Perspectives on Public Management and Governance, 2019, 301–313
doi:10.1093/ppmgov/gvz014
Article
Advance Access publication October 22, 2019

OXFORD

Article

Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration

Matthew M. Young*, Justin B. Bullock[†,‡], Jesse D. Lecy[§]

*Syracuse University; [†]Texas A&M University; [‡]Taiwan Institute for Governance and Communication Research; [§]Arizona State University

Address correspondence to the author at myoung10@syr.edu

Abstract

Public administration research has documented a shift in the locus of discretion away from street-level bureaucrats to "systems-level bureaucracies" as a result of new information communication technologies that automate bureaucratic processes, and thus shape access to resources and decisions around enforcement and punishment. Advances in artificial intelligence (AI) are accelerating these trends, potentially altering discretion in public management in exciting and in challenging ways. We introduce the concept of "artificial discretion" as a theoretical framework to help public managers consider the impact of AI as they face decisions about whether and how to implement it. We operationalize discretion as the execution of tasks that require nontrivial decisions. Using Salamon's tools of governance framework, we compare artificial discretion to human discretion as task specificity and environmental complexity vary. We evaluate artificial discretion with the criteria of effectiveness, efficiency, equity, manageability, and political feasibility. Our analysis suggests three principal ways that artificial discretion can improve administrative discretion at the task level: (1) increasing scalability, (2) decreasing cost, and (3) improving quality. At the same time, artificial discretion raises serious concerns with respect to equity, manageability, and political feasibility.

Introduction

Herbert Simon—notable as an intellectual giant in both public administration and artificial intelligence—argued that "decision-making is at the heart of administration" (2013). Advances in the field of artificial intelligence (AI) have profoundly changed the practice of decision making in organizations, thus posing fundamental changes to public administration as well. AI, similar to other information communication technologies (ICT), continues to augment and supplant human discretion within bureaucracies and governance entities (Bovens and Zouridis 2002; Busch and Henriksen 2018; Fountain 2001). New theoretical and empirical scholarship is needed to understand how public administration will evolve as a result of these technologies (Andrews 2018; Barth and Arnold 1999; Bullock 2019; Dunleavy et al. 2006; Gil-García, Dawes, and Pardo 2018).

We present a framework for public managers to determine appropriate implementation of artificial intelligence applications to enhance or replace human discretion in public organizations, resulting in the use of a special case of digital discretion that we term "artificial discretion" (AD). AD creates opportunities for more efficient and reliable government, as well as the potential to cause significant harm. With equal ease, artificial intelligence can be used for prosocial and equity-enhancing projects or to generate discriminative and regressive outcomes (Brundage et al. 2018). As a result, AI can be genuinely disruptive. It has already gained traction in local, state, and federal government agencies, but proper implementation of these technologies requires a clear governance framework. As administration tasks shift from humans to machine agents, public managers need tools to anticipate the impact of deploying these technologies and to assess how programs' costs and benefits might accrue to subpopulations differently.

Public managers and policymakers will be interested in how artificial discretion can improve outcomes from programs that currently rely on human administrative discretion. Using Salamon's (2002) framework for evaluating tools of governance, we develop a framework to compare artificial and human discretion based on the criteria of effectiveness, efficiency, equity, manageability, and political feasibility. We argue that AD is a governance tool whose optimal use differs according to the specificity of the task complexity of the task environment, that is, the degree of discretion needed to complete the task effectively.

Defining Artificial Intelligence and Artificial Discretion

The term "artificial intelligence" is used colloquially to describe many technologies that use data, algorithms, and hardware to identify relationships, make predictions, or accomplish semicomplex tasks. A more technically precise definition narrows AI applications to digital computers that either (1) possess and exhibit human-like intelligence or (2) are capable of rationally solving problems to achieve a goal (Poole, Mackworth, and Goebel 1998; Turing 1950). Scholars have differentiated between *Narrow AI*, which refers to systems designed for specific tasks involving one or more decisions (identifying faces in images or self-driving vehicles), whereas *Artificial General Intelligence (AGI)* refers to yet-unrealized and therefore theoretical systems that meet the definitional condition of *intelligence* across a broad set of cognitive activities (NTSC 2016a).

There are also architectural distinctions between AI systems. *Expert Systems* are built by software engineers who consult subject area experts to construct a set of rules that attempt to replicate and ultimately automate the expert's decision-making process. These systems require a precise understanding of the heuristics deployed by experienced humans to achieve performance commensurate with human experts. By contrast, *Machine Learning* approaches use training data to identify patterns between input data and outcomes to generate predictive models using probabilistic reasoning. They "learn" relationships from the training data and thus do not require strong theory about a domain or rigorous statistical techniques typically used in policy analysis and program evaluation. Instead, these predictive models leverage correlations without a precise understanding of underlying causal mechanisms. For example, some courts use machine learning–based systems to determine whether to remand or release individuals charged with crimes and to determine recidivism risk among prisoners seeking parole based on data from past cases by identifying characteristics that predict whether an individual will return to jail (Kleinberg et al. 2018). An expert systems approach would consist of a human-designed algorithmic replication of human decision rules identified in interviews with judges and legal scholars; machine-learning approaches generate their own algorithms based on data from prior cases.

For the purpose of this article, we define artificial intelligence as any domain-specific system using machine learning techniques to make rational[1] decisions pertaining to nondeterministic tasks. A deterministic task is one that can be performed accurately using a small number of rules, whereas nondeterministic tasks are characterized by too little or too much information, outcomes moderated by a variety of contingencies, or competing values that generate a frontier of optimal outcomes instead of a single optimal solution. This definition captures the characteristics commonly and tacitly attributed to current AI applications while not requiring the much more demanding standards applied to AGI systems. We avoid the use of human intelligence as a benchmark to account for the fact that AI has surpassed human capacity in some specific domains such as strategic games (e.g., go, chess) or medical diagnostics, but currently underperforms on other sensorimotor tasks such as perception and physical mobility (Russell and Norvig 2009).

We focus our analysis on administrative tasks in the public sector where AI may either augment or replace human discretion. Although many ICT tools such as databases or dashboards can improve decision making, the relationship of artificial intelligence to discretion is unique because of three of AI's design features: (1) it is built for automating learning and decision-making processes through abstract mathematical representation of problems; (2) it can utilize input data with speed and dimensionality that vastly outstrip human cognition; and (3) as more data become available, it can "learn" and adjust its behavior by updating its decision heuristics. By contrast, traditional ICT tools and expert systems are either passive vehicles for the generation, transmission, and storage of digital data or preconfigured representations of expert human

1 Here we use "rational" in the "bounded rationality" form employed by Simon (1991); all rational decisions are bound by limits to the information available to the decision maker and its capacity to utilize that information successfully. The field of artificial intelligence employs this concept of rationality as well (Russell and Norvig 2009).

decision-making processes that do not change without human intervention. Thus, we define artificial discretion (AD) as cases where artificial intelligence is used to augment or automate the exercise of administrative discretion. We include both commercial AI systems such as IBM's *Watson*, and ad hoc implementation of predictive models that augment tasks that require discretion.

Current research on "algorithmic" and "smart" governance focus on the potential and observed impacts of information technologies on government practices, but neither names artificial intelligence as its empirical focus nor considers the unique theoretical implications of these systems for administrative discretion and decision making (Anthopoulos and Reddick 2016; Janssen and Kuk 2016; Meijer 2018). Busch and Henriksen (2018) call attention to the impact of AI on digital discretion, but do not address the implications for public administration. We develop the construct of artificial discretion to fill this gap in the literature, articulate the potential for AD to significantly impact government practices in both good and bad ways, and provide a framework for evaluating the appropriateness of supplanting human discretion in public institutions.

The topic has some urgency given that labor advocates and academic researchers both argue that AI will begin to displace a significant number of service-sector workers (Frey and Osborne 2017; Lee 2016; NTSC 2016b). Unlike the displacement of low-skill human labor in manufacturing through automation, AI has the potential to target jobs that have typically been immune from automation because they require specialized cognitive skills. Law, for example, lends itself nicely to AI systems because legal databases and natural language processing algorithms can be used to model case law. Many bureaucratic or managerial tasks in the public and nonprofit sectors have similar characteristics as those in law, creating opportunities to automate specific tasks or wholly replace some public servants. There is no economic rationale for the public sector to favor human labor over machine labor (Alvarez-Cuadrado, Van Long, and Poschke 2018). The political and fiscal pressures to downsize government and "do more with less" will, if anything, promote labor substitution for practical or ideological purposes, potentially even when evidence suggests that this substitution neither lowers costs nor improves quality. For example, the privatization movement has been successful at marketizing public services and disposing public goods using ideological arguments, even when there is little empirical support for the practice (Ballard and Warner 2000; Savas and Savas 2000). In short, we can expect artificial discretion to become an increasingly important topic in the field of public administration.

Uses for Artificial Discretion in Governance

Artificial discretion presents an opportunity to improve on the status quo in government by addressing deficiencies that are ubiquitous in administrative decision making. These problems include (1) inaccurate predictions[2] on consequential discretionary tasks such as placing a child into foster care or granting a small business loan; (2) inconsistent quality of discretion through variance in accuracy across managers, or variance across time because of decision fatigue or emotional shocks[3]; (3) bias in discretion such as inconsistent citation rates in policing; (4) corruption that results when discretionary tasks are used to gain influence such as granting lucrative government contracts to political allies; and (5) the high labor costs of semi-routine bureaucratic tasks that require some discretion, such as processing applications to administer permits. AI offers opportunities to improve on each of these dimensions by being more accurate, more consistent in ways that reduce variance and bias, less corruptible by having a stronger relationship between data inputs and decisions, and more cost-effective by reducing labor costs associated with mundane tasks. The primary challenge in evaluating artificial discretion results from performance varying across these dimensions. If an AI system is half the cost of the current system, but 10% less accurate, is it better? What if overall accuracy improves but the system performs very poorly for a particular subpopulation?

Public managers will be interested in three principal uses of artificial intelligence in government: (1) creating structured data from unstructured inputs such as images, sensors, or text; (2) leveraging large and multidimensional data sets to identify patterns, generate new insights, or make accurate predictions; and (3) eliminating human components of administrative tasks through automation. Because the quality of decisions made by AI systems performing these tasks will determine performance in each domain, the topic of artificial discretion is central to public administration when AI is present. This section frames the discussion about which uses of AI are most appropriate based on the degree of discretion required for a given task.

AI is useful for generating structured data from unstructured inputs in task environments where large amounts of data are necessary to improve the administrative task, but the data-generation process is infeasible using humans because of cost or complexity.

2 Accuracy might be measured as a percent of correct predictions, but if the cost of false positives and false negatives differs significantly, it might be better captured by specificity or sensitivity of models.

3 Eren and Mocan (2018) find causal evidence of judges in juvenile courts imposing harsher sentences the week following an unexpected loss by their alma mater's football team.

Unstructured data might include real-time data generated from sensors or video, a situation where AI's scalability is particularly valuable. For example, AI has been used to read the text of public policies and develop taxonomies that simplify the laws (Rice and Zorn 2019). Another is the use of facial recognition software to monitor live-stream video feeds for real-time identification of known fugitives or terrorists (Viola and Jones 2004). AI can also generate insights using structured data to identify relationships in large and complex data sets, and using those discoveries to classify cases or predict outcomes. For example, Helsby et al. (2018) demonstrated that AI trained on law enforcement data identifies police officers who are likely to use excessive force in the future with greater accuracy—both in terms of minimizing false and maximizing true positives—than traditional approaches. In a more contentious example, the Department of Defense uses AI to identify enemy combatants based on patterns of behavior and communication (Lee 2018).

Third, AI can partially or fully replace human labor through automation of tasks involving discretion. The automation of labor requires a blending of real-time structured data generation from raw inputs such as sensors or text, the analysis of these data, and the ability to translate the decision into action. The action itself may originate from the AI system, or the decision can be transmitted to a separate system for execution. One commonly employed but often-overlooked example of this use of AI is the detection of malicious messages in email platforms. Modern systems fully automate that task using AI and quarantine messages before they hit your inbox.

Each of these categories of AI applications is potentially valuable for public sector organizations, but not every administrative task can be easily improved by AI. These tasks vary along several dimensions, including the amount, veracity, and completeness of information available to the agent at the time of a decision, the cost, risk, and reward associated with the decision, the amount of goal or value ambiguity, and the stakeholders involved.[4] Similarly, the quality of the decision will be shaped by the information processing capacity of the agent and the decision architecture used. Some tasks are individual, whereas others demand some degree of interpersonal communication and negotiation.

AI's appropriateness for a specific administrative task is a function of the task complexity, quality and availability of data, technical requirements, limitations of the AI system, risk and uncertainty associated with the task, and the political feasibility of using nonhuman solutions. These features of the task environment can be largely described by the degree of discretion that is necessary to perform the task well; thus, we use the concept of artificial discretion to organize the analysis of AI applications applied to public administration tasks. Table 1 illustrates the theoretical best-fit use of artificial discretion according to the degree of task discretion. It is crucial to note that these factors are time and place sensitive; we hold both constant here for the sake of theoretical clarity and brevity.

In cases where tasks require a relatively low level of discretion, automation is appropriate. The potential benefits include cost-savings, and eliminating situations where low-discretion tasks are dull and repetitive to human agents, leading to low-quality outputs and outcomes. Alternatively, automation can be used when the quality of discretion varies significantly across different human agents. Using AI as a decision-support tool can improve discretion when tasks are less structured, data are scarce, or outcomes are harder to define. In these cases, AI can increase the scope and quality of information available to the human agent or enhance the ability to explore scenarios or predict outcomes, thus augmenting human discretion.

High-discretion tasks are defined by poor data, uncertainty about the factors that lead to success, or tightly coupled systems that are difficult to model. Decision-support tools and predictive analytics are inappropriate in these scenarios because the problems are not well-enough defined or there is not enough data to model them effectively. The ability to generate large quantities of new data and the search for factors that best predict outcomes can vastly improve discretion in these scenarios. For example, Pentland (2014) describes how new smart badges were invented to collect the types of data that were necessary to effectively model collaboration in large organizations. In other instances, the administrative milieu is data rich, but tasks require high discretion because systems are complex and chaotic. Weather scientists, for example, have huge amounts of granular and accurate data, but weather systems destabilize quickly, making it difficult to predict extreme weather events or long-term forecasts.

Table 1. Potential Use of Artificial Discretion for Tasks by Degree of Discretion

Low Discretion	Medium Discretion	High Discretion
Automation	Decision-support tool, predictive analytics	New data generation, reduction of data complexity, relationship discovery

4 These factors that influence the type of task to be completed are similar in construct to Simon's (2013) idea of *decision premises*. Simon defined these decision premises "to refer to the facts and values that enter into the decision-fabricating process, a process that involves fact-finding, design, analysis, reasoning, negotiating, all seasoned with large quantities of 'intuition' and even guessing" (Simon 2013, 23).

In these three contexts human discretion becomes "artificial" when is it augmented or supplanted by AI that provides inputs into the discretionary tasks or replaces human discretion through automation. Some applications of artificial discretion generate clear, uncontested, and nonrivalrous public goods. For example, routing ambulances to shorten travel time by avoiding traffic benefits everyone without compromising other outcomes. But most solutions to public sector problems involve trade-offs that differentially affect specific interest groups or classes of people, creating winners and losers when policy changes. The "optimal" solution, which maximizes efficiency, might place an unfair burden on a neighborhood or minority group. A common use of discretion in bureaucratic decision making is to balance considerations of cost, efficiency, fairness, and justice, especially when these considerations are poorly defined in the policy or mandate (Gailmard and Patty 2012; Huber and Shipan 2002). As with any tool, determining whether or how the use of Artificial Discretion is appropriate requires careful consideration grounded in the implementation context. Without this consideration, implementation failures, along with unintended and negative consequences, are all but assured.

Key Task Dimensions for Artificial Discretion: Degree of Discretion and Contextual Level of Analysis

Two key task dimensions are important for understanding the potential utility of artificial discretion: the level of discretion required and the level of analysis (Busch and Henriksen 2018). Some tasks are straightforward, narrowly defined, and relatively constrained (e.g., issuing parking tickets), whereas others are informed by procedures but consist of a broader set of possible choices (e.g., law enforcement intervening in a domestic dispute). Many others are highly complex, open-ended, and largely unstructured, especially in the design phases of policies or programs. The level of analysis highlights the contextual factors that influence how tasks are completed. Some tasks can be completed by individual managers or in consultation with a limited number of stakeholders, whereas others are intraorganizational or take place at a superstructural level, such as multisector negotiations that occur while designing new policies. The more actors the process requires, the more constrained and interdependent discretion becomes. Thus, tasks can be described by whether they can be implemented at the individual level (micro), the team or intraorganizational level (meso), or the institutional and policy-setting level (macro). Table 2 organizes the classification of tasks along these two dimensions and provides illustrative examples of tasks for each combination.

The micro-level of analysis includes contextual factors across professionalization, computer literacy, decision consequences, information richness, and relational negotiations. Examples of low-discretion tasks at the micro-level include data entry and formulaic or menu-driven licensing or permitting for routine functions. Tasks requiring more professional training with serious individual and social consequences such as whether to place a child in foster care or to grant parole to a prisoner are examples of tasks where the agent is afforded more discretion. Tasks undertaken in situations of extreme uncertainty where the policy framework is not defined or with potentially immediate, life-or-death consequences require the most discretion.

Meso-level tasks shape and affect the organizational environment in which individual agents are embedded. Contextual factors at this level of analysis include formulation of organizational goals, formalization of routines, and interagency dependency. An example of a low-discretion meso-level task is the determination of when and how to use energy in facilities operations. More discretion is required for tasks such as hiring new personnel or determining performance management metrics for a department, whereas broad and ambiguous tasks such as organizational agenda and goal-setting require the most discretion.

Macro-level tasks are institutional in scope, including multiorganizational partnerships or negotiations that involve policy or goal-setting, and typically have broad consequences. The contextual factor at this level of analysis is the formulation of rules. This authority can be hierarchical (e.g., governments) or networked (e.g., supra-organizations or groups created to design and operate the institutional framework for a cross-sector collaborative effort). Low-discretion tasks at this level include the execution of well-defined statutory obligations, such as the Bureau of Labor reporting aggregate measures of economic activity. Formulating policies for well-understood issues is an example of macro-level tasks with additional discretion. Finally, macro-level tasks undertaken in response to shocks or crises have extremely high uncertainty, complexity, and impact, and thus high amounts of associated discretion. An example of this task type is the federal government's response to the September 11 attacks, which include the creation of a new Cabinet-level executive department, the USA PATRIOT ACT legislation, and the invasions of Afghanistan and Iraq.

An Evaluation Framework for Artificial Discretion

Artificial discretion presents both opportunities and challenges for the public sector. Potential positive outcomes include higher quality, lower cost, or novel government services, and increased consistency in the

Table 2. Matrix of Task Analysis by Level of Analysis and Degree of Discretion

Degree of Discretion	Low Discretion	Medium Discretion	High Discretion
Level of Analysis			
Micro (individual)	Data entry, issuing licenses or permits	Placing children in foster care, sentencing/parole	Emergency response
Meso (organizational)	Facilities operations	Hiring processes, performance management	Goal setting, strategic planning
Macro (institutional)	Statutory obligations	Policy formulation	Crisis response and management

exercise of discretion. Possible negative outcomes include the erosion of transparency and accountability, distance from tasks that might nurture learning and innovation, a significant risk of increasing inequality, and increases in the capacity for administrative evil (Bullock 2019; Gianfrancesco et al. 2018; O'Neil 2017). Rapid growth in AI capability means opportunities to deploy AD for many innovative purposes, but it also presents the challenge of anticipating areas in which the disruption it causes creates unacceptable harms (Bostrom 2014).

Salamon (2002) notes that the management challenges found in traditionally hierarchical service-providing public organizations have been exacerbated by modern governance arrangements. Collaborative or polycentric governance requires extensive planning and coordination, incurs higher transaction costs, and often has diffuse accountability. In these contexts, artificial discretion has the potential to achieve greater gains by reducing information blindness and decision fatigue experienced by administrators; however, it also has the potential for harm due to inappropriate or poorly designed applications, as well as cooptation of public goods by influential or strategic actors that use artificial discretion to limit the ability of public managers to exercise their own discretion. Salamon's (2002) framework for evaluating tools of governance is helpful for understanding the efficacy of artificial discretion.

Comparing Artificial and Human Discretion

Due to the centrality of decision making in organizations and its impact on performance, managerial effectiveness can be evaluated in part by assessing how well administrators employ the discretion afforded them. The implicit causal chain is that better use of discretion leads to better decisions, which lead to better organizational performance, and ultimately better outcomes (Bullock, Greer, and O' Toole 2019). Thus, technologies used to improve the exercise of discretion will assumedly have a positive impact on performance. The unit of analysis in this evaluative framework is the decision task (hereafter 'task'), a discrete instance where an agent uses the discretion afforded them to make a decision. The appropriate counterfactual to artificial discretion is nonaugmented human discretion exercised by a public administrator in the same task environment. There are three principal ways that artificial discretion can improve discretion at the task level: by increasing scalability, decreasing cost, and improving quality. Task scalability is a function of the number of tasks performable in a unit of time. Task cost is both the unit and marginal cost per task that an organization must pay. Task quality refers to success or failure rate over a set of tasks.

The exponential growth in computing power and information technologies beginning in the late 20th century presents clear opportunities to scale technical solutions to policy problems (Hilbert and López 2011). These gains created the necessary conditions for more organizational activities to become both data driven and networked. The combination of increased computational capacity, decreases in cost, and the growth of data reduced or removed many of the technical barriers that previously limited machine-learning variants of artificial intelligence (Kim 2016; McCorduck 2004). These conditions were necessary to support the design, implementation, and scaling of artificial discretion systems. As a result, artificial discretion enjoys a clear comparative advantage with respect to task scalability and cost. However, the relative quality of artificial versus human discretion is less clear because it requires a balanced assessment of task scalability, task cost, and task quality.

Task Scalability

Artificial Discretion is highly scalable because its machine learning–based architecture was developed specifically for serially processing vast quantities of high-dimensional data. The growth of data available to organizations has rapidly outpaced human capacity for exploiting the information; this mismatch of capacity and content is the defining characteristic of "Big Data" (Kitchin 2014; Meijer 2018). The per-task speed of AD outpaces the speed of human discretion. Even in circumstances where AD performs worse than human discretion in terms of task quality on average, it still produces superior performance. For example, some disaster response organizations use artificial discretion to categorize social media posts or process satellite images during periods of crisis to document where casualties and damage are greatest and more efficiently coordinate rescue operations. These AD systems can disambiguate and tag social media posts and identify specific items in satellite images faster, albeit less accurately, than humans. But the benefits of speed and scalability outweigh lost categorical accuracy for the outcome of interest: finding and saving victims (Meier 2015).

Task Cost

Artificial discretion is more cost-effective than human discretion under certain conditions. The fixed costs of design and setup for AD are usually nontrivial. Hardware systems need to be built. Machine-learning software needs to be architected and trained from past data. Custom databases or sensors often need to be created. Once created, though, the marginal cost of executing new tasks approaches zero, leading to dramatically lower unit costs over time. For programs that benefit from volume, AD's superior scalability makes cost-savings over human discretion likely, especially in the public sector where many workers still possess collectively bargained benefits. This cost differential is at the heart of the attention developments in AI have received in the past several years as a potential source of disruption in the labor market not just for unskilled labor, but for jobs that currently require advanced training and/or graduate degrees (Frey and Osborne 2017). In environments of fiscal austerity, there may be strong incentives to embrace this trade-off irrespective of its impact on organizational performance and governance outcomes. Care is needed to avoid overly preferencing cost reductions. Furthermore, possible applications for AD include tasks that are not currently feasible for governments. Such applications could lead to new programs instead of augmenting existing ones, complicating the cost-savings argument.

Task Quality

Task quality is where artificial discretion's comparative advantage is both more nuanced and fluid as the technology matures. In some cases, artificial discretion has surpassed human discretion with respect to task quality. For example, AD has replaced human agents for routine image identification tasks such as identifying the amount of money and signatures on handwritten checks, obviating the need for staff to verify the amount manually at the time of deposit. Another example of AD's demonstrated superiority is the improved rate of correctly identifying medical conditions such as cancer or genetic disorders under specific conditions (Gurovich et al. 2019). One of the most unambiguous conditions where artificial discretion outperforms humans in task quality is pattern recognition using high-dimensional data. This task is particularly important for extracting actionable information from "big" data when there is no appropriate theoretically informed model to draw on and is the capacity at the center of most current implementations of AI technologies. For other types of tasks, however, human agents obviously—and sometimes hilariously—outperform artificial discretion. Tasks such as navigating a complex physical environment or object recognition and identification are examples where human capabilities continue to exceed artificial systems.[5] Artificial discretion systems are also currently incapable of discerning meaning beyond objective functions, which can result in strange behavior (Bostrom 2003).

A well-known problem with administrative discretion is human bias, or cognitive limitations that prohibit an agent from making the correct decision (Kahneman 2011). Of concern—especially in the public sector—is bias that arises from prejudice against individuals based on their innate characteristics, religious beliefs, or other affiliations. Conceivably, the substitution of artificial discretion for human discretion could improve task quality by eliminating the explicit and implicit biases present in human agents. Empirically, however, the results are mixed at best. Even when artificial discretion systems include direct measures to reduce or eliminate the impact of sensitive characteristics such as race and ethnicity in its decision making, bias can be introduced through artifacts of past human biases embedded in the data required to "train" the system (Bellamy et al. 2018; Veale, Van Kleek, and Binns 2018). Bias can also occur because models are oversimplified and rely on sparse, incomplete, or inaccurate input data. For example, facial recognition programs trained on images of mostly white people perform well on white faces but very poorly when analyzing other skin tones (Buolamwini and Gebru 2018).

Evaluating Artificial Discretion

Public managers, policymakers, and researchers need criteria to weigh the normative benefits and costs of implementing artificial discretion ex ante and evaluating its impacts ex post. We argue that public sector use of artificial discretion should be evaluated according to the five criteria put forward by Salamon (2002): effectiveness, efficiency, equity, manageability, and political feasibility. For each criterion, we also examine how evaluation differs according to appropriate use of AD by the level of discretion required—automation for low-discretion tasks, decision support for medium, and discovery for high—as summarized

5 Artificial intelligence researchers refer to the counterintuitive fact that basic sensorimotor skills are extremely computationally intensive while high-order reasoning skills are not as "Moravec's Paradox."

in Table 1. These criteria allow researchers and practitioners to systematically examine the opportunities and threats artificial discretion poses for the public, public organizations, and governance institutions.

Effectiveness

The criterion of effectiveness refers to the degree of success or achievement an activity enjoys relative to its intended objective (Salamon 2002). For any task, it is important to carefully define both the intended objective and success or achievement relative to that objective. In task domains where more discretion is required, it is also correspondingly more challenging to carefully define objectives and success for a given task. Once a minimum threshold of effectiveness is met (e.g., can the system perform a given number of tasks at a designated success rate in a specified amount of time), evaluating effectiveness centers on comparisons against the opportunity costs of foregone alternatives (i.e., what is the most effective option from the set of options {A, B, C}).

For tasks with lower levels of discretion, artificial discretion will often dominate human discretion in terms of effectiveness. But in a subset of such cases, human agents may not be the appropriate unit of comparison. Whenever discretion is explicitly a function of task uncertainty (e.g., decision makers are given more discretion precisely because the level of uncertainty associated with the decision is relatively high), the need for discretion necessarily decreases as the level of uncertainty approaches zero (i.e., perfect information). In these cases, human-generated algorithmic automation—that is, expert system-based AI—is likely to be as or more effective than AD. For example, for a task that can be simplified to "If A, then B; else C," expert system automation is sufficiently effective so long as "A," "B," and "C" are sufficiently defined, and instances are preclassified as "A" or "not A." But in such cases where the classification of "A" or "not A" is not predetermined and has high uncertainty, artificial discretion is preferable.

As the level of discretion required for a task increases, human agents are likely to be more effective decision makers. In these circumstances, the most effective role of automated discretion is likely to be equivalent to intelligence amplification rather than automation. For example, AD may aid organizational goal formulation by evaluating complex unstructured data to identify previously unrecognized needs. For tasks requiring a moderate level of discretion, AD's effectiveness can be evaluated against the same task performed without a decision-support tool and with alternative tools such as human-designed algorithms or heuristics.

Tasks requiring the greatest amount of discretion are where artificial discretion may generate exponential gains in effectiveness because of its ability to reduce data dimensionality and volume into structures, patterns, and other information understandable to human decision makers. Under the strong assumption that the data inputs and system architecture are optimized with respect to the task, artificial discretion's use to amplify human decision makers' intelligence and situational awareness may prove to be its most effective use for public managers.

The criterion of effectiveness is perhaps the most challenging for evaluating because of uncertainty regarding future advances in machine learning, data generation, and other factors that determine the range of domain spaces where AD demonstrably outperforms human cognition. Variance across human agents is another important consideration for measuring AD's effectiveness. Artificial discretion need not outperform the best human agents or alternative decision-support tools in a decision-making environment to be more effective at a systemic level; once average task quality is sufficiently high, its scalability may make it dominant.

It is currently unclear at what level of discretion overall effectiveness related to the utilization of AD begins to decline and human discretion becomes likely to remain more effective. Organizations should thus tread carefully when experimenting with artificial discretion—especially with respect to its use for automation. Careful consideration should be given to determining an appropriate threshold for task quality, and this must first be met for any task for which artificial discretion replaces human discretion. Experimentation will be required to develop a clearer understanding about which specific task domains are more effectively executed by which form of discretion.

Efficiency

Efficiency balances outputs or outcomes against inputs and their associated costs. The most efficient choice need not be the most effective one; instead, it is the choice that maximizes the ratio of benefits to costs subject to a set of policy and political constraints (Salamon 2002). Although many technologies are characterized by large fixed costs of design and implementation and a low marginal cost of operation, one advantage of AD is that many of its fixed costs are comingled with other technological systems (e.g., telecommunications infrastructure for data input/output, etc.), thus making deployment cheaper (Brundage et al. 2018). However, large investments might be needed either in developing robust input data into the AI processes or in creating an organizational culture that can embrace new insights (Barton and Court 2012).

Unlike effectiveness, the efficiency of artificial discretion is not a function of task discretion. The high scalability and low unit and marginal costs of AD make it dominant to human efficiency for any discrete task that it can complete (Brundage et al. 2018; Husain 2018). This is likely to hold true even for tasks where human-designed algorithms are equally as effective because the total factor cost of human-designed algorithms includes the inputs of the software developers who write and/or modify the program. This dominance, however, belies the fundamental limitations—indeed, danger—of efficiency as a primary evaluative criterion: selecting a maximally efficient but Pareto inferior decision at best minimizes the welfare loss of required inputs, but the optimal choice would be to not make the Pareto inferior decision at all.

Potential improvements in efficiency from uses of AD are simultaneously a challenge as well as an opportunity for the public sector. The opportunities are sufficiently obvious; the challenges require attention. Furthermore, the challenges raised by artificial discretion's efficiency apply across all levels of task discretion and context. The ease of achieving economies of scale with artificial discretion has already led to suboptimal implementation outcomes with respect to both effectiveness and equity, in both the public and private sectors. Incentives to embrace efficient AD-driven systems may even be coercive in fiscally constrained environments, leading to trade-offs that can be punitive to marginalized populations (Kleinberg et al. 2018). Furthermore, it may be tempting to substitute easily measurable outputs for actual outcome objectives in the name of improving efficiency, potentially at the expense of effectiveness. This substitution may be even more tempting for tasks with higher levels of discretion, as the associated outcome objectives are likely more challenging to define. Finally, at higher levels of task context such as macro and meso when compared with micro, the substitution of outputs and outcomes for truly defined objectives can cause more widespread harm throughout a network.

Equity

Salamon (2002) argues for evaluating a tool's equity impacts along two dimensions: whether it provides equal treatment to all members of the public and whether it is redistributive in favor of disadvantaged subpopulations. Unequal treatment when high levels of discretion are present is well documented in decision tasks of high consequence within judicial and law enforcement systems (Schneider and Ingram 1993; Wade-Olson 2016). Thus, bureaucratic discretion sets a relatively low bar, across at least some domains, with respect to equity.

Public sector use of artificial discretion carries clear equity concerns for person- and place-based predictive policing. These issues are increasingly apparent as machine-learning systems trained on administrative data and implemented by criminal justice agencies have produced demonstrably biased decisions (Dressel and Farid 2018; Corbett-Davies and Goel 2018). Although popular media often refer to observed instances of these equity-harming outcomes of artificial intelligence as "Racist AI" or "Sexist AI" (see Dastin 2018; Lutz 2019; Zou and Schiebinger 2018), we argue that this perspective misses the mark. Observed inequality in artificial discretion can be attributed both to the biases of the developers and implementers in the human system and to inequalities generated from historic human decisions (made possible through the exercise of discretion) that are embedded in the data used to train the system. Given empirical evidence of structural and behavioral inequalities in public administration and governance, this suggests that there are significant challenges for implementing artificial discretion in the public sector with respect to optimizing equity for tasks at any level of discretion or context.

However, as with the criterion of effectiveness, properly evaluating artificial discretion's equity effects requires benchmarking it both against the normative value of maximal equality and AD's impact relative to alternatives. Appropriate alternatives for comparing equity effects are the same as for effectiveness: the set of human bureaucrats whose work would be automated for low-discretion tasks; the same task absent an AD-enabled support tool or made with non-AD tools for tasks requiring more discretion; and the same task without the use of artificial discretion for high-discretion tasks where AD's use is likely to be novel.

The second aspect of equity discussed by Salamon is the notion of redistribution of resources or government benefits to those who need them the most. Prior evidence shows that modern, data-driven systems have a tendency to exacerbate inequality (Eubanks 2018; Noble 2018; Selbst 2017). This arises because of differential access to new technologies, differences in ability to use technology, or differences in the voice needed to determine how technology is imposed on communities. As O'Neil (2017) notes, we see a disturbing pattern where human discretion is used for decisions regarding wealthy people, and artificial discretion is used for decisions regarding poor people. This implies that evaluating artificial discretion's equity effects must begin with a structural analysis of the proposed implementation, whether it is augmenting an existing task process or generating new ones.

Two characteristics of AD are especially important for evaluating its equity impacts. These are its dual-use nature, and its capacity to increase psychological distance and anonymity. The dual-use characteristics highlight that artificial discretion can, very broadly put, be aimed in any direction, toward any goal. Thus, artificial discretion can be employed in service of equity-enhancing or harming ends with equal ease. AD's facilitation of psychological distance and anonymity make it particularly dangerous in the service of administrators who seek to oppress or subjugate individuals or groups of people (Fennimore and Sementelli 2019).

Manageability

Salamon (2002, 24) notes that manageability, also referred to as implementability, "refers to the ease or difficulty involved in operating programs. The more complex and convoluted the tool, the more separate actors involved, the more likely it is to be difficult to manage." Artificial discretion presents an interesting challenge with respect to manageability. Although its use could eliminate some types of management issues in public administration, AD also intensifies other manageability challenges and introduces new ones.

A central if not always explicated normative argument for automating public sector tasks is that it represents the ideal form of a rational-technical public administration unfettered by moral hazard. In circumstances where expert system-based automation is not possible, artificial discretion-based automation still affords elected officials increased control by eliminating human bureaucrats. Not only does this automation reduce managerial complexity in terms of employee motivation and other organizational behavior challenges, it also eliminates the capacity for bureaucrats to passively or actively resist organizational or policy changes (O'Leary 2013). For tasks requiring more discretion such that complete automation is not feasible, AD can still provide checks against human-induced manageability issues through its use as a decision-support tool and/or information generator. Both uses of AD reduce task uncertainty, making it easier to standardize behavior and predict outcomes.

Even simple implementations of artificial discretion itself, however, create their own manageability challenges. Because it is so easy to scale, AD invites overambitious project or policy design, which can lead to overly or unnecessarily complex administrative systems. This is especially a concern with respect to AD's need for and facility with large volumes of high-dimensional data. The logic of using AD to generate new information about the state of the world and otherwise unobservable correlations compels policymakers and administrators to increase existing and deploy new data generating systems. These in turn necessarily increase overall system complexity, potentially further eroding manageability. Property rights around the AD systems themselves may also complicate their manageability. When AD capacity is supplied by private technology or service vendors via contract, it introduces both standard contract manageability issues and intellectual property protections that make system processes more secretive and thus difficult to evaluate (O'Neil 2017).

Moreover, artificial discretion's architecture introduces new manageability issues. The artificial neural network (ANNs) architectures that dominate the machine learning–based artificial intelligence landscape are often described as "black box" systems because their complexity makes it impossible to reverse-engineer the algorithm to understand the factors and weights used in its decision making *ex post* (Lipton 2016; Liu et al. 2017; Schmidhuber 2015). Relatedly, artificial discretion presents new manageability challenges with respect to system security and resilience. ANNs are susceptible to manipulation using "adversarial" data designed to force the system to make an incorrect decision (Brundage et al. 2018; Ilyas et al. 2018; Papernot and McDaniel 2018). The speed at which adversarial exploits are discovered is vexing in its own right: since 2014 a machine-learning system architecture known as generative adversarial networks pits offensive, or generative ANNs against defensive, or discriminatory ANNs, to recursively train the former to fool the latter or the latter to catch the former in the act (Goodfellow et al. 2014).

Legitimacy and Political Feasibility

The final criterion of legitimacy and political feasibility is defined as whether the general public and politically active agents deem a tool legitimate (appropriate for public action) and politically feasible (either enough political will exists to implement the tool or not enough political will exists to oppose it) (Salamon 2002). Of all the proposed criteria, this is the most mutable and difficult to generalize. The interaction of artificial discretion with democratic public administration further complicates the picture.

Exactly what constitutes a legitimate use of artificial discretion has yet to be determined across any level of task discretion or context. That said, the interaction of the previous evaluative criteria with current and future AD applications may make its adoption more or less palatable. In cases where using AD allows for outcomes that are both broadly considered prosocial and simply not possible via human cognition, it is easy to envision AD's acceptance as legitimate. In these instances, political feasibility is also likely to be high, but contingent on opportunity costs and fiscal constraints. For example, increases in effectiveness and efficiency benchmarks for the use of artificial discretion to generate new information in aiding disaster response teams are likely to increase the technology's perceived legitimacy and feasibility. This does not mean, however, that every organization tasked with disaster response and management will be able or willing to pay for these systems.

The political feasibility of artificial discretion's use to automate low-discretion tasks is highly volatile. Theories of public administration and management have heavily weighted efficiency and effectiveness since the field was founded; they remain a driving force in more recent and emergent frameworks as well (Dunleavy et al. 2006; Frey and Osborne 2017; Gil-García, Dawes, and Pardo 2018). At the same time, many low-discretion, automation-appropriate tasks in the public sector are currently performed by unionized or otherwise protected workers. Attempts to completely substitute capital for labor in these circumstances will almost certainly provoke political conflict. Even partial automation is likely to face resistance, as it is frequently perceived—often rightfully so—as a beachhead for future full-scale displacement of labor. Academic prognosis and empirical evidence of this friction and conflict have largely focused on private labor markets, likely due to the relative evisceration of collective bargaining and its benefits for labor security in that sector. One possible circumstance that could tilt political feasibility in favor of automation via artificial discretion, especially at the subnational level, is one or more future economic shocks that motivate elected officials to both cut costs via staff reductions and seek out efficient alternatives for necessary tasks.

Questions of artificial discretion's perceived legitimacy for both automation and as decision support for more discretion-intensive tasks are similarly complex. Research on "algorithm avoidance" suggests people tend to prefer human judgment[6] over an algorithm's, even after receiving unambiguous evidence of the latter's superior performance (Awad et al. 2018; Bonnefon, Shariff, and Rahwan 2016; Dietvorst, Simmons, and Massey 2015; Frick 2015). This aversion is not, however, absolute. On a basic level, the congruence between task and system appearance and behavior (whether physical, verbal, or written) increases acceptance and, by extension, legitimacy (Goetz, Kiesler, and Powers 2003). Framing systems employing artificial discretion in terms of the values of the human agents involved in their design can improve perceived legitimacy (Jago 2019). Similarly, providing some form of feedback mechanism that can modify system outputs, even to a minor degree, can facilitate acceptance of artificial discretion's replacement of human agency (Dietvorst, Simmons, and Massey 2016). Public managers looking to implement artificial discretion will need to incorporate these and related findings on human–machine interaction to maximize their potential legitimacy and feasibility.

Finally, it is important to note that to the extent legitimacy and feasibility are preconditions for public sector adoption of artificial discretion, organizational adoption of innovative technology follows a common pattern as the technology morphs over time (de Vries, Tummers, and Bekkers 2018). Adoption is often resisted when the technology is immature, as the technologies underlying artificial discretion currently are. However, as these tools become more commonplace and their value—real or perceived—becomes common sense wisdom, resistance can give way to a wave of mimetic adoption (DiMaggio and Powell 1983; Jun and Weare 2010). In the case of public sector adoption of artificial discretion, this phase transition could be accelerated by the technology's saturation of the private sector due to the ideological dominance of belief in private-sector efficiency.

Conclusion

Artificial discretion is an emergent domain that will be of great interest to public administration as discretion is enhanced and automated across government. Technological capabilities continue to increase in the level of task discretion for which artificial discretion can reasonably substitute for bureaucratic discretion (Grace et al. 2018). The level of task discretion guides this diffusion of artificial discretion across the micro, meso, and macro levels of governance tasks. This presents questions for scholars of governance as to how the governance tool of artificial discretion will affect the effectiveness, efficiency, and equity of governance, and for the manageability and legitimacy of the tool.

We provide a framework for defining, characterizing, and evaluating artificial discretion as a technology that both augments and competes with traditional bureaucratic discretion. We also utilize Salmon's tools of governance analysis to propose an evaluative framework for artificial discretion's use in the public sector, and describe the defining characteristics and key tool dimensions. We posit that artificial discretion arises in cases where artificial intelligence is used to improve or automate the exercise of administrative discretion in the form of decisions. We also identify three principal ways that artificial discretion can improve discretion at the task level: by increasing scalability, by decreasing cost, and by improving quality. We then examine the moderating role task discretion plays, along with the contextual factors likely to moderate the diffusion of artificial discretion as a substitute for human discretion. Finally, we examine the implications of the diffusion of artificial discretion across the evaluative criteria of effectiveness, efficiency, equity, manageability, and political feasibility.

Potential applications of artificial discretion in government are hard to predict. We know that low-discretion tasks can be automated, but AI is typically cost-effective when tasks are of high volume, so high fixed costs of new

6 Especially their own.

system design can be offset by savings from significant reductions in the marginal costs of tasks. But in cases of many emergent technologies, the challenge in predicting impact comes from new tasks that are made feasible, not the rationalization of existing tasks. For example, AI can be used to generate massive new data sets by automating the processing of images and sensors or by standardizing unstructured data such as social media posts or text. As a result, AI will affect both the nature of current discretion tasks and the constellation of tasks that governments adopt in the future.

Features of AI, such as accuracy, scale, and efficiency, offer significant advantages for artificial discretion over human discretion, but they pose significant challenges as well. AI algorithms can be difficult to understand and thus difficult to manage in a transparent and accountable manner. The growth of artificial discretion has potential to mirror the growth of third-party government, diminishing the organizational capacity of government offices and attenuating governance processes, leading to a submerged (and further hollowed out) state that has little accountability to citizens. Algorithms have well-documented biases that can be damaging to minority groups and poor populations, exacerbating the equity concerns of (and about) government. We offer the framework of artificial discretion to understand and anticipate the diffusion of AI in government, and assess its impact along several dimensions. It is not clear in the net how these changes in the boundaries between human and machine capabilities will affect public administration and governance. The field of public administration needs to grapple with these concerns before the outcomes are decided for us.

References

Alvarez-Cuadrado, Francisco, Ngo Van Long, and Markus Poschke. 2018. Capital-labor substitution, structural change and the labor income share. *Journal of Economic Dynamics and Control* **87** (February): 206–31.

Andrews, Leighton. 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.

Anthopoulos, Leonidas G., and Christopher G. Reddick. 2016. Understanding electronic government research and smart city: A framework and empirical evidence. *Information Polity* **21** (1): 99–117.

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. The moral machine experiment. *Nature* **563** (7729): 59.

Ballard, Michael James, and Mildred E. Warner. 2000. *Taking the high road: Local government restructuring and the quest for quality*. PhD diss., Ithaca, NY: Cornell University.

Barth, Thomas J., and Eddy Arnold. 1999. Artificial intelligence and administrative discretion: Implications for public administration. *The American Review of Public Administration* **29** (4): 332–51.

Barton, Dominic, and D. Court. 2012. Making advanced analytics work for you. *Harvard Business Review* **90** (10): 78–83.

Bellamy, Rachel K. E., Kuntal Dey, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, and Aleksandra Mojsilovic. 2018. AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. ArXiv Preprint ArXiv:1810.01943.

Bonnefon, Jean-François, Azim Shariff, and Iyad Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* **352** (6293): 1573–6.

Bostrom, Nick. 2003. Ethical issues in advanced artificial intelligence. In Schneider, Susan, ed. *Science fiction and philosophy: from time travel to superintelligence*. John Wiley & Sons, 2016.

———. 2014. *Superintelligence: Paths, dangers, strategies*. Oxford, UK: Oxford University Press.

Bovens, Mark, and Stavros 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.

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, *et al*. 2018. *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. https://arxiv.org/abs/1802.07228v1.

Bullock, Justin. 2019, forthcoming. Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*. doi:10.1177/0275074019856123.

Bullock, Justin B., Robert A. Greer, and Laurence J. O'Toole. 2019. Managing risks in public organizations: A conceptual foundation and research agenda. *Perspectives on Public Management and Governance* **2** (1): 75–87.

Buolamwini, J., and Gebru, T. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research* **81**: 1–15.

Busch, Peter André, and Helle Zinner Henriksen. 2018. Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity* **23** (1): 3–28.

Corbett-Davies, Sam, and Sharad Goel. 2018. The measure and mismeasure of fairness: A critical review of fair machine learning. ArXiv Preprint ArXiv:1808.00023 [Cs], July. http://arxiv.org/abs/1808.00023.

Dastin, Jeffrey. 2018. *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (accessed October 10, 2018).

de Vries, Hanna, Lars Tummers, and Victor Bekkers. 2018. The diffusion and adoption of public sector innovations: A meta-synthesis of the literature. *Perspectives on Public Management and Governance* **1** (3): 159–76.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* **144** (1): 114.

———. 2016. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* **64** (3): 1155–70.

DiMaggio, Paul J., and Walter W. Powell. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* **48** (2): 147–60.

Dressel, Julia, and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science Advances* **4** (1): eaao5580.

Dunleavy, Patrick, Helen Margetts, Simon Bastow, and Jane Tinkler. 2006. New public management is dead – Long live digital-era governance. *Journal of Public Administration Research and Theory* **16** (3): 467–94.

Eren, Ozkan, and Naci Mocan. 2018. Emotional judges and unlucky juveniles. *Applied Economics* **10** (3): 171–205.

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

Fennimore, Anne, and Arthur Sementelli. 2019. Accounting for the horsemen of the administrative apocalypse. *Public Integrity* **21** (2): 114–26.

Fountain, Jane E. 2001. *Building the virtual state: Information technology and institutional change*. Washington, DC: Brookings Institution Press.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* **114** (January): 254–80.

Frick, Walter. 2015. When your boss wears metal pants. *Harvard Business Review*. https://hbr.org/2015/06/when-your-boss-wears-metal-pants (accessed June 1, 2015).

Gailmard, Sean, and John W. Patty. 2012. *Learning while governing: Expertise and accountability in the executive branch*. Chicago and London: University of Chicago Press.

Gianfrancesco, Milena A., Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk. 2018. Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine* **178** (11): 1544–7.

Gil-García, J. Ramón, Sharon S. Dawes, and Theresa A. Pardo. 2018. Digital government and public management research: finding the crossroads. *Public Management Review* 20 (5): 633-646.

Goetz, Jennifer, Sara Kiesler, and Aaron Powers. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003, 55–60. doi:10.1109/ROMAN.2003.1251796.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. "Generative adversarial nets." *Advances in Neural Information Processing Systems* **27**: 2672–2680.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 2018. When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research* **62**: 729–54.

Gurovich, Yaron, Yair Hanani, Omri Bar, Guy Nadav, Nicole Fleischer, Dekel Gelbman, Lina Basel-Salmon, *et al*. 2019. Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine* **25** (1): 60.

Helsby, Jennifer, Samuel Carton, Kenneth Joseph, Ayesha Mahmud, Youngsoo Park, Andrea Navarrete, Klaus Ackermann, *et al*. 2018. Early intervention systems: Predicting adverse interactions between police and the public. *Criminal Justice Policy Review* **29** (2): 190–209.

Hilbert, Martin, and Priscila López. 2011. The world's technological capacity to store, communicate, and compute information. *Science* **332** (6025): 60.

Huber, John D., and Charles R. Shipan. 2002. *Deliberate discretion? The institutional foundations of bureaucratic autonomy*. Cambridge, UK: Cambridge University Press.

Husain, Amir. 2018. *The sentient machine: The coming age of artificial intelligence*. New York, NY: Scribner.

Ilyas, Andrew, Logan Engstrom, Anish Athalye, and Jessy Lin. 2018. Black-box adversarial attacks with limited queries and information. ArXiv Preprint ArXiv:1804.08598 [Cs, Stat], April. http://arxiv.org/abs/1804.08598.

Jago, Arthur S. 2019. Algorithms and authenticity. *Academy of Management Discoveries* **5** (1): 38–56.

Janssen, Marijn, and George Kuk. 2016. The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly* **33** (3): 371–7.

Jun, Kyu-Nahm, and Christopher Weare. 2010. Institutional motivations in the adoption of innovations: The case of e-government. *Journal of Public Administration Research and Theory* **21** (3): 495–519.

Kahneman, Daniel. 2011. *Thinking, fast and slow*. New York, NY: Macmillan.

Kim, Young Mie. 2016. Algorithmic opportunity: Digital advertising and inequality in political involvement. *The Forum* **14** (4): 471–84.

Kitchin, Rob. 2014. Big data, new epistemologies and paradigm shifts. *Big Data & Society* **1** (1): 1-12.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. Human decisions and machine predictions. *The Quarterly Journal of Economics* **133** (1): 237–93.

Lee, Kristin. 2016. *Artificial intelligence, automation, and the economy*. Whitehouse.Gov. https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy (accessed December 20, 2016).

Lee, Peter. 2018. *The US army is developing unmanned drones that can decide who to kill*. Business Insider. https://www.businessinsider.com/the-us-army-is-developing-unmanned-drones-that-can-decide-who-to-kill-2018-4 (accessed April 13, 2018).

Lipton, Zachary C. 2016. The mythos of model interpretability. ArXiv Preprint ArXiv:1606.03490.

Liu, Weibo, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. 2017. A survey of deep neural network architectures and their applications. *Neurocomputing* **234**: 11–26.

Lutz, Eric. 2019. *China has created a racist A.I. to track Muslims*. The Hive. https://www.vanityfair.com/news/2019/04/china-created-a-racist-artificial-intelligence-to-track-muslims (accessed April 15, 2019).

McCorduck, Pamela. 2004. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. 25th anniversary update. Natick, MA: A.K. Peters.

Meier, Patrick. 2015. *Digital humanitarians: How big data is changing the face of humanitarian response*. New York, NY: Routledge. doi:10.1201/b18023.

Meijer, Albert. 2018. Datapolis: A public governance perspective on 'smart cities'. *Perspectives on Public Management and Governance* **1** (3): 195–206.

Noble, Safiya Umoja. 2018. *Algorithms of oppression: How search engines reinforce racism*. New York, NY: NYU Press.

NTSC. 2016a. *Preparing for the future of artificial intelligence*. Washington, DC: National Science and Technology Council, Executive Office of the President.

———. 2016b. *Artificial intelligence, automation, and the economy*. Washington, DC: National Science and Technology Council, Executive Office of the President. http://search.proquest.com/docview/1850703740 (accessed October 2, 2018).

O'Leary, Rosemary. 2013. *The ethics of dissent: Managing guerrilla government*. Thousand Oaks, CA: Cq Press.

O'Neil, Cathy. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown.

Papernot, Nicolas, and Patrick McDaniel. 2018. Deep K-nearest neighbors: Towards confident, interpretable and robust deep learning. ArXiv Preprint ArXiv:1803.04765.

Pentland, Alex. 2014. *Social physics: How good ideas spread-the lessons from a new science*. New York, NY: Penguin.

Poole, David, Alan Mackworth, and Randy Goebel. 1998. *Computational intelligence: A logical approach*. Oxford, UK: Oxford University Press.

Rice, Douglas R., and Christopher Zorn. 2019. Corpus-Based Dictionaries for Sentiment Analysis of Specialized Vocabularies. *Political Science Research and Methods*. Cambridge, UK: Cambridge University Press, 1–16.

Russell, Stuart, and Peter Norvig. 2009. *Artificial intelligence: A modern approach*, 3rd ed. Upper Saddle River, NJ: Pearson.

Salamon, Lester M. 2002. *The tools of government: A guide to the new governance*. Oxford, UK: Oxford University Press.

Savas, Emmanuel S. 2000. *Privatization and public-private partnerships*. Chatham House, NJ: Chatham House.

Schmidhuber, Jürgen. 2015. Deep learning in neural networks: An overview. *Neural Networks* **61**: 85–117.

Schneider, Anne, and Helen Ingram. 1993. Social construction of target populations: Implications for politics and policy. *American Political Science Review* **87** (2): 334–47.

Selbst, Andrew D. 2017. Disparate impact in big data policing. *Georgia Law Review* **52**: 109.

Simon, Herbert A. 1991. Bounded rationality and organizational learning. *Organization Science* **2** (1): 125–34.

———. 2013. *Administrative behavior*. New York, NY: Simon and Schuster.

Turing, Alan M. 1950. Computing machinery and intelligence. *Mind* **59** (236): 433–60.

Veale, Michael, Max Van Kleek, and Reuben Binns. 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 440:1–440:14. CHI '18. New York, NY: ACM.

Viola, Paul, and Michael J. Jones. 2004. Robust real-time face detection. *International Journal of Computer Vision* **57** (2): 137–54.

Wade-Olson, Jeremiah. 2016. Race, staff, and punishment: Representative bureaucracy in American state prisons. *Administration & Society*. doi:10.1177/0095399716667156.

Zou, James, and Londa Schiebinger. 2018. AI can be sexist and racist – It's time to make it fair. *Nature* **559** (7714): 324.