

# Managing Algorithms for Public Value

Friso Selten, Utrecht University, The Netherlands

Albert Meijer, Utrecht University, The Netherlands

## ABSTRACT

Public organisations increasingly rely on machine learning algorithms in performing many of their core activities. It is therefore important to consider how algorithms are transforming the public sector. This article aims to clarify this by assessing algorithms from a public value perspective. Based on a discussion of the literature, it is demonstrated that algorithms are generally expected to strengthen organisational performance on a first cluster of values related to the ability to be effective and efficient (sigma values). At the same time, the use of algorithms is linked to negatively affect a second cluster of values that involves fairness and transparency (theta values). In the current academic debate, little attention is given to an important third cluster of values: the ability of organisations to be adaptive and robust (lambda values). This discussion highlights that algorithms invoke public value opportunities, but also public value risks and trade-offs. This article therefore presents five principles for managing algorithms from a public value perspective.

## KEYWORDS

Algorithms, Machine Learning, Management Principles, Public Administration, Public Value

## INTRODUCTION

The literature on the use of algorithms within the public sector is dense. The use of algorithms is described to improve governance effectiveness and efficiency (Wirtz et al., 2019; Meijer & Grimmeliikhuijsen, 2020), to reshape accountability relations and alter how transparency is provided (Peeters, 2020; Giest & Grimmeliikhuijsen, 2020), to transform public organisational structures (Lorenz et al., 2020), and to change how street-level decisions are made (Young et al., 2019; Busuioc, 2020). Moreover, algorithms impact all organisational levels; they require choices at the managerial level, as do they affect the work of frontline professionals (Lorenz et al., 2020; Meijer et al., 2021; Bullock et al., 2020).

Authors examining the implementation of algorithms repeatably, but not always explicitly, consider how algorithms are used to build or erode public values. This makes sense, public values are: “citizens collective expectations in respect to government and public services” (Moore, 1995). They are the most important units to assess the performance of public organisations because these values are what constitutes their legitimacy (Jørgensen & Bozeman, 2007). The analysis of a new public policy instrument, such algorithms, therefore always involves illustrating how they affect the creation public values. A direct discussion on how algorithms influence *public value* creation – a concept that is employed in this article to describe the grand total of all public values that are created by a public organisation or government – is however missing. This is the objective of this article. Using Hood’s (1999) classical framework of three public value clusters, this article makes explicit how algorithms

DOI: 10.4018/IJPADA.20210101.0a9

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

impact the ability of governments and public organisations to create public value. Building off this discussion, which shows that algorithms generate public value trade-offs, guidance is provided on how to manage these dilemmas. In doing so, this article's contribution to our academic knowledge about algorithms in the public sector is threefold: (1) it present public value as an umbrella for discussing the impact of algorithms on the public sector performance, (2) it describes how the development and use of algorithms in the public sector affects the creation of three different types of public values, and (3) it establishes principles to provide guidance for managing algorithms to create public value.

The information in this paper about the management of algorithms in the public sector is also relevant for practitioners. Algorithms fundamentally change the role that public managers play in the policy and decision making process (Van der Voort et al., 2019). Policy makers are however unprepared for the challenges they face when working with algorithms (Agarwal, 2018). The insights presented in this article can therefore support practitioners in performing this new role.

## **DEFINING PUBLIC SECTOR ALGORITHMS**

To understand the importance of discussing how algorithms impact the creation of public value, it is crucial to understand how these technologies differ from previous digital innovations in the public sector. After all, the use of digital tools in the public sector is not a modern invention (Kitchin, 2017). The first wave of digitalisation of governance practices, beginning roughly at the start of this century, was focused on standardizing simple decision making procedures (Bovens & Zouridis 2002; Margetts & Dunleavy, 2013). ICT systems and artificial intelligence in these days were used to solve problems that could easily be quantified and represented in formal rules and repetitive tasks (Buffat, 2015; Goodfellow et al., 2016). Recent advances in artificial intelligence enabled algorithms to infer on increasingly difficult tasks, problems that need to be solved intuitively like recognizing spoken words or faces in images (Meijer & Grimmelikhuijsen, 2020; Goodfellow et al., 2016; LeCun et al., 2015). Public organisations are increasingly relying on artificial intelligence tools to perform these kind of complex tasks. Algorithms are for example being used in the development of climate policies (Valle-Cruz et al. (2019), urban planning (Ibrahim et al., 2020), policing (Meijer & Wessels, 2019), the detection of fraud (Zouridis et al., 2020) and a whole range of other policy domain (comprehensive overviews can be found in Valle-Cruz et al. (2019) and Engin & Treleaven (2019). This new wave of digitalisation, with artificial intelligence and algorithms at its core, implies a fundamental shift in how public organisations function and the role of public managers have in the decision- and policy-making process (Van der Voort et al., 2019; Meijer & Grimmelikhuijsen, 2020; Buffat, 2015; Bullock, 2019; Peeters, 2020).

In this paper, we explicitly focus on the technology that is enabling this kind of artificial intelligence: machine learning algorithms. Machine learning approaches autonomously learn and update sets of mathematical rules and procedures from historical data to calculate a particular probabilistic prediction of a real world outcome on new input data (LeCun et al., 2015, p. 436; Busuioc, 2020). The learning procedure of machine learning algorithms can formally be defined as: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Mitchell, 1997). However, while algorithms are 'self-learning' they are not 'self-built'. The development and use of algorithms, especially in a public context, is at all stages influenced by human decision making. Humans decide what task T is, how experience E is generated and determine how performance P is evaluated (Boyd, & Crawford, 2012; Diakopoulos, 2014). These choices are influenced by financial, political, legal and institutional pressures (Kitchin, 2014, p. 185). As a result, depending on the choices, preferences, and goals of the developers and policy makers, two algorithms that are designed to perform the same task can vary widely in their outcomes (Shepperd et al., 2014). This leads Kraemer et al., (2011) to conclude that algorithms are not neutral but value laden.

Furthermore, in most cases algorithms are not making decisions autonomously – they are not autonomous agents. Rather, algorithms are used as an instrument to inform, assist and enhance human decision makers (Buffat, 2015; Chen et al., 2019; Peeters & Widlak, 2018; Busuioc, 2020). Policy makers base their decisions on all kind of information, such as experience, expert opinions and the political climate in which they operate (Kitchin, 2014). Algorithms are an additional instrument in this elaborate decision making toolbox. However, humans do not process information rationally. They are subject to a bounded rationality and cognitive biases (Simon, 1947; Kahneman, 2011; Battaglio et al., 2019). The use of algorithms may therefore lead to the introduction of automation bias in public decision-making processes (Busuioc, 2020). Automation bias refers to the mechanism that decision makers have the tendency to display an overreliance in the validity and rationality of automated predictions (Lyell & Coiera, 2017). Automated decision systems might therefore not reduce decision making errors, but only lead to the emergence of different types of errors (Skitka & Mosier, 1999). Furthermore, algorithms have a political dimension. Substantiating a decision using an algorithm can be beneficial because they are perceived to produce objective facts; numbers, probabilities and statistics and therefore provide strong arguments to substantiate a political preference (Meijer & Thaens, 2018; Boyd & Crawford, 2012). The human-machine interaction – the interaction between the decision maker and the algorithm – is therefore one of the most important determinants of the impact algorithms have on the creation and enforcement of public policies (Young et al., 2019; Bullock, 2019). These interactions are influenced by the design of the algorithm itself, but also by psychological processes and organisational practices (Binns, 2020; Peeters, 2020; Meijer & Grimmelikhuijsen, 2020; Meijer et al., 2021).

To summarize, the development of and outcomes produced by machine learning algorithms in the public sector are part of a complex decision making process (Diakopoulos, 2014, Kitchin, 2017). Because of their ability to self-learn and solve increasingly complex tasks, machine learning approaches have unique – distinct and divergent from other digital innovations in the public sector - characteristics in how they alter public organisational functioning and public decision making (Buffat, 2015; Lorenz et al., 2021; Busuioc, 2020). Reviewing how algorithms affect public decision-making processes is therefore important. Central in this discussion should be the notion that the core task of governments is to create public value (Jørgensen & Bozeman, 2007; Bozeman, 2012, pp. 175-177).

## THE IMPACT OF ALGORITHMS ON PUBLIC VALUE CREATION

The importance of investigating the capacity of governments and public organisations to create public values, builds on the notion that citizens have “inalienable rights” that government cannot violate (Bozeman, 2007, p. 177). The combination of these rights is what constitutes public value (Bozeman, 2012, p. 13). Citizens however do not have a uniform, collective expectation of the public sector. Public organisations should therefore not serve specific goals, but society as a whole (Moore, 1995). To create public value they need to “create or contribute the common good and to the public interest” (Jørgensen & Bozeman, 2007, p. 361).

Positioning public value vis-à-vis private and ethical values helps to clarify the significance of this concept for governance practices. Private values are created through transactions that produce economic benefits and can be assessed through efficiency and effectiveness (Harrison et al., 2012). Evaluating the public value organisations create requires complementary measures: not only effectiveness and efficiency matters but also fairness, accountability, being unbiased, transparency, social inclusivity, and paying attention to individual well-being (Twizeyimana & Andersson, 2019; Harrison, et al., 2012). These are ethical considerations, they are concerned with what is right or wrong (Bannister & Connolly, 2014). A discussion of public value however goes beyond right or wrong thinking; public value takes more practical principles as a starting point. For example, transparency and accountability are needed to prevent unethical behaviour such as corruption, but also to regulate, coordinate and control public organisations (Da Cruz et al., 2016; Schillemans, 2016). When public

value is not created this has negative effects that go beyond the fact that it is morally wrong. In that case we speak of a public value failure, which can lead to a decrease in functioning and legitimacy of the public organisation (Jørgensen & Bozeman, 2007; Bozeman, 2012, p. 12). Creating public value is thus essential for public organisations, and it entails balancing both private and ethical values.

In this paper, the discussion of the impact of algorithms on public value will be guided by the three clusters of administrative values outlined by Hood (1991). *Sigma-type* values refer to organisations attempting to be efficient and effective. Pursuing this type of values requires to have a lean organisation and to accomplish goals using as little resources as possible. *Theta-type* values describe organisations principles of being honest and fair. This is accomplished by preventing distortion, inequity, bias, and abuse of office. Finally, *lambda-type* values refer to the reliability and robustness of governance practices. These are strengthened when organisations prepare for crisis situations, have back-up systems, diversify their operations, and acquire more capacity and resources than strictly necessary. Public organisations should not strive to create only one of these values. Only when all three types of values are cherished and a surplus arises when weighing values against each other, public value is created.

In the next sections of this article, the impact that machine learning algorithms as a policy instrument have on the creation of these three values is discussed. Next, guidance is provided for how public managers can device a public value surplus when adopting and implementing algorithms.

## Sigma Values

Sigma values allude to principles of efficiency and effectiveness. These are the values that are regularly used to promote e-government applications (Maciejewski, 2017; Wirtz et al., 2019; Meijer & Grimmelikhuijsen, 2020). Modern algorithms are able to complete increasingly complex tasks and are in some instances even approaching or exceeding human performance (Domingos, 2012; Goodfellow et al., 2016). Algorithms analyze data much faster than humans and in this data detect otherwise undiscovered patterns. Furthermore, they constantly learn and update their rules and procedures (Young et al., 2019; Peeters; 2020). This creates opportunity for a whole array of policy domains to become partially automatized, e.g. reviewing immigration applications, service delivery and surveillance (Chun 2007; Zheng, et al., 2018; Power, 2016). Furthermore, when proven effective an algorithm can cheaply be applied on many cases at the same time to substantially stimulate governance efficiency (O'Neill, 2016, p. 42).

Algorithms at the same time threaten the effectiveness and efficiency of public sector organizations. Civil servants can become overconfident in the results produced by an algorithm and disregard other important decision instruments such as local knowledge, experience and common sense (O'Neill, 2016; Fry, 2018; Bolander, 2020). In addition, new algorithmic outputs can also be used as merely an instrument to pursue pre-defined (political) goals (Van der Voort et al., 2019). Finally, while an algorithm can be effective in theory, it is hard to establish the effects and effectiveness of an algorithm when it is being used in practice. Algorithms can become a self-fulfilling prophecy: "If policy-makers expect ex ante to find more crime among group A than group B, then it is possible that they will find this expectation validated ex post, but only because they have spent more time looking for crime among members of group A than among members of group B" (Chiao, 2019, p. 128).

In spite of these possible drawbacks, sigma-type values are the reason why many public sector organizations feel that they need to explore the potential of algorithms.. Algorithms can exceed human performance to make decision making more effective and can efficiently do this on many cases at the same time (Bullock, 2019).

## Theta Values

Theta values comprise of values such as fairness, transparency and accountability (Hood, 1991). The biggest critiques on the use of algorithms by public organisations focusses on violations of these type of values. Algorithms are linked to increase unfair outcomes and to introduce biases in

the decision making process (O'Neill, 2016; Mittelstadt et al., 2016; Crawford & Schultz, 2014). For example, algorithms used to predict the risk of reoffending have been proven to be almost twice as likely to mistakenly flag black than white defendants (Crawford & Calo, 2016). These issues are mostly caused by the fact that the underlying data the algorithms are trained on already contain these biases. Machine learning algorithms will learn and reinforce these existing structure – a problem that is especially problematic because of the above outlined abilities of algorithms to become a self-fulfilling prophecy and to be highly scalable (Willis et al., 2007; Ferguson, 2012; Diakopoulos 2014; O'Neil, 2016; Andrews, 2019).

Underlying all these elements is one main challenge; algorithms are in many instances not transparent. This lack of transparency makes it difficult to distinguish whether an algorithm is producing fair and accurate results or is actually discriminatory. In the current literature there are mainly three causes presented for this lack of transparency: internal, external and process opacity. External opacity means that many algorithms do not reveal their internal mechanisms. A key reason for external opacity is intellectual property protection but even when an algorithm is open-source, it is externally opaque for a majority of people since only a small part of the population can understand an algorithm's functioning by looking at the code (Lepri, et al., 2018). Yet, even when input data and the internal mechanisms are disclosed and reviewed by experts, algorithms can be non-transparent due to internal opacity (Oswald, 2018). Algorithms learn how to perform tasks by themselves, but what they learn and how they function is often not interpretable for humans (Guidotti, et al., 2018). Internal opacity refers to this fact that algorithms will always contain a "degree of unavoidable complexity when solving complex tasks" (Burrell, 2016, p. 5). In addition to these two technology-related forms of opacity, process opacity means that is unclear how exactly algorithms are used by policy makers (Saunders et al., 2016). Even when an algorithm is working very well, if it is used incorrectly it can still produce unfair outcomes. Because of their inherent opaqueness, public decision makers are unable to control algorithmic outcomes (Peeters, 2020). This can result in the introduction of new types of biases in public decision-making processes (Deley & Dubois, 2020; Busuioc, 2020).

These three types of opacity; external, internal and process, create a situation that makes governance less transparent and distorts accountability relations, i.e. it hinders an organisation's ability to create theta-type values. At the same time, it should be noted that algorithms according to some are evaluated against unreasonably high standards. Human decision making is also not neutral and is also subject to transparency problems and biases (Fry, 2018, pp. 7176, Zerilli et al., 2019b; Young et al., 2019). Algorithms analyse all available information and apply the similar rules to similar situations (Young et al., 2019; Binns, 2020). From this perspective, if we succeed in developing well-functioning and transparent algorithms, these can even make governance more accountable and rule out adverse effects of human decision making e.g. prevent regulatory arbitrage, remove human bias, and reduce inconsistency in decision-making (Halachmi, & Greiling, 2013; Mittelstadt et al., 2016; Crawford & Calo, 2016; Wieringa, 2020, Bolander 2020; Chiao, 2019). Doing so, algorithms can also become a tool to improve trust in the functioning of public organisations. For example, a recent study by Miller and Keiser (2020) demonstrates that minority groups in certain cases regard automated policing systems as being more fair than human police officers.

## **Lambda Values**

The advantages of algorithms to sigma-type values and disadvantages to theta-type values are extensively described in existing literature. Much less attention is given to how the use of algorithms affects values belonging to this third set of lambda-type values. These are the ability of organisations to be robust and adaptive. This while algorithms potentially have a big impact on this set of public values.

Algorithms recognize and learn patterns in data that humans are not able to observe (Domingo, 2012; Lecun et al., 2015). Doing so, they can foresee and predict certain risks long before these are observable by humans. For example, algorithms have been used to predict flu-trends based on social media data (Achrekar, et al., 2011; Dugas et al., 2013). Despite this ability to predict unforeseen

events, algorithms are not predicting the future. They only acquire predictive capacities by detecting and learning patterns from historical data and assume these patterns are recurring. Algorithms thus by nature only function if patterns repeat. Their predictive capacities diminish in situations where this is not the case. In crises that completely alter the status-quo they are bound to fail (O'Neill, 2016; Binns, 2020).

Furthermore, when tasks are becoming completely automatized this is demanding an extremely stable system. However, algorithms, like all software, are at the threat of aging. The older digital tools become, their performance reduces, reliability decreases and costs to upkeep rise (Parnas, 1994). Even the most stable system will therefore require some form of human oversight. However, when an automated system is performing a task, people will no-longer be able to maintain the skills to perform the task themselves. This can cause large problems when vital tasks are automatized and the algorithm fails (Fry, 2018, pp. 133-134). Automation is “most dangerous when it behaves in a consistent and reliable manner for most of the time” (Banks et al. 2018b, p. 283, c.f. Zerilli et al., 2019a). These factors combined pose serious threat for a breakdown of essential administrative functions in a society that is driven by algorithms.

Last, algorithms hinder the adaptive capacity of state functioning. It is widely understood that complex public decision making requires flexibility and adaptation. The value of having a certain amount of freedom and discretion in decision making is vital for the work of civil servants (Lipsky 1980, p. 14; Tummers & Bekkers, 2014). Algorithms do not allow for such discretion. The goal of computers is to learn general rules from which they cannot divert based on inherent local knowledge and discretion (Dominigo, 2012; Binns, 2020).

The effect of algorithms on the creation of lambda-type values is – since machine learning is still a relatively new phenomenon in the public sector – generally not constituting an immediate risk. However, the above discussion highlights that, when algorithms are becoming a more established instrument within public organisations, considering how these technologies affect lambda-type values is essential.

## **Algorithmic Value Trade-Offs**

Digital technologies have long been regarded as making public organisations function more efficient (Dunleavy, et al., 2006; Margetts & Dunleavy, 2013; Wirtz et al., 2019), improve frontline service delivery (Zheng et al., 2018) and making governance more transparent and accountable (Halachmi, & Greiling, 2013). The above discussion shows this traditional view on ICT applications needs revision with the emerging adoption of machine learning algorithms in the public sector.

The discussion of the literature highlighted that algorithms are often being implemented to make governments and public organisations function more efficient and effective, the so-called sigma-type values. Simultaneously, they are mostly seen as having a negative impact on theta-type values, they produce unfair outcomes, are not transparent and cannot be held accountable. Reality, as always, was demonstrated to be more nuanced. Algorithms do not immediately make organisations more efficient, but neither does it make them automatically biased and unfair.

Interestingly, the risks that algorithms pose to lambda-type values are relatively little debated in the literature. However, with the wider adoption of algorithms these threats to adaptivity and robustness of essential state functions are just as important as the other two sets of values. A challenge for public organisations is to pay careful attention to these different values as to manage algorithms for public value.

Even though it seems desirable to maximize scores on all values, the literature highlights that governing in the public sector will require trade-offs between values. This leads to dilemmas for public organisations when deciding about the use of algorithms. Questions such as the ones posed by O'Neill (2016, p. 95): “[are] we as a society [...] willing to sacrifice a bit efficiency in the interest of fairness [?]”, and Young et al. (2019, p. 303): “If an AI system is half the cost of the current system, but 10% less accurate, is it better?”, are therefore extremely relevant.. It is the task for public

managers to find a balance between all public values. Do they prioritize lambda-, theta- or sigma-type values? Citizens expect efficiency, transparency and robustness, but it might not always be possible to realise all these values at the same time. Managing the adoption and implementation of algorithms thus entails balancing these expectations. This management of public expectations is what Bozeman (2012, p. 176) refers to as the management of publicness.

## MANAGING THE PUBLICNESS OF ALGORITHMS

The discussion in the previous section highlighted that algorithms induce value trade-offs in public decision-making processes. Drawing on traditional public administration literature in this section guidance is provided to balance these trade-offs by presenting five principles for managing the publicness of algorithms. These principles describe the implications of seeing algorithms as *a policy instrument*, emphasize the importance of *having multiple critical audiences*, express why algorithms need to *be reliable*, demonstrate the importance of *empirical evaluation*, and describe why the publicness of algorithms should collectively be managed *with citizens*. The goal of these principles is to stimulate the development of algorithms that are able to create public value.

### Principle 1: Manage Algorithms as Tools in Organizational Practice

Algorithms are not merely a technology: they are a policy instrument that is used by civil servants within a complex decision-making process (Kitchin, 2014; Van der Voort et al., 2019; Meijer & Grimmelikhuijsen, 2020; Veale, et al., 2018). Drawing on Simon's (1972) concept of 'bounded rationality' this has important implications for how we observe algorithms from a public value perspective. The bounded rationality theory implies that even when an algorithm is functioning correctly, they can produce adverse effects because the information they provide is not used correctly. Human decision makers are subject to cognitive biases (Kahneman, 2011), algorithms can therefore introduce new types of bias in public decision-making processes (Skitka & Mosier, 1999; Busuioc, 2020; Peeters, 2020). Thus, it is crucial not only to focus on the features of algorithms as such but also on how these features are expressed in organizational practice and how algorithms affect decision-makers.

To stimulate the creation of sigma-type values of algorithms and reduce the chance of adverse effects on theta-type and lambda-type values, it is therefore important that people who are deciding about the use of algorithms or working with them, are enabled to assess the quality of the information the algorithm provides. Policy makers need to be informed about how the algorithm is developed and its real world effects to be able to use them effectively and prevent automation biases (Saunders et al., 2016; O'Neill, 2016; Lyell, & Coiera, 2017; Veale et al., 2018; Zerilli et al., 2019a; Weller, 2019).

### Principle 2: Account for Algorithms to Multiple Audiences

Accountability and transparency are traditional requirements to promote trust in public institutions (Harrison & Sayogo, 2014). However, providing transparency and accountability of algorithmic functioning is difficult, due to internal, external and process opacity (Burrell, 2016, p. 5; Goad & Gal, 2018; Lepri et al., 2018). For this reason, Kemper and Kolkman (2019) introduce the notion that a critical audience is needed to make algorithms transparent and accountable. Who these critical audiences are in the public sector can be identified using Bovens' (2007) five accountability relations. The five accountability relations are: political (e.g. ministerial responsibility), legal (judges), administrative (auditors, inspectors and controllers), professional (professional peers) and social (interest groups, charities and civil society). See Wieringa (2020) a complete overview of these accountability relations in the context of algorithms, here a distinction is made between what are defined to be technical audiences and non-technical critical audiences.

First, accountability towards the technical audiences, which are the administrative, professional and legal accountability audiences. Accountability to these audiences implies auditing of the algorithm

by experts to test if it is functioning correctly and function in accordance with the law. Making technical accountability possible requires providing transparency about input and output data, have data management protocols, make the algorithmic code open source and work on techniques to make algorithmic functioning more explainable (Kemper & Kolkman, 2018; Burrell, 2016; Datta et al., 2016; Goodman & Flaxman, 2017; Wieringa, 2020). This is important because it can help point out potential flaws in the algorithm, such as coding errors or biases (Weller, 2019). While this type of accountability (partly) solves external and internal opacity problems, it does not tackle the process opacity issue.

Second, non-technical audiences are the political and social accountability audiences. These non-technical audiences, political parties but also civil servants and citizens for example, are not necessarily able to obtain the technical knowledge to understand algorithmic functioning. For these audiences the use of algorithms creates a principal-agent relationship. Algorithms are developed by a team of programmers, but are used in practice by civil servants who have little understanding about how the algorithm functions and makes predictions (Meijer, 2018; Van der Voort et al., 2019). Similarly, citizens are affected by a decisions that are made using an algorithm but are not able to understand how these decisions were constructed (O'Neill, 2016, p. 8). Accountability thus also needs to be rendered to society as a whole. It is important that citizens broadly understand and become comfortable with the strengths and limitations of an algorithm to overcome fear of the unknown (Weller, 2019).

### **Principle 3: Ensure the Reliability of Algorithms over Time**

Research on government transparency shows that more transparency does not automatically lead to increased trust in government functioning (Cucciniello et al., 2017). Providing the wrong type of transparency can have adverse effects and make people unnecessary critical about the use of algorithms by public organisations (Ananny & Crawford, 2018; Weller, 2019). Building on Deley and Dubois (2020) who describe that trustworthiness of algorithms can only be provided by demonstrating their reliability, it is therefore argued that public organisations must do precisely this – demonstrate the reliability of the algorithms they use. This is what Grimmelikhuijsen and Meijer (2014) refer to as transparency of policy outcomes or effects.

Reliability is tied to all three value clusters. It entails that algorithms make public organisations function more effective (sigma-type values) and that decisions are not leading to adverse, unfair or discriminatory outcomes (theta-type values). Reliability also logically refers to accounting for lambda-type values. Clear back-up procedures for when the algorithm breaks are needed and a detailed strategy for maintaining the reliability of algorithms over time is needed to ensure the production of public value over time. Ultimately to enhance trust in algorithms also requires to demonstrating how they affect decision-making and policy outcomes. This is only possible through proper evaluation of the real-world effects of algorithms. The next principle presented will therefore describe how this aspect of reliability can be demonstrated.

### **Principle 4: Evaluate Algorithms to Demonstrate Causal Mechanisms**

Providing reliability over outcomes, and thus make accountability of algorithms possible demands evaluation by all critical audiences that are described by Bovens (2007). This requires that these audiences are enabled to monitor the performance of algorithms that are used as policy instruments (Grimmelikhuijsen & Meijer, 2014). However, as described in the discussion about sigma-type values, using only technical evaluations is not sufficient to discern all effects algorithms have (Fry, 2018; Chiao, 2019). Building on public administration research that aims to discern the causal mechanisms of policy interventions, implies this is only possible using empirical evaluation (James, et al., 2017).

Evaluating the effects of the implementation of a new policy intervention, which we defined algorithms to be, is difficult. To test the effect of any policy intervention we need to know what would have happened when the intervention would not have been implemented. This is what within experimental research is referred to as a as the ‘counterfactual’ (James et al., 2017). The recent



increase in popularity of experimental studies within public administrations provides opportunity for demonstrating this counterfactual and thus showing the causal effects algorithms have when implemented in a public organisation (Bouwman, & Grimmelikhuijsen, 2016). The use of field experiments is especially suitable to test the effects of algorithms (Hansen & Tummers, 2020). Field experiments include a careful monitoring of the implementation of an algorithm and comparing this to similar situations where the algorithm is not used. This type of experiments enables policy makers and researchers to discern the advantages and adverse effects the implementation of an algorithm has in practice. Simultaneously, testing the effects of an algorithm in a real world setting therefore requires careful consideration costs, ethics and validity (Hansen & Tummers, 2020). The implementation of algorithms in certain policy domains such as national security or a social safety net do not permit the same type of learning as for example increasing production efficiency (Bozeman, 2007, pp. 179-180).

### **Principle 5: Establish Publicness with Citizens**

Finally, it is important to explicitly discuss the public value dilemmas that have been presented in this article with citizens. In this discussion it is important to remember that not all value trade-offs are equal. There is a distinction between what Tetlock et al. (2000) define to be ‘secular values’ and ‘sacred values’.

Secular values are values that can be expressed monetary. This is what Tetlock et al., (2000) refer to as being a routine trade-off. A routine trade-off is for example a trade-off between the cost of developing an algorithm and the administrative efficiency gains it is expected to deliver. Sacred values are values that are hard to express monetary such as safety, security, health and justice (Tetlock et al., 2000; Tetlock, 2003). Trade-offs between secular and sacred values are what Tetlock et al (2000) refer to as being taboo trade-offs. No matter how big the monetary advantages an algorithm might be able to deliver, if it negatively impacts these sacred values citizens will not be willing to accept the use of algorithms (Tetlock et al, 2000).

For the use of algorithms, many of the theta- and lambda-type values (e.g. discrimination, biases, risk of collapse) might fall in the sacred value category. This means that many of the public value trade-offs civil servants will be dealing with are not routine, but will be taboo trade-offs. This urges everyone who is developing or working with algorithms to consider the trade-offs they produce about fundamental issues such as; “human life (what price access to medical care?), justice (what price access to legal representation?), preserving natural environments (what price endangered species?)” (Tetlock, 2003, p. 320). These are questions that are also important for the implementation of algorithms. Algorithms are used to design health, justice and environmental policies (Valle-Cruz, 2019). When do the advantages algorithms bring outweigh their negative conquests - which values are secular and which are sacred? Most importantly, answers to these questions need to be formulated together with citizens, because public value is the collective expectation of citizens (Moore, 1995). To develop algorithms that create public value it is thus required that research is conducted to get insights in the opinions of citizens on the trade-offs between the different values algorithms invoke.

## **CONCLUSION**

This article set out to describe the use of algorithms from a public value perspective. This perspective contributes three important insights to the role of algorithms in the public sector. First, it was established that a variety of debates on algorithms in the public sector can be understood from the perspective of public value. Second, it was demonstrated how algorithms affect the creation of three different types of public value. Third, it provided five principles that can be used to manage the publicness of algorithms.

The impact on public value creation was discussed using Hood’s (1991) distinction between sigma-, theta- and lambda-type values. An analysis of the literature indicated that the positive effects

of algorithms are mainly framed as providing sigma-type values; efficiency and effectiveness. The adverse effects are primarily referring to theta-type values; algorithms might produce unfair outcomes and are often not transparent. The discussion about the effect of algorithms on these two value clusters in this article showed that these effects are in reality not such a simple dichotomy. Furthermore, algorithms have impact on so-called lambda-type values. They can be used to predict risks, but at the same time severely hinder robustness of state functioning. Breakdown can occur when aging or inflexibility of algorithms leads to the failure of an algorithm that performs essential tasks. From a public (sigma-type) value perspective just disregarding the advantages algorithms can bring to public sector functioning is not possible. However, sigma-type value gains might come at a cost of theta-type (unfair, biased and discriminatory practices) and lambda-type values (less resilience and robustness) (O'Neill, 2016; Mittelstadt et al., 2016; Fry, 2018; Parnas, 1994). Consequently, using but also not using algorithms requires public managers to make public value and taboo trade-offs.

Based on the public administration literature five principles for balancing these trade-offs and thus for managing the publicness of algorithms were presented. First, it was described that algorithms should not be seen as a replacement of the policy maker, but mostly as one of the multiple instruments that can be used to inform them. This implies that they are part of a complex decision making process and that the information they provide is processed through to cognitive biases (Simon, 1972; Kahneman, 2011). This led to the second principle; accountability should be rendered to different audiences (Bovens, 2007; Kemper & Kolkman, 2018). Here, mainly the distinction between technical and non-technical audiences was emphasized. The third principle described that, especially to make accountability for non-technical audiences possible, the focus should not be necessarily on making the internal mechanisms transparent but on making algorithms reliable (Deley & Dubois, 2020). To accomplish this, transparency of policy outcomes and effects is important (Grimmelikhuijsen & Meijer, 2014). The fourth principle described that this type of transparency can only be provided using empirical (experimental) evaluation (James et al., 2017). The last guiding principle most directly reflects on the management of the public value dilemmas that algorithms invoke. Here, it was demonstrated that algorithms can create taboo trade-offs and that balancing these trade-offs can only be done by society as a whole – in a dialogue between developers, politicians, policy makers and citizens (Tetlock et al., 2000). While grounded in literature – the proof is in the pudding. Future studies should aim to research how these principles can successfully be applied by public organisations.

To conclude, this article demonstrated that public values should at the centre of the discussion about the implementation of algorithms in the public sector. This changes our perspective on what algorithms are, with whom they should be developed, why they need to be reliable, how their usage should be evaluated and how the publicness of algorithms can be managed. These principles are important because the development of public sector algorithms should always be aimed at creating public value.

## REFERENCES

- Achrekar, H., Gandhe, A., Lazarus, R., Ssu-Hsin, Yu., & Liu, B. (2011). Predicting flu trends using twitter data. *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 702–707. doi:10.1109/INFCOMW.2011.5928903
- Agarwal, P. K. (2018). Public administration challenges in the world of ai and bots: Public administration challenges in the world of ai and bots. *Public Administration Review*, 78(6), 917–921. doi:10.1111/puar.12979
- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. doi:10.1177/1461444816676645
- Andrews, L. (2019). Public administration, public leadership and the construction of public value in the age of the algorithm and ‘big data’. *Public Administration*, 97(2), 296–310. doi:10.1111/padm.12534
- Banks, V. A., Plant, K. L., & Stanton, N. A. (2018). Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016. *Safety Science*, 108, 278–285. doi:10.1016/j.ssci.2017.12.023
- Bannister, F., & Connolly, R. (2014). ICT, public values and transformative government: A framework and programme for research. *Government Information Quarterly*, 31(1), 119–128. doi:10.1016/j.giq.2013.06.002
- Battaglio, R. P. Jr, Belardinelli, P., Bellé, N., & Cantarelli, P. (2019). Behavioral public administration *ad fontes*: A synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. *Public Administration Review*, 79(3), 304–320. doi:10.1111/puar.12994
- Bengio, Y., Janda, R., Yu, Y. W., Ippolito, D., Jarvie, M., Pilat, D., Struck, B., Krastev, S., & Sharma, A. (2020). The need for privacy with public digital contact tracing during the COVID-19 pandemic. *The Lancet. Digital Health*, 2(7), e342–e344. doi:10.1016/S2589-7500(20)30133-3 PMID:32835192
- Bolander, T. (2020). Correction to: What do we lose when machines take the decisions? *The Journal of Management and Governance*, 24(2), 557–557. doi:10.1007/s10997-020-09507-z
- Bouwman, R., & Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014: A systematic literature review and ways forward. *International Journal of Public Sector Management*, 29(2), 110–131. doi:10.1108/IJPSM-07-2015-0129
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447–468. doi:10.1111/j.1468-0386.2007.00378.x
- Bovens, M., & Zouridis, S. (2002). From street- level to system- level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174–184. doi:10.1111/0033-3352.00168
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878
- Bozeman, B. (2007). *Public values and public interest: Counterbalancing economic individualism*. Georgetown University Press.
- Bozeman, B., & Bretschneider, S. (1994). The ‘publicness puzzle’ in organization theory: A test of alternative explanations of differences between public and private organizations. *Journal of Public Administration: Research and Theory*, 4(2), 197–224.
- Buffat, A. (2015). Street-level bureaucracy and e-government. *Public Management Review*, 17(1), 149–161. doi:10.1080/14719037.2013.771699
- Bullock, J., Young, M. M., & Wang, Y.-F. (2020). Artificial intelligence, bureaucratic form, and discretion in public service. *Information Polity*, 1–16.
- Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. *American Review of Public Administration*, 49(7), 751–761. doi:10.1177/0275074019856123

- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 205395171562251. doi:10.1177/2053951715622512
- Busuioc, M. (2020). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*. 10.1111/puar.13293
- Chen, T., Ran, L., & Gao, X. (2019). AI innovation for advancing public service: The case of China’s first Administrative Approval Bureau. *20th Annual International Conference on Digital Government Research on - Dg.o 2019*, 100–108. doi:10.1145/3325112.3325243
- Chiao, V. (2019). Fairness, accountability and transparency: Notes on algorithmic decision- making in criminal justice. *International Journal of Law in Context*, 15(2), 126–139. doi:10.1017/S1744552319000077
- Chun, A. H. W. (2007, July). Using AI for e-Government Automatic Assessment of Immigration Application Forms. In AAAI (pp. 1684-1691). Academic Press.
- Coglianesi, C., & Lehr, D. (2016). Regulating by robot: Administrative decision making in the machine-learning era. *Geological Journal*, 105, 1147.
- Cordella, A., & Bonina, C. M. (2012). A public value perspective for ICT enabled public sector reforms: A theoretical reflection. *Government Information Quarterly*, 29(4), 512–520. doi:10.1016/j.giq.2012.03.004
- Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, 538(7625), 311–313. doi:10.1038/538311a PMID:27762391
- Crawford, K., & Paglen, T. (2019). *Excavating AI: The politics of images in machine learning training sets*. Excavating AI.
- Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *BCL Rev.*, 55, 93.
- Cucciniello, M., Porumbescu, G. A., & Grimmelhuijsen, S. (2017). 25 years of transparency research: Evidence and future directions. *Public Administration Review*, 77(1), 32–44. doi:10.1111/puar.12685
- Da Cruz, N. F., Tavares, A. F., Marques, R. C., Jorge, S., & de Sousa, L. (2016). Measuring local government transparency. *Public Management Review*, 18(6), 866–893. doi:10.1080/14719037.2015.1051572
- Datta, A., Sen, S., & Zick, Y. (2016). Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. *2016 IEEE Symposium on Security and Privacy (SP)*, 598–617. doi:10.1109/SP.2016.42
- de Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology*, 31(4), 525–541. doi:10.1007/s13347-017-0293-z PMID:30873341
- Deley, T., & Dubois, E. (2020). Assessing trust versus reliance for technology platforms by systematic literature review. *Social Media + Society*, 6(2). doi:10.1177/2056305120913883
- Diakopoulos, N. (2014). *Algorithmic accountability reporting: On the investigation of black boxes*. 10.7916/D8ZK5TW2
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. doi:10.1145/2844110
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78–87. doi:10.1145/2347736.2347755
- Dugas, A. F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., & Rothman, R. E. (2013). Influenza forecasting with google flu trends. *PLoS One*, 8(2), e56176. doi:10.1371/journal.pone.0056176 PMID:23457520
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2006). New public management is dead—Long live digital-era governance. *Journal of Public Administration: Research and Theory*, 16(3), 467–494. doi:10.1093/jopart/mui057
- Engin, Z., & Treleaven, P. (2019). Algorithmic government: Automating public services and supporting civil servants in using data science technologies. *The Computer Journal*, 62(3), 448–460. doi:10.1093/comjnl/bxy082

- Ferguson A. G. Policing P. Suspicion R. (2012). 62 Emory Law Journal 259 (2012). Available at SSRN: <https://ssrn.com/abstract=2050001>
- Fry, H. (2018). *Hello world: How to be human in the age of the machine*. Doubleday.
- Giest, S., & Grimmelikhuijsen, S. (2020). Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective. *Information Polity*, 25(4), 409–417. doi:10.3233/IP-200010
- Goad, D., & Gal, U. (2018). Understanding the impact of transparency on algorithmic decision making legitimacy. In U. Schultze, M. Aanestad, M. Mähring, C. Østerlund, & K. Riemer (Eds.), *Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology* (pp. 64–79). Springer International Publishing. doi:10.1007/978-3-030-04091-8\_6
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press. <https://www.deeplearningbook.org/>
- Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision- making and a “right to explanation”. *AI Magazine*, 38(3), 50–57. doi:10.1609/aimag.v38i3.2741
- Grimmelikhuijsen, S. G., & Meijer, A. J. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. *Journal of Public Administration: Research and Theory*, 24(1), 137–157. doi:10.1093/jopart/mus048
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 1–42. doi:10.1145/3236009
- Halachmi, A., & Greiling, D. (2013). Transparency, e-government, and accountability: Some issues and considerations. *Public Performance & Management Review*, 36(4), 572–584. doi:10.2753/PMR1530-9576360404
- Hansen, J. A., & Tummers, L. (2020). A systematic review of field experiments in public administration. *Public Administration Review*, 80(6), 921–931. Advance online publication. doi:10.1111/puar.13181
- Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., Hrdinova, J., & Pardo, T. (2012). Open government and e-government: Democratic challenges from a public value perspective. *Information Polity*, 17(2), 83–97. doi:10.3233/IP-2012-0269
- Harrison, T. M., & Sayogo, D. S. (2014). Transparency, participation, and accountability practices in open government: A comparative study. *Government Information Quarterly*, 31(4), 513–525. doi:10.1016/j.giq.2014.08.002
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19. doi:10.1111/j.1467-9299.1991.tb00779.x
- Ibrahim, M. R., Haworth, J., & Cheng, T. (2020). Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities (London, England)*, 96, 102481. doi:10.1016/j.cities.2019.102481
- James, O., Jilke, S. R., & Van Ryzin, G. G. (2017). Causal inference and the design and analysis of experiments. In O. James, S. R. Jilke, & G. G. Van Ryzin (Eds.), *Experiments in Public Management Research* (pp. 59–88). Cambridge University Press. doi:10.1017/9781316676912.005
- Jørgensen, T. B., & Bozeman, B. (2007). Public values: An inventory. *Administration & Society*, 39(3), 354–381. doi:10.1177/0095399707300703
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information Communication and Society*, 22(14), 2081–2096. doi:10.1080/1369118X.2018.1477967
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE Publications.
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information Communication and Society*, 20(1), 14–29. doi:10.1080/1369118X.2016.1154087
- Kraemer, F., van Overveld, K., & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology*, 13(3), 251–260. doi:10.1007/s10676-010-9233-7

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. doi:10.1038/nature14539 PMID:26017442
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy & Technology*, 31(4), 611–627. doi:10.1007/s13347-017-0279-x
- Lorenz, L., Meijer, A., & Schuppan, T. (2021). The algocracy as a new ideal type for government organizations: Predictive policing in Berlin as an empirical case. *Information Polity*, 26(1), 71–86. doi:10.3233/IP-200279
- Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: A systematic review. *Journal of the American Medical Informatics Association: JAMIA*, 24(2), 423–431. doi:10.1093/jamia/ocw105 PMID:27516495
- Maciejewski, M. (2017). To do more, better, faster and more cheaply: Using big data in public administration. *International Review of Administrative Sciences*, 83(1, suppl), 120–135. doi:10.1177/0020852316640058
- Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: A quasi- paradigm for government on the Web. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1987). 10.1098/rsta.2012.0382
- Meijer, A. (2018). Datapolis: A public governance perspective on “smart cities”. *Perspectives on Public Management and Governance*, 1(3), 195–206. doi:10.1093/ppmgov/gvx017
- Meijer, A., & Grimmelikhuijsen, S. (2020). Responsible and Accountable Algorithmization, How to generate citizen trust in governmental usage of algorithms. In *The Algorithmic Society*. Routledge.
- Meijer, A., Lorenz, L., & Wessels, M. (2021). Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems. *Public Administration Review*. 10.1111/puar.13391
- Meijer, A., & Thaens, M. (2018). Urban technological innovation: Developing and testing a sociotechnical framework for studying smart city projects. *Urban Affairs Review*, 54(2), 363–387. doi:10.1177/1078087416670274
- Meijer, A., & Wessels, M. (2019). Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration*, 42(12), 1031–1039. doi:10.1080/01900692.2019.1575664
- Miller, S. M., & Keiser, L. R. (2020). Representative bureaucracy and attitudes toward automated decision making. *Journal of Public Administration Research and Theory*. 10.1093/jopart/muaa019
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 205395171667967. doi:10.1177/2053951716679679
- Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard University Press.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). Crown.
- Osborne, S. P. (2006). The new public governance? *Public Management Review*, 8(3), 377–387. doi:10.1080/14719030600853022
- Oswald, M. (2018). Algorithm-assisted decision-making in the public sector: Framing the issues using administrative law rules governing discretionary power. *Philosophical Transactions - Royal Society. Mathematical, Physical, and Engineering Sciences*, 376(2128), 20170359. doi:10.1098/rsta.2017.0359 PMID:30082305
- Parnas, D. L. (1994, May). Software aging. In *Proceedings of 16th International Conference on Software Engineering* (pp. 279-287). IEEE.
- Peeters, R. (2020). The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity*, 25(4), 1–16. doi:10.3233/IP-200253
- Peeters, R., & Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture – The case of the Dutch civil registry’s master data management system. *Government Information Quarterly*, 35(2), 175–183. doi:10.1016/j.giq.2018.02.003

- Power, D. J. (2016). "Big Brother" can watch us. *Journal of Decision Systems*, 25(sup1), 578–588. 10.1080/12460125.2016.1187420
- Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: A quasi- experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology*, 12(3), 347–371. doi:10.1007/s11292-016-9272-0
- Schillemans, T. (2016). Calibrating Public Sector Accountability: Translating experimental findings to public sector accountability. *Public Management Review*, 18(9), 1400–1420. doi:10.1080/14719037.2015.1112423
- Shepperd, M., Bowes, D., & Hall, T. (2014). Researcher bias: The use of machine learning in software defect prediction. *IEEE Transactions on Software Engineering*, 40(6), 603–616. doi:10.1109/TSE.2014.2322358
- Simon, H. A. (1947). *Administrative behavior: A study of decision-making processes in administrative organizations*. Simon & Schuster.
- Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, 1(1), 161-176.
- Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51(5), 991–1006. doi:10.1006/ijhc.1999.0252
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324. doi:10.1016/S1364-6613(03)00135-9 PMID:12860191
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78(5), 853–870. doi:10.1037/0022-3514.78.5.853 PMID:10821194
- Tummers, L., & Bekkers, V. (2014). Policy implementation, street-level bureaucracy, and the importance of discretion. *Public Management Review*, 16(4), 527–547. doi:10.1080/14719037.2013.841978
- Twizeyimana, J. D., & Andersson, A. (2019). The public value of E-Government – A literature review. *Government Information Quarterly*, 36(2), 167–178. doi:10.1016/j.giq.2019.01.001
- Valle-Cruz, D., Alejandro Ruvalcaba-Gomez, E., Sandoval-Almazan, R., & Ignacio Criado, J. (2019). A review of artificial intelligence in government and its potential from a public policy perspective. *20th Annual International Conference on Digital Government Research on - Dg.o 2019*, 91–99. doi:10.1145/3325112.3325242
- Van der Voort, H. G., Klievink, A. J., Arnaboldi, M., & Meijer, A. J. (2019). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly*, 36(1), 27–38. doi:10.1016/j.giq.2018.10.011
- Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–14. doi:10.1145/3173574.3174014
- Weller, A. (2019). Transparency: Motivations and challenges. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Müller (Eds.), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (Vol. 11700, pp. 23–40). Springer International Publishing. doi:10.1007/978-3-030-28954-6\_2
- Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18. doi:10.1145/3351095.3372833
- Willis, J. J., Mastrofski, S. D., & Weisburd, D. (2007). Making sense of compstat: A theory- based analysis of organizational change in three police departments. *Law & Society Review*, 41(1), 147–188. doi:10.1111/j.1540-5893.2007.00294.x
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, 42(7), 596–615. doi:10.1080/01900692.2018.1498103
- Young, M. M., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: A framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance*. Advance online publication. doi:10.1093/ppmgov/gvz014

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019a). Algorithmic decision-making and the control problem. *Minds and Machines*, 29(4), 555–578. doi:10.1007/s11023-019-09513-7

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019b). Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology*, 32(4), 661–683. doi:10.1007/s13347-018-0330-6

Zheng, Y., Yu, H., Cui, L., Miao, C., Leung, C., & Yang, Q. (2018, April). SmartHS: An AI platform for improving government service provision. *Thirty-Second AAAI Conference on Artificial Intelligence*.

Zouridis, S., van Eck, M., & Bovens, M. (2020). Automated discretion. In T. Evans & P. Hupe (Eds.), *Discretion and the Quest for Controlled Freedom* (pp. 313–329). Springer International Publishing. doi:10.1007/978-3-030-19566-3\_20