

University of Ottawa

**Applying Artificial Intelligence to Public Sector Decision Making**

Major Research Paper

Will Alexander

8785120

PAP7998

Dr. David Brown

December 19th 2022

	2
<b>Introduction</b>	<b>3</b>
<b>Artificial Intelligence and Public Sector Organizations</b>	<b>6</b>
An Early History of AI and Organizational Theory	6
A Recent History of AI and Organizational Theory	8
The Contemporary Canadian Public Sector's Take on AI	11
Technostructure and Technological Rationality	13
The Technical Knowledge-Policy Knowledge Rift	14
Human-AI Interdisciplinarity	16
Section Summary	17
<b>Algorithmic Governance</b>	<b>18</b>
Big Data and Automation	18
Automated Decision-Making Systems	19
And the Humans?	21
Governance by Algorithm	22
Section Summary	23
<b>Theoretical Model</b>	<b>23</b>
Decisions	24
Limits	25
AI-Human Dimorphism	27
Humans	28
ADM Systems	29
Efficiency and Decision	31
Section Summary	32
<b>Implications</b>	<b>32</b>
Organizational Transformation	34
The Technical Knowledge Gap	36
The Efficiency/Ethics Compromise	41
Section Summary	46
<b>Conclusion</b>	<b>47</b>
<b>Notes</b>	<b>50</b>
<b>References</b>	<b>53</b>

## Introduction

Data and data analysis are central to decision making in governance. In accordance with the predictions of Digital Era Governance (DEG) put forth by Dunleavy et al. during the Web 2.0 era, the governance climate of the post-New Public Management age has been characterized by evolving IT systems and changes in information networks (Dunleavy et al. 2006). The rate at which technology evolves has accelerated as it has diffused through all layers of government. In this new environment, older models of understanding technology rapidly fade in relevance as pressure mounts to digitize and to automate legacy systems and procedures across the board (Ojo 2019).

A more recent development in the current wave of DEG is the rising prominence and popularity of artificial intelligence (AI) systems. Straddling the technological boundary between big data and automation, today's AI systems differ from their predecessors in that they are capable of 'learning.' Where a conventional system would be limited to a static set of functions operating at a consistent quality of output, a 21st century AI system is unique in that it can actively improve its prediction, classification, and evaluation capacities.

I use the term 'learning' somewhat loosely because there is a tendency across the literature to treat all AI-systems as though they are identical, or as though they are designed for the sole purpose of directly emulating human cognitive abilities, but that is not the case. AI does not learn in the same way that humans learn. In truth, the current generation of AI systems are complex statistical models with a wide variety of functions and applications (Knox 2018). These misconceptions about the nature of AI likely stem from the role that science fiction has had in framing AI in the public consciousness (Perez 2019) and the fact that anthropomorphizing AI tends to increase its acceptability (Pelau et al. 2021) while making it simpler for the digital

layperson to understand — ignoring the nuances of AI technology in favour of treating it as an inscrutable, immutable whole.

Recent public administration literature has shown that there is a need to understand how contemporary governance will have to adapt in response to AI (Zuiderwijk 2021; Newell 2015; Veale 2019; Gomes de Sousa 2019; Alexopoulos 2019). Benefits, such as efficiency or accuracy of decision-making, are often claimed and are positioned opposite drawbacks like workforce substitution (Valle-Cruz 2020; Wirtz et al. 2020) or ethical concerns like bias (Newell 2015; Wirtz et al. 2020), lack of accountability, and opaqueness (Alexopoulos 2019).

However, a common ‘technology adoption’ narrative in the literature implicitly treats AI systems as incidental to a larger independent governance process. As Höchtl et al. remark, “[i]n most studies, technology is viewed as complementary to existing organizational and administrative practices but is hardly viewed as capable of triggering changes in the actual structure of decision-making processes,” despite the fact that technology has a history of precipitating complete organizational transformation (Höchtl et al. 2016, p.147). Roy describes this process of technology-spurred change in terms of the Schumpeterian trilogy: as beginning with the automation of existing systems, followed by a search for new areas to innovate, before eventually transforming the institution entirely (Roy 2008). Broadly speaking, public organizations appear to already be in the process of a new phase of AI-fueled transformation.

If AI is integrated into broader governance structures it will become a part of that very process to which it is presently being compared at arm’s length. Automation and integration of AI into decision-making processes will impact how governance is understood to function and how solutions are evaluated by public servants. This will recursively redefine how the aforementioned issues are perceived, in a cyclical process of reinvention and redefinition.

Current administration literature on AI generally focuses on specific issues involving AI, ignoring the broader structural picture and rarely stopping to consider the nature, capabilities, and limitations of the technology itself. In response to these shortcomings I will develop a new, adapted bounded rationality model to demonstrate the current deficiencies in both academic and government guidelines regarding the most recent wave of DEG: that of AI-based automated decision-making systems. This model will use ‘decision’ as its building block to frame both human and AI elements as comparable actors that make choices limited by their nature in public sector organizational structure.

In light of these thoughts, this paper asks: ‘How will organizational structure adapt to changes brought on by AI, and what effect will these changes have on the guiding principles of public organizations?’ This inquiry will be directed by a modernized model of bounded rationality framed by the dimensions of efficiency and ethics.

This paper contributes to the greater body of public administration research by offering an alternative framework for understanding the current transformation in technology that has begun to reconceptualize the decision-making process of an emerging age of advanced learning systems. I argue that traditional methods of thinking about governance alongside older generations of technology are limited when discussing AI-technologies because they make a critical oversight: they ignore the unique properties of AI technology, and consequently overlook some of the ways that AI can transform governance institutions at the deepest levels.

This paper will demonstrate that framing AI using bounded rationality has advantages for public administration analysis and critique. Using this theoretical model, it will argue that an imminent AI transformation will shift the balance of organizational power towards technocrats, and will obligate the government to choose a path forward based on the level of ethical risk it is

willing to accept to improve organizational efficiency. The degree of forthcoming structural change (as well as the future of public sector AI) will depend on the policies and mitigation measures that the government establishes now.

This paper is arranged in the following way: first, a review of the history of AI, a background on the relationship between technology and government, and an overview of some of the key issues from the literature on contemporary algorithmic governance. This will be followed by a theoretical model adapted from the works of Herbert Simon and then, finally, a discussion of how AI-based decision-making systems will affect public sector structure and decision making.

## Artificial Intelligence and Public Sector Organizations

### **An Early History of AI and Organizational Theory**

Artificial intelligence is a lot older than many assume and, as a practice, it has been intertwined with organizational theory since its inception. The field of AI research is built on a core idea that dates back to antiquity: ‘Can an artificial entity *think* as humans do?’ (Crevier 1993, p.2). While this question is metaphysical in nature and beyond the scope of this paper, we can derive from it that it is the *desire* to create a facsimile of the human mind that has driven much of the progress in developing AI.

Many of the ideas touted today as paradigm-shifting have actually been explored in depth many times before but were limited by the technology of their times, fading back into obscurity as their spotlight waned. Over the last half-century, AI has gone through a few cycles of boom and bust. A new leap forward in AI research would generate excitement, funding, and interest until the technology itself could not keep pace with the grand aspirations that had brought it to

prominence. Interest would then overcorrect into disillusionment and cynicism and resulting in an “AI Winter” that would last until the next great stride in technological progress, at which point the cycle would recommence (Crevier 1993).

My hope for this section is to provide some context and in the same stroke dispel some of the more concerning misconceptions about AI technology by providing a genealogy of the shared history of organizational decision making theory and AI systems. I will explore Herbert Simon’s legacy with the benefit of hindsight and an understanding of the upper limit of technological capacity today.

This story begins with the ‘Logic Theorist,’ a computer program designed by Herbert Simon, Al Newell, and Cliff Shaw in 1956 largely considered to be the first AI system. The program made use of the rudimentary technology of the time to automatically apply a simple heuristic, or series of rules, in order to prove a hypothesis. For Simon, the decision making process was really just a systematic search through a finite number of options — Newell and Shaw used this same process as a framework to program their system (Crevier 1993).

In this sense, the Logic Theorist was also the first AI-powered automated-decision making system. Under Simon’s theory of bounded rationality, a decision maker in an organization makes choices by reviewing readily available options and selecting the best one according to the limits of their judgement, defined as “satisficing” — it was this concept that formed the basis of both Simon’s theories on organizational behaviour and his Logic Theorist AI system (Simon 1956).

As proof of the program’s potential they used the Theorist to successfully prove a number of theorems from Russell and Whitehead’s *Principia Mathematica*. One of the proofs was even done “more elegantly” than in the original (Crevier 1993, p.46). Their work was presented at the

1956 Dartmouth Conference where the term “Artificial Intelligence” was coined, leading to the birth of the field.

There are two important takeaways from this account: Firstly, the Logic Theorist proved that an information processing system could “satisfice” much like humans, and could even make better decisions in some cases, and secondly it established the rule-based heuristic as central to this process of automating decision making which would persist for the remainder of the century (Gugerty 2006).

### **A Recent History of AI and Organizational Theory**

The computers of the mid-20th century were underpowered and expensive and, as a result, interest in AI had largely faded by the 1970s (Crevier 1993). Computers did not yet have the capacity to imitate some of the more complex abilities of the human mind like speech or text recognition, nor were they able to process and adapt using ‘known’ information like humans can. McCorduck remarks that researchers working on speech-recognition in the 70s came to realize that “[k]nowledge [...] is mainly dynamic and not static. The symbols that stand for knowledge are entities with a functional property” and the AI of the time was not capable of applying the ‘functions’ of that knowledge (McCorduck 2005, p. 313). The heuristic model first used by Simon’s team, with its static rules, was not sufficiently advanced to emulate the higher abilities of the human mind. Solving the riddle of advanced AI required a system that could operationalize knowledge in the same way that humans can — a system that could take data and transform it into knowledge, or *learn*, autonomously, which was a problem that lay beyond the reach of that era’s technology.

The 1980s gave rise to the New Public Management (NPM) movement in public administration characterized by widespread decentralization, stifling the flow of information within government (Ojo 2019). NPM reorganized government into “discrete corporate hierarchies,” siloing departments and increasing reliance on private sector contracts for things like IT services (Dunleavy et al. 2006, p. 227). This is also when automated-decision making systems began to gain traction as tools to assist in government work. Hadden, writing in the late 80s, noted that expert systems had potential to increase the accuracy and speed of decisions, as well as be a “means for providing a justification for decision” (Hadden 1989, p. 207).

Another AI Winter hit in the 90s, but by the 2000s interest had renewed again in earnest and it has been accelerating ever since, in particular due to a recent leap forward in technological progress — the great advances made in the field of Machine Learning (ML). ML and AI are not synonymous, although they are closely related. In the words of Mitchell, “Machine Learning is the study of computer algorithms that improve automatically through experience” (Mitchell 1997, p. xv). In the context of the present analysis, ML is a version of Simon’s original 1956 heuristic with the upgraded functionality of dynamic, adaptable rules.

The field of ML began with the old problem of teaching a computer to learn, but quickly evolved to focus on “classification and regression tasks.” Langley points out that recently ML as a whole has shifted primarily towards statistics, borrowing heavily from probability theory (Langley 2011, p. 277), and is now effectively the study of extremely complicated adaptable mathematical models (Knox 2018).

Deep learning is a term for the overlap of AI and ML — the application of the dynamic statistical models from ML to AI to teach AI to ‘learn,’ thereby solving the riddle that had stumped researchers 25 years earlier. This is the leap ahead that has brought AI back to the fore.

Thanks to advances made in deep learning over the last decade AI has finally had massive success in “natural language understanding, particularly topic classification, sentiment analysis, question answering and language translation” (LeCun et al. 2015, p. 436).

Between its inception in the 1950s and today, the guiding principle of artificial intelligence has evolved from “can an artificial entity think?” to “can an artificial entity do things that previously only humans could?” Contemporary AI systems are no longer designed to reproduce the human mind, but to reproduce specific human capabilities. AI systems, properly defined, are machines capable of carrying out a particular task that previously could only be done by the human brain — learning and reasoning, among others.

Today, AI has become extremely profitable for businesses (Chen et al. 2019), but the NPM of the 1980s has led to an imbalance in expertise favouring the private sector (Wirtz 2020). Deciding to outsource technological talent 30 years ago has led to a drought in today’s public sector IT expertise.

Outlook on new technology has historically been stretched between idealistic futurologists and diehard cynics, with examples dating back to the industrial revolution and earlier. The best perspective likely lies somewhere in the middle — as Floridi notes, AI is “neither a miracle nor a plague” but a new challenge for humanity to come to terms with (Floridi 2020, pg. 2). For Talby, “there are no accurate machine learning models, only agreeable ones” and “agreeableness” is largely dependent on the observer’s ability to understand what is actually happening under the hood of an AI system (Talby 2019).

Concern is warranted, but much of the current concern stems from a poor understanding of the subject matter. There are obvious efficiency benefits that have been starved for attention and there are also large gaps in public administration scholars’ review of ethical concerns, but in

order for both to be properly analyzed and assessed we must first firmly establish how we understand AI in relation to the public sector.

### **The Contemporary Canadian Public Sector's Take on AI**

In Canada, there is earnest government interest in getting ahead of the curve on AI. The Treasury Board's Policy on Service and Digital section 4.1.2.1.3 mandates investigating the potential of "using emerging technologies and the implications and opportunities of doing so for the Government of Canada" (Treasury Board Secretariat 2019). Literature on emerging technologies for government purposes ranges tonally between cautious optimism and heavy skepticism. The most recent 'AI spring' is still relatively new to the world of public administration (Ben Rjab et al. 2019), so it comes as no surprise that there is trepidation among institutions to jump on the AI bandwagon (Kuziemski & Misuraca 2020).

The Government of Canada has begun to lay the groundwork for future work in this sector. In 2021 the government invested \$208 million in research over ten years through several institutes (Innovation, Science and Economic Development Canada 2022) and certain organizations, such as the Canadian Treasury Board Secretariat and the European High-Level Expert Group on AI have put forth guidelines for designing public sector AI systems (Treasury Board Secretariat 2021; European Commission 2021); however these methods are characteristic of Roy's concept of outdated modes of controlling new technology (Roy 2008).

Particularly telling of the government's discomfort and unfamiliarity with this area of technology is the TBS's Algorithmic Impact Assessment Tool, which contains primarily yes/no questions, such as "will the algorithmic process be difficult to interpret or to explain?" and which only four departments or agencies had filled out as of early 2022 (Longo 2022, p.387).

Hildebrandt remarks that misperceptions about how algorithms function is responsible for much of the distrust currently directed towards AI (Hildebrandt 2021).

There are several independent factors that complicate the government's ability to make sense of the AI technologies recommended to them as 'helpful' or 'useful.' The mystification of AI (McKelvey & MacDonald 2019), the conflation of numerous technologies with vastly separate functionality under the same name (Gomes de Sousa 2019), and the imbalanced distribution of knowledge favouring the private sector (Wirtz et al. 2020) have all put the government at an early disadvantage

In order to address these factors with the benefit of a more technical perspective, it is easier to begin by focusing on what AI systems are not: At the current level of technology, AI systems are not direct replacements for human beings. Today, AI research is more interested in reproducing specific human abilities than creating direct human substitutes.

AI systems are also not yet fully independent, self-generating, nor are they magic. Parameters must still be set, criteria established, categories delineated, and data procured (Anastasopoulos 2018; Hildebrandt 2021), and they are mostly used as support systems for data-driven decision-making (Alexopoulos 2019). The specifics of algorithmic governance will be elaborated below, but the mere threat of low organizational understanding of AI is important to consider early on because it entails the risk that poorly-informed managers may be taken advantage of by "AI snakeoil salespeople" (Longo 2022) or simply make poor decisions about AI integration in the public sector, leading to situations reminiscent of the Phoenix pay system debacle.

Within the broader scope of government, current measures like macro-level horizontal guidelines are equivalent to a manager laying out ground rules for a new employee before having

a position ready for them, or even knowing anything about them. There is strong need for a comprehensive exploration of how (and where, and even *if*) these systems should actually exist within the meso- and micro-structures of government, in addition to guidelines.

### **Technostructure and Technological Rationality**

This section will address the foundational considerations that must be made for a valid technological understanding of the AI-government situation, the most important of which is that a model must incorporate the unique features of AI technology that distinguish it from its predecessors. The urge to reuse the cookie-cutter ‘tech-adoption’ methods that have worked for past innovations like the emergence of information-storage systems, wireless communication, and Web 2.0 must be avoided (Ojo 2019). Importantly, understanding this new generation of technology requires technical knowledge in addition to policy experience.

This matter can largely be divided into the following three related concerns: a lack of IT expertise at the apex of public sector organizations, a risk of legitimacy loss in the public sector, and a widening explainability gap between technical experts and policy experts.

Dunleavy et al. describe modern government IT as a bicameral system with a “support structure” and a “technostructure.” The support structure is responsible for the upkeep of existing systems and makes up the bulk of IT staff in government, while the role of the technostructure is to investigate government processes and evaluate them for efficiency (Dunleavy et al. 2006). However, Clarke et al. and others remark that governments are typically ill-equipped to take on the process of thoroughly evaluating the efficiency of their technological backbone (Clarke et al. 2017; Valle-Cruz 2020; Wirtz 2020). In particular, the higher pay and better working conditions offered by the private sector has created a knowledge gap in the public sector technostructure

(Wirtz 2020). The result is a less-adaptable public service and lower IT reflexive capacity in government.

However, acquiring a stronger technostructure in government is not necessarily a cure-all either. If the bar for entry to adjust organizational structure is advanced systems knowledge, then that authority is only available to technical experts. If more decisional authority skews towards these experts, another issue arises: loss of government legitimacy. A larger and more powerful technostructure is less accountable to both elected officials and the general public unless there are strong mechanisms in place to keep them in check (Newell 2015). Roy makes the case that this puts democratic principles at odds with bureaucratic power (Roy 2008).

As government changes due to technology, it is natural that ways of thinking about government will change as well. Haque cites Marcuse to explain that “the larger implication of technology is not primarily limited to technology, but to faith in technological rationality” (Haque 2015). This faith translates to a transformation in the way that power is distributed in the public sector — responsibility for governance could shift further towards those with more technical know-how as, in a more technologically dense organization, they would be seen as better-suited to bear the brunt of decision making in a hyper-rationalist value paradigm. Kernaghan argues that this makes them also bear moral responsibility for the systems that they design (Kernaghan 2014), although the underlying question of ensuring accountability within a technological system fundamentally alien to core public servants remains unanswered.

### **The Technical Knowledge-Policy Knowledge Rift**

The relationship between technical experts and public servants is central to this issue, and it is an old problem in public administration, illustrated by Hodgetts in 1955’s *Pioneer Public*

*Service* where he used the example of the tribulations that occurred during the 19th century construction of parliament to emphasize the importance of communication and common understanding between those with policy knowledge and those with technical knowledge (Hodgetts 1956). AI does, however, have one unique attribute that complicates this already tenuous relationship — a particularly troubling quirk of modern AI is that the most sophisticated ML methods exceed what technical experts can properly explain to laypeople, or even understand themselves (Yampolskiy 2019). This is the ‘explainability problem,’ and it couples with Haque’s ‘faith in technological rationality’ problem to evoke a future where the government will reach a crossroads forcing it to choose between having blind faith in technology in return for improved returns on efficiency or forgoing the advantages of AI to preserve institutional accountability.

The situation is somewhat more complicated than a simple binary decision. Regarding the technician/policy expert schism, a middle ground has been proposed. Roy has encouraged collaboration, engagement, and empowering leadership to understand the issues of AI (Roy 2008) and Clarke et al. have stressed the importance of having “more Canadian public administration scholars ‘on the case’ with richer linkages to specialists in related disciplines and professions, such as computer and data sciences” (Clarke et al. 2022, p.461).

However, there are many reasons why bridging the gap will be difficult. Dunleavy et al. observe that government IT specialists work separately from the general public service, elites in government are typically generalists with little knowledge of the finer workings of technology, and the routineness of IT processes tends to distract from the incremental changes they have on organizational structure (Dunleavy et al. 2006). The larger process of transformation at work

behind the scenes is nestled deep within a multi-layered hierarchical decisional framework that requires a nuanced understanding of organizational decision making to properly unpack.

### **Human-AI Interdisciplinarity**

This challenge of integrating AI into public sector organizations will require expertise in multiple fields of knowledge (Valle-Cruz 2020; Wirtz 2020; Clarke et al. 2017; Dwivedi et al. 2021); while Newell and Shaw had backgrounds in computer science, Simon's was in organizational theory and public administration (Crevier 1993). The field of AI began at the convergence of computer science and organizational theory, and it is probable that the government of tomorrow will require not just expertise in both fields, but expertise in how the two can work together. The best place to begin is by highlighting a few ways that AI technology will spur change in organizational thinking: firstly in specific instances, and secondly as a broader transformative influence.

For considering individual cases, it is essential to remember that most AI are no longer designed to reproduce the entire human mind, and so modern AI systems should not be considered human-analogous or universal human substitutes — AI systems are, by nature, better at certain tasks than humans and worse at others. It is the specific capabilities that they are designed to mimic that they excel at.

The role of contemporary AI is therefore not always a “bureaucrat substitute” and the role of the techno-bureaucrat cannot be restricted purely to “tool user” or “technological overseer.” While these classic, general models of public administration are useful in certain contexts, when the question at hand concerns AI technology a comprehensive model must give due

consideration to how that technology can change the way that things are done within an organization.

This leads to the second consideration: that technology will gradually shape the way that entire organizational structures operate over time. Weber argued long ago that technology is an ‘iron cage’ — that the emphasis on rationalization caused by technology will also cement institutional dependence on it. This process is described by Schroeder and Ling as “routinization,” where technology becomes structurally embedded in the way that organizations are run due to their ubiquity in society (Schroeder & Ling 2014). For Roy, this process of automation, innovation, and reinvention eventually results in widespread institutional reform, wherein AI could transform the way that government works through a process of gradual normalization (Roy 2008). While an organization tries to get AI to adapt to it through Impact Assessment Tools and policy initiatives, AI will also subtly change how things are done as it evolves and propagates.

### **Section Summary**

Change is at hand, both technological and organizational. The public sector does not have the experts to handle the change, and even if they did, power would shift towards them and trouble the fundamental principles of accountability and legitimacy that public sector organizations are built on. The need for a new model is not merely a recommendation but an inevitability if AI integration continues on its current path. This paper will frame this change as the struggle between technical superiority — efficiency — and ethical consistency within bounded rationality.

A final note is that the route forward for Canadian bureaucracy is still largely undetermined. To what degree automated technologies will be incorporated into government, how organizations will prepare as the future inches closer, and what reciprocal changes will be precipitated by AI can be estimated but, at present, remain to be decided.

## Algorithmic Governance

### **Big Data and Automation**

There are many ways to look at AI in the context of administrative theory because AI is neither a unitary technology nor is governance a single undertaking. AI, as a category of technology, has not emerged from the aether but is rather an umbrella term for a host of distinct but related techniques in computing (Valle-Cruz 2020).

Interest in AI may be the most recent trend in public administration research, but it is also the most recent innovation in a long sequence of continuous developments in computer and data science (Zalnieriute et al. 2019). It is hard to limit AI to a singular definition or frame it in a single context because the technology is heterogenous and constantly evolving (Kuziemski & Misuraca 2020). AI has a multitude of functions, and worldwide has been applied in different ways across almost all primary functions of government (Gomes de Sousa 2019).

Within the broader government-technology landscape, the leading edge of AI integration is located at the nexus of big data and automation (McKelvey and MacDonald 2019). Big data is a term referring to modern systems of data that differ from traditional data in three ways: they contain substantially higher volumes of data (meaning exponentially larger datasets), they have higher velocity (meaning that information travels faster within the system) and they have a greater variety of data (meaning that data exists in a greater diversity of formats) (Redden, 2015,

Höchtel et al. 2016). Big data itself is generally considered to be a reliable part of the public sector reporting process (Janssen et al. 2020).

Automation is a process that reduces human involvement in a task by delegating work to an automatic system. Automation typically lends itself better to simple tasks, but recent innovations in ML have allowed AI systems to take on tasks of increasing complexity such as classifying hate-speech (Tang and Dalzell 2019), identifying fake news (Khan et al. 2021), and synchronizing personal records across databases (Ansolabehere and Hersh 2017). AI was noted early on to be particularly suited to automate the ‘routine but complex tasks’ that make up a considerable portion of overall government workload (Hadden 1989; Valle-Cruz 2020).

### **Automated Decision-Making Systems**

The confluence of big data and automation for governance is where AI-assisted automation can help interpret and mobilize government information — deep learning is currently the best means available to unlock the operational potential of big data (Anastasopoulos 2018). In government, data networks are often fragmented across multiple departments and use mismatched standards, which has historically caused problems when trying to draw meaningful conclusions (Janssen et al. 2020) or providing supporting evidence for policy solutions (Alexopoulos, 2019; Valle-Cruz 2020). AI systems can process this data more efficiently than human workers, allowing new solutions to be produced faster, and unlocking new pathways for governance going forwards.

AI includes computer vision (interpreting images and video), language processing, and advanced robotics — among other techniques — all highly specialized in both their function and application (Kuziemski & Misuraca 2020). Computer vision and language processing may be

used in future government services (and indeed have been used in processing mail since the 1990s) (Mehr 2017); the Canadian Treasury Board Secretariat’s “Responsible AI” whitepaper categorizes these aids as ‘service delivery,’ comprising search enhancement systems, intelligent chatbots, and automated support. They are the natural successors to the current automated systems already in use by the Canadian government (Treasury Board Secretariat 2018).

AI also allows for the automated interpretation of data, and has been variously described as a supporting/expert system (Alexopoulos et al. 2019; Zalnieriute 2019), a way to simulate human intellect (Valle-Cruz 2020) and a classification/prediction machine (Anastasopoulos and Whitford 2018) — all functions that can be categorized under ‘policy design and risk response,’ the TBS’s second grouping for AI systems.

This second subset of government AI exists within the broader semantic category of Automated Decision-Making (ADM) systems, which are systems that process large quantities of data via an algorithm to produce a data-informed output (Araujo et al. 2020); examples of this include automating immigration and employment service application processes (Kuziemski & Misuraca 2020). While ADM systems predate AI-technology, recent breakthroughs in ML have vastly expanded their potential.

ML breaks down largely into supervised and unsupervised models, under which there are more specialized (and increasingly complex) approaches such as neural networks, classification trees, and support vector machines (Anastopoulos et al. 2018; Knox 2018).<sup>1</sup> These last few techniques are some of the most accurate methods available today, but they are also the most complicated (Yampolskiy 2019).

## **And the Humans?**

The ideal level of human involvement in ADM systems is still a matter of debate (Araujo et al. 2020), as are ethical concerns like the future of legitimacy and accountability (Zalnieriute et al. 2019), and systematic discrimination (Redden 2015). AI systems have, however, been noted to make better decisions (Araujo et al. 2020), have the potential to accelerate and refine the policy cycle (Höchtel et al. 2016; Valle-Cruz, 2020), and could transform government processes in order to adapt to a more wholly-digital age (Roy 2008). Overall, considerable efficiency benefits are expected, alongside risks of ethical losses in the realms of accountability and transparency (Newell 2015; Di Vaio 2022).

There are reasons to believe that old paradigms will need to be adapted or replaced as AI becomes more important in both the public and private worlds (Roy 2008; Ojo 2019). The rise of AI-powered ADM systems presents a host of possibilities and challenges for the future of public administration research. In the field of public administration there is a lack of technical knowledge on AI (Valle-Cruz 2020), leading to technical aspects being overlooked in research (Gomes de Sousa 2019). A new way of understanding advanced technology is a prerequisite to making sense of AI for the field of public administration.

If we view the whole of government as a system for making decisions (as will be explored in the following section) then the very objective of work is changing fundamentally. Bounded rationality dictates that humans will make the best decision constrained by their means, and AI has the potential to enlarge those bounds by a substantial margin.

After designing the Logic Theorist, Simon predicted that the “the range of problems [AI] can handle will be coextensive with the range to which the human mind has been applied” (Simon 1957, quoted in Crevier 1993, p.1), and many of the old limitations detailed in the 70s

are no longer concerns for today's systems (Simon 1973). Bearing this in mind, it is likely that the role of the human will evolve in accordance as AI systems become more proficient in the types of work better suited to their nature.

### **Governance by Algorithm**

“Algorithmic governance” is a term that has emerged recently to describe an organization in which algorithmic systems have become ubiquitous. Under this model, the human and the algorithm are equally relevant and essential to the functioning of the state. While no one has yet claimed the title of ‘first algorithmic state’ or been framed as such in the literature, the prevailing direction of public organizations tends to favour technological innovation (Ojo 2019).

AI is already involved in increasingly sophisticated public sector functions, and has begun to embed itself deeply into government operations (Engstrom and Ho 2020). Most of the literature gives the government of tomorrow the role of “regulatory actor” or at best “facilitator” (Kuziemski & Misuraca 2020, p.3), inferring a future where we become entirely reliant on AI (Valle-Cruz 2020).

This age of algorithm could constitute the next phase of DEG, or a new era of governance altogether. However, these arguments do err slightly on the side of technological determinism as there remain many obstacles in the way of widespread AI implementation, not least of which are the ethical concerns that will be reviewed in the penultimate section of this paper. Accordingly public distrust of algorithmic systems is still high, but Spatola and Macdorman note that attitudes about new technologies become more favourable as they become more common (Spatola and Macdorman 2021). The effects of Weber's “routinization,” through their pervasive use by the private sector, could eventually normalize AI's presence and influence across society at large.

## Section Summary

From Simon's original algorithmic system to today, the idea of machines governing humans has grown from a faint possibility to a budding reality. The key arguments from this section are the following: AI have numerous functions that are useful to the public sector, one of which is their capacity to automate decision making. In this case, they are called ADM systems, and their appeal stems from their ability to condense massive volumes of data into actionable fragments of knowledge far more effectively than humans are able to. Output data can be reviewed and accurate recommendations produced quickly and automatically, albeit with considerable risks to transparency and accountability.

AI and humans have different abilities and limits, and a likely future is one in which a mainstay of public administration research will be the relationship between AI systems and human bureaucrats.

## Theoretical Model

A public sector organization is a complex information processing system for making decisions. This section will lay the groundwork for a public sector organizational model derived from Herbert Simon's theory of bounded rationality updated to today's standard of technology — the foundation of this section is built on Simon's *Administrative Behavior* (1947), *Models of Man* (1957), *The Sciences of the Artificial* (1969), and ensuing work on the subject.

Within the proposed framework, an organization can be thought of as a system that transforms input into output. Input can be broadly understood in terms of money (transmuted into work hours and resources), information, and nebulous social factors. The output in Simon's model is dynamic — while in the private sector an organization's objective is almost always to

maximize profit, in the public sector goals generally align with the ever-impalpable ‘public interest.’ Money fits neatly into an empirical framework, but public sector goals often have incommensurable units that can only be compared through the construction of indices. Output across varying government activities is consequently distorted by a layer of abstraction, as any attempt to evaluate an objective’s attainment must first pass through a weighing of values.

A deconstruction of the processes that occur between the input and the output stages, framed within the context of the AI technologies discussed previously, is the objective of the current section. We will begin with an anatomy of the ‘decision’ within an organizational context and build upwards to construct a framework capable of explaining the changes produced by a shift towards algorithmic governance.

## **Decisions**

A decision is a conclusion reached after deliberation, and it is this model’s fundamental unit. Everything from a junior bureaucrat's choice to press a key on their keyboard to a new government-wide mandate is considered a decision within this framework. Organizational processes are complex and involve many decisions about what should be spent where, how much attention should be given to what issues, what goals should be prioritized over others, and so forth, but all decisions have two common dimensions.

Firstly, a decision made by a rational actor is composed of factual and value elements. Factual elements are positive, verifiable claims about the world and how it operates. The return delta on a change in tax policy, for example, can be calculated and proven. Value elements are the basis for evaluating competing objectives that use different units, where a value judgement is necessary in order to adapt a decision to fit with an empirical model. Electing to fund project X

over project Y because the purpose of the former aligns more closely with organizational objectives would be an example of a value-driven decision. In the case of government, organizational values must be defined concretely so that the relative merits of a decision can be assessed alongside factual elements. Most decisions involve both factual and value elements, although different contexts will necessarily demand greater or lesser amounts of fact and value.

The second dimension worthy of note is that every decision occupies a distinct rank within the broader means-ends hierarchy of an organization. For Simon, the only difference between a purpose and a process is degree; the purposes of those working at the lower strata of an organization become part of the processes of those at higher levels. Broadly speaking, value-laden decisions affecting the grand direction of an organization are made near the top of a hierarchy, while fact-based decisions are made in the lower strata. In today's organizations, ADM systems are much more prominent in low-level roles performing tasks such as interacting with customers via chatbots, screening programs for candidate eligibility, and deducing facts and correlations from big data analysis, but they are slowly beginning to ascend the hierarchy (Treasury Board Secretariat 2018, Larsson 2021, Redden 2018).

## **Limits**

Factual elements in a decision were originally explained by Simon to be limited by the extent of an actor's knowledge and information, described as "the mass of knowledge that human minds can accumulate and apply" and constrained by "how rapidly knowledge can be assimilated" (Simon 1947, p.46). In one of his later articles, Simon expanded on this limitation in the context of computational capacity — the limits to factual evaluation by 1973 were the technical capacity of data processing and storage systems, the scarcity of human attention

vis-à-vis overwhelming volumes of information, and *crucially* the capacity to discern the relevancy of data to any given decision (Simon 1973). However, as evidenced by the current state of AI, the potential of computational power and ability has grown to surpass these limitations over the last half-century.

Simon's later work considered the mind and computer to be comparable information processing systems, but Dick (2015) remarks that the modern AI system is incisive in ways that far exceed the limits of human capabilities. The two have become entirely different subjects of study with different rules. A sophisticated ADM system can easily overcome all three of Simon's 1973 limits, but only at a level of abstraction representing formal, rather than material, parameters about the world. This creates what amounts to a new limit to rational decision making unique to computational systems: The extent to which the complexity of the world can be accurately translated into data.

Value elements in a decision are limited by the purposes engendered to them by personal and organizational goals. Value has a vertical component — the organization as a whole will have priorities that are parcelled down the chain-of-command by way of intermediary goals. While an organization's mandate might prioritize ideals like equity or accountability, at the lower levels intermediary values manifest themselves as diversity initiatives, quotas, or transparent reporting procedures.

Automated systems do not possess an innate sense of value. At the current standard of technology everything from the foundationally categorical nature of language to the implicit bias in datasets can shift the final product of an algorithmic system (Hildebrandt 2021). Any perceived value-bias in a system is necessarily caused, at some earlier point, by human intervention, but this does not mean that ADM systems are necessarily value-neutral in terms of

their influence. As mentioned earlier, their mere presence exerts pressure towards rational empiricism. Together, these limits form the outer bounds of options available for decisions made within an organization.

### **AI-Human Dimorphism**

At this stage of analysis we will take a moment to consider the primary actors in our model: human beings and ADM systems. The human mind and the algorithmic system are fundamentally different in how they operate, which severely limits any scholarship that treats them as interchangeable or that carelessly subordinates one to the other.

Humans possess a sense of value acquired cumulatively over a lifetime of experience, with various abilities and information of varying relevance to any task. ADM systems, on the other hand, tend to be highly specialized and are indisputably better at processing and making sense of large volumes of information.

The AI-as-tool analogy is useful in certain contexts, but here it runs the risk of mischaracterizing technology as nothing more than a ‘thing-to-be-used,’ ignoring technology’s fluid nature and tendency to reciprocally affect its user. As Haque remarks, technology is not only a tool “but an organized enframing of the human mind” (Haque 2015, p.23). Just as the blade itself incites to deeds of violence, technology makes up the processes of government that in turn determine how government acts.

One of the pillars of this analysis is that technology and humans must be considered subjects of equal interest interacting with each other within organizational structure. In this model humans are unseated as the ‘assumed worker’ — as the atom of the government

workforce. Weighing the relative merits of humans and computer systems as equal participants in a system is the final piece of this model.

## **Humans**

Humans are the original nodes of organization. Possessing a diversity of experiences and skills, humans are uniquely capable of addressing issues that lie outside the boundaries of rational empiricism. However, humans have woefully finite attention spans and limited processing power. When confronted with the sea of Big Data, the human brain is rendered practically obsolete. Humans are also known to be somewhat parochial in their decision making, as their options tend to be determined by their recent experiences and immediate context.

The human is still essential to the process of governing. Simon described the human agent in the technological age as “more and more an observer, moderator, maintenance man, and repairman for a nearly autonomous process” (Simon 1973, p.269), where the human would oversee value judgements using the ends produced by algorithmic systems. The public organization’s purpose, as noted at the beginning of this section, is to process information into meaning — meaning that still exists solely in relation to human subjectivity.

In summary, humans are better suited to make value judgments but have reached their upper limit as information-processing agents. However, determining the higher purposes of a public organization is still, and for the foreseeable future, the exclusive purview of human agents. This arrangement is perfectly sensible given that the public sector is ultimately beholden to people, and not to machines.

## ADM Systems

ADM systems are a more recent addition to organizational structure, having been in use in governments since the 1990s for simple language-processing tasks (Mehr 2017). Perhaps the greatest benefit of AI is that it “listens more than it speaks” (Anastopoulos 2018, p.1), simultaneously addressing Simon’s concerns about humans’ finite attention span and finite processing power — and addressed specifically by AI’s capacity to quickly discern the relevance of any data to a given decision, extending the limits of knowledge and solving problems exceeding human cognitive limits. Machines are infinitely better at making sense of the lower levels of the information pyramid.

ADM systems are, however, not wholly independent. An algorithm must be calibrated in order to produce an output, and the choice of design is ultimately just as important as the choice to use an ADMS in the first place. A supervised machine learning algorithm, for instance, is a predictive model based on training data (Anastopoulos 2018). If that data is skewed in any way this bias will manifest itself in the output, which can perpetuate pre-existing societal biases. By way of illustration, if training data for a recidivism-assessing algorithm contains historical decisions that were prejudiced against a certain group, future results will continue to target that group even if the model contains no parameters explicitly screening for them (Angwin et al. 2016).

This is an example of an ADM system exhibiting what amounts to unconscious bias — however the AI’s bias was present from the outset, imbibed from training data. Perilously, that data and resulting bias would not be easily visible within the architecture of the algorithm itself, but only through its output. The explainability problem is to blame here. Without a clear way to measure the bias in a system (because that bias comes from data so complex that only AI can

understand) ADM systems have free rein to make what are, effectively, their own value judgments.

Countermeasures exist to limit AI from inadvertently making value-judgements. In some arrangements a human is involved, like human/society-in-the-loop or evidence-informed (rather than evidence-based) policymaking (Wirtz 2020). In these cases, having a human deeply involved with the algorithm's operation “provides trust that somebody would bare [sic] the consequence of mistakes, and thus have incentive to minimize those mistakes” (Rahwan 2018, p.7).

Head has stressed the importance of ‘rigour’ in decision making with mind to the relative position of ADM system output in the evidence hierarchy (Head 2013, p.17), and Engstrom and Ho have called for an “internal law of administration” to guarantee accountability (Engstrom & Ho 2020, p.14) but ultimately bias is a necessary precondition for how an algorithm functions (Hildebrandt 2021) and therefore mitigation may be the best approach.

An AI system can only optimize for criteria that it is aware of, so determining what constitutes viable evidence is the key to a successful ADM system. This mode of algorithmic governance (for better or worse) does not create omniscient truth-telling machines or remove the human element of decision making, but merely obfuscates human value judgements and then magnifies them.

This is why an essential part of the algorithmic governance puzzle is the choice of parameters, the review process for training data, and other mitigation and accountability measures. When value-indices are employed as criteria in lieu of the values themselves, then desired ends risk being sacrificed by systems who maximize for an abstraction of their true goal.

## **Efficiency and Decision**

In this input/output decision-centered model, an organization takes information, money, and social factors as input and transforms them into decisions. An organization has a finite amount of money and work hours to distribute, and depending on the quality of decisions made, these resources can be distributed more or less effectively in pursuit of the organization's goals.

Simon describes efficiency as “the ratio between input and output” and explains that it must be a guiding criterion in administrative decision making (Simon 1947). In an ideal universe, if all decisions were made to optimize for factual and value elements a government could produce the most output with the least input. For Simon, good administration is behaviour that is realistically adapted to meet its goals, and the more adapted the behaviour, the better the output. Optimizing for efficiency with ADM systems is simpler for decisions that have more fact than value, which prompts a reevaluation of the role of the human in government.

If factual components of government decision making can increasingly be delegated to ADM systems then the role of the human will naturally skew towards value judgements and the overall efficiency of the system will increase. However, factual and value elements do not exist in isolation, and a human will make better value judgments if they are more knowledgeable about the factual processes that underlie their decisions.

Value and factual components of a decision must ultimately be considered together. An ideal human-AI governing machine is one in which there is harmony, but a balanced relationship is not something that will emerge unprompted.

## Section Summary

A public sector organization can be understood as a machine that makes decisions. A decision is a choice made between available options after considering factual and value elements. Within a hierarchical organization, decisions made at lower levels tend to involve more fact while those at higher levels tend to involve more value.

Rational actors who make decisions generally have their options limited by various factors. Previously, humans making decisions were limited by their ability to make sense of complex information. Now, computers who make decisions are limited by the extent to which abstracted data is faithful to reality, and by their inability to fully comprehend human values.

Generally speaking, humans are better at making value judgements and computers are superior when it comes to making factual judgements, but the two can make decisions together to arrive at an acceptable compromise. Under algorithmic governance, this arrangement may cause organizations to fundamentally rethink what a ‘good decision’ is. The balance of power between human and machine will be fundamental to decision making under algorithmic governance.

## Implications

For the sake of clarity, we will now run through the arguments made thus far before building on them to discuss their consequences:

AI has come a long way in the last half-century, having emerged originally as an attempt to create sentient machines. Contemporary AI are advanced mathematical models, combining the heuristic “satisficing” approach pioneered by Simon in the 1950s and the learning potential of ML to create algorithmic systems that are highly proficient at particular tasks. AI systems have begun to overtake humans in terms of performance, previously solely for logic-based tasks but

increasingly for tasks requiring adaptive reasoning like recognizing language and interpreting data.

Digital-era governance has provided governments with massive quantities of data on a variety of subjects that has been very difficult for humans to put to proper use. AI has opened the door to automate the interpretation of this data through the use of AI-enabled ADM systems. These systems can sift through large quantities of data to make recommendations, but require some level of human involvement to provide good training data and to delineate how data should be interpreted. The diffusion of such systems is pushing us closer to an age of ‘algorithmic governance’ which will fundamentally redefine the rules of organizational structure. Technology generally begins as a tool before ultimately spurring organizational transformation.

To make sense of this process, we draw on the work of Herbert Simon to create a new model for a new age. In this framework, the government is a decision-making machine where every decision involves factual and value elements. Factual elements are generally better left to ADM systems while value judgements are better suited to humans, but most decisions will require some degree of both. Decisions also change depending on where they are made within a hierarchy — lower-level decisions involve more fact while higher-level decisions have more value.

Understanding how AI works is necessary to understand how AI-informed decisions will be made. The management of governing organizations will need to have some understanding of how advanced systems operate in order to properly direct the future of the public sector, because much of the information made available to them under algorithmic governance could be provided by processes involving AI systems. ADM systems will also gradually become more prevalent at higher strata of organizations as technology improvements expand the types of work

that they can perform on par with humans. However, public sector technostructures were liquidated decades ago under NPM, so there may be rough waters ahead.

AI-based systems have the potential to boost the efficiency of an organization, but there is cause for considerable concern in the realm of ethics. The remainder of this section will discuss the implications of this analysis.

### **Organizational Transformation**

The first thread that this section will deal with is the notion of organizational transformation, described by Roy as a process heralded by automation and innovation (Roy 2008). Right now, mainstream public administration literature is still dealing with the “technology adoption” narrative which, for Ojo, comes from an earlier era that studied explainable systems that existed in isolated contexts. As we approach algorithmic governance, these theories will fade in relevance as the need to replace them with an updated paradigm emerges (Ojo 2019, p.169).

The theoretical model from the previous section shows that the best way to understand algorithmic governance is by imagining ADM systems and humans as two different types of workers who interact within a system where they are both useful. Unlike the original conception of artificial intelligence from the 1950s, recent research has demonstrated that humans and AI are fundamentally different creatures (Dick 2015). AI tends to be better at repetitive empirical tasks, and humans at work involving intimate knowledge of human values and needs.

An organization’s values are fed down the chain of command. In a democratic society, they are generated by elected officials and, by the time they arrive at the lower levels of hierarchy, they have taken on the form of discrete objectives such as mandates or quotas. When

decisions are made at these lower levels, they are made by workers who satisfice — who select the best available option based on a combination of factual and value elements. Under any form of algorithmic governance, it is reasonable to assume that ADM systems will contribute to these decisions, but only up to the point that abstract values can be reified into data. AI's influence will work its way back up the hierarchy until it eventually begins to redefine the entire organization through processes of routinization and reinvention.

The methods used by an organization have a self-referential effect on that organization's overall values. Means are not value-neutral and changing what qualifies as a 'good' decision within an organization will change what that organization values, as well as how it carries out its affairs. The most important variable in this process is the degree of allowed ADM involvement. It is natural to think that if public sector use of AI systems is banned outright then that influence will be zero, while if their use is encouraged influence will increase proportionally.

While satisficing, a better decision can always be made as long as there are still limits to decision-making. The question is then "how can we arrive at that better decision?" and the answer, given our previous analysis, is to "push the limits to their extremes in every situation." In the case of algorithmic governance, this is done by finding the optimal balance between human judgement and ADM system analysis where both provide maximum value for any given task. It is this 'golden mean,' dependent on innumerable social and technical factors, that will produce the best available decision under algorithmic governance.

The breakpoints of this human-AI partition of responsibility and its golden mean will change as time goes by. In the era where Simon wrote his later work, limits in computation technology made it so that almost all tasks landed squarely on the 'humans should do it' side,

while today the process of determining the ideal balance between robot and human is considerably more complex.

Where once computers were seen as merely a more efficient means to store and retrieve information, they are now on the verge of eclipsing human usefulness in many tasks involving processing information and recommending action. It is not inconceivable that, in the future, AI may inch closer to areas of judgement that modern theorists consider sacred to humanity.

A sound plan to balance human and AI elements within a transforming public sector organization may result in numerous benefits, while a poorly conceived strategy could result in discord, efficiency losses, or ethical strife.

### **The Technical Knowledge Gap**

The second implication of this analysis is that there are dangers to having poor AI literacy inform public sector AI strategy. In the realm of value judgements, AI systems are best thought of as ethical catalysts: an automated system may not have any predispositions or bias in a vacuum, but implementing a poorly conceived system may compound pre-existing injustices at previously inconceivable scales (Hildebrandt 2021). Although AI systems are now widely in use by businesses around the world, “the private sector has not stepped up to lead regarding” ethical concerns, leading Fatima et al. to conclude that the public sector should take charge with their own approaches (Fatima et al. 2020, p.147).

The realm of ethics extends beyond human elements in organizations. While the potential technically exists to control for all instances of bias within a decision-making process by modifying its issue-specific ADM system, the contravening risks are equally imposing. There are

also caveats regarding bias and transparency, like the emergence of an upper limit of explainability for new unsupervised ML systems.

The explainability problem is the following: There is a tradeoff between explainability and efficiency in algorithmic systems. As alluded to earlier, there is an inverse relationship between how well an ADM system works and how easily it can be explained to someone unfamiliar with AI systems, while the most effective methods — neural networks, classification trees, and vector machines — can be complete black boxes even to professional ML programmers (Yampolskiy 2019). In these cases, simplifying their actual workings is not just difficult, but impossible. AI that reason beyond human limits cannot explain their reasoning within human limits. It would be like a human trying to explain calculus to a chimpanzee.<sup>2</sup>

If factual elements are instead entrusted to AI experts, something must be done to guarantee that these experts will conform to value judgments that have been democratically formulated. In order for public sector management to ensure that AI-based decisions are made with ethical standards in mind, they must either have personal knowledge of how AI systems operate (as well as their ethical limits) or have some alternative way of guaranteeing that public sector value judgements are respected, otherwise a dangerous amount of power accumulates on the side of technology.

This is where the ‘faith in technological rationality’ concern becomes a real problem. Doing things without knowing why is very dangerous for the public sector. Newell remarks that it was this kind of blind faith in AI recommendations that was partially responsible for the 2007-08 financial crisis (Newell 2015). Financial institutions and senior managers trusted what their algorithmic systems told them without knowing why, causing them to take on risk without

knowing it. For unexplainable AI, this myopia removes the ability to question why an ADMS-recommended decision was even made.

The mystification of AI mentioned in the introduction has also drummed up fear that human agency is at risk of being removed entirely. This is a misconception. Volition has not been removed or fully-mechanized but transformed into a language that requires a significant amount of digital literacy to understand (Yampolskiy 2019). This is where the usefulness of mitigation measures from the ADM section of the model — the human/society-in-the-loop model, the codified ‘internal law’ dictating how accountability translates through AI systems, and the assignment of a deliberate weight to AI evidence based on its level of transparency — shine through.

The human/society-in-the-loop model delegates ultimate decisional authority to a “human operator who thus controls the actions of AI to achieve the common goals of the stakeholders involved” (Wirtz 2020, p.819). Under ‘internal law,’ public servants “proactively design and then maintain systems that are more transparent and auditable” as a matter of principle (Engstrom 2020, p.14). Head argues that a more critical look at the value of different types of evidence is necessary in order to “demystify the authority of social statistics.” Embedding a warning into all ADMS-generated evidence reminding policymakers of its origin and inherent dangers would shift the ethical onus onto whoever is responsible for the deliberation process (Head 2013, p. 401).

All of these methods prescribe roles for human actors in the ADMS process, thereby ensuring accountability. If an AI system fails, responsibility falls on those with decisional authority in the Wirtz model, on those who designed the ADM system in the Engstrom model, and on those who overvalued AI evidence in the Head-based model. These measures ensure

ethical compliance by using humans to limit AI influence, even in the cases of unexplainable algorithms.

Ironically, this is once again a manifestation of bounded rationality, but in this case the limits exist only for value judgements and apply solely to ADM systems. In a situation where an unexplainable algorithm arrives at a particular solution X to a policy problem, a policy analyst may not be able to fully assess the value implications of that solution or ensure accountability. They could instead go with their own solution that would almost certainly be less effective than solution X, but would be much more likely to align with their desired values. Equally, they could apply a mitigation measure to arrive at a middle ground where accountability is preserved, but at some cost to overall efficiency.

What we have here is ostensibly a catch-22 — optimizing for facts sacrifices value and possibly accountability, but prioritizing for value loses out on potential efficiency gains. There may, however, be an enforceable maximum degree of allowed risk for public sector organizations if good practices are put in place to limit the powers of ADM systems.

Determining what can be fed into an ADM system and how to review what that system outputs is also important for controlling the dangers of AI. There are two conclusions that we can draw from our prior analysis to inform potential solutions. Firstly, when an algorithm is designed, parameters are set and training data is imbibed. Veale remarks that, right now, elected officials and senior public servants have no input in this process, leading to ambiguity in how their values are translated into machine-informed decisions (Veale 2019). The technical knowledge-policy knowledge gap is just too vast to accurately transmute prescribed values into technical choices about what training data should be selected or what weight should be ascribed to certain values.

However, if the public sector's technocracy, versed in both the workings of policy and technology, were rebuilt it may be capable of bridging the gap between the humans dictating organizational values and the machines proficient in determining fact — in other words, by making technocrats ambassadors between the understandable limits of AI and the value judgements of human beings, it may be possible to reliably translate value into ADM system design. This, of course, would require rebuilding in-house technical expertise or finding some way to fill that need through the private sector.

Secondly, accountability can persist through an algorithmic system even though, at the moment, it may be difficult to track. This is because accountability takes on a whole new dimension under algorithmic governance. There have been historical cases where parties guilty of ethical offences attempt to 'hide behind the AI veil' to avoid responsibility, be it the earlier example of AI-assessed recidivism bias, facial-recognition systems 'autonomously' breaching privacy, or AI policing systems overtargeting low-income neighbourhoods; but again, there were deliberate decisions made by humans in these processes to select training data and to use these systems without a thorough review process (Hildebrandt 2021; Nzobonimpa 2022). Well thought-out mitigation measures can keep accountability alive under algorithmic governance.

Choosing to design an algorithm a certain way is as much a policy decision as choosing to set a budget at a particular amount, or designing a trade policy to stem the flow of a particular good, or prioritizing one issue over another. The only difference is that today's algorithmic system designers are not necessarily beholden to any ethical standard, as their work is not explainable to mainstream public servants.

A comprehensive and dependable way of enforcing AI best practices could significantly reduce the risk of rogue AI systems. However, creating those best practices requires knowledge

of what makes an ‘ethical AI system,’ which necessitates some technical knowledge. This is where we return to the earlier notion of an interdisciplinary workforce with an equal interest and understanding of both human and AI components of organization. This is the best way to optimize for efficiency while remaining true to organizational values.

As Simon explains: “it is the guarantee of due process, not the guarantee of infallibility, that democratic institutions require” (Simon 1973, p.276). A method to reliably evaluate whether AI decisions have been made responsibly is likely the best solution to bridge the AI technical knowledge gap.

### **The Efficiency/Ethics Compromise**

The third and final implication of this analysis is that future public sector organizations who intend to employ ADM systems will need to clearly lay out their desired efficiency improvements and the concordant maximum level of ethical risk they are willing to take on. It is best if these expectations are clearly established before widespread AI implementation becomes a major issue. If not, structures hastily put in place by those with low technical knowledge risk permanently defining public sector attitudes as they are cemented through mechanisms like path dependence and institutional inertia.

Due to the unique nature of how ADM systems operate, optimizing for efficiency can cause ethical dilemmas, but adopting a low-risk strategy will have a high opportunity cost in terms of efficiency. I hesitate to call this a zero-sum game as we are dealing with value and not fact, but the same gain/loss principles apply. Some strategy or plan is necessary now to lay the foundation for what an acceptable balance between ethics and efficiency should look like for a

public sector organization under algorithmic governance. The government needs a well-researched approach to AI risk-management.

Within our theoretical model, efficiency is defined as the ratio of output from one option available to a decision maker over the output from alternatives (Simon 1947). Given that the public sector is ultimately beholden to the public, and getting the most return per tax dollar is a timeless public interest, efficiency should be a high priority value.

However, input and output in public sector organizations have different units. Money, work hours, and social factors go in and a mixed bag of projects, initiatives, funding, and regulations come out. Some of this output is more valuable to certain people or groups than others, and for some the efficiency gains may not be worth the risk of automated discrimination or any of the other complications that algorithmic governance entails. Likewise, the mere risk of workforce substitution and increased pressure towards a rational empirical worldview may be enough to dissuade many from accepting any risk of ethical peril. It does not take a cynic to realize that the world of automation is ripe for abuse.

Efficiency is, without a doubt, the most significant benefit to government under algorithmic governance. AI are unmatched in their accuracy, speed, and potential to relieve burden from overworked employees. There is also a finite amount of work hours available to the public sector system, and having AI take on laborious, repetitive, or mind-numbing tasks could free up time for humans to focus on more invigorating work.

The ethical dimension is harder to pin down, but within this analysis it largely corresponds to the quality of value judgements made within an organization. A public sector organization should operate according to its values, and good value judgements, under bounded rationality, are judgements that prioritize values desirable to the organization. ‘Good’ is a relative

term, but in the public sector a ‘good decision’ will align with the public interest. However, values exist in a relative space and typically have incommensurable units so it may, for instance, be better for a public sector organization to make a decision yielding a high return of less-desirable goal X over a negligible return of more-desirable goal Y, simply because a lot of X was judged a better way to meet that organization’s goals. For that reason, workers must be able to make decisions according to the relative degree that any given option will further their organization’s goals based on their understanding of the relationship between real output and value output.

AI throws a wrench into this already-complicated equation. As explained previously, value is generally parcelled down the chain of command in reified forms such as quotas or mandates. Humans have lived experience in social environments and are typically capable of intuiting the true intent behind an intermediate value. A human, for example, would be able to tell that an inclusive hiring policy was made for the sake of workplace equity, but an ADM system will naturally try to game the objective, leading to any number of unexpected outcomes depending on how the value, ‘equity,’ was parameterized and subsequently construed by the AI. When value-indices are employed as criteria in lieu of values themselves, ends deemed valuable by an organization are at risk of being sacrificed for ends that are more tangible to an ADMS — the system will naturally prioritize form over substance. This risks ushering in a dangerous transformation in the way that policy is formed and framed.

Within a bounded rationality framework, the quality of an AI’s decision (vis-a-vis fact and value) also directly reflects the quality of information used to justify that decision. The data used to train an algorithm is central to determining that algorithm’s future biases, but it is very difficult to discern from looking at the data what ethical risks it may cause. On the micro-level,

there is the important dimension of high variability across specific-case implementation (Veale 2019): just because something works somewhere once, there is no way of telling if it will work everywhere all the time. Howlett asserts that “policies can succeed or fail in numerous ways” (Howlett 2009), and merely having a macro-level factual basis for a decision is only a small piece of a greater picture.

ADM systems have only one way of understanding the world — the way in which they have been told to differentiate data. Bias is often cited as a reason to be wary of AI, but bias is present in all algorithmic systems. In fact, bias is necessary for all learning algorithms to function — allowing them to adapt their internal heuristic by changing the relative importance assigned to different data values or parameters. Hildebrandt remarks that bias is “a generative and inevitable precondition” for AI, and notes that ultimately it is a human operator who determines how an algorithm should recognize distinctions between data, and what methods it should prefer over others to do so (Hildebrandt 2021, p.47). Designing an algorithm to be biased a certain way requires a degree of technical skill and an understanding of how algorithmic systems work which, as we have seen previously, is beyond the reach of the average bureaucrat. Not all algorithmic bias is inherently harmful, but being able to tell the difference between necessary and harmful biases (and being able to correct a system accordingly) is necessary for ethical algorithmic governance.<sup>3</sup>

We have established that increased integration of AI will eventually put ethics and efficiency at odds. This framework has classified workers in the public sector as being either ‘human’ or ‘ADM system’ and has broken down decision into ‘fact’ and ‘value.’ Finally, we have ‘ethics’ and ‘efficiency’ — competing public sector macro-values with different benefits.

Today, people do most public sector work because they are readily available for hire, structures exist to facilitate their roles and responsibilities, and humans have been organizing society at a satisfactory level for quite some time. However, AI systems have arrived in the mainstream, and they are better at making decisions that rely more on fact than value than their human counterparts.

To ensure optimal government performance, the person/ADMS division of labour should occur at the equilibrium point in the human-ADMS spectrum where a given type of work benefits maximally from both human and AI involvement. This point will vary depending on whether the task at hand requires more insight into fact or value. Simon observed in the 70s that “[t]he key to the successful design of information systems lies in matching the technology to the limits of attentional [or human] resources” (Simon 1973, p.274) and, to an extent, that is the challenge of the ethics/efficiency compromise — to match work to those most suitable to do it.

However, this precipitates what Kuziemski and Misuraca call a “a tragic double bind” where the temptation to boost the public sector’s efficiency can also make it more difficult to control the dangers of public sector algorithms. The dilemma is “to govern algorithms, while governing by algorithms” (Kuziemski & Misuraca 2020, p.1). While there is no ‘correct’ answer to this problem, it is in the public sector’s best interests to be proactive when dealing with AI as a measured approach is more likely to succeed in the long-term. Rahwan suggests using AI to evaluate AI (Rahwan 2018), and Di Vaio et al. bring up the ambidexterity theory of innovation, which encourages managing today’s challenges while simultaneously increasing analysts’ capacity to prepare for tomorrow’s AI (Di Vaio et al. 2021).

At some point a compromise is necessary between efficiency and ethics. It will also become necessary to clearly draw the line between ‘human work’ and ‘AI work,’ and to be

prepared to adjust it or maintain it as AI technology improves. This struggle boils down to a series of value judgements that all require a substantial understanding of the relative risks and usefulness of AI.

The question at the center of the ethics/efficiency issue is neither the Fordistic “how can we maximize output?” nor the utopian “how can we avoid all risk?” but rather a more moderate “how can we design a better organization that can more readily meet our needs?”

### **Section Summary**

Applying AI to public sector decision making is not a simple task. This section has built on the premises and model established previously to draw out three related implications that algorithmic governance poses for the future of the public sector.

Increasing algorithmic influence in government will eventually lead to organizational transformation. This process is best understood as moving from a human-centered understanding of government operations to a human-machine hybrid vision. By shifting the center of gravity away from humans and towards a more balanced ‘social machine’ approach, it becomes easier to understand the effects of increased algorithmic integration and to weigh the shift in values that this transformation will entail.

The technical knowledge gap is a problem that arises from a lack of in-house technical expertise coupled with path dependence and institutional inertia. A lack of digital literacy in the higher strata of government makes it difficult to maneuver effectively in a new technological context, which increases the risk of AI debacles and poor decision-making for matters concerning new technology.

Finally, the efficiency/ethics compromise is a macrocosm of the AI-in-government issue. As most decisions have both fact and value elements, the question of whether people or ADM systems should be responsible for decision making will determine whether more weight is given to fact or value. Our model states that the government is a decision-making machine; however, between the input and output stages are a complex series of processes and relationships that are all worthy of careful consideration for their effects on the ADM system/human split on decision-making responsibility.

## Conclusion

In this paper I have attempted to create a model to further public administration understanding of AI, and to forecast what a future under algorithmic governance may entail for public sector decision making. This paper has addressed the question “how will organizational structure adapt to changes brought on by AI, and what effect will these changes have on the guiding principles of public organizations?” The answer has been complex and multifaceted.

Under algorithmic governance, organizational structures could undergo widespread transformation and see a shift in their constellation of guiding ideals. As AI is increasingly integrated into government, AI systems may take on a larger role in determining the future of the public sector, troubling conventional notions of what a ‘good decision’ means as well as having profound effects on traditional notions of efficiency and ethics.

The field of AI began at the convergence of computer science and organizational theory, and it is probable that, going forward, governments will require not just expertise in both fields but expertise in how the two fields work together. In this new age, strategies put in place by those

with low technical knowledge and older, ‘technology adoption’ methods in mind risk forming permanent fixtures that limit potential future benefits.

Conversely, while AI has many tangible benefits, there are also appreciable dangers. Accountability is a major issue. Deep-learning algorithms tend to be black boxes where a given output cannot be justified or explained by its ADMS. In these cases, implementing a policy without an explainable justification beyond ‘the algorithm said so’ should be a serious cause for concern. The real world is also infinitely more complicated than even the most advanced computer, and AI technology is by no means a silver bullet to any problem, let alone all of them. In light of these concerns, it would be best if a thorough plan were designed by those with both policy knowledge and cutting-edge technological understanding in anticipation of increased AI presence in government.

There is the further caveat that we often do not truly appreciate the downsides of a technology until we actually start using it. AI systems are a potential solution to upcoming labour shortages, but also present the risk of job losses that often occur when substituting capital for labour. AI winters have also happened in the past, and there is no guarantee that this most recent wave of popularity will meet all the expectations that have been set out for it.

There are other concerns: the colossal carbon emissions that come with running massive databases and computer systems, the difficulty of implementing AI systems across a heterogeneous public sector, and the million-dollar-question of how the government should handle private sector AI regulation. Moreover, this analysis is founded squarely in the rational-empirical framework meaning that it is constrained by all the limitations of such a practice of thought.

The public sector is approaching a crossroads. When Keynes suggested a new paradigm for macroeconomics in 1936, he wrote that “[t]he difficulty lies, not in the new ideas, but in escaping from the old ones, which ramify, for those brought up as most of us have been, into every corner of our minds” and, to an extent, that is the challenge facing the public sector today (Keynes 1936). In my estimation human exceptionalism, path-dependence in institutions, and techno-apprehension have led to a widespread slow uptake of technologies in government that have been hugely beneficial in the private sector. However, the future of what algorithmic governance will (or will not) look like is still to be decided.<sup>4</sup>

## Notes

[1] A neural network is an adaptable model that uses nodes called ‘neurons’ to map input/output relationships between variables or to identify patterns in a dataset. A classification tree is a predictive model that sorts data from a ‘root’ based on an adaptable heuristic, and refines it further through its ‘branches’ and ‘leaves.’ The ‘random forest’ is one popular submethod of the decision tree model. A support vector machine is a method for classification and regression analysis that uses geometric planes in a lower dimension to the dataset, called ‘hyperplanes’ to split data into groups. For further reading see *Machine Learning: A Concise Introduction* (2018) by Steven Knox.

[2] The comprehensibility problem is a subset of the explainability problem (although I use ‘explainability’ for both cases in this paper for the sake of simplicity). It describes AI systems that are not just unexplainable, but that have also exceeded the absolute upper bounds of human comprehension. Herbert Simon, in *Sciences of the Artificial*, describes the process of empirical discovery by quoting the Dutch physicist Simon Stevin — *wonder, en es gheen wonder*, or, “wonderful, but not incomprehensible” (Simon 1969, p.1). However, nowadays, some discoveries made by AI systems are well and truly incomprehensible (Yampolskiy 2019).

[3] A final point on ethics is that most of the problems caused by AI are just exaggerated versions of problems that already exist out in the world. As much as I have argued that AI is transformative in this paper, for ethical quandaries and systemic injustice it begins to feel a little like “plus ça change, plus c’est le même” syndrome. Systemic problems don’t appear out of nowhere — rather, they evolve to fit their medium. Whether we operate on the deterministic

assumption that algorithmic governance will happen and prepare to face the new forms of injustice that will arise or try to stop AI in its tracks now, the government will continue to face the same underlying problems that have persisted for centuries albeit in a new, more threatening form.

[4] Out of personal interest, I decided to let an AI chatbot respond to my research question (after I had finished writing the paper, of course). This is the response that OpenAI's ChatGPT gave me:

Artificial intelligence (AI) has the potential to bring about significant changes to organizational structures, particularly in terms of how work is carried out and how decisions are made. These changes can have significant impacts on the guiding principles of public organizations.

One way in which AI may change organizational structure is through the automation of certain tasks or processes. This could lead to a shift in the roles and responsibilities of employees, as some tasks may no longer be necessary or may be performed by AI systems. In such cases, it may be necessary to reevaluate the roles and responsibilities of employees to ensure that they are aligned with the needs of the organization.

Another potential impact of AI on organizational structure is the increased use of data analytics and decision-making tools. These tools may be used to analyze large amounts of data and make recommendations or decisions based on that analysis. This could lead to a shift in the way that decisions are made within the organization, with more emphasis being placed on data-driven approaches.

In terms of the guiding principles of public organizations, the use of AI may require a reevaluation of the values and goals of the organization. For example, if AI is used to automate certain processes, the organization may need to consider the implications for job security and the potential for AI to disrupt existing workflows. Additionally, the use of AI for decision-making may raise questions about accountability and transparency, as the decision-making process may become less visible to employees and external stakeholders.

Overall, the adoption of AI is likely to bring about significant changes to organizational structures and the guiding principles of public organizations. It will be important for organizations to carefully consider the potential impacts of these changes and to adapt their structures and principles as needed to ensure that they continue to meet the needs of their stakeholders. (Open AI)

Not bad.

## References

- Alexopoulos, Charalampos, et al. “How Machine Learning Is Changing e-Government.” Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance, ACM, 2019, pp. 354–63, <https://doi.org/10.1145/3326365.3326412>.
- Anastasopoulos, L. Jason, and Andrew B. Whitford. “Machine Learning for Public Administration Research, With Application to Organizational Reputation.” Journal of Public Administration Research and Theory, vol. 29, no. 3, 2019, pp. 491–510, <https://doi.org/10.1093/jopart/muy060>.
- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias: There’s software used across the country to predict future criminals. and it’s biased against blacks. ProPublica, May 23. Retrieved October 30th, 2022. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Ansolabehere, Stephen, and Hersh, Eitan D. “ADGN: An Algorithm for Record Linkage Using Address, Date of Birth, Gender, and Name.” Statistics and Public Policy, vol. 4, no. 1, 2017, pp. 1–10, <https://doi.org/10.1080/2330443X.2017.1389620>.
- Araujo, T., et al. “In AI We Trust? Perceptions About Automated Decision-Making by Artificial Intelligence.” AI & Society, vol. 35, no. 3, 2020, pp. 611–23, <https://doi.org/10.1007/s00146-019-00931-w>.
- Ben Rjab, Amal, and Sehl Mellouli. “Artificial Intelligence in Smart Cities: Systematic Literature Network Analysis.” Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance, ACM, 2019, pp. 259–69, <https://doi.org/10.1145/3326365.3326400>.

- Chen, Tao, et al. "AI Innovation for Advancing Public Service: The Case of China's First Administrative Approval Bureau." Proceedings of the 20th Annual International Conference on Digital Government Research, ACM, 2019, pp. 100–08, <https://doi.org/10.1145/3325112.3325243>.
- Clarke, Amanda, et al. "Understanding Governance in the Digital Era: An Agenda for Public Administration Research in Canada." Canadian Public Administration, vol. 60, no. 4, 2017, pp. 457–75, <https://doi.org/10.1111/capa.12246>.
- Crevier, Daniel. AI : the Tumultuous History of the Search for Artificial Intelligence. Basic Books, 2nd ed.,— [archive.org/details/aitumultuoushist00crev](https://archive.org/details/aitumultuoushist00crev).
- Dick, Stephanie. "Of Models and Machines: Implementing Bounded Rationality." Isis, vol. 106, no. 3, 2015, pp. 623–34, <https://doi.org/10.1086/683527>.
- Di Vaio, Assunta, et al. "Data Intelligence and Analytics: A Bibliometric Analysis of human–Artificial Intelligence in Public Sector Decision-Making Effectiveness." Technological Forecasting & Social Change, vol. 174, 2022, p. 121201–, <https://doi.org/10.1016/j.techfore.2021.121201>.
- Dunleavy, Patrick and others, 'The Theory of Modern Bureaucracy and the Neglected Role of IT', *Digital Era Governance: IT Corporations, the State, and e-Government* (Oxford, 2006; online edition, Oxford Academic, 3 Oct. 2011), <https://doi.org/10.1093/acprof:oso/9780199296194.003.0001>, accessed 29 Oct. 2022.
- Dwivedi, Yogesh K., et al. "Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy." International Journal of Information Management, vol. 57, 2021, p. 101994–, <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>.

- Engstrom, David Freeman and Ho, Daniel E., Artificially Intelligent Government: A Review and Agenda (March 9, 2020). Big Data Law (Roland Vogl, ed., 2020), Available at SSRN: <https://ssrn.com/abstract=3551549>
- European Commission, High-Level Expert Group On Artificial Intelligence. “Ethics Guidelines For Trustworthy AI” 2021 (Online). Available: [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=60419](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419)
- Fatima, Samar, et al. “National Strategic Artificial Intelligence Plans: A Multi-Dimensional Analysis.” *Economic Analysis and Policy*, vol. 67, 2020, pp. 178–94, <https://doi.org/10.1016/j.eap.2020.07.008>.
- Floridi, Luciano. “AI and Its New Winter: From Myths to Realities.” *Philosophy & Technology*, vol. 33, no. 1, 2020, pp. 1–3, <https://doi.org/10.1007/s13347-020-00396-6>.
- Gomes de Sousa, Wesley, et al. “How and Where Is Artificial Intelligence in the Public Sector Going? A Literature Review and Research Agenda.” *Government Information Quarterly*, vol. 36, no. 4, 2019, p. 101392–, <https://doi.org/10.1016/j.giq.2019.07.004>.
- Gugerty, Leo. “Newell and Simon’s Logic Theorist: Historical Background and Impact on Cognitive Modeling.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, 2006, pp. 880–84, <https://doi.org/10.1177/154193120605000904>.
- Hadden, Susan G. “The Future of Expert Systems in Government.” *Journal of Policy Analysis and Management*, vol. 8, no. 2, 1989, pp. 203–08, <https://doi.org/10.2307/3323379>.
- Haque, Akhlaque. *Surveillance, Transparency, and Democracy : Public Administration in the Information Age*. The University Alabama Press, 2015.

- Head, Brian W. "Evidence-Based Policymaking - Speaking Truth to Power?" *Australian Journal of Public Administration*, vol. 72, no. 4, 2013, pp. 397–403, <https://doi.org/10.1111/1467-8500.12037>.
- Hildebrandt, Mireille. "The Issue Of Bias: The Framing Powers Of Machine Learning" *Machines We Trust : Perspectives on Dependable AI*. Edited by Marcello Pelillo and Teresa Scantamburlo, The MIT Press, 2021.
- Höchtel, Johann, et al. "Big Data in the Policy Cycle: Policy Decision Making in the Digital Era." *Journal of Organizational Computing and Electronic Commerce*, vol. 26, no. 1-2, 2016, pp. 147–69, <https://doi.org/10.1080/10919392.2015.1125187>.
- Hodgetts, J. E. *Pioneer Public Service : an Administrative History of the United Canadas, 1841-1867*. University of Toronto Press, 1955.
- Howlett, Michael. "Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada." *Canadian Public Administration*, vol. 52, no. 2, 2009, pp. 153–75, [https://doi.org/10.1111/j.1754-7121.2009.00070\\_1.x](https://doi.org/10.1111/j.1754-7121.2009.00070_1.x).
- Innovation, Science and Economic Development Canada. *Pan-Canadian Artificial Intelligence Strategy 2022* (Online). Available: <https://ised-isde.canada.ca/site/ai-strategy/en>
- Janssen, Marijn, et al. "Data Governance: Organizing Data for Trustworthy Artificial Intelligence." *Government Information Quarterly*, vol. 37, no. 3, 2020, p. 101493–, <https://doi.org/10.1016/j.giq.2020.101493>.
- Kernaghan, Kenneth. "Digital Dilemmas: Values, Ethics and Information Technology." *Canadian Public Administration*, vol. 57, no. 2, 2014, pp. 295–317, <https://doi.org/10.1111/capa.12069>.

- Keynes, John Maynard. *The General Theory of Employment, Interest, and Money*. Springer International Publishing, 1936 (2018 ed.), <https://doi.org/10.1007/978-3-319-70344-2>.
- Khan, Ali, et al. "The Anatomy of 'Fake News': Studying False Messages as Digital Objects." *Journal of Information Technology*, vol. 37, no. 2, 2022, pp. 122–43, <https://doi.org/10.1177/02683962211037693>.
- Knox, Steven W. *Machine Learning : a Concise Introduction*. John Wiley & Sons, Inc., 2018.
- Kuziemski, Maciej, and Gianluca Misuraca. "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings." *Telecommunications Policy*, vol. 44, no. 6, 2020, pp. 101976–101976, <https://doi.org/10.1016/j.telpol.2020.101976>.
- Langley, Pat. "The Changing Science of Machine Learning." *Machine Learning*, vol. 82, no. 3, 2011, pp. 275–79, <https://doi.org/10.1007/s10994-011-5242-y>.
- Larsson, Karl Kristian. "Digitization or Equality: When Government Automation Covers Some, but Not All Citizens." *Government Information Quarterly*, vol. 38, no. 1, 2021, p. 101547–, <https://doi.org/10.1016/j.giq.2020.101547>.
- LeCun, Yann, et al. "Deep Learning." *Nature (London)*, vol. 521, no. 7553, 2015, pp. 436–44, <https://doi.org/10.1038/nature14539>.
- Longo, Justin. "When Artificial Intelligence Meets Real Public Administration." *Canadian Public Administration*, vol. 65, no. 2, 2022, p. 384–, <https://doi.org/10.1111/capa.1465>.
- McCorduck, Pamela. *Machines Who Think : a Personal Inquiry into the History and Prospects of Artificial Intelligence*. 2nd ed., A.K. Peters, 2004.

- McKelvey, Fenwick Robert, and Maggie Macdonald. “Artificial Intelligence Policy Innovations at the Canadian Federal Government.” *Canadian Journal of Communication*, vol. 44, no. 2, 2019, pp. PP43–PP50, <https://doi.org/10.22230/cjc.2019v44n2a3509>.
- Mehr, Hilla. “Artificial Intelligence for Citizen Services and Government.” Ash Center for Democratic Governance and Innovation, Aug. 2017. Available: [https://ash.harvard.edu/files/ash/files/artificial\\_intelligence\\_for\\_citizen\\_services.pdf](https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf)
- Mitchell, Tom M. (Tom Michael). *Machine Learning*. McGraw-Hill, 1997.
- Newell, Sue, and Marco Marabelli. “Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of ‘datification.’” *The Journal of Strategic Information Systems*, vol. 24, no. 1, 2015, pp. 3–14, <https://doi.org/10.1016/j.jsis.2015.02.001>.
- Nzobonimpa, S. (2022). Algorithmes et intelligence artificielle : une note sur l’état de la réglementation des technologies utilisant la reconnaissance faciale automatique au Canada et aux États-Unis. *Revue Gouvernance / Governance Review*, 19(2), 99–114. <https://doi.org/10.7202/1094078ar>
- Ojo, Adegboyega, et al. “A Realist Perspective on AI-Era Public Management.” *Proceedings of the 20th Annual International Conference on Digital Government Research*, ACM, 2019, pp. 159–70, <https://doi.org/10.1145/3325112.3325261>.
- Open AI. “ChatGPT.” *OpenAI*, <https://chat.openai.com/chat>.
- Pelau, Corina, et al. “What Makes an AI Device Human-Like? The Role of Interaction Quality, Empathy and Perceived Psychological Anthropomorphic Characteristics in the Acceptance of Artificial Intelligence in the Service Industry.” *Computers in Human Behavior*, vol. 122, 2021, p. 106855–, <https://doi.org/10.1016/j.chb.2021.106855>.

- Perez, Timothy J. *Everything I Know About AI Came from the Movies: How AI Science Fiction Narratives Drives Selection of News Frames*. ProQuest Dissertations Publishing, 2019.
- Rahwan, Iyad. "Society-in-the-Loop: Programming the Algorithmic Social Contract." *Ethics and Information Technology*, vol. 20, no. 1, 2018, pp. 5–14, <https://doi.org/10.1007/s10676-017-9430-8>.
- Redden, Joanna. "Democratic Governance in an Age of Datafication: Lessons from Mapping Government Discourses and Practices." *Big Data & Society*, vol. 5, no. 2, 2018, p. 205395171880914–, <https://doi.org/10.1177/2053951718809145>.
- Roy, Jeffrey. "Beyond Westminster Governance: Bringing Politics and Public Service into the Networked Era." *Canadian Public Administration*, vol. 51, no. 4, 2008, pp. 541–68, <https://doi.org/10.1111/j.1754-7121.2008.00041.x>.
- Schroeder, Ralph, and Rich Ling. "Durkheim and Weber on the Social Implications of New Information and Communication Technologies." *New Media & Society*, vol. 16, no. 5, 2014, pp. 789–805, <https://doi.org/10.1177/1461444813495157>.
- Simon, Herbert A. *Administrative Behavior; a Study of Decision-Making Processes in Administrative Organization*. Macmillan Co., 1947.
- Simon, Herbert A. "Applying Information Technology to Organization Design." *Public Administration Review*, vol. 33, no. 3, 1973, pp. 268–78, <https://doi.org/10.2307/974804>.
- Simon, Herbert A. *Models of Man : Social and Rational : Mathematical Essays on Rational Human Behavior in a Social Setting*. John Wiley & Sons Inc., 1957.
- Simon, Herbert A. *The Sciences of the Artificial*. M.I.T. Press, 1969.

- Spatola, Nicolas, and Karl F. Macdorman. "Why Real Citizens Would Turn to Artificial Leaders." *Digital Government* (New York, N.Y. Online), vol. 2, no. 3, 2021, pp. 1–24, <https://doi.org/10.1145/3447954>.
- Talby, D. (2019, June 12). *Council post: There are no accurate machine learning models, only agreeable ones*. *Forbes*. Retrieved December 16, 2022, from <https://www.forbes.com/sites/forbestechcouncil/2019/06/12/there-are-no-accurate-machine-learning-models-only-agreeable-ones/?sh=79fc5ed43466>
- Tang, Yiwen, and Dalzell, Nicole. "Classifying Hate Speech Using a Two-Layer Model." *Statistics and Public Policy*, vol. 6, no. 1, 2019, pp. 80–86, <https://doi.org/10.1080/2330443X.2019.1660285>.
- Treasury Board Secretariat. *Directive on Automated Decision-Making 2021* (Online). Available: <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592>
- Treasury Board Secretariat. *Policy on Service and Digital 2019* (Online). Available: <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32603&section=html>
- Treasury Board Secretariat. *Responsible Artificial Intelligence in the Government of Canada 2018* (Online). Available: [docs.google.com/document/d/1Sn-qBZUXEUG4dVk909eSg5qvfbpNIRhzIefWPtBwbxY/e](https://docs.google.com/document/d/1Sn-qBZUXEUG4dVk909eSg5qvfbpNIRhzIefWPtBwbxY/e)  
dit#
- Valle-Cruz, David, et al. "Assessing the Public Policy-Cycle Framework in the Age of Artificial Intelligence: From Agenda-Setting to Policy Evaluation." *Government Information Quarterly*, vol. 37, no. 4, 2020, p. 101509–, <https://doi.org/10.1016/j.giq.2020.101509>.

- Veale, Michael, and Irina Brass. "Administration by Algorithm? Public Management Meets Public Sector Machine Learning." IDEAS Working Paper Series from RePEc, 2019, <https://doi.org/10.31219/osf.io/mwhnb>.
- Wirtz, Bernd W., et al. "The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration." *International Journal of Public Administration*, vol. 43, no. 9, 2020, pp. 818–29, <https://doi.org/10.1080/01900692.2020.1749851>.
- Yampolskiy, Roman V. "Unexplainability and Incomprehensibility of AI." *Journal of Artificial Intelligence and Consciousness*, vol. 07, no. 02, 2020, pp. 277–291., [doi:10.1142/s2705078520500150](https://doi.org/10.1142/s2705078520500150).
- Zalnieriute, Monika, et al. "The Rule of Law and Automation of Government Decision-Making." *Modern Law Review*, vol. 82, no. 3, 2019, pp. 425–55, <https://doi.org/10.1111/1468-2230.12412>.
- Zuiderwijk, Anneke, et al. "Implications of the Use of Artificial Intelligence in Public Governance: A Systematic Literature Review and a Research Agenda." *Government Information Quarterly*, vol. 38, no. 3, 2021, p. 101577–, <https://doi.org/10.1016/j.giq.2021.101577>.