

Algorithmic Pollution: Understanding and Responding to Negative Consequences of Algorithmic Decision-Making

Olivera Marjanovic, Dubravka Cecez-Kecmanovic, Richard Vidgen

► **To cite this version:**

Olivera Marjanovic, Dubravka Cecez-Kecmanovic, Richard Vidgen. Algorithmic Pollution: Understanding and Responding to Negative Consequences of Algorithmic Decision-Making. Working Conference on Information Systems and Organizations (IS&O), Dec 2018, San Francisco, CA, United States. pp.31-47, 10.1007/978-3-030-04091-8_4. hal-02083582

HAL Id: hal-02083582

<https://hal.inria.fr/hal-02083582>

Submitted on 29 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Algorithmic Pollution: Understanding and Responding to Negative Consequences of Algorithmic Decision-making

Olivera Marjanovic¹ Dubravka Cecez-Kecmanovic² Richard Vidgen²

¹ University of Technology Sydney, Sydney, Australia

² University of New South Wales, Sydney Australia
olivera.marjanovic@uts.edu.au

Abstract. In this paper we explore the unintended negative social consequences of algorithmic decision-making, which we define as “algorithmic pollution”. By drawing parallels with environmental pollution, we demonstrate that algorithmic pollution is already here and causing many damaging, unrecognised and yet-to-be understood consequences for individuals, communities and a wider society. Focusing on transformative services (i.e., services that transform human lives, such as social support services, healthcare, and education), we offer an innovative way of framing, exploring and theorizing algorithmic pollution in the contemporary digital environment. Using sociomateriality as a theoretical lens, we explain how this type of pollution is performed, how it is spreading and who is responsible for it. The proposed approach enables us to articulate a preliminary set of IS research challenges of particular importance to the IS community related to living with and responding to algorithmic pollution, together with an urgent call for action. Our main practical contribution comes from the parallels we draw between the environmental protection movement and the newly created socio-material environment that needs protecting from the spread of algorithmic pollution.

Keywords: Algorithmic Pollution, Sociomaterial Environment, Negative Consequences, Algorithmic Justice.

1 Introduction

Algorithmic decision-making is playing an increasing role in our private lives, in business, and in social worlds. Indeed, algorithms are now used to make “big decisions about people’s lives” from job applications, job performance, provision/denial of social services, insurance, medical benefits and financial services [1]. While we acknowledge the various reported benefits of algorithms, for example in medicine, science, agriculture [2-5], in this paper we focus on the growing evidence of various harmful effects of algorithmic decision-making for society and for individuals working and living in a contemporary digital data environment [6].

We are particularly concerned with the negative consequences of automated algorithmic decision-making (i.e., decision-making with no human intervention) in the so-

called transformative services. These are services that transform human lives, such as social support services, education, healthcare and aged care [7-9]. The consequences of these algorithms are far-reaching, unpredictable and potentially hurtful to individuals, their families, community groups and society at large [1, 10, 11]. Moreover, these consequences are propagated and amplified by the widespread and often invisible processes of datafication, where individual's data are harvested, consolidated and constructed through often unknown, unregulated and un-auditable processes, and perpetually re-constructed and reused in future unpredictable contexts [12-15]. Consequently, these negative consequences are extending well beyond the original service provision. Yet, they are hard to detect and to prove unjustified or unjust, and even harder to completely reverse.

In this paper we contribute to a growing body of knowledge on the negative consequences of automated algorithmic decision-making in transformative services by providing an innovative way of framing these consequences as “algorithmic pollution”. By using a well-developed concept of pollution (or environmental pollution), we conceptualize these harmful consequences as a new type of pollution, one that pollutes the social environment where we work and live. Compared to environmental pollution, which is recognized and regulated, we demonstrate that algorithmic pollution is unrecognized, unregulated, and rapidly spreading – yet it is masked by myths of objectivity, neutrality, efficiency and superiority of algorithms over human decision-makers.

The purpose of this paper is two-fold: (i) to articulate and bring our collective attention to a new phenomenon here-termed “algorithmic pollution” and (ii) to propose a sociomaterial theoretical grounding for its exploration and collective action. Based on a literature review we identify the widespread and widening consequences of algorithmic decision-making on individuals (e.g., citizens, customers, patients, employees), organizations and a wider society. Of particular interest are the unintended negative consequences of automated algorithmic decision-making, which require the urgent attention of researchers and social action. Specifically, we aim to:

1. Define “algorithmic pollution” by building upon the concepts of “environmental pollution” and demonstrate that it is appropriate for framing negative unintended consequences of algorithmic decision-making;
2. Propose a sociomaterial theorization of algorithmic pollution in order to explain how this type of pollution is performed, how it is spreading and who is responsible for it;
3. Identify and articulate new research challenges for the IS community related to living with and responding to “algorithmic pollution”.

Our research makes several research and practical contributions. First, we articulate a new concept of algorithmic pollution, thus expanding the existing bodies of knowledge in algorithmic decision-making as well as in environmental pollution. Second, we offer strong evidence, collected from multidisciplinary literature, why algorithmic pollution is a growing problem that is of concern to the IS discipline, requiring urgent attention. Third, we propose a preliminary approach to studying algorithmic pollution which discloses the hidden and uncertain **performing** of algorithms in socio-

material environments that would enable researchers to examine the nature of algorithmic pollution and how the damage is done. Fourth, we identify and articulate a set of IS research challenges that call for our urgent attention.

Our main practical contribution comes from the parallels we draw between environmental protection and the sociomaterial environment that needs protecting from the spreading of algorithmic pollution. This line of thinking opens up future opportunities for building upon lessons from the environmental protection movement [16], including environmental justice and impact assessment frameworks. We argue that they need to be expanded to include the sociomaterial environment and algorithmic pollution.

The paper is organized as follows. In the next section we describe our research focus, which is at the intersect of transformative services, algorithms and datafication. Using the related literature from different disciplines, we then provide evidence of negative consequences of algorithmic decision-making in transformative services. Having set the context for our research, we then introduce the concept of pollution and use it to frame the negative consequences of algorithmic decision-making as “algorithmic pollution”. We then proceed to identify different mechanisms that contribute to the generation and spreading of algorithmic pollution. By seeing algorithms as actors in complex and emerging sociomaterial assemblages, we then propose a preliminary approach to studying how algorithmic pollution is performed in the first instance and how and why it is spreading. Based on the proposed approach, we identify a set of IS research challenges and discuss a possible way forward.

2 Research focus: algorithmic decision-making in transformative services

Our research into algorithmic decision-making is situated at the intersection of transformative services, algorithms and datafication (Figure 1). We are particularly concerned with algorithmic decision-making in the context of transformative services where machines make important decisions about people but with little or no human judgement or intervention. This section sets the scope and foundations for our research by introducing the key concepts, as follows.

Transformative services

Compared to transactional services (e.g., in the retail sector such as online commerce), transformative services are those services that transform human lives by having a direct impact on the well-being of individuals, communities and the wider ecosystem [8, 9]. Prominent examples of transformative services include healthcare, social support services, education, financial services and law enforcement [9]. Compared to other service-related research, the emerging multidisciplinary field of transformative services research (TSR) brings the impact of service outcomes on human well-being to the forefront [8]. Indeed, “Improving well-being through transformative services” was identified as one of the top research priorities of the emerging multidisciplinary field of Services Sciences [7].

The key aspects of human well-being such as discrimination, marginalisation, and disparity in adverse conditions in population groups, as well as the issues of inclusion and access to services [8] are of particular interest to TSR. These are the same aspects that are now impacted by algorithmic decision-making. Hence, our motivation to situate this research in the context of transformative services.

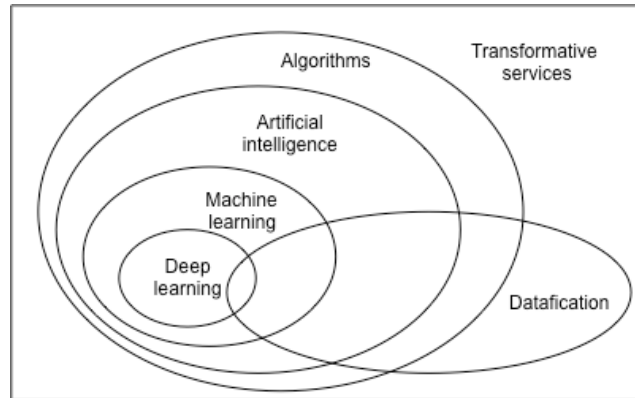


Fig. 1. The algorithmic decision-making research setting (drawing on Chollet [17])

Algorithms

An algorithm is often understood to be “a set of instructions for how a computer should accomplish a particular task” [1, p.1]. The Computer Science community, including widely used textbooks, often quote Kowalski’s [18] definition: Algorithm = Logic + Control (e.g., Kitchin [19]). Here “Logic” specifies what is to be done to solve a particular well-defined problem, while “Control” describes how the logic should be implemented under different scenarios. Thus, in Figure 1, the general class of algorithms is shown in the outer ellipse.

Within the general class of algorithms, we focus on a particular subclass – artificial intelligence. While terms are often used interchangeably, we follow Chollet [17] to differentiate between artificial intelligence (AI), machine learning, and deep learning. AI has roots in the 1950s enthusiasm for building a machine that could think. Typically, the approach was to hard-code lots of rules, which worked well for well-defined problems, such as chess, but less well with the sorts of ill-defined problems that humans are good at – such as recognizing images and speech and understanding and engaging in argumentation. Traditional approaches (also known as symbolic AI) take rules plus data as input and provide answers as the output (for example, some predictive policing applications may use this approach).

Machine learning turns this on its head and works by taking data and answers as the inputs with the output being rules (using neural networks). Thus, a machine learning algorithm is trained through examining lots of examples in which it detects patterns. Machine learning is empirical and is more engineering-oriented than statistical, i.e., the inner workings may be black-boxed and the test of a ‘good’ algorithm is in its predictive

performance rather than its theoretical correctness. While machine learning is an undoubtedly powerful technique, it is not without dangers, as evidenced by the failure of Google flu, which failed as a result of model over-fitting, relying on unreliable data, and failing to update the model to take account of underlying changes [20]. Deep learning is a subclass of machine learning in which multiple layers of data transformations are made to give an increasingly meaningful representation of the data. Deep learning has been successfully applied, despite only becoming a prominent technique in the early 2010s, in notoriously difficult domains such as image and speech classification, digital assistants (e.g., Google Now and Alexa), autonomous driving, and natural language processing [17].

Explaining how a deep learning application comes to the conclusions it does is certainly an issue for traceability and justification in decision-making [21]. However, Rahimi, an AI researcher at Google, argues that the problem is larger, that machine learning is a form of alchemy in which researchers do not know why some algorithms work and others don't and have no rigorous criteria for choosing one AI architecture over another. Rahimi distinguishes between a machine learning application that is a black box and "an entire field that has become a black box" (Rahimi quoted by Hutson [22]). The use of AI in general (and deep learning in particular) poses significant issues for understanding decision-making.

Datafication

Algorithms work based on data and also produce data. However, rather than focusing solely on data, we expand our focus to *datafication*, a process of turning everything and everyone into data [2, 6]. While datafication may have positive effects, such as new types of jobs and more value for customers [23], there is growing evidence of its negative and unintended social consequences [6, 12-14]. This is because datafication will "unavoidably omit many features of the world, distort others and potentially add features that are not apparent in the first instance" [24, p.384]. Moreover, datafied phenomena are further propagated, processed, reused and distorted [14], having the performative power to re-shape the organizational and social worlds in unpredictable and undesirable ways [25].

Datafication is at the core of algorithmic decision-making as algorithms are applied to 'datafied individuals' (i.e., individuals represented by an always limited set of attributes and the corresponding data values). Additional data about individuals may be acquired from third parties who applied their own datafication processes in unknown ways and contexts. Algorithms also contribute to further datafication of individuals as their outcomes create new data also attached to these individuals (e.g., customer scoring or customer rating). These customer scores may be combined with more data and fed into other algorithms further down the "datafication chain".

When used in transformative services, algorithmic decision-making without human intervention, combined with ongoing datafication processes, often results in negative consequences for human well-being, as discussed in the next section.

3 The consequences of algorithmic decision-making in transformative services

Algorithmic decision-making is making its way into transformative services at “a breathtaking pace”, often in the name of innovation and progress [10, p.11]. Consequently, “algorithms driven by vast troves of data, are the new power brokers in society [26, p.2] that already “have control of your money market funds, ... your retirement accounts, ... they will decide your chances of getting lifesaving organs” [5, p.214].

However, in spite of the reported enthusiasm, there is growing evidence of unfair, unjustified and discriminatory effects of these algorithms for individuals and wider communities [1, 11, 27]. Eubanks [10] offers a vivid example of devastating effects of automated eligibility systems implemented in a particular type of transformative services (social support services):

“Across the country, poor and working-class people are targeted by new tools of digital poverty management and face life-threatening consequences as a result. Automated eligibility systems discourage them from claiming public resources that they need to survive and thrive... Predictive models and algorithms tag them as risky investments and problematic parents. Vast complexes of social services, law enforcement, and neighborhood surveillance make their every move visible and offer up their behavior for government, commercial, and public scrutiny” [10, p.11].

Law enforcement is another prominent example of transformative services using algorithmic decision-making, most notably in predictive policing. Celebrated as the new era of “data-driven scientific decision-making”, predictive policing has been increasingly used to inform decisions such as arresting people or determining the length of their sentence, based on the calculated probability of them committing future crimes [28]. This practice relies on datafication of individuals, including a variety of personal as well as other data that individuals may not have any control over, such as past data on “gang districts” or post codes. Consequently, individuals are assigned “risk scores” that are subsequently used for decision-making in a variety of contexts. The practice is spreading with more and more departments embracing the scientific approach to policing. For example, Ferguson [28] reveals that almost 400,000 Chicago citizens now have an official police risk score:

“This 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” [27, p.1].

In addition to citizens not being aware of their risk scores, Ferguson [28] warns that “[c]urrently there is no public oversight of the police data, inputs or outputs, so communities are left in the dark unable to audit or challenge any individual threat score”

(p.1). Consequently, the “profound benefits” of predictive policing continue to be celebrated in the popular press. For example, in a controversial Financial Times article Gilian Tett (a respected financial reporter) offers the following support for predictive policing: “After all, ...the algorithms in themselves are neutral” [29, p.1].

Similar examples are reported in other types of transformative services including education, healthcare, social support and disability services (see for example, [1, 10, 11, 27, 30]). In particular, Caplan and colleagues [1] warn, “there is a need for greater reflection on models of power and control, where the sublimation of human decision-making to algorithms erodes trust in experts” (p.6). There is a growing number of critical studies of algorithmic decision-making emerging across a range of disciplines including sociology, public policy, communications and media studies, political studies, and increasingly information systems (see for example, [6, 12, 13]). We contribute to the existing body of knowledge by proposing a novel approach to framing the problem of negative consequences of algorithmic decision-making as *algorithmic pollution*.

4 Algorithmic pollution

Having outlined some of the negative consequences of automated algorithmic decision-making in transformative services we now frame the problem situation using the lens of pollution. The pollution concept is suitable since it allows us to see algorithms as a force for good while recognizing that there are both intended and unintended negative consequences that affect individuals, communities and society. It would be hard to argue that electricity is not a good thing, but if its generation causes pollution in the form of global warming then remedial action and regulation are needed. We note that using the concept of pollution as a framing device is not new: for example, there has been research defining crime as pollution (CAP) [31].

4.1 Defining algorithmic pollution

According to a broad definition “Environmental pollution is the discharge of material, in any physical state, that is dangerous to the environment or human health” [32]. (Encyclopedia.com). Similarly, the Encyclopedia Britannica defines environmental pollution as: “the addition of any substance (solid, liquid, or gas) or any form of energy (such as heat, sound, or radioactivity) to the environment at a rate faster than it can be dispersed, diluted, decomposed, recycled, or stored in some harmless form” [33].

When algorithms produce negative consequences on individual or collective well-being, and when such consequences cannot be effectively detected and eliminated or dealt with, we argue that they ‘contaminate’ our sociomaterial environment. We call this phenomenon **algorithmic pollution** and define it as follows:

Algorithmic pollution denotes the presence of unjustified, unfair, discriminatory, or other harmful consequences of algorithmic decision-making for individuals, groups, organizations, sections of the population, the economy, or society at large.

While our definition of algorithmic pollution accords with the view of pollution as harms suffered by living organisms caused by exposure to pollutants [31], the scientific definition of pollution is concerned with the presence of chemicals in the environment at a concentration above their normal background level that perturb the environment in a way that is harmful [34]. Harrison [34] also points out that not all instances of pollution involve the addition of chemicals to the environment; pollution can also be caused by adding to naturally occurring phenomena, such as light and noise. The scientific definition suggests that there is some background level of pollution that is acceptable. Since algorithms are probabilistic then some error in the form of false positives and false negatives is unavoidable, i.e., there will always be some level of background algorithmic pollution. The challenge is to detect when the outcomes of algorithmic decision-making exceed a threshold and result in systematic injustice (individual decisions) and larger, systemic social injustice (emergent societal effects). While there must inevitably be a substantial element of subjectivity in this judgment for algorithms it is also present with environmental pollution where acceptable limits (e.g., parts per million) have to be established to calibrate a ‘normal’ background level.

Drawing on Lynch and colleagues [31], we note that the pollution literature distinguishes between an ‘end-pipe’ (e.g., a factory chimney) and the process that generates the pollution (e.g., a manufacturing process). Further, end-pipes are a stationary source of pollution, while pollution itself is mobile and it is not always possible to know the source of pollution from monitoring the environment. Pollution can be generated as point source pollution (PSP) and nonpoint source (NPSP). While a factory chimney is a PSP, motor vehicles, which pollute the environment by generating particulates and nitrogen dioxide while in standing traffic, are an example of NPSP.

4.2 Generation and spreading of algorithmic pollution

For algorithmic pollution to occur there must be data (input) to feed the algorithms and the algorithms must lead to decisions (output). Algorithms only have the capability to produce pollution when they are used to automate decisions based on datafied individuals, i.e., data, algorithms, and decisions are always present and implicated in algorithmic pollution (even when this is not readily apparent). Using this production metaphor and concepts from the environmental pollution literature, we identify necessary elements that work together in the case of AI and machine-learning algorithms:

Datafication: algorithms are constructed using ‘datafied’ individuals i.e., individuals represented by limited number of attributes that have been chosen as relevant. As they are never complete, such a datafication practice is bound to create the so-called “representational” harm [35]. This type of harm is further intensified with organisations (including governments) increasingly acquiring data from “data scorers and brokers” with individuals already datafied (e.g., a person’s score) through some unknown processes [15, 27].

Production: in the case of machine learning, algorithms then learn from the data produced as a result of datafication to create rules. However, using past data to predict future behaviour is sometimes out-dated or inappropriate. For example, data collected

on historical “gang districts” are now used by police for predictive policing even though these districts may be no longer representative [28]. As O’Neil [11] 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” (p.1). Therefore, an algorithm might be designed in such a way that it discriminates against certain individuals, or it might learn from data to discriminate.

End-pipes: algorithms are embedded in processes in transformative services of all types in order to make decisions with little or no human intervention. In the language of environmental pollution, this is the ‘end-pipe’ of algorithmic pollution, i.e., the point where decisions are made and enacted. Some of these decisions may result from point stationary pollution (PSP), where the decisions are generated by an identifiable organization, process, and algorithm. Other decisions are likely to have characteristics of non-stationary pollution (NPSP), being generated by multiple algorithm sources working collectively (for example decision generated by networks of inter-communicating algorithms).

Consequences: algorithmic decision-making in the transformative services has a direct impact on individuals, for example when a person is approved or declined for a loan, a medical operation, or parole. Algorithmic pollution is further propagated and amplified by highly interconnected systems of algorithms (NPSP). How these individual algorithms interact is not only invisible to those affected, but often to those who constructed them and deployed them [19, 36]. For example, a person with a poor credit card history is very likely to have difficulties finding employment; without a job they are very unlikely to repay their debt, which in turn will further impact on their credit card history [27]. Teachers being scored by algorithms as non-performing will face difficulties finding future employment, which in turn will limit their ability to improve their score [11]. In environmental pollution this is unlikely to be the case as, for example, radiation and water pollution do not augment each other. Algorithmic pollution, however, emerges from complex interactions that cannot be predicted and possibly cannot be traced back to root decisions, i.e., a cause and effect logic may be insufficient when addressing complex algorithmic pollution.

Feedback loops: the algorithmic decision-making outcomes themselves lead to the generation of further data about individuals that can be harvested by machines and fed back into the algorithm building process. Not only can humans be removed from the decision-making process; they can also be removed from the algorithm building process as machines learn from data that was generated by machine-made decisions in the first place. Without some form of intervention these feedback loops may result in self-perpetuating vicious cycles, which may, ultimately, lead to deep rifts in the fabric of society. These feedback loops are thus unique to algorithmic pollution in that the consequences can affect the means of production in a direct and automated manner through datafication and machine learning, aided and amplified by data brokers and data scorers [27]. The end result is pollution itself becoming a pollutant, producing new forms of pollutions. This is a new digital phenomenon, not present in environmental pollution.

4.3 Implications

Algorithmic pollution is not only spreading but also intensifying. While in other cases of pollution humans and/or instruments are capable of detecting pollution (e.g., through sight, smell, or by radiation reading), algorithmic pollution remains hidden. Indeed, algorithmic effects are hidden and as such very hard to detect and prove. A customer of a health insurance company offers a vivid example:

“The insurance company repeatedly told me that the problem was the result of a technical error, a few missing digits in a database. But that’s the thing about being targeted by an algorithm: you get a sense of a pattern in the digital noise, an electronic eye turned towards you, but you can’t put your finger on exactly what’s amiss” [10, p.5].

Moreover, it is very difficult (if possible at all) for an individual to fight the effects of algorithms. In fact, “bad inferences” about people are fast becoming “a larger problem than bad data because companies can represent them as “opinions” rather than fact. A lie can be litigated, but an opinion is much harder to prove false” [27, p.32]. To make matters worse, algorithmic decision-making continues to be celebrated as superior to human-judgment [10, 11, 37]. The cult of “science” continues as “Technocrats and managers cloak contestable value judgements in the grab of “science” [27, p10]. Inevitably, and often unknowingly, they also contribute to the creation and propagation of pollution.

As algorithms now proliferate into various sectors of the economy and society they fundamentally reconfigure power relations and transform the ways sectors operate, without public awareness or broader understanding of their consequences. Given the potentially complex outcomes of algorithmic pollution, focusing on and opening up the ‘black box’ of how algorithms work (the production stage), is unlikely to be a successful strategy for algorithmic regulation. The lack of transparency, the hidden acting of algorithms, and the presence of feedback make the study of algorithmic pollution particularly challenging.

5 Theorizing algorithmic pollution

Contemporary forms of knowing, in particular those advanced by computer and communication technology research, artificial intelligence, data analytics and data science, are focused on the translation of problems (business, market, political, scientific) into codified knowledge and ultimately codes of algorithms. A considerable body of literature thus focuses on translations: first, the translation of tasks (or problem solutions) into formal statements or instructions (pseudo-code); and second, the translation of the pseudo-code into a source code [19, p. 17]. The first translation is particularly critical as it essentially codifies existing knowledge about a problem at hand in such a way that all possible conditions relevant for solving it (e.g., variables in a decision-making model) and for generating decisions are taken into account. Algorithms that have errors

in the codified knowledge (logic or control or both) produce wrong or problematic outcomes [38] and can thus be identified as and shown to be pollutants.

However, algorithmic pollution occurs more often due to negative unintended consequences of carefully designed and seemingly correct algorithms. As examples above illustrate, the execution of algorithms involves mobilization and use of diverse data sources (e.g., shopping history data, prescription drugs, medical and wellness data, social media records, search engine logs/history, voting preferences, credit card transactions and many more, often sold by data brokers who consolidate data per individuals). These data are prone to errors, are uncertain, and contain a significant amount of noise [15]. Based on such data, algorithms calculate scores, make decisions or produce other outcomes (risk scores for individuals, loan or insurance approvals/declines, short listing of job applicants) that become effective in concrete sociomaterial practices: individuals are targeted as criminal suspects based on risk scores; bank clients are declined loans; applicants are not short listed. Such effects can be damaging while typically not being justified [10, 27].

The consequences of the algorithmic execution thus depend to a large extent on numerous heterogeneous actors enrolled in complex and uncertain sociomaterial assemblages. These consequences are rarely predictable and detectable at the design and coding stage. While understanding the reasoning embedded in the code of an algorithm and how outcomes (decisions) are calculated and derived from inputs, is important and necessary, it is far from being sufficient to comprehend what an algorithm is actually doing, on what grounds its decisions are produced and whether they are correct, fair, just, and justified. As algorithms become actors who “form a complex and at various times interpenetrating sociomaterial assemblage[s] that [are] diffused, distributed, fragmented, and, most importantly, “liquid” [39, p. 18-19]. Their effects are produced by the sociomaterial assemblages. To understand algorithmic pollution, we argue, requires empirical attention to the workings of individual algorithms or systems of algorithms and the examination *within* their complex, diffused, fragmented, liquid and often interpenetrating sociomaterial assemblages [40]. This proposition opens many conceptual and methodological questions.

The notion of an algorithm as “Logic + Control” [18] or a set of instructions for completing a particular set of tasks [1], implies a self-sufficient non-human object with specified functions with defined inputs, a complete set of conditions and action possibilities (outcomes). When its code (a set of instructions) is executed an algorithm becomes the actor. In other words, the execution of code constitutes an algorithm as an actor. Code execution is a particular materialization of relations that form sociomaterial assemblages (mobilize and enact heterogeneous actants – servers, various data sets, other algorithms, things and human actors as objects of knowledge). It is through these relations that particular outcomes (decisions, recommendations) are achieved (credit scores calculated; crime suspect regions demarcated; terrorists predicted). How to study these enfolding relations becomes the key methodological issue, an issue that is yet further complicated by the ability of machines to learn from data without human input and, due to the use of deep learning, may be unable to fully account for the “Logic” that is being used to automate decisions [18,19].

Through repeated execution algorithm codes keep performing their objects of knowledge and thus enact different entities (things and people) into being. They perform new distinctions and new categories of customers/clients/citizens often based on unchecked and error prone data, from dubious and unreliable sources [15]. Mackenzie and Vurdubakis [41] similarly note:

“the conduct of code, we might say, its execution, is a fraught event, and analysis should be brought to bear on the conditions and practices that allow code to, as it were, access conduct. While the knowledge or form that lies at the heart of the code promises completeness and decidability, the execution of code is often mired in ambiguity, undecidability and incompleteness” (p. 7).

Importantly, “ambiguity, undecidability and incompleteness” are kept hidden, buried in the complexity of code execution and largely unrecognized. Even the designers and users of algorithms (especially those with learning capabilities) do not understand the intricacies of code execution and cannot explain the resulting outcomes which are nevertheless actioned undoubted and undisputed.

This brief discussion suggests that to understand the actual doings of algorithms we need first to examine and dig deeper into the relational unfolding of code execution (a digital life of code) and the emergence of sociomaterial assemblages (the intra-acting in Barad’s [42] sense), including the mobilization of numerous actors (data sets, companies, internet, data analytics technologies) and the *performing* of subjects and objects. The relational unfolding of a code however is hidden behind the interface and the traces of code execution are typically not provided.

When the algorithmic outcomes become enacted they continue the ‘doing’ by being enrolled in sociomaterial practices of the users (for example, in a police department or in a bank loan approval process). Entangled within these practices algorithmic outcomes have a life of their own, reconfiguring users’ action domains while becoming constitutive of subjects (suspect citizens or risky clients) and objects (high criminality regions). Algorithmic outcomes thus perform what Barad [42] calls agential cuts, making particular subjects and objects in the image of their calculation. For instance, the subjects (citizens, clients, job applicants) become performed as particular calculated figures (e.g., citizens equated with their risk scores), the power of which comes from an unquestioned authority of algorithms [43].

These performative effects of algorithms are not only taken for granted, they are celebrated as unbiased, objective and thus fair [11]. That they are based on unchecked and uncertain data sets, often breaching basic human rights, and thus unjustified, unethical, and potentially illegal, is conveniently hidden and kept far from the public eye. Algorithmic pollution, we might tentatively conclude, is carefully covered up, systematically hidden, and deceptively dressed up as technological progress.

These initial ideas, conceptualizations and methodological issues only scratch the surface of unprecedented methodological complexity of investigating algorithmic pollution. Uncovering it, revealing how algorithms pollute various sociomaterial environments and what injustices and damages are done to people and communities, will be an uphill battle for anybody who dares to research and report on it.

In this battle we can learn from ANT scholars who have studied similar reconfigurations of sociomaterial practices and the performing of subjects and objects in a variety of contexts (see for example, [44-47]). What is however different and new in algorithmic doing and performing is the particular discursive-calculative-digital nature of algorithmic outcomes and how they come into being in code execution. For this reason, ANT or other field methodologies would need to be adapted or reinvented to be delicately attuned to diffused, fragmented, uncertain and hidden sociomaterial practices of algorithmic doings and reality making.

6 A research agenda for algorithmic pollution

As we have shown, in addition to investigating the design of algorithms (that has been extensively studied in the literature, see [18, 43], researching and understanding algorithmic pollution requires empirical examination of interrelated sociomaterial practices of two key aspects: algorithm deployments and ongoing datafication processes. Both aspects offer new research challenges, as discussed in this section.

Algorithm deployment includes on the one hand, coding, execution of code and emergence of a sociomaterial assemblage that produce algorithmic pollution, and on the other, the enactment of algorithmic outcomes (decisions, recommendations) that reconfigures users' practices and performs subjects and objects. The need to better understand how these sociomaterial practices of algorithm deployment contribute to algorithmic pollution, leads us to a number of research questions.

For example, what is the relationship between coding practices and algorithmic pollution? We envisage the design of new frameworks and methods that could be used to guide developers to bring to the surface potential sources of algorithmic pollution, including datafication practices. An important step in this direction is the IEEE Global Initiative on Ethics of Autonomous and Intelligence Systems [48]. We envisage the challenge of translating the proposed generic principles into practical approaches in a particular context, taking into account the complexities of transformative services. These yet-to-be-discovered practices, could include for example moral imagining [49].

The challenge of making visible the process of emergence of a sociomaterial assemblage that produces algorithmic pollution opens yet more research questions. For example: How might we effectively recognize, report and mitigate the effects of algorithmic pollution throughout society? Who should be doing it? How might we recognise and deal with different types of polluters? How can we educate governments, organisations and society at large to look beyond the current hype of algorithmic neutrality, efficiency and superiority and comprehend the urgency of dealing with algorithmic pollution?

In framing a response to algorithmic pollution, we propose to turn our attention and learn from the established field of environmental justice, which is defined as follows:

“Environmental justice is a social movement seeking to address the inequitable distribution of environmental hazards among the poor and minorities. ... From a

policy perspective, practicing environmental justice entails ensuring that all citizens receive from the government the same degree of protection from environmental hazards and that minority and underprivileged populations do not face inequitable environmental burdens.” [50, p.1].

We propose that algorithmic pollution is added to the environmental dimensions already identified in the environmental justice movement, as *algorithmic justice*. As previously illustrated, there is already strong evidence that the hazards of algorithms are distributed inequitably. As Eubanks [10] explains, these new algorithmic systems for “automating inequality” have the most destructive and deadly effects on the poor and the underprivileged. Even more, “[t]he widespread use of these systems impacts the quality of democracy for us all” [10, p.12].

The environmental justice agenda is reflected in the Toronto Declaration [51], which is calling for governments and companies to ensure that algorithms respect basic principles of equality and non-discrimination. Yet, as our paper demonstrates, this is going to be very difficult to implement in practice, as today’s algorithms are so “deeply woven into the fabric of social life that, most of the time, we don’t even notice we are being watched and analysed” [10, p.5].

A possible way forward could be found by simultaneously looking forward at the emerging developments in AI and algorithmic decision-making and looking back at the history of the environmental movement. For example, a possible way of providing more visibility into the inner working of algorithms would be to store the trace of automated algorithmic decisions using a blockchain, as suggested by Schmelzer [52]. This has the benefit of the decision audit trail being stored in a way that can be shared, cannot be tampered with, and is not owned by the algorithm producer or deployer. A similar idea might be applied to the problem of spreading pollution by creating an audit trail for the ongoing datafication of individuals in transformative services.

By considering the history of environmental protection, we could also learn, for example, about the ways in which traditional pollution has been addressed. We could then expand or redevelop the existing frameworks and methods, such as Environmental Impact Assessment (EIA) [53] and market-based controls [31], to include algorithmic pollution. Finally, by researching the inner working of government environmental protection agencies we may identify new opportunities for policy development and possible establishment of similar agencies for algorithmic justice.

7 Conclusion and an urgent call for action

In his work titled “Love Your Monsters”, Latour turns his attention to technology “to protect the planet from ecological crisis” [54, p1.]. We argue that a new type of crisis is already here, caused by a new type of technologically-induced pollution that we identify and name “algorithmic pollution”.

Inspired by the observed parallels with environmental pollution, in this paper we articulate a new type of widespread, hidden, largely unregulated and evidently harmful “algorithmic pollution”. Focusing on the transformative services, we offer evidence

collected from multidisciplinary literature why this pollution is a growing problem that requires our urgent attention. We also offer a preliminary approach to studying algorithmic pollution that discloses the hidden and uncertain performing of algorithms in sociomaterial environments, which would enable researchers to examine the nature of algorithmic pollution and how the damage is done. This enables us to identify and articulate a set of IS research challenges that call for our urgent attention. By drawing parallels between environmental protection and the need to protect the observed socio-material environment that is now affected by algorithmic pollution, we open up new opportunities for a practical contribution.

We recognize that algorithms undoubtedly have the potential to provide society with significant benefits (e.g., healthcare, driverless vehicles, fraud detection). Therefore, this paper is not a treatise against algorithms. Far from it. It is however explicitly and consciously against algorithmic pollution. Recalling the words of the poet Ella Wheeler Wilcox – ‘to sin by silence, when we should protest’ – we, IS researchers, should raise our voices and enact our professional responsibility.

Building upon the fundamental principle of environmental justice that “all people deserve to live in a clean and safe environment free from industrial waste and pollution that can adversely affect their wellbeing” [50, p.1], we conclude this paper with a claim that all people equally deserve to live in an environment free and safe from algorithmic pollution. If algorithms are our future, as many claim, then understanding, fighting against and preventing algorithmic pollution, may save our collective dignity and humanity.

8 References

1. Caplan, R., Donovan, J., Hanson, L. and Matthews, J.: Algorithmic accountability: A primer. Prepared for the Congressional Progressive Caucus. April 18. 2018. Data & Society, Washington, DC. USA. https://datasociety.net/wp-content/uploads/2018/04/Data_Society_Algorithmic_Accountability_Primer_FINAL-4.pdf, last accessed 2018/08/21.
2. Cukier, K., Mayer-Schonberger, V.: *Big data: A revolution that will transform how we live, work and think*. Houghton Mifflin Harcourt, Boston. USA (2013).
3. Davenport, T.: *Big data @ work: Dispelling the myths, uncovering the opportunities*. Harvard Business Review Press, London (2014).
4. Perkins, A.: May to pledge millions to AI research assisting early cancer diagnosis. *The Guardian* May 21, 1-2 (2018).
5. Steiner, C.: *Automate this: How algorithms came to rule our world*, Penguin, NY, USA (2012).
6. Newell, S., Mirabelli, M.: 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 (2015).
7. Ostrom, A.L., Parasuraman, A., Bowen D.E. et al.: Service research priorities in a rapidly changing context”, *Journal of service research* 18(2), 127-159 (2015).
8. Anderson, L., Ostrom, A.L.: Transformative Service Research – Advancing Our Knowledge About Service and Well-Being. *Journal of Service Research* 18(3), 243-249 (2015).
9. Anderson, L., Ostrom, A. L., Corus, C. et al.: Transformative service research: An agenda for the future, *Journal of Business Research* 66(8), 1203-1210 (2013).

10. Eubanks, V.: Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press, New York (2017).
11. O'Neil, C.: Weapons of math destruction: How big data increases inequality and threatens democracy. Penguin Random House. NY. USA (2016).
12. Galliers, R.D., Newell, S., Shanks, G., Topi, H.: 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 (2017).
13. Markus, L.: Datafication, organizational strategy, and IS research: What's the score?," *Journal of strategic information systems* 26(3), 233-241 (2017).
14. Marjanovic, O. and Cecez-Kecmanovic, D.: 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 (2017).
15. Clarke R.: Risks inherent in the digital surveillance economy: A research agenda, *Journal of information technology* (forthcoming) (2018).
16. EPA: EPA – United States Environmental Protection Agency, <https://www.epa.gov/laws-regulations/summary-national-environmental-policy-act>, last accessed 2018/08/21.
17. Chollet, F.: Deep Learning with R. Manning Publications (2018).
18. Kowalski, R.: Algorithm = Logic + Control. *Communications of the ACM* 22(7), 424-436 (1979).
19. Kitchin, R.: Thinking critically about and researching algorithms, *Information, Communication & Society* 20 (1), 14–29 (2017).
20. Lazer, D, Kennedy, R, King, G., Vespignani, A.: The parable of Google flu: Traps in big data analysis, *Science* 343(6176): 1203–1205 (2014).
21. Voosen, P.: The AI detectives: as neural nets push into science, researchers probe back. *Science* 357(6346): 22-27 (2017).
22. Hutson, M.: AI researchers allege that machine learning is alchemy. *Science*. 3 May 2018 <http://www.sciencemag.org/news/2018/05/ai-researchers-allege-machine-learning-alchemy>, last accessed 2018/05/05.
23. Loebbecke, C., Picot, A.: 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 (2015).
24. Lycett, M.: Editorial: 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems* 22(4), 381-386 (2013).
25. Gitelman, L.: "Raw data" is an oxymoron. MIT Press (2013).
26. Daikopoulos, N.: Algorithmic accountability reporting: On the investigation of black boxes. A Tow/Knight brief. Tow Center for Digital Journalism, Columbia Journalism School. <http://towcenter.org/algorithmic-accountability-2/>, last accessed 2018/08/21.
27. Pasquale, F.: *The Black box society: The secret algorithms that control money and information*, Harvard University Press, Cambridge, Massachusetts (2015).
28. Ferguson, A. G.: The police are using computer algorithms to tell if you're a treat, *Time*, Oct. 3, 2017, 1-2. (2017).
29. Tett, G.: Mapping crime – or stirring hate?," *Financial Times*, Opinion, Aug (2014).
30. Terhune, C.: They know what's in your medicine cabinet, *Business Week*, July (2008).
31. Lynch, M., Kimberly L., Stretesky, P., Long, M., Jarrell, M., Ozymy, J.: Crime as pollution? Theoretical, definitional and policy concerns with conceptualizing crime as pollution. *American Journal of Criminal Justice* 40, 843–860 (2015).
32. Online Encyclopaedia, www.encyclopedia.com, last accessed 2018/08/21.
33. Encyclopedia Britannica, www.britannica.com/science/pollution-environment, last accessed 2018/08/21.

34. Harrison, R. M.: (Ed.). An introduction to pollution science. London: Royal Society of Chemistry (2006).
35. Reisman, D., Schultz, J., Carwford, K. and Whittaker, M.: Algorithmic impact assessments: A practical framework for public agency accountability, *AI Now*. (2018).
36. Seaver, N.: *Knowing Algorithms*. *Media in Transition* 8, Cambridge, MA. <http://nickseaver.net/papers/seaverMIT8.pdf>, last accessed 2018/08/21.
37. McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., Barton, D.: Big data: The management revolution, *Harvard Business Review* (90:10), 61-67. (2012).
38. Drucker, J.: Performative materiality and theoretical approaches to interface, *Digital Humanities Qtrly*, 7 (1). <http://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html>, last accessed 2018/05/01.
39. Introna, L.: Algorithms, governance, and governmentality: On governing academic writing, *Science, Technology & Human Values* 41(1), 17-49 (2016).
40. Shotter, J.: Understanding process from within: An argument for 'witness'-thinking. *Organization Studies*, 27(4), 585-604 (2006).
41. Mackenzie, A., Vurdubakis, T.: Code and coding in crisis: Signification, performativity and excess, *Theory, Culture & Society*, 28 (6), 3- 23 (2011).
42. Barad, K.: *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press, Durham, NC. (2007).
43. Beer, D.: The social power of algorithms, *Information, Communication & Society* 20 (1), 1-13 (2017).
44. Latour, B.: *Reassembling the social: An introduction to Actor-Network-Theory*. Oxford University Press, Oxford. (2005).
45. Law J.: *After method: Mess in social science research*. Routledge, London (2004),
46. Law, J., Hassard, J.: (eds.) *Actor network theory and after*, Blackwell and the Sociological Review, Oxford and Keele (1999).
47. Law, J., Lien, M. E.: Slippery: Field notes in empirical ontology, *Social Studies of Science*, 42(3), 363-378 (2012).
48. IEEE: The IEEE Global initiative on ethics of autonomous and intelligent systems, IEEE Standards 2018, <http://standards.ieee.org>, last accessed 2018/08/21.
49. Wrhane, P.: Mental models, moral imagination and systems thinking in the age of globalization, *Journal of Business Ethics* 78, 463-474 (2008).
50. Arney, J.: Environmental Justice. *Encyclopaedia Britannica*, <https://www.britannica.com/topic/environmental-justice>, last accessed 2018/05/16.
51. Bandom, R.: New Toronto declaration calls on algorithms to respect human rights. *The Verge*, 16 May (2018).
52. Schmelzer, R.: Combination of blockchain and AI makes models more transparent. *TechTarget*, May 2018, <https://searchenterpriseai.techtarget.com/feature/Combination-of-blockchain-and-AI-makes-models-more-transparent>, last accessed 2018/08/21.
53. Gilpin, A.: Chapter 1: EIA approaches. *Environmental Impact Assessment (EIA): Cutting edge for the twenty-first century*, Cambridge University Press, Cambridge, UK 1-15. (1995).
54. Nordhaus, T., Shellenberger, M.: The monsters of Bruno Latour, *Breakthrough Journal*, Spring (2012).