. This in turn, may empower them to engage in important conversations, especially when dealing with third parties selling their algorithmic solutions and data (i.e., datafied individuals) to governments and other providers of social services.

DOI: 10.1177/02683962211010356

Third, inspired by prior research in social pollution (Sarine, 2012), we see the need for education that goes beyond the content (e.g., what are the algorithms, what they can and cannot do), to include broader societal context, including existing systemic and other structural injustices. This is also echoed by an emerging stream of FAT-ML researchers such as Benjamin (2019), Gebru (2019, 2020), and Crawford (2019).

Forth, to be effective, education about algorithmic pollution needs to empower, not just inform. It needs to empower knowledge workers and anyone working *in collaboration with algorithms* to deal with new societal moral and ethical issues arising from the ongoing tensions created by competing goals of different legitimate stakeholders, from the position of responsibility for welfare of others, personal integrity and compassion for our shared humanity. It also needs to empower citizens and other societal stakeholders for an ongoing society-wide dialogue about algorithmic pollution.

The previous discussion leads to a number of research questions related to education about algorithmic pollution: How might we design, implement, and evaluate education about algorithmic pollution? Who should be involved? Who should be responsible for its implementation? How to design effective pedagogical methods and practices to educate a wide range of stakeholders about fairness from the perspective of social justice and human rights? How to 'educate to empower' knowledge-workers, managers, citizens and other stakeholders for an ongoing society-wide dialogue about algorithmic pollution? How to make this education embedded, contextualised and 'living', to enable co-existence with ever-changing algorithms (Schultze, et al. 2018) such that human agents are responsible, and in charge?

Finally, the five transformative actions of prevention, detection, redress, mitigation and education taken together not only open up new research challenges for IS scholars, they also invite us to reflect on 'the how' of doing research on algorithmic pollution. We emphasise the importance of conducting such research from foundations of care and compassion for our shared humanity. In doing so, we join Raman and McClelland's (2019) call for bringing compassion into IS research. In this case, both when researching algorithmic pollution and also when participating in transformative actions to prevent, detect, redress, mitigate, and educate about algorithmic harm.

6. Concluding remarks and a call for action

By focusing on the unintended harmful societal effects of automated algorithmic decision-making in the context of transformative services, we have put forward a claim - and provided evidence - that these harmful effects constitute a new type of widespread, hidden, and largely unregulated digital pollution, which we name algorithmic pollution. Building upon well-established research on social pollution, we recognise algorithmic pollution as a new kind of social pollution and offer a theoretical explanation of how it is performed. By using the term algorithmic pollution in a non-metaphorical sense, we foreground harms performed by automated algorithmic decision-making as a new type of largely invisible, wide-spread social digital pollution. We thus make it visible and raise public awareness of its dangers, calling for urgent action.

Our main theoretical contributions come from: (1) *critical insights* into, and a *transformative redefinition* of the harmful effects of AADM as algorithmic pollution; (2) novel theoretical explanation of how algorithmic pollution is performed in sociomaterial environments, using a critical performative approach; (3) proposed transformative actions of prevention, detection, redress, and mitigation of, and education about, algorithmic pollution; and (4) identification of associated future research challenges for the information systems (IS) community. Our main practical contributions include: (1) drawing public awareness and recognition of a new type of digital social pollution; (2) enabling broader understanding of how algorithmic pollution is performed and how its consequences spread; and (iii) motivating different actors to engage in transformative actions.

DOI: 10.1177/02683962211010356

It is important to reiterate the positive intentions of those organisations wishing to innovate transformative services through AI and AADM. These services are often critical for the wellbeing of the most vulnerable members of our society. Ultimately, by making algorithmic pollution visible, we aim to help organisations adopting algorithmic decision-making in their transformative services to better achieve their positive intentions of improving the wellbeing of service users. We thus recognize that algorithms undoubtedly have the potential to provide society with significant benefits (e.g., healthcare, education, fraud detection). Therefore, this paper is not a treatise against algorithms. Far from it. As Wachter-Boettcher (2017:11-12) points out "[i]t's not that digitizing the world is inherently bad. But the more technology becomes embedded in all aspects of life, the more it matters whether that technology is biased, alienating, or harmful."

By deliberately using the word 'pollution' to name this new phenomenon, we aim to make it easier for all of us (research communities, policy makers and the general public) to relate to these effects (both intellectually and emotionally) in order to understand the seriousness of the current situation. If algorithms are our future, then understanding, and continually looking for new ways to prevent, detect, redress and mitigate algorithmic pollution as a new kind of social pollution, may help us to maintain and even improve our individual and collective wellbeing, as well as our humanity.

References

1. ACLU (2016) *Statement of Concern About Predictive Policing*, Report. American Civil Liberties Union. Available at: https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice (accessed 14 Nov 2020).
2. Alarconn A, Zeide E, Wikelious K et al. (2014) *Data & Civil Rights: Education Primer*, Data and Society Research Institute. Available at: https://datasociety.net/output/data-civil-rights-education-primer/ (accessed 18 Nov 2020).
3. Alston P (2019a) *Report of the Special rapporteur on extreme poverty and human rights*. Report. UN General Assembly. Report id A/74/493. Available at: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/312/13/PDF/N1931213.pdf?OpenElement (accessed 24 Oct 2020).
4. Alston P (2019b) *Report of the Special rapporteur on extreme poverty and human rights*: *Visit to the United Kingdom of Great Britain and Northern Ireland.* Report. UN General Assembly. Report id A/HRC/41/39/Add.1. Available at: https://undocs.org/en/A/HRC/41/39/Add.1 (accessed 24 Oct 2020).
5. Alvesson M and Deetz S (2000) *Doing Critical Management Research*. London: SAGE.
6. Anderson L and Ostrom AL (2015) Transformative service research – Advancing our knowledge about service and well-being. *Journal of Service Research* 18(3): 243-249.
7. Anderson L, Ostrom AL, Corus C et al. (2013) Transformative service research: An agenda for the future. *Journal of Business Research* 66(8): 1203-1210.
8. Anderson S, Nasr L and Rayburn SW (2018) Transformative service research and service design: synergistic effects in healthcare. *The Service Industries Journal* 38(1-2): 99-113.
9. Australian Human Rights Commission (2019) *Human Rights and Technology Discussion Paper*. Available at https://tech.humanrights.gov.au/sites/default/files/2019-12/TechRights_2019_DiscussionPaper.pdf (accessed 30 Oct 2020).
10. Baer T and Kamalnath V (2017) *Controlling machine-learning algorithms and their biases*. Report. McKinsey & Company. Available from: https://www.mckinsey.com/business-functions/risk/our-insights/controlling-machine-learning-algorithms-and-their-biases (accessed 20 Oct 2020).
11. Barad K (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press.
12. Beer D (2017) The social power of algorithms. *Information, Communication & Society* 20 (1): 1–13.
13. Bellamy R et al. (2018) *AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias*. IBM Research. Available from: https://arxiv.org/pdf/1810.01943.pdf (accessed 20 Nov 2020).
14. Benjamin B (2019) *Race After Technology*. Cambridge: Polity Press.
15. Bergson H (1911) *Creative Evolution*. Trans. by A. Mitchell. New York: Henry Holt & Co.
16. Bhattacharyya G, Gabriel J and Small S (2002) *Race and Power: Global Racism in the Twenty-first Century*. London: Psychology Press.
17. Blocker C and Davis B (2019) CFP: JSR Special issue: Transformative service research and unintended consequences: Helping without harming. *Journal of Service Research.* Available at: https://www.servsig.org/wordpress/2019/02/cfp-jsr-transformative-service-research/ (accessed 20 Dec 2020)

DOI: 10.1177/02683962211010356

18. Bohme R (2019) Response to Clarke: Empirical research useful, also in the age of surveillance risks, *Journal of Information Technology*, 34(1): 93-95.
19. Brandom R (2018) New Toronto Declaration calls on algorithms to respect human rights. *The Verge*, 16 May 2018.
20. Broad E (2018) *Made by Humans: The AI Condition*. Melbourne: University of Melbourne Press.
21. Bucher T (2018) *Neither black nor box. If ..Then: Algorithmic power and politics*. Oxford University Press. USA.
22. Butler O (2017) I made my shed the top rated restaurant on TripAdvisor. *Vice*, 07 Dec 2017. Available at https://www.vice.com/en_au/article/434gqw/i-made-my-shed-the-top-rated-restaurant-on-tripadvisor (accessed 21 July 2020)
23. Caplan R, Donovan J, Hanson L and Matthews J (2018) *Algorithmic accountability: A primer*. Prepared for the Congressional Progressive Caucus. April 18. Data & Society, Washington, DC. USA. Available at: https://datasociety.net/wp-content/uploads/2018/04/ Data_Society_Algorithmic_Accountability_Primer_FINAL-4.pdf. (accessed 21 July 2020).
24. Carney T (2018) The new digital future for welfare: Debts without legal proofs or moral authority?. *UNSW Law Journal Forum* 41(1):1-16.
25. Cecez-Kecmanovic D (2016) From substantialist to process metaphysics – Exploring shifts in IS research, Chapter 1 in Introna L, Kavanagh D, Kelly S, Orlikowski W and Scott S (eds). Beyond Interpretivism? New Encounters with Technology and Organization. Springer, 35-57.
26. Cecez-Kecmanovic D (2019) The resistible rise of the digital surveillance economy: A call for action. *Journal of Information Technology* 34(1): 81-83.
27. Cecez-Kecmanovic D, Galliers RD, Henfridsson O, Newell S and Vidgen R (2014) The sociomateriality of information systems: Current status, future directions. *Management Information Systems Quarterly* 38(3): 809-830.
28. Celis LE, Huang L. et al. (2018) Classification with fairness constraints: A Meta-algorithm with provable guarantees. FAT 2019 Conference on Fairness Accountability Transparency, January 29-31, Atlanta, GA USA, ACM NY, pp.1-10. Available at https://doi.org/10/1145/3287560.3287586 (accessed 21 July 2020).
29. Cheney-Lippold J (2018) *We Are Data: Algorithms and the Making of Our Digital Selves*. NY: New York University Press.
30. Chouldechova A (2017) Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data* 5: 153–63.
31. Cino JG (2018) Deploying the secret police: The use of algorithms in the criminal justice system, *Georgia State University Law Review* 34(4):1073-1102. Available at: https://readingroom.law.gsu.edu/gsulr/vol34/iss4/6. (accessed 21 July 2020).
32. Clarke R (2019a) Risks inherent in the digital surveillance economy: A research agenda. *Journal of Information Technology* 34(1): 59-80.
33. Clarke R (2019b) Future-oriented research agendas and competing ideologies: Responses to commentaries on 'The digital surveillance economy'. *Journal of Information Technology* 34(1): 96-100.
34. Crawford K (2019) Halt the use of facial-recognition technology until it is regulated. *Nature*. 572:563. Available at: https://www.nature.com/articles/d41586-019-02514-7 (accessed 24 Oct 2020).
35. Danaher T and Gallan A (2016) Service research in health care: Positively impacting lives. *Journal of Service Research* 19(4):1-5.
36. Data & Society (n.d) Available at: https://datasociety.net (accessed 21 July 2020)
37. Datactive (n/d) *The Politics of Data According to Civil Society.* Available at: https://data-activism.net/ (accessed 28 Oct 2020).
38. Davenport TH (2016) Rise of the strategy machines. *MIT Sloan Management Review* 58(1):13-16.
39. Davenport TH and Kirby J (2016) Just how smart are smart machines? *MIT Sloan Management Review* 57(3): 21—25.
40. David M (2019) AI and the Illusion of Human-Algorithm Complementarity. *Social Research: An International Quarterly* 86(4): 887-908.
41. Dencik L, Hintz A and Cable J (2019) Towards data justice: Bridging anti-surveillance and social justice activism. in Bigo, D., Isin, E., Ruppert, E (eds). *Data Politics: Worlds, Subjects*, pp. 167-186. Rights. Routledge.
42. Dencik L, Hintz A, Redden, J and Trere E (2019) Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society* 22(7): 873-881.
43. Dencik L, Redden J, Hintz A and Warne H (2019) The 'golden view': data-driven governance in the scoring society. *Internet Policy Review* 8(2): 1-24. Available at: https://policyreview.info/articles/analysis/golden-view-data-driven-governance-scoring-society (accessed 21 July 2020).
44. Deville J and van der Velden L (2016) Seeing the invisible algorithm: The practical politics of tracking the credit trackers. in Amoore, L. and Piotukh, V. (eds), *Algorithmic Life: Calculative Devices in the Age of Big Data*, pp.87-106. Routledge, London and NY.
45. Diakopoulos N (2013) *Algorithmic accountability reporting: On the investigation of black boxes*. Tow Center for Digital Journalism, Columbia Journalism School. Available at: http://towcenter.org/algorithmic-accountability-2/ (accessed 21 July 2020).
46. Douglas M (1966) *Purity and Danger: An Analysis of Concepts of Pollution and Tabo*. New York: Routledge.
47. Douglas M (1969) Pollution. in Sills D L (ed). *International Encyclopedia of the Social Sciences*. XII: 336-341.

DOI: 10.1177/02683962211010356

48. Douglas M (1975) Pollution in Douglas M. *Implicit Meanings: Selected Essays in Anthropology*, pp. 47–59. London: Routledge & Kegan Paul.
49. Dunham C R (2017) Third generation discrimination: The ripple effects of gender bias in the workplace. *Akron Law Review*. 51(1): 55-98. Available at: http://ideaexchange.uakron.edu/akronlawreview/vol51/iss1/2 (accessed 15 Nov 2020).
50. Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. NY: St. Martin's Press.
51. European Commission (2019) *Ethics Guidelines for Trustworthy AI*. Independent High-Level Expert Group on Artificial Intelligence. Available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. (accessed 15 July 2020).
52. Fedorova A and Menchikova M (2014) Social pollution factors and their influence on psychosocial wellbeing at work. International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM2014, SGEM2014 Conference Proceedings, Book 1, Vol. 2, pp. 839-846. DOI: 10.5593/SGEMSOCIAL2014/B12/S2.107
53. Ferguson AG (2017) The police are using computer algorithms to tell if you're a treat. *Time*. Oct. 3 1-2.
54. Fraser N (2008). Abnormal justice. *Critical Inquiry* (34), Spring, 393-422.
55. Froomkin AM (2015) Regulating mass surveillance as privacy pollution: Learning from environmental impact statements. *U. ILL. Law Review* 2015(5):1713 -1790.
56. Gal U, Jensen T and Stein C (2020) Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. *Information and Organization* 30: 1-15.
57. Galliers RD, Newell S, Shanks G and Topi H (2017) Datafication and its human, organizational and societal effects: The strategic opportunities and challenges of algorithmic decision-making. *Journal of Strategic Information Systems* 26(3): 185-190.
58. Gangadharan SP and Niklas J (2019) Decentering technology in discourse on discrimination. *Information, Communication & Society* 22(7): 882-899.
59. Gebru T (2019) Race and gender. In Dubber, M.D., Pasquale, F. and Das. S. (eds). *Oxford Handbook on AI Ethics*. pp. 1-27. Oxford University Press, Available at: arXiv:cs.CY/1908.06165 (accessed 24 Oct 2020).
60. Gebru T (2020) *Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision*. CVPR2020. June 19. Available at: https://sites.google.com/view/fatecv-tutorial/home (accessed 24 Oct 2020).
61. Gherardi S (2019) Practice as a collective and knowledgeable doing. Working Paper Series. Collaborative Research Center 1187 Media of Cooperation. Universitat Siegen. Available at: http://wp-series.mediacoop.uni-siegen.de (accessed 15 July 2020).
62. Gillingham P (2019) Decision support systems, social justice and algorithmic accountability in social work: A new challenge. *Practice* 31(4):277-290.
63. Gitelman L (2013) *"Raw Data" is an Oxymoron*. Cambridge MA: MIT Press.
64. Gurses S Gangadharan S and Venkatasubramanian S (2019) Critiquing and Rethinking Accountability, Fairness, and Transparency. Report. Our Data Bodies Project US. Available online: https://www.odbproject.org/2019/07/15/critiquing-and-rethinking-fairness-accountability-and-transparency/ (accessed 23 Oct 2020).
65. Gutierrez M (2018) *Data Activism and Social Change*, Cham Switzerland: Palgrave MacMillan.
66. Haladyna TM and Nolen SB (1991) Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher* 20(5): 2-7.
67. Hope A (2008) Internet pollution discourses, exclusionary practices and the 'culture of over-blocking' within UK schools. *Technology, Pedagogy and Education* 17(2): 103-113.
68. ICO (2018) What is automated individual decision-making and profiling. UK Information Commissioner's Office. 1-23. Available from https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-profiling/#id2
69. Inbar Y and Pizarro DA (2014) Pollution and purity in moral and political judgment. In Wright J and Sarkissian H (eds.) *Advances in Experimental Moral Psychology*: Affect, Character, and Commitment. Continuum Press. 111-129
70. Introna L (2016) Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology & Human Values* 41(1): 17-49.
71. Introna L (2019) Performativity and sociomaterial becoming: What technologies do. In Webb SA (Ed.), *The Routledge Handbook of Critical Social Work*. London: Routledge.312-323.
72. Jarrahi MH (2018) Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making, *Business Horizons* 61(4): 577-586.
73. Johnson K (2020) Ruha Benjamin on deep learning: Computational depth without sociological depth is 'superficial learning'. *AI Weekly*. April 29. Available from: https://venturebeat.com/2020/04/29/ruha-benjamin-on-deep-learning-computational-depth-without-sociological-depth-is-superficial-learning/ (accessed on 30 Oct 2020).
74. Keddell E (2019) Algorithmic justice in child protection. *Social Sciences* 8(281):1-22.
75. Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20 (1): 14–29.
76. Law J (2004) *After Method: Mess in Social Science Research*. London: Routledge.

DOI: 10.1177/02683962211010356

77. Leidner DE (2019) No risk, no reward. *Journal of Information Technology* 34(1): 84-86.
78. Licklider J C R (1960) Man-Computer symbiosis, *IRE Transactions on Human Factors in Electronics* HFE-1: 4-11. Available at: https://groups.csail.mit.edu/medg/people/psz/Licklider.html (accessed 04 Dec 2020).
79. Loebbecke C and Picot A (2015) Reflection on societal and business model transformation arising from digitization and big data analytics: A research agenda. *Journal of Strategic Information Systems* 24(3): 149-157.
80. Lycett M (2013) Editorial: 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems* 22(4): 381-386.
81. Mann, M. (2020). Technological politics of automated welfare surveillance: Social (and data) justice through critical qualitative inquiry. *Global Perspectives* 1(1), pp. 1-12.
82. March ST (2019) Alexa, are you watching me? A response to Clarke, 'Risks inherent in the digital surveillance economy: A research agenda'. *Journal of Information Technology* 34(1): 87-92.
83. Marda V (2019) Introduction. *2019 Global Information Society Watch: Artificial Intelligence: Human Rights, Social Justice and Development*. Association for Progressive Communication (APC). Article 19. NY: APC Publishing.
84. Marjanovic O and Cecez-Kecmanovic D (2017) Exploring the tension between transparency and datafication effects of open government IS through the lens of Complex Adaptive Systems. *Journal of Strategic Information Systems*. 26(3): 210-232.
85. Markus L (2015) New games, new rules, new scoreboards: the potential consequences of big data. *Journal of Information Technology* 30(1):58-59.
86. Markus L (2017) Datafication, organizational strategy, and IS research: What's the score?", *Journal of Strategic Information Systems* 26(3): 233-24 1.
87. Metcalf J, Moss E, Watkins EA, Sigh R and Elish MC (2021) Algorithmic impact assessments and accountability: The co-construction of impacts. ACM Conference on Fairness, Accountablity, and Transparency (FAccT'21), March 3-10, 2021, Virtual Event, Canada, ACM. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3736261 (accessed 3 Feb 2021).
88. Mikalef P, Popovic A, Lundstorm, JE, Conboy, K (2019) Call for Papers: Special issue on the Dark side of analytics and artificial intelligence. *European Journal of Information Systems* n/n:1-3.
89. Morozov E (2013) *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist.* London: Allen Lane.
90. Mosaic (2019) Available at: https://www.experian.co.uk/business/marketing/ segmentation-targeting/mosaic/ (accessed 4 July 2020).
91. Muller M (2015) Assemblages and actor-networks: Rethinking socio-material power, politics and space, *Geography Compass*. 9(1): 27-41.
92. Murphy M (2017) Ginni Rometty on the end of programming. *Bloomberg Businessweek*. Sept 20. Available at: https://www.bloomberg.com/news/features/2017-09-20/ginni-rometty-on-artificial-intelligence (accessed 21 Dec 2020)
93. Myers M and Klein S (2011) A set of principles for conducting critical research in IS, *Management Information Systems Quarterly* 35(1):17-36.
94. Nagle JC (2009) The Idea of Pollution, 43 University of California Davis Law Review 1 Available at: https://scholarship.law.nd.edu/law_faculty_scholarship/313 (accessed 21 Dec 2020)
95. Newell S and Mirabelli M (2015) Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datafication. *Journal of Strategic Information Systems* 24(1):3-14.
96. Noble SU (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York University Press.
97. Noriega-Campero A, Garcia-Bulle B, Cantu LF, Bakker MA, Tejerina L, Pentland A (2020) Algorithmic targeting of social policies: fairness, accuracy, and distributed governance. In: FAT* '20: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. January 2020: 241–251. Available at: https://doi.org/10.1145/3351095.3375784 (accessed 20 Nov 2020).
98. Norman K (2004) Equality and exclusion: 'racism' in a Swedish town. *Ethnos* 62(2):204-228.
99. NotMyDebt (n/d) #NotMyDept: Confused & concerned about your Centrelink debt? Available at: https://data-activism.net/ (accessed 20 Nov 2020).
100. O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. NY: Penguin Random House.
101. Orlikowski WJ (2007) Sociomaterial practices: Exploring technology at work. *Organization Studies* 28(9): 1435-1448.
102. Orlikowski WJ, Scott SV (2008) Sociomateriality: Challenging the separation of technology, work and organization. *The Academy of Management Annals* 2(1): 433–474.
103. Ostrom AL, Parasuraman A, Bowen DE et al. (2015) Service research priorities in a rapidly changing context. *Journal of Service Research* 18(2): 127-159.
104. Paradis R., Demers M, Dion E, Tivendell J. et al. (2014) Interpersonal pollution in organisations: Exploring ethical leadership and the dark side of organizations. *Polish Journal of Social Science* 9(1):7-26.
105. Park S and Humphry J (2019). Exclusion by design: Interactions of social, digital and data exclusion. *Information, Communication & Society* 22(7): 934-953.

DOI: 10.1177/02683962211010356

106. Pasquale F (2015) *The Black box society: The secret algorithms that control money and information*. Cambridge MA: Harvard University Press.
107. Pietrulewicz B (2016) The interpersonal pollution, dark side of personality and its effect on group members' well-being, and on culture of unity in organizational context. *Journal for Perspectives of Economic Political and Social Integration* 22(1-2): 159-196.
108. Raman R and McClelland L (2019) Bringing compassion into information systems research: A research agenda and call to action. *Journal of Information Technology* 34(1): 2-21.
109. Redden J (2018) Democratic governance in an age of datafiction: Lessons from mapping government discourses and practices. *Big Data & Society* 5(2):1-15.
110. Reisman D, Schultz J, Carwford K and Whittaker M (2018) *Algorithmic impact assessments: A practical framework for public agency accountability*. AI Now Institute. Available at https://ainowinstitute.org/aiareport2018.pdf (accessed 10 July 2020).
111. Rosenblat A, Wikelius K et al. (2014a) *Data & Civil Rights: Health Primer*. Data and Society Research Institute. Available at: https://datasociety.net/output/data-civil-rights-health-primer/ (accessed 18 May 2020).
112. Rosenblat A, Wikelius K et al. (2014b) *Data & Civil Rights: Housing Primer*. Data and Society Research Institute. Available at: https://datasociety.net/output/data-civil-rights-housing-primer/ (accessed 18 May 2020).
113. Rosenblat A, Wikelius K et al. (2014c) *Data & Civil Rights: Employment Primer*. Data and Society Research Institute. Available at: https://datasociety.net/output/data-civil-rights-employment-primer/ (accessed 18 May 2020).
114. Rosenblat A, Wikelius K et al. (2014d) *Data & Civil Rights: Criminal Justice Primer*. Data and Society Research Institute. Available at: https://datasociety.net/output/data-civil-rights-criminal-justice-primer/ (accessed 18 May 2020).
115. Sarine LE (2012) Regulating the social pollution of systematic discrimination caused by implicit bias. *California Law Review* 100 (5):1359-1399.
116. Schultze U, Aanestad M, Mahring M, Osterlund C and Riemer K (2018) *Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology*. IFIP Advances in Information and Communication Technology Book Series. IFIP AICT 543. Cham: Springer.
117. Seaver N (2013) Knowing Algorithms. *Media in Transition 8*:1-12. Available at: http://nickseaver.net/papers/seaverMiT8.pdf. (accessed 21 May 2020).
118. Sherman, GD and Clore GL (2009) The color of sin: While and black are perceptual symbols of moral purity and pollution. *Psychological Science* 20(8): 1019-25.
119. Sloan R H and Warner R (2020) Beyond bias: Artificial intelligence and social justice. *Virginia Journal of Law and Technology*. Epub ahead of print 1 Feb 2020. Available at http://dx.doi.org/10.2139/ssrn.3530090 (accessed 10 Nov 2020).
120. Sun M and Gerchick M (2019).The scales of (algorithmic) justice: Tradeoffs and remedies. *AI Matters* 5(2):30-40.
121. Terzis P (2020) Onward for the freedom of others: marching beyond the AI ethics. In: *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. January 2020: 220–229 https://doi.org/10.1145/3351095.3373152
122. UN (2020) *Artificial Intelligence and Gender Equity: Key findings of UNESCO's Global Dialogue*. UNESCO. Paris. France. Available: https://en.unesco.org/AI-and-GE-2020 (accessed 30 Oct 2020)
123. US Government Accountability Office (2019) *Our new science, technology assessment, and analytics team*. Report. Available at: https://blog.gao.gov/2019/01/ 29/our-new-science-technology -assessment-and-analytics-team/ (accessed 18 Oct 2020)
124. Vesely-Flad RL (2017) *Racial purity and dangerous bodies: Moral pollution, black lives, and the struggle for justice*. Minneapolis: Fortress Press.
125. Wachter-Boettcher S (2017) *Technically Wrong: Sexist Apps, Biased Algorithms and Other Threats of Toxic Tech*. NY:W.W. Northon & Company.
126. Wakil K, Naeem MA, Anjum GA and Waheed A. (2019) A hybrid tool for visual pollution Assessment in urban environments. *Sustainability* 11(8): 2211.
127. Washington AL and Kuo R (2020) Whose side are ethics codes on?: power, responsibility and the social good. In: *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. January 2020: 230–240. https://doi.org/10.1145/3351095.3372844
128. Zalnieriute M, Moses LB and Williams G (2019) The rule of law and automation of government decision-making. *The Modern Law Review* 82:425-455.
129. Zavrsnik A (2019) Algorithmic justice: Algorithms and big data in criminal justice settings, *European Journal of Criminology*. Epub ahead of print Sept 2019. doi:10.1177/1477370819876762
130. Zavrsnik A (2020) Criminal justice, artificial intelligence systems, and human rights. *ERA Academy of European Law Forum* 20:567-583.
131. Zemel R, Wu Y et al. (2013) Learning Fair Representations. *Proceedings of the 30th International Conference on Machine Learning*. Atlanta, Georgia USA. 28:1-9.
132. Zuboff S (2015) Big other: Surveillance capitalism and the prospect of an information civilization. *Journal of Information Technology* 30: 75-89.

DOI: 10.1177/02683962211010356