Perspectives on Public Management and Governance, 2022, **5**, 232–242
https://doi.org/10.1093/ppmgov/gvac008
Advance access publication 28 March 2022
Article

Legitimacy of Algorithmic Decision-Making: Six Threats and the Need for a Calibrated Institutional Response

Stephan Grimmelikhuijsen*, Albert Meijer*

*Utrecht University School of Governance, The Netherlands
Address correspondence to the author at s.g.grimmelikhuijsen@uu.nl.

Abstract

Algorithmic decision-making in government has emerged rapidly in recent years, leading to a surge in attention for this topic by scholars from various fields, including public administration. Recent studies provide crucial yet fragmented insights on how the use of algorithms to support or fully automate decisions is transforming government. This article ties together these insights by applying the theoretical lenses of government legitimacy and institutional design. We identify how algorithmic decision-making challenges three types of legitimacy—input, throughput, and output—and identify institutional arrangements that can mitigate these threats. We argue that there is no silver bullet to maintain legitimacy of algorithmic government and that a multiplicity of different institutional mechanisms is required, ranging from legal structures and civic participation to closer monitoring of algorithmic systems. We conclude with a framework to guide future research to better understand the implications of institutional design for the legitimacy of algorithmic government.

Introduction

Machine-learning techniques are increasingly used to support or automate decision-making processes in government (Margetts and Dorobantu 2019; Vogl et al. 2020). For instance, police forces all across the globe are implementing predictive policing systems to predict where patches of high-impact crimes are likely to occur (Meijer and Wessels 2019) and welfare distribution is being supported by algorithms that use a broad set of classifiers and big data to predict fraudulent behavior (Zouridis et al. 2020). On the one hand, such machine-learning algorithms promise better government decision-making (Kahneman, Sibony, and Sunstein 2020) and eventually a more effective and efficient government (e.g., Pencheva et al. 2020), yet at the same time there is concern about the impact of their use on the balance of power and public values.

Many scholars now worry whether the rise over government algorithmic decision-making could erode state legitimacy. Legitimacy is critical to governments as it provides a so-called "reservoir of goodwill", as a basis for the acceptance of government decisions (Easton 1975). Some authors highlight possible threats of algorithmic decision-making for democratic rule: a strong focus on efficient and accurate algorithms may give rise to a democratically unchecked rule of experts (e.g., Sætra 2020; Zouridis et al. 2020). Also, scholars have linked these legitimacy concerns to a misalignment of values. Government algorithmic decision-making strongly emphasizes a culture of technical rationality, which emphasizes effectiveness and efficiency over ethical or normative concerns (Young et al. 2021). Finally, scholars in public administration are concerned about how algorithms might harm core governance principles such as public accountability and transparency (Busuioc 2020; Giest and Grimmelikhuijsen 2020). Indeed, scarce empirical research into this topic has already indicated that algorithmic decision-making systems are perceived as less legitimate than systems in which humans are involved (Starke and Lünich 2020).

These various studies yield important yet scattered insights on how algorithmic decision-making could threaten government legitimacy. In this article, we will use the concept of government legitimacy as a lens to systematically take stock of these threats to three components of legitimacy: input, throughput, and output-based legitimacy. In other words, we will discuss how algorithmic decision-making lacks democratic decision-making and oversight (threatening input legitimacy), violates with procedural and legal requirements (threatening throughput legitimacy) and potentially produces outcomes that misalign with public values (threatening output legitimacy).

Our systematic analysis of threats will serve as the basis for proposing adjusted and new mechanisms to that algorithmic decision-making is aligned with democratic, procedural, and public values. These mechanisms do not focus on technological features but are institutional arrangements. According to Scharpf (1997, p. 38) institutions concern "systems of rules that structure the course of actions that a set of actors may choose". These rules may be formal and explicit and codified in legislation, but can also be informal rules-in-use by actors involved in the development and implementation of algorithmic decision-making. The institutional mechanisms, we will argue, are needed to respond to the various legitimacy threats. In short, we will discuss the following questions in this article:

1. In what way can the use of algorithms by government organizations threaten the legitimacy?
2. What institutional mechanisms are needed to safeguard the legitimacy of the use of algorithms by government organizations?

By analyzing threats and institutional mechanisms through a legitimacy lens, we aim for two contributions to the current debate. First, we offer a more precise conceptualization of existing debates on algorithmic decision-making in government by distinguishing threats for *input*, *throughput,* and *output* legitimacy, and by structuring debates along these lines. Second, we not only analyze threats but also offer strategies in all these three areas to counter legitimacy threats posed by algorithmic government. The rise of algorithms constitutes an important change in the organizational environment of governments. Organizational theorists, then, describe the term "strategy" broadly as the major decisions needed for organizations to "maintain an effective alignment with its environment." (Miles et al. 1978, p. 547). This is a complex process and in this article we highlight that no silver bullet exists; instead a wide range of interventions are required to maintain legitimacy. This entails reassessing the existing methods of safeguarding the legitimacy of the government. In our conclusion, we will describe how institutional design can provide a theoretical lens to better understand how we can get "the institutions right" (Ostrom 1990) in order to maintain government legitimacy in the algorithmic age.

The article is structured as follows. We begin by defining the concept of algorithmic government based on public administration and computer science literature. The literature centers on the idea of algorithmic government as a new method of organization around the use of an algorithm and not just the use of a new tool within the existing organization. Following that, we discuss the concept of legitimacy, distinguishing between three routes towards legitimate government: input, throughput, and output legitimacy. We use this distinction to discuss the various threats to the legitimacy of algorithmic government and then to identify strategies to strengthen its legitimacy. We end with conclusions and connect the legitimacy strategies to the idea of institutional design and calibration to ensure the legitimacy of algorithmic government.

Algorithmic Government: Complexity, Opacity, and Interdependence

In recent years, we have seen that these machine-learning algorithms are being introduced to support not just routine but also knowledge-intensive government tasks (Young et al. 2019; Zouridis et al. 2020). Machine-learning algorithms can be described as algorithms that can learn without being explicitly programmed (Samuel 1959). This is where the crucial difference lies with the previous forms of technological developments: knowledge-intensive tasks are being taken over by modern algorithms with a certain degree of artificial intelligence. Currently, the use of machine-learning algorithms is still limited, but it is expected that these technologies will play a more important role in government for the years to come (Vogl et al. 2020).

Governments have become highly dependent on regular and machine-learning algorithms for the execution of its core tasks. This shows the contours of what we argue to be *algorithmic government*. Algorithmic government can be thought of as the most recent manifestation of the broader phenomenon of digital government development (Meijer and Grimmelikhuijsen 2020). Digital government involves the use of modern information and communication technologies to support government in all of its facets (West 2005). Here, we define algorithmic government as:

> the use of machine-learning algorithms for a range of government processes, such as decision-making, service provision, and policymaking.

In this article, we predominantly focus on machine-learning algorithms in government. Since there are various studies in public administration referring to machine learning, algorithms, or Artificial Intelligence (e.g., Peeters 2020; Young et al. 2019), it is important to be precise in our definition and typology on what technologies are available and how they differ from "traditional" algorithms.

Artificial Intelligence is an umbrella term for many types of algorithms, of which machine-learning techniques are one. Algorithms are simply calculation rules, and the use of algorithms is nothing new compared to "ordinary" forms of human decision-making or automated decision-making processes (Hill 2015). Machine-learning algorithms, then, are different from traditional statistical modeling as there is no formalization or a priori theorization of relationships between variables (Athey and Imbens 2019). A further discussion of the various types of machine-learning is presented in Box 1.

Box 1: Four Types of Machine-Learning

Machine-learning algorithms come in various shapes which need be understood understand their characteristics and consequences for government decision-making. Here we outline a couple of commonly distinguished types of machine learning (cf. Guidotti et al. 2018).

Supervised learning: A human predetermines the relevant categories and labels in a training dataset. Based on this training data, the algorithm learns which categories are associated with which outcome. A supervised algorithm will be fed with new, larger, datasets which helps to make more and more accurate predictions. An example is an AI assistant (chatbot) that text typed by citizens in the chat to classify if this citizen should file a report with the police for online sales fraud (Odekerken and Bex 2020). Categories that are used to predict whether to report or not are number of days after initial transaction, payment status, and the reputation of the online retailer. Because of the human supervision it is often relatively easy to track model performance and error.

Unsupervised learning: there is no human involved in a priori categorizing or classifying the training data for the algorithm. The algorithm analyzes unlabeled training data and seeks for patterns and clusters in the data. A well-known example is the recommendation algorithms used by YouTube and Netflix. The algorithm learns to detect patterns and clusters in viewing behavior and uses this to provide recommendations. Since there is no pre-labeling it is relatively hard to track model performance and error (Mohri et al. 2018).

Reinforcement learning: here an algorithm is directly rewarded (or punished) for making a correct or incorrect prediction. The algorithm learns by maximizing rewards over the course of various actions and iterations. Reinforcement learning is often used in robotics (Kober et al. 2013) and has most gained

fame by exceeding human performance in complex games such as AlphaGo (Silver et al. 2016).

Deep learning: one of the more "mysterious" forms of machine learning is through deep learning, which is used to analyze large unstructured data, such as images and video. Through various hidden layers in a network of nodes, the algorithm continuously learns to distinguish and recognize different features of an image. For instance, one layer in the network learns to recognize the whiskers of the cat, the other layer the contours of its paws, and another the color of its tail. The logic of these layers does not follow the reasoning logic of humans (Burrell 2016). Deep learning is often combined with reinforcement or unsupervised learning. An example of a real-world application of deep learning is face-recognition software.

In the remainder of the section, we highlight three common characteristics of these machine-learning algorithms that shape algorithmic government and which generate news threats to government legitimacy: complexity, opacity, and interdependence. We need to understand these characteristics to appreciate why algorithmic decision-making calls for new institutional arrangements.

Complexity relates to both the complex technological structures of unsupervised machine learning. A common usage of this kind of algorithm is to aid human judgments on how probable a certain outcome is based on profiling and detected patterns and correlations (e.g., Zarsky 2016). Making such predictions requires more complex algorithms and more complex (big) data to be sufficiently accurate. This is a difference with how "regular" algorithms were used in government in which algorithms were programmed by human beings in an IF-THEN decision structure. For instance, the information system of the Education Executive Agency in the Netherlands determines whether or not a student is eligible using IF-THEN decision trees. In contrast, newly introduced algorithms use much more complex formulae and more complex data to predict the risk of some undesired behavior, such as recidivism (Kleinberg et al. 2017).

Opacity is partially the result of this complexity, and partly the result of deliberate policy. Opacity relates to the continuous adaptation of machine-learning algorithms based on supervised or unsupervised learning. Algorithms that evolve through unsupervised learning, and especially this that use neural networks (deep learning) is particularly difficult to understand what the decisive variables are for producing a certain outcome (Burrell 2016). While opacity is related to the complex technological structure, sometimes it is a deliberate choice to restrict transparency. For instance, private companies might restrict access to protect property rights and government might limit access out of fear that subjects "game the system" (Mittelstadt et al. 2016). We will discuss this in more detail as a threat to government legitimacy in the section on "Throughput legitimacy".

Finally, *interdependence* relates to the fact that machine-learning algorithms use a range of different datasets, not only from one's own organization but also from other organizations, in addition to open datasets. This means that the design of external data sets influences the outcome of the algorithm (Cicirelli et al. 2019; Zouridis et al. 2020). This not only further impedes opacity but can also give rise to numerous questions about liability and responsibility for the final decisions made. Furthermore, external dataset may have been collected for a different purpose than what it is eventually used for in machine learning, complicating matters further.

Overall, in this section we highlighted that machine-learning algorithms are increasingly used by government to aid decision making and service delivery and in doing so we witness the emergence of algorithmic government. We have highlighted that algorithmic government is characterized by complexity, opacity, and interdependence. These features transform the nature of decision making in government and they generate, as we will argue below, a series of threats to government legitimacy. Before discussing these implications, we first define and conceptualize government legitimacy.

Defining and Conceptualizing Government Legitimacy

Legitimacy is a crucial concept, which ultimately involves accepting the authority of government. In this article, we have adopted Suchman's much-used definition (1995, 574): *"Legitimacy is a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions"*.

This definition highlights a few key elements of legitimacy. First, legitimacy is "generalized": it oversteps perceptions of a single incident. This means that legitimacy is not built overnight but takes time to emerge. Second, Suchman refers to "perception or assumption" in his definition. In this article, we focus on this subjective side of legitimacy meaning that while legitimacy may flow from properties of government organizations, it is essentially about how government processes and actions are perceived by the citizenry. So, legitimacy is generated if the actions of government are deemed appropriate or desirable, in other words, when government actions are in line with the shared norms and values of a community.

Political scientists highlight various ways in which this "generalized perception" can be achieved. The traditional model of Easton (1965) on the functioning of the political system describes this process of legitimacy creation. Easton argues that the political system in a fundamental sense entails converting inputs (citizens' preferences) to outputs (policy). This conversion takes place in government organizations using organizational processes (throughput).

Scharpf (1999) and Schmidt (2013) demonstrate how this model can be used to understand the various routes to legitimacy: input, throughput, and output legitimacy (see table 1). Scharpf (1999) argues that input legitimacy relates to the acceptance of authority based on an appropriate democratic process offering room for an open debate and elections that have proceeded in the correct manner, and that output legitimacy relates to the acceptance of authority based on the results achieved by governments, such as safety, prosperity, and peace. Schmidt (2013) adds that throughput legitimacy should be understood as a separate route, by correctly translating the input into policy processes. We will use these three paths to legitimate government to identify the threats to the legitimacy of algorithmic government and to provide insights into the strategies for strengthening legitimacy.

The traditional route to legitimacy proceeds through input legitimacy. Scharpf (1999: 7–21) states that this relates to the

extent to which the starting point of government processes meets citizens' preferences. Democratic processes therefore constitute the core of input legitimacy. First and foremost, this concerns the quality of representative democracy, in other words, the functioning of the electoral system and Parliament. The basic idea is that fair elections and a properly functioning Parliament contribute to the legitimacy of the political system. Furthermore, direct democracy (e.g., participation) is equally important. The idea is that better participation opportunities strengthen the legitimacy of policy. Based on this route, we argue that the legitimacy of algorithmic government increases the better citizens' preferences have been translated through democratic processes into the design and the use of an algorithm by a government organization.

As stated above, based on his analysis of the European Union, Scharpf (1999) has added another route to legitimacy: output legitimacy. This form of legitimacy ultimately centers on the ability of the political system to solve citizens' problems and to generate desirable results. Scharpf (1999) emphasizes that many political scientists primarily analyze the quality of democracy but adds that, ultimately, what the political system actually delivers matters a great deal to citizens. He points out that the European Union therefore has a certain legitimacy due to its contribution to peace and safety in Europe. The same argumentation can be used to look at other political systems. Legitimacy, then, depends on what an authoritative institute ultimately delivers. Based on this route, we argue that the legitimacy of algorithmic government increases when the outcomes of the use of an algorithm contribute to the realization of values that citizens consider important.

Finally, Schmidt (2013) states that throughput—the manner in which citizens' preferences are translated into policy—is also important for legitimacy. She refers to the government's "black box," the gap between input and output that is often ignored by political scientists. She states that not just the accuracy but also the transparency of the throughput, the receptiveness to participation, and the accountability for organizational processes are crucial to the legitimacy of the political system (Schmidt and Wood 2019). Based on this route, we argue that the legitimacy of algorithmic government increases the more the manner in which outcomes are achieved through the algorithm meet the requirements imposed on this by citizens, elected representatives, and constitutional institutions.

These three routes to legitimacy will be used to analyze systematically the relevant conditions for the legitimacy of algorithmic government and, on that basis, what threats may exist. In this article, we identify threats based on an analysis of relevant literature. Next, we apply our own analytical reasoning to match each threat with a strategy for mitigating this threat.

Input Legitimacy: Threats and Strategies for Democratic Algorithms

A threat to the input legitimacy of government arises because the relationship between the political mandate—the direction provided by the political leaders to the executive government organization—and the use of algorithms in government is not clear. The public administration literature pays considerable attention to the connection between politics and administration, in which government becoming detached from the political leaders is deemed a high risk (Demir and Nyhan 2008). The core idea is that the translation of citizens' preferences into administrative action takes place in this connection: a majority in Parliament results in support for a political leader, who then provides direction to the implementation of government policy. However, due to the technologically complex nature of algorithms, there is a risk that an adequate translation will not take place here and that algorithms will only be defined based on technical and administrative considerations rather than on the basis of a political process. We first discuss the two main threats and subsequently accompanying strategies to remedy them (table 2).

Table 1. Routes to the Legitimacy of Algorithmic Government

Route	Legitimacy of Algorithmic Government
Input legitimacy	The legitimacy of algorithmic government increases if citizens' preferences have been translated well through democratic processes into the design and the use of the algorithm.
Throughput legitimacy	The legitimacy of algorithmic government increases through the manner in which outcomes are achieved, such as by adhering to legal and fair process requirements.
Output legitimacy	The legitimacy of algorithmic government increases if the outcomes of the use of an algorithm contribute to the realization of values that citizens consider important.

Table 2. Summary of Threats and Strategies Concerning Input Legitimacy of Algorithmic Government

Threats to Input Legitimacy	Strategies to Strengthen Input Legitimacy
1. Erosion of democratic control on algorithmic decision-making: ◦ Implicit political decisions by algorithm developers remain "under the radar". ◦ Privatization of decision-making with outsourced algorithms. ◦ Algorithms are continuously evolving, outside of democratic oversight. 2. Limited responsiveness of algorithmic decision-making: ◦ No civic participation in algorithmic design.	1.Strengthening democratic control on algorithmic decision-making: ◦ Strengthening political sensitivity of public servants in charge of algorithmic design. ◦ Strengthening the democratic control on the purchase of third-party (commercial) algorithms. ◦ Using explainable AI (XAI) to explain how algorithms evolve. 2. Strengthening responsiveness of algorithmic decision-making: ◦ Increasing civic participation in the design process and in the monitoring of algorithms

The first threat is the erosion of democratic control on algorithmic decision-making. The first part of this problem is that political oversight is lacking both in the executive branch of government (e.g., responsible minsters in the administration) and in the legislative branch (e.g., members of Parliament). The lack of knowledge in both branches of government makes it extremely difficult for them to verify whether the algorithms that have been developed actually meet the preferences of citizens (Köning and Wenzelburger 2021). Citizens can, for example, elect representatives because they want proper supervision to ensure the honest and correct provision of welfare assistance. However, if welfare assistance is a largely automated process based on algorithms, elected officials cannot fulfill this task.

In addition, politically loaded (value based) decisions are made implicitly by developers and programmers of algorithms. In his classic work on the significance of information systems for legal practices in the public sector, Lessig (1999) had already stated that legal requirements are implemented implicitly in computer programs and that the use of these programs has implications for the decisions of government. He describes this as "Code is Law", and this means that the developer assumes the role of the regulator. At the same time, those who develop or technically manage the development of computer programs are often insufficiently aware of the fact that the technical rules they program have a political-administrative meaning (Zouridis et al. 2020).

Furthermore, democratic control is further undermined when government purchase algorithms from commercial third parties who refuse to make their algorithm accessible to protect their intellectual property (Brayne 2020, p. 135; Mittelstadt et al. 2016). Eventually this leads to further outsourcing of government decision-making to private actors outside of democratic control. Finally, the new generation of machine-learning algorithms is not static but is dynamic and therefore continuously evolving. Even if a machine-learning algorithm was being monitored at the start it can evolve over time outside of traditional democratic control (Burrell 2016; Desai and Kroll 2017; Lepri et al. 2018). This problem is compounded by the dependence of the organization on external datasets.

The second threat for input legitimacy regards limited responsiveness to the needs of citizens in algorithmic development. In policy areas such as urban planning, citizens' participation in decision making is formally organized and public consultation meetings are held. In the Netherlands, participation and client councils exist that ensure citizens can participate in decision-making affecting them. (Michels and De Graaf 2010). In the United States, town hall meetings are organized to organize civic participation on hot button local issues. Such mechanisms for participation hardly exist with regard to governmental usage of algorithms, even though such algorithms directly affect various groups of citizens—an example is the SyRI algorithm for the detection of welfare fraud (Van Schendel 2019). König and Wenzelburger (2021, p. 3) highlight that algorithmic decision-making systems allow for "linking decision parameters to aggregate outcomes of decision-making ex ante", which prevents values and voices from being heard early on in the development new systems. Eventually this leads to a technocratic focus and as such a deficit in terms of input legitimacy.

We propose two main strategies: strengthening democratic control and responsiveness. To strengthen democratic control, we first need to improve the political sensitivity in the design of algorithms. The political sensitivity of public servants relates to the ability to distinguish between situations in which administrative decision-making suffices and situations in which political decision-making is needed ('t Hart and Wille 2006). Administrative decision-making suffices where situations are deemed routine and where interventions can be made based on the existing frameworks. However, political decision-making is necessary in matters that require a new assessment of values. In developing predictive policing systems, questions on bias, for example, are highly sensitive and need political decision-making. For the administrative management of the design processes, it is important to be able to identify which aspects of algorithms are so sensitive that they must be submitted—in a comprehensible manner—to political decision-makers.

We also need to intervene on the other side of the politics-administration dichotomy: strengthening capacity to exercise political oversight of algorithms, especially when commercial algorithms are considered. The technological complexities compound the "classic" difficulties of political oversight over policy implementation in large organizations or administrative systems (Kaufman 1960; Pressman and Wildawsky 1984). The political involvement in highly invasive algorithmic systems was extremely low for a long period of time, because the structure of the system was regarded as an implementation issue. It is important to strengthen the capacities of political leaders such as ministers or city managers so that they have the ability to identify and manage politically sensitive issues, such as privacy, discrimination, or data ownership. This requires at least basic knowledge of technical processes as well as support to facilitate political oversight over the actual implementation of algorithms in government processes.

Furthermore, as a response to the continuous evolvement of algorithms outside of democratic control there could be demands about the explainability of machine-learning algorithms. For instance, machine-learning algorithms can be programmed in a way they can explain based on what data and which decision rules certain predictions were reached, this is called explainable AI (xAI). xAI refers to the explainability of specific decisions, but can also refer to how a model works and develops (e.g., Miller 2019). Such transparency allows us to check the algorithm to judge if the model work based on desired parameters.

Finally, to counter the second threat to input legitimacy (lack of responsiveness in algorithmic design), there needs to be more attention for citizen participation in developing algorithmic systems in government (König and Wenzelberger 2021). The reasoning behind this is largely the same as that of participation in spatial plans or client participation in education, healthcare, or housing (Michels and De Graaf 2010). Through participation in the design process of algorithms, citizens will be able to indicate immediately where there are sensitive issues and what the main concerns are. Sharing the code on platforms such as Github, which recently occurred during the development of the Dutch coronavirus tracking app, CoronaMelder, can contribute to this. This participatory process led to an app with strong privacy safeguards. Direct participation may prevent a technocratic focus when new systems are developed and implemented (König and Wenzelberger 2021).

Throughput Legitimacy: Threats and Strategies for Procedural Fairness

Throughput legitimacy is based on the strength and fairness of procedures for translation democratic input into policy processes (Schmidt 2013). A core element for throughput legitimacy is that it contributes to fair procedures in government decision-making. Such procedures are crucial in bestowing legitimacy to a decision-maker. Tyler (2006) have developed the concept of procedural fairness and throughout the past decades many studies have shown that people's perceived fairness of decision-making procedures affects their overall trust in authority and decision outcomes. Central to procedural justice theory is the relation between how authorities use their power and how subordinates assess their claims of power. When decision-making power is exercised in a way that is procedurally fair, it is more likely that the decision is acceptable and the decision-making authority is trusted (Sunshine and Tyler 2003; Tyler and Huo 2002). Algorithmic government threatens this aspect of procedural fairness by potential infringement of citizens' privacy and a poor translation of legal procedural requirements into code. Not only does algorithmic government threaten procedural fairness, there is also a risk for checks and balances on government decision-making. Furthermore, there is a threat that it is no longer clear who is responsible for algorithmic decisions or algorithmically supported decisions (table 3). We will explain these threats in more detail below.

First, a frequently discussed threat concerns the infringement of citizens' privacy by algorithms, especially their right to protection of personal data (Young, Katell, and Krafft 2019). The application of algorithms requires a lot of data from different sources, which are connected and processed to, for instance, calculate risk scores. A large amount of such data has never been gathered for this purpose. On top of that, the combinations of different datasets may reveal new insights into individuals (Mergel, Rethemeyer, and Isett 2016). Furthermore, individuals often do not know what data exactly are processed and in what manner. Citizens' privacy may even be harmed if anonymized data are used, because the combination of data derived from different data sets in turn can lead to the unique identification of individuals.

In addition, algorithmic decision-making threatens procedural fairness because legal requirements that apply to government decision-making might be circumvented when algorithms are used. These legal requirements—primarily the requirements concerning fundamental rights and other public values, such as the right to non-discrimination and the right to legal protection (Zarsky 2016)—are crucial to the lawful functioning of algorithms. Algorithms are designed by technical experts who often have limited knowledge of legal requirements, which poses the risk of developing a technically accurate algorithm that nonetheless fails to comply with basic human rights (Livingston and Risse 2019). A complicating factor is that legal requirements themselves can be ambiguous and unclear, which makes the conversion to an algorithmic rule complex or, in certain cases perhaps, fundamentally impossible.

The second overarching issue concerns the lack of checks and balances to track algorithms used in government decision-making because they lack basic transparency (Giest and Grimmelikhuijsen 2020). For instance, a recent study in Dutch government agencies shows that decision rules in algorithms as well as programmers' assumptions are often invisible to both the person who makes a decision with the assistance of an algorithm and the controlling bodies (Zouridis et al. 2020). The obscurity of complex systems is an important point in discussions on information systems in the public sector (Meijer 2009). With regard to algorithms, it is specifically emphasized that the decision-making rules of algorithms are continuously evolving through machine learning and as a result are also no longer known by the designers (Lepri et al. 2018). Furthermore, commercial parties holding intellectual property rights may preclude publication of the code or model (Mittelstadt et al. 2016). Finally, algorithmic transparency may concern technical aspects of how the algorithm works but could also relate to organizational transparency, for example, policies and safeguards on how algorithms are used and developed at the organizational level (Meijer and Grimmelikhuijsen 2020).

Another element that limits checks and balances is the blurring of responsibility in the complex relations arising from the use of algorithms (De Fine Licht and De Fine Licht 2020; Meijer 2009). You could say that the person who uses the algorithm in a decision remains responsible, but this becomes more complicated if this person is not able to fully fathom the algorithm. The question then arises whether the developer of the system remains responsible. Furthermore, the system continues to evolve on the basis of machine-learning processes. Perhaps those who "train" the algorithm should then remain responsible, or those who manage the data sets used to train the algorithm, or a combination of the two. In view of the complexity, it is sometimes even argued that the algorithms should be regarded as autonomous actors and that therefore they bear responsibility. In addition, the dependence on datasets of other organizations for training algorithms in effect means that actors outside of the realm of the organization who are responsible for external datasets directly affect the algorithm. In these complex relations, responsibilities become blurry, to the extent we risk that ultimately no one can really be held responsible (Bovens and Zouridis 2002; Meijer 2009).

We propose two strategies to mitigate threats to procedural fairness (table 3). The first is to conduct a Data Protection Impact Assessment (DPIA) in cases where algorithms play a consequential role in government decision-making. A DPIA is an organizational tool that is used to identify the privacy risks associated with a data processing operation beforehand so that better decisions can be made as to whether and how an algorithm can be used (Bu-Pasha 2020). Under the European Union's overarching legal framework (the General Data Protection Regulation), conducting a DPIA is sometimes legally required when organization "systematically and extensively evaluate personal aspects relating to natural persons based on automated processing, including profiling, and on which they base decisions that produce legal effects concerning natural persons" or "process special categories of personal data on a large scale or personal data relating to criminal convictions."

One way to improve the translation of legal requirement into algorithmic decision-making is to implement a human rights impact assessment or even a broader, general legal assessment of algorithmic government. Especially algorithms that are used in impactful decisions such an assessment can be valuable. Human rights impact assessment have been developed and used in many areas in government, such as in public

health policies (Gostin and Mann 1994). The legal assessment should be carried out when planning the use of an algorithm by a government organization. The assessment comprises a series of questions about specific legal requirements concerning the storage and public availability of data, and more broadly on the compliance of human rights. These questions are answered and result in a report providing an account of whether the algorithm and the manner in which it is used within the government organization complies with the legal requirements that may be imposed on it. The implementation of such assessments could potentially be supported by means of an independent knowledge center or center of expertise.

A logical response to the second main threat (check and balances) is to increase algorithmic transparency. On the one hand this can be done by increasing access to the data, model, and code (Giest and Grimmelikhuijsen 2020). The average citizen will not be able to do that much with accessibility, but external experts will be able to utilize it for a critical analysis. At the same time, full accessibility may not always be desirable due to privacy concerns or the risk that citizens will "game the system" (Mittelstadt et al. 2016). A second proposed component of algorithmic transparency is explainability. Explainability means that—even if a machine-learning algorithm is used in decision-making—it must be made clear that the substantial reasons for a decision are understandable and correct (Tutt 2017). We want to emphasize that algorithmic transparency does not relate solely to technical aspects of the algorithm but also to the manner in which it is used in the organization. De Fine Licht and De Fine Licht (2020) emphasize that the purpose is to make clear which decision has been made, based on which arguments and who is responsible for making it. Young, Katell, and Krafft (2019: 2) state that ways should be found to explain why algorithmic systems should be "legible" for policy makers and stakeholders. The expectation is that greater transparency creates possibilities for both decision makers and external parties to verify the algorithm. At the same time, we acknowledge that the Webs of dependency between organizations providing datasets for the algorithm make realizing comprehensive transparency complicated.

Finally, to further strengthen checks and balances of algorithmic decision-making, we need to clarify the responsibilities relating to algorithmic government (Busuioc 2020). Where do the responsibilities lie: policy-makers, street-level decision-makers, the algorithm developer, the suppliers of datasets, the party that maintains the algorithm, etc.? When implementing an algorithm in an organization, the responsibilities for the use and the outcomes of the algorithm should be clearly assigned and recorded in a document. The clarification of responsibilities increases accountability while reducing the probability of blame shifting between parties.

Output Legitimacy: Threats and Strategies for Public Value Creation

A threat to output legitimacy arises when the use of an algorithm fails to lead to the realization of values that citizens consider important. Output legitimacy is threatened if the use of algorithms results in outcomes regarded as poor performance by citizens, such as discrimination or the absence of human contact (table 4).

Here we identify two elements that contribute to the threat of inefficient and ineffective algorithmic decision-making. First, use of algorithms in government decision-making requires large upfront investments in expertise and technological infrastructure that leads to uncertain efficiency gains. Governments have a history for failed ICT projects in the past (Anthopoulos et al. 2016; Dada 2006) and there is no reason to believe that machine-learning projects are any different. The expectations for the use of algorithms are high, especially in terms of the effectiveness and efficiency of government processes, but firm evidence for the contribution is not clearly provided (Meijer and Wessels 2019). It

Table 3. Summary of Threats and Strategies Concerning Throughput Legitimacy of Algorithmic Government

Threats to Throughput Legitimacy	Strategies to Strengthen Throughput Legitimacy
1. Algorithmic decision-making does not meet standards of procedural fairness: ◦ Infringement of citizens' privacy ◦ Poor translation of legal requirements in algorithms 2. Checks and balances of algorithmic decision-making insufficient: ◦ Low level of transparency of algorithms ◦ Blurring of responsibilities related to algorithmic decision-making	1.Building safeguards for good algorithmic governance ◦ Data Protection Impact Assessment ◦ Legal assessment of algorithmic government 2. Building safeguards of checks and balances of algorithmic decision-making: ◦ Transparency of algorithmic government ◦ Clarifying responsibilities of algorithmic decision-making

Table 4. Summary of Threats and Strategies Concerning Throughput Legitimacy of Algorithmic Government

Threats to Output Legitimacy	Strategies to Strengthen Output Legitimacy
1. Algorithmic decision-making is ineffective and inefficient ◦ Algorithmic systems are costly and rarely deliver ◦ Cost and benefits shift over time due to gaming and function creep 2. Algorithmic decision-making leads to undesirable outcomes ◦ Bias in algorithmic systems ◦ Algorithms limit possibility for human contact	1.Increase attention for effectiveness and efficiency of algorithmic decision-making ◦ Compulsory cost–benefit analysis of the algorithm ◦ Periodic audits of the use of algorithms 2. Prevent undesirable outcomes ◦ Create guidelines to ensure the right to human contact ◦ Critical thinkers in development teams and algorithmic impact assessment

often seems that a trend is being followed and that there is a deep-seated belief in the technological possibilities rather than a clear business case.

An issue that also threatens the effectiveness and efficiency of algorithmic decision-making is that initial cost-benefit analyses may shift over time due to the adaptivity of both government and society. Here we distinguish *gaming* and *function creep*. Gaming may occur on the side of society. People may start to respond strategically that get more favorable outcomes from the system, or start the use their own smart apps to dodge government rules (Mittelstadt et al. 2016). For instance, motorists attempt to dodge speed checks using smart apps. A related issue is that machine-learning systems in government are highly receptive to function creep. In other words, these systems are eventually being used for other purposes than what they were set out to do. For instance, the implementation of a fingerprint-database for asylum seekers in the European Union was set out to prevent asylum seekers to go "asylum hopping" in various EU Member States. However, step by step the database was becoming used to also track illegal immigration and to uncover terrorist activities (Balzacq 2008; Broeders 2011). Although this example is from the "pre machine-learning era", the possibility of self-learning algorithms and more data will only increase opportunities for function creep.

Next to threatening efficiency there is a real risk that algorithmic decision-making creates undesirable outcomes. While some herald the potential of machine-learning algorithms to make government services more equitable and efficient (Pencheva et al. 2020), they has been heavily criticized for producing biased and even discriminatory predictions because of biased model parameters and/or biased data (Eubanks 2018). Often, human biases are consciously and unconsciously automated and integrated in automated decision-making. The types of bias referred to in the literature include the following (Jackson 2018; Williams, Brooks, and Shmargad 2018; Zarsky 2016): a focus on specific target groups (in some cases based on ethnic profiling), a focus on specific areas (in some cases precisely where certain groups of people live), and a focus on past performance rather than on the current and future situation (due to which demographic changes are insufficiently incorporated). Bias can arise as a result of using skewed data to train the algorithm (often data already containing a bias), selective data (and hence the bias already occurs when selecting the data and is inherited when training the algorithm), and incorrect analyses based on the data (in which certain patterns are misinterpreted). Algorithms are often trained with datasets from other organizations and these patterns of interdependence mean that biases that originate outside of an organization can be introduced into the algorithm.

The second force that could lead to undesirable outcomes is that the increasing use of algorithms often undermines the quality of human contact and human input in decision-making (Bovens and Zouridis 2002). Some scholars have even warned for "robotic bureaucracy", in which organizations increasingly use automated systems and automated replies to handle contact with humans (Bozeman and Youtie 2020). In its most far-reaching form, the algorithm will replace human beings, and citizens will only be dealing with an algorithm rather than a human being. This already applies to decisions that are adopted for large groups of people, such as student grants and loans or traffic fines (Zouridis et al. 2020). And even if there is a "human in the loop," the discretion of the human decision-maker may be limited by an algorithmically generated advice (Peeters 2020). The risk of automation bias occurs in this process: a human being relies too heavily on the algorithm, even in the face of erroneous predictions (Goddard et al. 2011).

A first mechanism to counter these threats to efficiency and effectiveness is to carry out mandatory cost-benefit analyses of algorithms. The analysis could form a component of the DPIA (see Section on Throughput Legitimacy). The use of algorithms is often driven by the big promise of technology, but it is essential to conduct a realistic assessment of such promises prior to implementing algorithms as well as of the financial and non-financial costs of the use of the algorithm. Experts in various fields and stakeholders should be involved in the cost-benefit analysis to ensure, above all, that the potential undesirable side effects are also clearly identified.

Related to these, periodic audits could strengthen output legitimacy. The purpose of the audit is to verify whether the use of the algorithm still produces the desired outcomes and precludes the undesirable outcomes. Preferably an external party should perform a periodic audit of the functioning of the algorithms, whether they serve the purpose for which they are deployed, whether the human dimension is taken into consideration, and other aspects (Tutt 2017).

Furthermore, to reduce undesirable consequences of algorithmic decision-making include ways to minimize bias in the development of the algorithm (Baer 2019). To minimize the bias, it is important to involve critical opponents from diverse backgrounds in project teams, who can constantly identify the specific forms of bias that can occur. Moreover, when using algorithms, it is vital to measure various potential forms of bias. One technical way to address this issue is to use machine learning tools that are trained to detect and highlight potential biases (Mehrabi et al. 2019).

Furthermore, government may want to set out more detailed rules on the right to human contact. Particularly in cases where algorithms are used to make decisions on individual citizens, it is vital that citizens are not only referred to a Web site for an explanation of a decision but can actually communicate with a human decision-maker. Scholars argue, based on existing legal requirements in EU countries, that citizens have a right to human intervention to contest automated decisions (Bayamlıoğlu 2021; Wachter and Mittelstadt 2019).

The use of machine-learning algorithms in the public sector may still be limited, it is increasing rapidly and could have a significant impact on the functioning of government organizations (Vogl et al. 2020). Algorithmic government potentially offers various possibilities for strengthening the effectiveness of the execution of tasks, but given the specific characteristics of the technology, the complexity, opacity, and interdependence, it also raises fundamental questions about the legitimacy of government.

Calibrating Institutional Design for Algorithmic Government

A general message is that there is no silver bullet to safeguarding legitimacy of algorithmic government and that a multiplicity of different institutional arrangements is required. In other words, redesign of institutions is needed to ensure legitimate government algorithmic decision-making. However, as Ostrom already noted over 30 years ago "'getting

the institutions right' is a difficult, time-consuming, conflict-invoking process" (Ostrom 1990, p. 14). In this final section, we provide a broad outline of this process of institutional design and redesign and how, at an abstract theoretical level, we can get there.

Institutional design can be described as making intentional changes in institutional characteristics, which should be differentiated from slow and emergent and unconscious changes in institutions that always occur (Ostrom 1990). According to Klijn and Koppenjan (2006, p. 149), institutional design is the "deliberate attempt to change the set of rules that structures interactions" between actors. Such deliberate attempts can be aimed at changing formal or informal rules that together alter the institutional make-up of government algorithmic decision-making (see Klijn and Koppenjan 2006) (figure 1). There is some overlap between the terms "institutional design" and "strategy"—which we described as making decisions to adapt organizational internal structures to a changing environment (Miles et al. 1978). Here, we use theory on institutional design as a theoretical lens to better understand how organizational strategies can help to "get the institutions right" in order to maintain government legitimacy.

1. Changing formal and informal access composition or access rules. Allowing a broader or more limited set of actors. A different set of actors will affect their interactions and possibly outcomes (see also under 2 and 3). In the case of algorithmic decision-making access rules include providing access to civic participation, allowing the voices and values of non-technical experts to be given a place in the development and implementation of algorithmic systems. Also formalizing and strengthening democratic control is a way of involving elected officials in this process.
2. Changing formal and informal interaction rules. Influencing interactions between actors in a sustainable way. For instance, by introducing procedural standards to facilitate interaction/conflict resolution. Applied to algorithmic decision-making this includes strategies to change legal safeguards and transparency allowing for different—fairer—procedures.
3. Changing formal and informal outcome rules. This institutional design feature involves changing evaluation criteria and pay-off structures. Applied strategies include to the use of cost-benefit analyses and periodic audits of algorithmic systems. Including such evaluation methods and connecting these evaluations to reward systems that not only focus on the production of technical efficiency but also public value.

Our framework can be used as a point of departure for empirical research on algorithmic government. A first set of questions concerns the complex relations all the institutional mechanisms that we have discussed. We need in-depth empirical research that takes a "life cycle approach" (Huang and Chiu 2018). In other words, we need research that looks at how various threats and strategies (input-throughput-output) interact and influence government algorithmic decision-making. Empirical research needs to embrace this complexity so we can truly understand the impact of institutional design on algorithmic decision-making.

A second set of questions concerns the numerous interconnections between algorithms that are used in intra- and inter-organizational settings. In the examples given in this article, and more broadly in the literature, there is one organization using an algorithm for a clear purpose, such as predictive policing (e.g., Meijer and Wessels 2019). In reality, however, there often is a multiplicity of interconnected algorithms that influence each other and that transcend organizational boundaries (Cicirelli et al. 2019). Indeed, institutional mechanisms need to be calibrated to facilitate appropriate and accountable *collaborative governance* of public, civic, and private entities connected through algorithmic decision-making (cf. Emerson, Nabatchi and Balogh 2012). Empirical research is needed to assess whether the institutional mechanisms indeed contribute appropriate and accountable collaborative governance.

A third set of questions focuses on how the institutional set-up affect the behavior of individual employees. Will improved institutional design ensure eventually improve how bureaucrats interact with individual citizens and how street-level bureaucrats themselves interact with complex systems? (e.g., Peeters 2020). A specific question could focus on whether transparency requirements or the right to human ensure that street-level bureaucrats use algorithmic decision support systems in a critical and thoughtful manner? Policymakers and their use of big data have become more common when making important policy decisions (Van de Voort et al. 2019). A specific question here could be whether individual policymakers

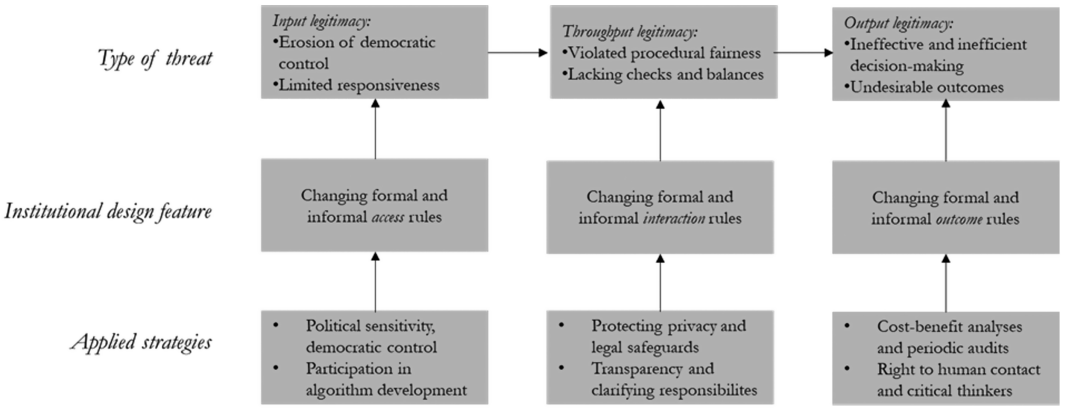

Figure 1. Overview of threats, institutional design features, and applied strategies.

and developers consider the input of civic stakeholders in algorithmic decision-making if they are required to do so.

In addition to the academic value of our framework, we would like to highlight that it can also be a useful tool for practitioners. From a practical perspective, this framework can be used to detect legitimacy deficits of specific algorithmic decision-making systems. For instance, when none of the deficits can be adequately addressed—an algorithm is proprietary and complicated but has great impact—there is a strong case to avoid algorithmic decision-making altogether. However, if legitimacy deficits can be appropriately addressed algorithmic decision-making can in fact be an effective and legitimate governance approach.

Practitioners need to continuously balance between legitimacy concerns on the one hand while at the same time preventing excessive administrative burdens and maintaining flexibility for innovations. One way to address this balance is to calibrate institutional arrangement based on the degree of risk of an algorithm (Krafft, Zweig, and König 2022). For instance, an algorithm that affects somebody in a minor way, such as adjustment of the amount of a one-time benefit, may be subject to lighter arrangements than algorithms that support life-altering decisions, such as sentencing (Bannister and Connolly 2020). In addition, some of the more burdensome tasks involved with monitoring algorithmic systems can be positioned with an independent algorithmic regulatory (Tutt 2017). Altogether, calibrating institutional design for algorithmic government is a complex and multifaceted endeavor, but one that we need to undertake to ensure its legitimacy.

Funding

This work is part of the ALGOPOL research project and has received financial support by the Netherlands Organization for Scientific Research (NWO, grant number 406.DI.19.011).

References

Anthopoulos, L., C. G. Reddick, I. Giannakidou, and N. Mavridis. 2016. Why e-government projects fail? An analysis of the Healthcare.gov website. *Government Information Quarterly* 33 (1): 161–73.

Athey, S., and G. W. Imbens. 2019. Machine learning methods that economists should know about. *Annual Review of Economics* 11: 685–725.

Baer, T. 2019. *Understand, Manage, and Prevent Algorithmic Bias: A Guide for Business Users and Data Scientists*. New York, NY: Apress

Balzacq, T. 2008 The policy tools of securitization. Exchange, EU foreign and interior policies, *Journal of Common Market Studies* 46 (1): 75–100.

Bannister, F., and R. Connolly. 2020. Administration by algorithm: A risk management frame-work. *Information Polity* 25 (4): 471–490.

Bayamlioglu, E. 2021. The right to contest automated decisions under the General Data Protection Regulation: Beyond the so called "right to explanation". *Regulation and Governance* 1–21. doi:10.1111/rego.12391

Bovens, M., and S. Zouridis. 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): 174–84.

Bozeman, B., and J. Youtie. 2020. Robotic bureaucracy: Administrative burden and red tape in university research. *Public Administration Review* 80 (1): 157–62.

Brayne, S. 2020. *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York, NY: Oxford University Press.

Broeders, D. 2011. A European 'border' surveillance system under construction. In *Migration and the new technological borders of Europe*, ed. H. Dijstelbloem en A. Meijer, 40–67. Houndsmills, Basingstoke and Hampshire: Palgrave.

Bu-Pasha, S. 2020. The controller's role in determining "high risk" and data protection impact assessment (DPIA) in developing digital smart city. *Information and Communications Technology Law* 29 (3): 391–402.

Burrell, J. 2016. How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data and Society* 3 (1): 1–12.

Busuioc, M. 2020. Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*. doi:10.1111/puar.13293

Cicirelli, F., A. Guerrieri, C. Mastroianni, G. Spezzano, and A. Vinci, eds. 2019. *The Internet of Things for smart urban ecosystems*. Cham: Springer.

Dada, D. 2006. The failure of E-government in developing countries: A literature review. *The Electronic Journal of Information Systems in Developing Countries* 26 (1): 1–10.

De Fine Licht, K., and J. de Fine Licht. 2020. Artificial intelligence, transparency, and public decision-making. *AI & Society* 35 (4): 917–926.

Desai, D. R., and J. A. Kroll. 2017. Trust but verify: guide to algorithms and the law. *Harvard Journal of Law & Technology (Harvard JOLT)* 31 (1): 1–64.

Demir, T., and R. C. Nyhan. 2008. The politics–administration dichotomy: An empirical search for correspondence between theory and practice. *Public Administration Review* 68 (1): 81–96.

Easton, D. 1965. *A Systems Analysis of Political Life*. New York: John Wiley.

Easton, D. 1975. A re-assessment of the concept of political support. *British Journal of Political Science* 5 (4): 435–457.

Emerson, K., T. Nabatchi, and S. Balogh. 2012. An integrative framework for collaborative governance. *Journal of Public Administration Research and Theory* 22 (1): 1–29.

Eubanks, V. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Giest, S., and S.G. Grimmelikhuijsen. 2020. Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective. *Information Polity* 25 (4): 409–17.

Goddard, K., A. Roudsari, and J. C. Wyatt. 2011. Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association* 19 (1): 121–7.

Gostin, L., J. M. Mann 1994. Towards the development of a human rights impact assessment for the formulation and evaluation of health police. *Health and Human Rights* 1 (1): 58–81

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. 2018. A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)* 51 (5): 93.

Hart, P.'t, and A. C Wille. 2006. Ministers and top officials in the Dutch core executive: Living together, growing apart? *Public Administration* 84: 121–46.

Hill, R. K. 2015. What an algorithm is. *Philosophy & Technology* 29 (1): 35–59.

Huang, M. C., and Y. P. Chiu. 2018. Relationship governance mechanisms and collaborative performance: A relational life-cycle perspective. *Journal of Purchasing and Supply Management* 24 (3): 260–73.

Jackson, J.R. 2018. Algorithmic bias. *Journal of Leadership, Accountability and Ethics* 15 (4): 55–65.

Kahneman, D., O. Sibony and C. Sunstein. 2020. *Noise. Flaws in Human Judgement*. New York, NY: HarperCollins Publishers.

Kaufman, H. 1960. *The Forest Ranger: A Study in Administrative Behavior*. Baltimore: Johns Hopkins University Press.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. 2017. Human decisions and machine predictions. *The Quarterly Journal of Economics* 133 (1): 237–93.

Klijn, E. H. and J. F. Koppenjan. 2006. Institutional design: changing institutional features of networks. *Public Management Review* 8 (1): 141–60.

Kober, J., J. A. Bagnell, and J. Peters. 2013. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research* 32 (11): 1238–74.

König, P. D., and G. Wenzelburger. 2021. The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it. *Technology in Society* 67: 101688.

Krafft, T. D., K. A. Zweig, and P. D. König. 2022. How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regulation & Governance* 16 (*1*): 119–136.

Lepri, B., N. Oliver, E.Letouze, A. Pentland, and P. Vinck. 2018. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology* 31 (4): 611–27.

Lessig, L. 1999. *Code and Other Laws of Cyberspace*. Basic Books.

Livingston, S., and M. Risse. 2019. The future impact of artificial intelligence on humans and human rights. *Ethics & International Affairs* 33 (2): 141–58.

Margetts, H., and C. Dorobantu. 2019. Rethink government with AI. *Nature* 568: 163–65.

Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. 2019. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)* 54 (6): 1–35.

Meijer, A. J. 2009. Complex responsibilities: An empirical analysis of responsibilities and technological complexity in Dutch immigration policies. *Public Management Review* 11 (6): 771–90.

Meijer, A., and M. Wessels. 2019. Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration* 42 (12): 1031–39.

Meijer, A. and S. Grimmelikhuijsen 2020. Responsible and accountable algorithmization: How to generate citizen trust in governmental usage of algorithms. In *The Algorithmic Society: Technology, Power and Knowledge*, ed. Schuilenburg, M. and Peeters, R.. London: Routledge.

Mergel, I., R. K. Rethemeyer, and K. Isett. 2016. Big data in public affairs. *Public Administration Review* 76 (6): 928–37.

Michels, A., and L. De Graaf. 2010. Examining citizen participation: Local participatory policy making and democracy. *Local Government Studies* 36 (4): 477–91.

Miles, R. E., C. C. Snow, A. D. Meyer, and H. J. Coleman Jr. 1978. Organizational strategy, structure, and process. *Academy of Management Review* 3 (3): 546–62.

Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267: 1–38.

Mittelstadt, B. D., P. Allo, M. Taddeo, S. Wachter, and L. Floridi. 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society* 3 (2): 1–21.

Mohri, M., A. Rostamizadeh, and A. Talwalkar. 2018. *Foundations of Machine Learning*. Cambridge, MA: MIT Press.

Odekerken, D., and F. Bex. 2020. Towards transpaent human-in-the-loop classification of fraudulent web shops. In *Legal Knowledge and Iinformation Systems: JURIX 2020: The Thirty-Third Annual Conference*, ed. S. Villata, J. Harašta, and P. Křemen, 239–42). (Frontiers in Artificial Intelligence and Applications; Vol. 334). Amsterdam, NL: IOS Press. doi:10.3233/FAIA200873

Ostrom, E. 1990. *Governing the Commons*. Cambridge: Cambridge University Press.

Peeters, R. 2020. The agency of algorithms: Understanding human-algorithm interaction in administrative decision-making. *Information Polity* 25 (4): 507–22.

Pencheva, I., M. Esteve, and S. J. Mikhaylov. 2020. Big Data and AI–A transformational shift for government: So, what next for research? *Public Policy and Administration* 35 (1): 24–44.

Pressman, J. L., and A. Wildavsky. 1984. *Implementation: How great expectations in Washington are dashed in Oakland; Or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation*. Berkeley: University of California Press.

Sætra, H. S. 2020. A shallow defence of a technocracy of artificial intelligence: Examining the political harms of algorithmic governance in the domain of government. *Technology in Society* 62: 101283.

Samuel, A. L. 1959. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development* 3 (3): 210–29.

Scharpf, F. W. 1997. *Games Real Actors Play: Actor-Centered Institutionalism in Policy Research*. Boulder, CO: Westview Press.

Scharpf, F. W. 1999. *Governing in Europe: Effective and democratic?* Oxford: Oxford University Press.

Schendel, S. van. 2019. The challenges of risk profiling used by law enforcement: Examining the cases of COMPAS and SyRI. In *Regulating New Technologies in Uncertain Times. Information Technology and Law Series*, ed. Reins L., vol 32. The Hague: T.M.C. Asser Press. https://doi-org.proxy.library.uu.nl/10.1007/978-94-6265-279-8_12.

Schmidt, V. A. 2013. Democracy and legitimacy in the European Union revisited: Input, output and 'throughput'. *Political Studies* 61 (1): 2–22.

Schmidt, V., and M. Wood 2019. Conceptualizing throughput legitimacy: Procedural mechanisms of accountability, transparency, inclusiveness and openness in EU governance. *Public Administration* 97 (4): 727–40.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529: 484–89.

Starke, C., and M. Lunich. 2020. Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy* 2: e16-1–e16-17.

Suchman, M. C. 1995. Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review* 20 (3): 571–610.

Sunshine, J., and T. R. Tyler. 2003. The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review* 37 (3): 513–48.

Tutt, A. 2017. An FDA for algorithms. *Administrative Law Review* 69 (1): 83–124.

Tyler, T. R. 2006. *Why People Obey the Law*. Princeton, NJ: Princeton University Press.

Tyler, T. R., and Y. J. Huo 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. New York, NY: Russell Sage Foundation.

Van der Voort, H. G., A. J. Klievink, M. Arnaboldi, and A. J. Meijer. 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.

Vogl, T. M., C. Seidelin, B. Ganesh, and J. Bright. 2020, Online First. Smart technology and the emergence of algorithmic bureaucracy: Artificial intelligence in UK local authorities. *Public Administration Review*. https://doi-org.proxy.library.uu.nl/10.1111/puar.13286.

Wachter, S., and B. Mittelstadt. 2019. A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review* 2019 (2): 494–620.

West, D. M. 2005. *Digital Government: Technology and Public Sector Performance*. Princeton, NJ: Princeton University Press.

Williams, B. A., C. F. Brooks, and Y. Shmargad. 2018. How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy* 8:78–115.

Young, M., M. Katell, and P. M. Krafft. 2019. Municipal surveillance regulation and algorithmic accountability. *Big Data & Society* 6 (2): 1–14.

Young, M. M., J. Himmelreich, J. B. Bullock, and K. C. Kim. 2021. Artificial intelligence and administrative evil. *Perspectives on Public Management and Governance* 4 (3): 244–58.

Zarsky, T.Z. 2016. The trouble with algorithmic decisions. An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values* 41 (*1*): 118–32.

Zouridis, S., M. van Eck, and M. Bovens. 2020. Automated discretion. In *Discretion and the Quest for Controlled Freedom*, ed. T. Evans and P. Hupe, 313–29. Cham: Palgrave Macmillan.