4-13-2022

Assessing Automated Administration

Cary Coglianese
University of Pennsylvania Carey Law School

Alicia Lai
University of Pennsylvania

Assessing Automated Administration

Cary Coglianese and Alicia Lai

ABSTRACT

To fulfill their responsibilities, governments rely on administrators and employees who, simply because they are human, are prone to individual and group decision-making errors. These errors have at times produced both major tragedies and minor inefficiencies. One potential strategy for overcoming cognitive limitations and group fallibilities is to invest in artificial intelligence (AI) tools that allow for the automation of governmental tasks, thereby reducing reliance on human decision-making. Yet as much as AI tools show promise for improving public administration, automation itself can fail or can generate controversy. Public administrators face the question of when exactly they should use automation. This paper considers the justifications for governmental reliance on AI along with the legal concerns raised by such reliance. Comparing AI-driven automation with a status quo that relies on human decision-making, the paper provides public administrators with guidance for making decisions about AI use. After explaining why prevailing legal doctrines present no intrinsic obstacle to governmental use of AI, the paper presents considerations for administrators to use in choosing when and how to automate existing processes. It recommends that administrators ask whether their contemplated uses meet the preconditions for the deployment of AI tools and whether these tools are in fact likely to outperform the status quo. In moving forward, administrators should also consider the possibility that a contemplated AI use will generate public or legal controversy, and then plan accordingly. The promise and legality of automated administration ultimately depends on making responsible decisions about when and how to deploy this technology.

KEYWORDS: autonomous systems, automation, artificial intelligence, AI, information technology, e-government, digital government, machine learning, constitutional & administrative law

Assessing Automated Administration

Cary Coglianese† and Alicia Lai††

Public administration is susceptible to a series of well-documented sources of error and bias that stem from human cognition as well as from group decision-making. At times, these limitations of human judgment have unfortunately led to major tragedies and troublesome inefficiencies. To overcome the limitations of human decision-making, while concurrently responding to pressing governance needs in the face of limited resources, public administrators are beginning to explore new possibilities offered by artificial intelligence (AI) (Coglianese & Ben Dor, 2021; Engstrom et al., 2020).[*]

AI tools promise to increase the accuracy, consistency, and speed of governmental decisions and task performance. But they also present their own set of legal, management, and political issues. When, if at all, should government agencies use automated AI tools to substitute for existing human-based processes?

This paper assesses this question, offering a framework for deciding whether and when administrative agencies should use AI tools that replace or significantly augment human decision-making. These tools are themselves not perfect, but the relevant question confronting public administrators is whether they do better than human decision-making at specific tasks. Assessing the promise of automated administration will ultimately depend on making concrete, relative assessments of whether specific AI tools can improve the status quo. Nevertheless, it is possible to offer some general considerations to account for when making specific assessments.

We begin by presenting a general case for using automated AI tools in the public sector. We next discuss the core legal issues that are likely implicated by governmental reliance on AI tools. These tools can and do raise a range of legal issues, but we conclude that there currently exist no intrinsic legal barriers to the use of AI under the prevailing administrative law doctrines in the United States. As long as public administrators approach the development and deployment of these tools with due care and responsibility, the use of AI can comfortably fit within prevailing law. We conclude by offering guidance for the ultimate challenge for public administrators: deciding when these tools should replace human decision-making in the performance of specific tasks, and then determining how to design and implement these tools in a responsible manner.

---

† Edward B. Shils Professor of Law and Professor of Political Science, University of Pennsylvania; Director, Penn Program on Regulation.
†† Judicial Law Clerk, United States Court of Appeals for the Federal Circuit.

The opinions set forth in this paper are solely those of the authors and do not represent the views of any other person or institution. This paper is forthcoming in Justin Bullock et al., eds., *Oxford Handbook on AI Governance*.

I. The Case for Automating Administration

Governmental tasks today depend largely on human decision-making. As a result, many of government's failings—from major tragedies to minor inefficiencies—can be traced to limitations on human decision-making. Even when humans act with good intentions, their decisions can be prone to a wide array of well-documented cognitive biases, physical limitations, external pressures, and basic missteps. Additional dysfunctionalities come into play when groups of humans make collective decisions, as they often do in government.

Consider, for example, a few cases of major governmental breakdowns that stemmed in substantial part from limitations in individual and collective decision-making:

- When forced to confront a pandemic in early 2020, the U.S. federal government found itself ill-prepared and fumbling in its response due to a variety of cognitive, bureaucratic, and political factors. For years prior, governmental leaders tended to discount the risks of and under-prepare for a major pandemic in the United States (Fisher, 2020; Lewis, 2021; Slavitt, 2021). Even after the outbreak emerged, government officials and many Americans found themselves prone to downplaying risks and resisting social distancing, masking, and vaccinations (Halpern, Truog & Miller, 2020).

- The 1986 Space Shuttle Challenger exploded on liftoff, resulting in the deaths of all seven crew members, due to a series of individual and ground decision-making deficiencies in deciding whether to launch in cold weather (Report of the Presidential Commission, 1986; Vaughan, 2016). Group pressures among NASA managers and Thiokol representatives play a pivotal role in the decision, as did the fatigue that had set in for the key decision-makers (Forrest, 2005).

- With the Bay of Pigs mission in 1961, human and organizational factors led to over a hundred deaths in the failed attempt by the United States to overthrow the Castro government in Cuba (Wyden, 1979). Decision-makers had succumbed to a sunk-cost fallacy and President Kennedy and his advisors had asked few questions about risks and back-up options and failed to invite alternative perspectives (Neustadt & May, 1986).

In raising these debacles across the decades, we do not necessarily to claim that AI would have prevented them but instead to show the shortcomings of human decision-making that can play a significant role in affecting public administration.

Unfortunately, these examples are not nearly as exceptional as they are prominent (Light, 2014; Schuck, 2014). To these examples, we can add a list of delays, biases, errors, and inconsistencies that afflict other, more routine decisions made by individual actors within government. A legion of frailties and foibles show that human decision-making is far from perfect. At an individual level, the range of shortcomings can be grouped into at least six categories:

1) *Memory, fatigue, and aging*. Human working memory is limited in its capacity (Cowan, 2001). Fatigued individuals are less alert, slower to react, experience lapses in memory,

and generally have difficulty processing information (Alhola, 2007). Over time, aging causes the human brain to shrink in size and its memory to decline (Peters, 2006).

2) *System neglect*. People can overemphasize signals relative to the underlying system which generates the signals, such as in financial sectors (Kremer et al., 2011).

3) *Loss aversion.* People tend to dislike losses far more than they like corresponding gains of the same magnitude (Freund & Özden, 2008).

4) *Hindsight and availability bias*. People increase their estimates of probabilities of past events when retroactively considering an event (Fischhoff, 1975). Recall is subject to availability bias, giving prominence to hazards that are more cognitively available through recent and memorable instances of harm (Tversky & Kahneman, 1974).

5) *Confirmation bias*. People tend to search for and favor information that confirms their existing beliefs, and they tend to ignore or discount information that is inconsistent with or challenges those beliefs (Lord et al., 1979).

6) *Racial and gender biases*. People are subject to implicit biases in making judgments whenever race, gender, and other similar characteristics are involved (Pascalis, 2005; Eberhardt, 2019).

These various limitations of individual judgment ultimately affect governmental decision-making in many instances on a regular basis. For example, judges' criminal sentencing decisions have been shown to vary systematically based on the race of the defendant (Rehavi & Starr, 2014). Human processing of social security disability benefits produces results that differ markedly from one administrative official to another (TRAC, 2011). Elected politicians tend to process information selectively when evaluating the performance of public and private schools, emphasizing the information that tends to confirm their prior beliefs (Baekgaard et al., 2019).

Although it may often be thought that group decisions bring added wisdom, once people "put their heads together," it is far from clear that group decision-making always fares better than individual decision-making. Research generally shows that "[i]f some bias, error, or tendency predisposes individuals to process information in a particular way, then groups exaggerate this tendency" (Hinsz, Tindale & Vollrath, 1997). Moreover, collective decision-making brings its own distinctive pathologies and limitations. One of these pathologies, known as groupthink, follows from a psychological drive for consensus that suppresses dissent and the appraisal of alternatives (Janis, 1972). Another related pathology, social loafing, stems from reductions in motivation and accountability (Kogan & Wallach, 1967) and manifests as members of groups become more likely to slack off (Latane et al., 1979). This corresponds with the well-known problem of collective action, in which individuals have an incentive to free-ride on the production of any collective good (Olson, 1965). In addition, even if all members of a group are active and well-motivated, their collective decisions may lead to incoherent outcomes due to the cycling of preferences among the group's different members (Arrow, 1950). It was not without reason that Otto von Bismark quipped that the making of laws reminded him of the making of sausages. Some research indicates that half of all decisions made within organizations end up in failure (Nutt, 1999).

The many limitations of human decision-making—both individual and collective—almost singlehandedly make the case for governmental use of AI tools. Although automated alternatives to human judgment cannot eliminate every human error, AI tools can be designed to "perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement" (NIST, 2021). They offer the prospect for improving the accuracy and consistency of a variety of governmental tasks, as well as reducing delays and improving administrative efficiencies (Lai, 2021; Sunstein et al., 2021).

Studies have shown that algorithms can more accurately recall memorized content than humans can (Panigrahi, 2018). They can also do better than humans in terms of predictive accuracy (Agrawal, Goldfarb & Gans, 2018). Research indicates that, if applied to bail decisions, machine learning algorithms could be expected to reduce crime rates substantially even with no change in the rate of jailing—or, alternatively, they could reduce jailing rates up to 42 percent with no change in crime rates (Kleinberg et al., 2018). If used by environmental regulators to select industrial facilities to inspect for compliance with water pollution regulations, machine learning algorithms could help increase the identification of violators by as much as 600 percent (Hino, Benami & Brooks, 2018).

Automated systems may also promote greater uniformity of decision-making, especially for tasks that currently necessitate repeated judgments involving multiple individuals, such as decisions concerning taxation, immigration status, or benefits eligibility. The alternative to automation—namely, the training and overseeing of human decision-makers—can be costly or ineffective (or both). But a uniform algorithmic system that applies nationwide may feasibly achieve more consistent results across individual cases (Shrestha, 2019). Furthermore, when errors arise or biases are discovered, automated systems that rely on a common algorithm may be easier to modify and fix than to re-train a large number of human decision-makers.

In an important sense, the case for AI is already being made by its widespread adoption and use in the private sector (McKinsey Global Institute, 2018). Digital algorithms are being shown to outperform experts in performing certain medical diagnoses, making mortgage lending decisions, and placing bets on sporting events (Tschandl, 2019; Gates, 2010; Pretorius & Parry, 2016).

Public administrators are also realizing that AI holds promise. Federal, state, and local agencies have begun using automated tools for a variety of administrative tasks (Engstrom et al., 2020; Bray, 2014). The Bureau of Labor Statistics (BLS), for example, currently uses an AI system to categorize reports of workplace injuries submitted by over 200,000 businesses (Chenok & Yusti, 2018; BLS, 2019). The Food and Drug Administration uses AI for real-time tracking and reporting of microbial sources in foodborne outbreaks (FDA, 2011). The Federal Communications Commission has used an AI tool to sort and analyze the 22 million public comments it received in connection with its net neutrality rulemaking (Bray, 2014).

Although the case for AI-driven automation appears quite strong, AI tools are themselves far from perfect. They cannot be expected to address all deficiencies or errors in the public sector, nor will they be devoid of new sources of error of their own, especially if their forecasts are based on learning from data which already contains human biases. Moreover, when it comes to the application of AI in the public sector, a key question arises about their legality. It is one thing, after all, for AI tools to automate the selection and display of video choices on Netflix, but another to take the place of human judges and public administrators making decisions that affect people's lives. Prior to deciding whether to use algorithms to automate governmental decision-making, public administrators would do well to consider whether that use might pose any constitutional or administrative law roadblocks.

II. The Legality of the Automated State

The initial set of questions thus arises over whether government has the legal authority to substitute machines for humans. AI tools have prompted a series of legal concerns that largely stem from their black-box nature. Many of these legal concerns also arise, of course, with governmental decisions based on human judgment. Yet today these concerns are targeted toward algorithms, with particular worry expressed about the potential for machine learning to discriminate, obscure, and rob human officials of autonomy and discretion. In deciding whether to automate tasks using AI, then, public administrators will likely confront five major sets of legal issues about accountability, procedural due process, transparency, privacy, and equality. We conclude that none of these legal issues should act as any insuperable barrier to the responsible deployment of AI in the public sector, even at times in ways that substitutes automated decision-making for human judgment.

Accountability. The first question is whether a government agency replacing human-based processes with AI-driven automated systems would deprive individuals of a right to an accountable decision-maker (Busuioc, 2020). This concern is lessened when humans remain "in the loop" in AI systems—that is, where the systems only point humans to possible options and then the decision as to whether to take those options remains in human hands. A state supreme court, for example, has sustained the use of an algorithm rating tool by a criminal judge in the sentencing process because the scores generated by the algorithm were not the "determinative factor" in the sentence imposed by the judge (State v. Loomis, 2016).

Even human-out-of-the-loop systems, in which the outcomes of AI analysis fully replace human judgment, may still satisfy legal accountability principles. Under current constitutional law principles in the United States, for example, the nondelegation doctrine requires that a delegation of lawmaking powers must be constrained by an "intelligible principle" as to the basis for these systems' legal decisions. Is it possible that a delegation of decisional authority to an AI-based rulemaking system could run afoul of the nondelegation doctrine?

We think not, for two reasons. First, although the Supreme Court may well reinvigorate the nondelegation doctrine in the years ahead, the doctrine has in reality never placed a major constraint on administrative government and seems unlikely to do so in the future. The Court long ago determined, for example, that the legislative articulation of goals as vague as "public interest, convenience, and necessity" satisfy the intelligible principle test.

Second, and more importantly, any exercise of rulemaking authority by an automated system based on AI would be driven by what would of necessity be a highly intelligible principle, as machine learning algorithms can only work if their goals are stated clearly and with mathematical precision. Thus, even under a potentially reinvigorated nondelegation doctrine in the years ahead, automation using AI tools could not only pass muster but will arguably make administration more accountable, not less, given the precision with which machine-learning algorithm's objectives must be specified (Coglianese 2021).

A related accountability issue arises when algorithms that are used by governments are designed and operated by private entities. In such instances, they could run afoul of the *private nondelegation doctrine*, which disfavors the outsourcing of governmental decision-making to nongovernmental entities. Although intuitively this doctrine might constrain governments in relying on contractors to develop AI systems, the principal rationale of the private nondelegation doctrine—avoidance of corruption—simply does not fit the context of machine learning. Private delegation has only been disfavored by the courts when private entities are likely to make decisions based on their own narrow self-interest instead of the broader public interest. Algorithms, however, are programmed to optimize objectives defined by the operators. As long as those operators act responsibly and are held accountable to the public through active governmental input and oversight, then the algorithms themselves pose no risk of corruption. AI-driven automation may even result in more accountable and faithful digital agents than human officials (Coglianese & Lehr, 2017).

Procedural due process. A second legal question is whether the principle of procedural due process requires a human decision, such that automation cannot lawfully be used to take humans out of the loop (Citron, 2008). Procedural due process typically calls for government decision-makers who will listen, serve as neutral arbiters, and render reasoned judgments. Yet as much as these usual requirements might seem intuitively to call for a human decision, under prevailing constitutional principles what is demanded of procedural due process is supposed to be assessed using a balancing test comprising three factors: (1) the affected private interests; (2) the risk of decision-making error; and (3) the government's interests concerning fiscal and administrative burdens (Mathews v. Eldridge, 1976). AI-driven automated decisions would likely pass muster quite easily under this balancing test if an automated system would reduce errors and operate more efficiently. Indeed, due process might even eventually demand that government rely on algorithmic tools to achieve fairer and more consistent decisions than humans can deliver.

Moreover, despite the black-box nature of machine learning, due process expectations would likely be satisfied if automation is developed in a manner open to scrutiny (Coglianese & Lehr, 2019). Affected interests could be afforded the opportunity to interrogate design choices made by architects of these algorithms, and key design choices could be vetted through open processes with input from advisory committees, peer reviewers, and public comments or hearings.

Transparency. A third concern, transparency, stems from what has come to be considered a hallmark of best practices in public administration. One best practice—which has been called "fishbowl transparency" (Coglianese, 2009)—would imply that information about the design and operation of an algorithmic system be made publicly available and any key decision-making meetings about the algorithm should also be open to the public. Yet fishbowl transparency is not absolute. Laws such as the Freedom of Information Act contain a variety of exceptions to required disclosures of public information, such as for law enforcement strategies or trade secrets. Some automated systems could legitimately fall under such exceptions.

An additional form of transparency—"reasoned transparency" (Coglianese, 2009)—demands that administrators provide reasons for their decisions. The use of machine learning algorithms would seem problematic from the standpoint of reason-giving because their outputs cannot be intuitively understood or easily explained. Yet human decision-making can also be a black box, and human officials are not expected to furnish complete explanations of their decisions. Under prevailing reason-giving doctrine, administrators that rely on automation can likely satisfy their reason-giving obligations by explaining in general terms how their algorithms were designed to work and what data sources they are using (Coglianese & Lehr, 2019). Courts already defer to administrators' expertise in cases involving complex machinery or mathematical analyses, so it can be expected that courts will likely assume a similar deferential approach to evaluating an agency's reasons for decisions based on machine learning.

Furthermore, new technological advances continue to increase the interpretability of machine learning models, opening up the algorithmic black box (Selbst & Barocas, 2018; Kearns & Roth, 2019). With time, administrators will be able to do even more to make sense to all who are affected by algorithms of the logic behind their decisions.

Privacy. Because machine learning algorithms thrive on large quantities of data, they raise a fourth category of legal concerns centered on privacy. Even when processed data includes sensitive or personal information, any privacy concerns should be as manageable as any of those related to large administrative data systems that exist today which do not drive AI systems. Agencies already routinely handle an array of personal information, such as names, Social Security numbers, and biometrics. Protecting these data while concomitantly advancing agency goals is a task many agencies already have experience addressing.

Reasonably well-settled legal standards guide U.S. agencies in addressing privacy concerns. The Privacy Act of 1974 limits how agencies can collect, disclose, and maintain personal information. The E-Government Act of 2002 requires agencies conduct privacy impact assessments when developing technology that implicates privacy concerns. Admittedly, growing calls exist for the federal government to adopt new data privacy laws, akin to the European Union's General Data Protection Regulation (GDPR) or to privacy laws adopted in some states, such as California. To the extent such new laws are adopted, they could necessitate greater attentiveness to privacy protections, but so far neither GDPR nor California's privacy law has created an insuperable barrier to the use of artificial intelligence.

Machine learning algorithms do present a distinctive privacy concern in their ability to combine seemingly disparate, non-sensitive data to yield predictions about personal information, such as sexual orientation or political ideology. In the private sector, retailers have been able to use predictive analytics to infer private details from seemingly benign data, and then they have used those highly accurate inferences for targeted advertising. When it comes to the public sector, the worry is that government officials could do the same, but for ill purposes. Still, such officials already are constrained from using this reverse-engineering potential of learning algorithms in ways adverse to individuals' interests because current law prohibits the "abuse of discretion" by government officials (Appel & Coglianese, 2021). As long as federal agencies use AI tools responsibly, there should seem to be no intrinsic privacy law impediment to that use.

Equal protection. Finally, automated systems raise legal concerns about equality. Biases in the existing data on which machine learning algorithms train may simply become perpetuated, or perhaps even exacerbated, in AI-driven systems. Even when demographic details on race or gender are excluded from training datasets, algorithms might "find" a pattern based on these variables using the other variables in the dataset. Some research has documented the potential for racial and gender discrepancies in AI-based systems used for employment hiring (Bertrand & Mullainathan, 2004) and medical decision-making (Obermeyer et al., 2019). A widely circulated report published in ProPublica showed signs of racial bias in the (non-learning) algorithms in a system known as COMPAS meant to provide an objective measure of recidivism (Angwin et al., 2016).

Bias obviously exists with human decision-making, and indeed human bias will often be an underlying source of bias in AI-driven systems when they are trained on data that accumulated in the past from human judgments (Mayson, 2019). But bias is also a concern with machine learning algorithms, especially when the underlying training data are already biased. As a legal matter, when bias is intentional, it will clearly offend constitutional equality protections—whether in a system driven by humans or machines. But absent some independent showing of such intentional animus in the underlying objectives established for an AI system, it will be difficult for litigants to show unlawful discrimination with AI just from its outputs.

One challenge for individuals in protected classes who are adversely affected by an AI system will be to demonstrate that they were discriminated *on the basis* of race (Coglianese & Lehr, 2017). This is due to the inscrutability of machine learning algorithms; they do not assign weights to or permit causal inferences about specific variables. Even when individuals within a protected class are treated worse than others, it is possible that the algorithm led to better outcomes for that class overall than would have been in a counterfactual situation—or it might even be that any demographic variables in the dataset on which the algorithm trained were not actually influential at all in producing the outcomes. Individuals claiming algorithmic injustice must show that the government's decision involved real reliance on a suspect classification. Such claims will face an uphill climb.

Another obstacle to sustaining an equal protection challenge to machine learning systems will be the lack of "categorical treatment" in any adverse decisions produced through machine learning (Coglianese & Lehr, 2017). The Supreme Court has disapproved of administrative decisions on equal protection grounds when the government has afforded a categorical preference or disadvantage to certain classes. Such categorical treatment is unlikely ever to arise with automated administration because an algorithm's objective function will be defined in terms of some class-neutral outcome, even if underlying data contains human bias. As a result, we may expect few equal protection challenges to algorithmic administration to trigger heightened scrutiny. Furthermore, even if a court did apply heightened scrutiny to a machine learning system, this might not lead it to find a violation of equal protection. After all, when administrators rely on machine learning systems, they often do so to advance the kinds of compelling state interests which will withstand heightened scrutiny.

Summary. The five legal issues potentially implicated by governmental use of machine learning—accountability, procedural due process, transparency, privacy, and equal protection—do not appear to present any insurmountable barriers to the use of AI under prevailing law. These legal issues already arise with respect to governmental decisions based on human judgment or more conventional types of analytic tools. If the courts tolerate a status quo grounded on human biases and frailties, then they are unlikely to object categorically to automated systems that rely on digital algorithms, especially when these automated systems can be shown to overcome human limitations. Obviously, the law may always change in ways that could lead to greater scrutiny of AI tools, as new legal rulings and precedents may lead the law to evolve alongside the increased use of algorithmic decision-making in government (Deeks, 2019). Indeed, some local jurisdictions have in recent years adopted restrictions on the use of AI-based facial recognition software by law enforcement personnel. Yet is also quite possible that legal change in the future could work in favor of the adoption of AI tools. In some cases, we might expect that responsible use of algorithmic tools will *enhance* principles of administrative justice and eventually the use of these tools by government come to be legally *required* (Coglianese & Hefter, 2021). This would be especially so if data scientists continue to make progress in developing new algorithmic techniques that successfully address the various fairness, transparency, accountability, and other concerns that have been raised about the use of AI tools(Berk & Kuchibhotla, 2021; Kearns & Roth, 2019; Mullainathan 2019).

Nevertheless, in concluding that prevailing administrative law and constitutional law principles pose no intrinsic legal bars to the use of AI, we do not deny that some specific instances of automation will likely be challenged in court in the future. Already algorithmic tools deployed by governments have found themselves subject to legal challenge (Coglianese & Ben Dor, 2021; Yoo & Lai, 2020). If administrators fail to exercise due diligence in their development of particular automated systems, they may well face court decisions disapproving of their use or finding that such use violates due process or is "arbitrary or capricious." As a result, when it comes to adopting AI in specific contexts for specific tasks, public administrators will need to take care in deciding whether to replace human decision-making with an automated system and then, when they do, in how they design and operate such systems. The key will be for public administrators to act responsibly when choosing automated administration.

III. Choosing Automated Administration

Decisions about the use automation will need to be guided by a careful assessment of whether AI tools can lead to better processes and outcomes. Such evaluative judgments cannot be meaningfully made in the abstract but must be made within specific contexts and with respect to specific tasks. In some cases, machine learning will prove more beneficial than human decision-making, while other times it will not. Despite the inherent context-specific nature of choices about automation, it is possible to outline relevant factors that ought to be considered. In general, the criteria for deciding whether to shift an existing human-driven process to one driven by artificial intelligence will fall into three categories: preconditions for the successful use of automation, the overall value of any such use, and considerations related to justifying automation so that it meets with acceptance by courts, legislative overseers, and the public.

A. Preconditions for Use

To use machine learning tools effectively, a public administrator will need to ensure at least four preconditions for such use are met. Where the preconditions for artificial intelligence cannot be sufficiently met, automation should be avoided. The following four conditions are necessary but not sufficient for shifting from a human-driven process to automation.

1. *Adequate resources*. Administrators will need both the budgetary and human capital resources needed to build, test, maintain, and oversee an automated system. These systems require expertise, some of which could reside within a government agency but which could also be met through private contractors. Government agencies will also need adequate digital computing resources, such as data storage and digital processing capabilities. Both the stored data and computing processors will need adequate cybersecurity protections.

2. *Goal clarity and precision*. Machine learning algorithms must be programmed to optimize a mathematically defined objective. Vague goals, such as fairness or reasonableness, cannot meet the mathematical precision needed. Fairness would need to be defined, such as by specifying that favorable outcomes generated by an automated system are to be distributed across racial and gender groups in ways that are proportionate to the distribution in an applicant pool or in society overall. In addition, any tradeoffs between values must be specified precisely. If the question is how much forecasting accuracy should be sacrificed for proportionality in outcomes, the tradeoff will need to be specified mathematically. In making such tradeoffs, agencies will need to have sufficient statutory direction and social input about how to define such tradeoffs.

3. *Data availability*. Machine learning works by identifying patterns within large quantities of data. If large quantities of relevant data are not available, then machine learning will not be an option for automating an agency task. Data may be lacking for a variety of reasons. In some cases, they may exist, but only in paper form. Or they may be practically unavailable if disparate digital datasets are unable to be combined because they lack a common identifier to link data about specific businesses or individuals.

More conceptually, data may be unavailable due to an insufficient number of repeated events around which a common set of variables could be collected. It is one thing to draw on data of individuals' DNA, for example, to determine whether a DNA sample from a crime scene matches that of a defendant. It is another to have a dataset that could determine whether a defendant was driving a red Corvette through the intersection of Sixth and Main Streets on July 23rd (Rigano, 2018). It will obviously be difficult to find a large set of data for cases or tasks that are truly *sui generis*.

4. *External validity*. The data available for training a machine learning algorithm must fit with and be applicable to the population that will be subjected to the outcomes of an automated

system. If relevant factors change more quickly than an algorithm can be replenished with new data and retrained, then the algorithm can become be "brittle"—in other words, lack external validity (Cummings, 2020). A machine learning algorithm used for economic forecasting, for example, might work well under relatively stable economic circumstances but fail to generate accurate forecasts of employment during an unprecedented pandemic-induced recession. Of course, when circumstances become unusual, human judgment can also become brittle, so the need for external validity on the part of AI systems will not demand perfection but will instead require greater external validity than existing human-based alternatives.

B. The Value of Use

When the four preconditions above are met, AI-driven automation will be a plausible substitute for human judgment. None of these preconditions needs to be perfectly satisfied. But if they are not even minimally met in a given use case, it will not make sense to contemplate the use of automation.

In deciding whether to go forward, public administrators should determine whether improvement over the status quo will be attained. Although the specific criteria for determining such improvement will depend on the specific use contemplated (Young, Bullock & Lecy, 2019), at least three general sets of impacts warrant consideration.

1. *Task performance*. Current human-based systems center on tasks, so one set of criteria should be guided by the objectives underlying those tasks: Would automation complete public administrators' tasks more accurately? Would it reduce the time it takes to complete these tasks? Would it be less costly, needing fewer FTEs or other resources? Would automation yield a greater degree of consistency in outcomes? Each of these questions can be asked in light of the current purpose of the human-driven system in use. The overarching idea will be to determine whether automation would help public administrators better achieve their objectives.

2. *User or beneficiary impacts*. It will also be vital to attend to the effects of automation on the applicants, beneficiaries, or other individuals and businesses who use or are directly affected by a specific system. These impacts may already be part of the task objectives, but if not they should be carefully considered. How does the system treat those directly affected by its outputs? Do some portion of users suffer disproportionate adverse effects? Do they feel like the system has sufficiently served them well? In answering these questions, it is not necessary that an automated system be perfect—just better than the status quo. If the status quo of human-based systems necessitates that members of the public wait hours on the telephone to speak to someone who can assist them, a machine learning chatbot might be much better in relative terms, even if not ideal. In this regard, it is notable that the online clearinghouse, eBay, deploys a fully automated dispute resolution software system that resolves disputes so satisfactorily that customers with disputes are actually more inclined to return to eBay than those who never have a dispute (Barton & Bibas, 2017).

3. *Societal impacts*. An administrator should also consider broader societal effects: How would automation indirectly affect others? Will remaining errors have broader consequences? Errors might not have huge ramifications with, say, an automated customer service system used for answering commonly asked questions. But the ramifications to society overall could be profound for an AI system that automatically determines who receives commercial aircraft pilot licenses. Again, the key question to ask will be whether any broader societal effects would be better or worse than similar effects under the status quo.

In characterizing the value of an automated system across all three sets of impacts—that is, task performance, direct impacts, and indirect impacts—it is important to be as specific as possible about the *degree of* improvements (or performance declines) resulting from a shift to automation. Automation will not necessarily fare equally well on each of these impacts. Decisions will need to be made about which impacts are more important. If automation proves less costly but slightly less accurate than the status quo, how important is accuracy for the use case at hand? How consequential are any errors that might remain with an automated system? It might be one thing for the U.S. Postal Service to tolerate some modest number of additional mistakes in letter sorting if automation lowers the costs of mail sorting dramatically, but it would be another matter altogether to make that same kind of tradeoff with respect to identifying safety risks on offshore oil rigs.

C. Justifying the Use

Decisions to automate governmental tasks can pose risks of controversy and conflict. Agency officials should consider these risks and seek to manage them by building a justification for their use of automation. The justification will need to inform and satisfy legislative overseers, courts, and members of the public. The need for justification—and the risk of controversy—will likely be affected by two factors: (1) the degree to which automation overrides human decision-making, and (2) the stakes associated with the automated system's use.

Automation could provide just an *input* into human decision-making, as in the use at the heart of *State v. Loomis* (2016). Or automation could provide a *default* output that stands unless overridden by a human. Or it could make the final *decision* subject only to judicial review—essentially, a human-out-of-the-loop system. All things being equal, new uses of machine learning that only provide inputs into agency actions will be less likely to be create controversy or be challenged in court than uses that create defaults or decisions (Coglianese & Lehr 2017).

In addition, the lower the stakes associated with the task performed by automation, the lower the risk of controversy or litigation. Among the lowest conceivable stakes will be purely internal staff uses within a public agency, such as when its computer support staff uses a machine learning algorithm as part of a chat bot that answers calls for password resets on office computers. That chatbot could even be designed to work with humans out of the loop, responding to requests entirely on its own, and yet controversy will be highly unlikely (Brown, 2016). On the other hand, automated systems that are involved in the processing of applications for economically valuable licenses or permits by private businesses will have substantial stakes—and thus will pose risk of controversy or even litigation—even if they provide only inputs into decisions made by human officials.

When AI systems create defaults or decisions over high-stakes matters, public administrators can anticipate controversy to ensue or legal challenges to be filed. In these cases, they will need to take pains to demonstrate that they have engaged in careful planning and testing of these automated systems. Public notice and public engagement will often be appropriate in the development and design of automated systems in these circumstances, both to build confidence that may head off controversy and to reassure overseers that such systems have been designed and are being used responsibly.

Validation of automated systems will also be vital in justifying their use. Administrators should make validation data available to the public. They may also consider subjecting the design and performance of automated systems to external peer review or third-party auditing. Pilot programs could be established to run the algorithm in tandem with human decision-makers for a period of time to observe how it will operate in practice. Once the digital system has replaced a human-driven system, validation efforts should continue to occur early in its use before any irreversible loss of human capital. Future upgrades to the digital system will benefit from continued validation that each iteration improves on the one before—or at least does not present unacceptable side effects.

When contracting out for technical support and services in developing a machine learning system, public administrators should scrutinize with care the sale pitches that contractors make (Dekel & Schurr, 2014), aware of the reality that their agencies, more than the contractors, will be held responsible for problems that arise with an automated system. In crafting contracts for AI services, administrators will do well to account for their need to access and disclose sufficient information about the algorithm, underlying data, and validation results to satisfy transparency norms (Coglianese & Lampmann, 2021). Overall, administrators should ensure that the procurement process will result in transparent outcomes for public administration, even if it involves proprietary software that would otherwise shield the algorithms from scrutiny.

Finally, public administrators should approach automation with humility, caution, and care, recognizing that AI systems are themselves creatures of human decision-making that can be prone to blind spots and biases. Because frailties of human judgment affect all human decisions—including decisions about whether and how to design and use automated systems—public administrators should remain vigilant. Such vigilance will also help in reassuring courts, legislative overseers, and the public.

CONCLUSION

A future of governance driven by AI-based automation naturally gives rise to concerns by judges, policymakers, and members of the public about how such use of digital tools will affect the efficacy, fairness, and transparency of governmental processes. But the status quo already relies on human “algorithms” that are far from perfect. If algorithmic automation can usher in a new public administration that—at least for specific uses—achieves better results at constant or fewer resources than current processes based on human decision-making, then both government administrators and the public would do well to support the use of artificial intelligence.

We conclude that the current legal system creates no insuperable obstacles to such public sector use of automation as a substitute for, or an augmentation of, current decision-making by human officials. Choices about whether and when to use AI tools will thus rest firmly on the shoulders of public administrators. These choices should be made with care, taking into due consideration the preconditions for the use of AI, the value of such use in terms of improving the status quo, and the need for public administrators to engage the public and ensure sufficient justification of their decisions about automation. Responsible, well-informed human decisions about the design and deployment of automation will shape whether AI is permitted to overcome the limitations of human decision-making and improve governmental performance.

NOTE

[*] Although no universally accepted definition of AI exists, for the purposes of this paper we focus on systems that rely on the use of machine learning algorithms, both supervised and unsupervised. Such algorithms, which themselves come in a variety of forms, are increasingly making possible the automation of a range of previously human-centered activities, such as with the driving of automobiles and the reading of x-rays. These kinds of algorithmic tools are sometimes described with a variety of phrases, including “big data,” “predictive analytics,” “deep learning,” “reinforcement learning,” “neural networks,” “random forests,” and “natural language processing.” In the context of public administration, the use of such tools would constitute one manifestation of what has been described variously as smart or data-smart government (Goldsmith & Crawford, 2014; Noveck, 2015) or moneyballing government (Nussle & Orszag, 2015). For discussion of the relevant characteristics of the type of machine learning algorithms we have in mind, see Coglianese & Lehr (2019), Anastasopoulos & Whitford (2018), and Coglianese & Lehr (2017).

REFERENCES

Agrawal, A., Goldfarb, A. & Gans, J. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*.

Alhola, P. et al. (2007). Sleep Deprivation: Impact on Cognitive Performance. *Neuropsychiatric Disease & Treatment*, 3, 553.

Anastasopoulos, L.J. & Whitford, A.B. (2018). Machine Learning for Public Administration Research, With Application to Organizational Reputation. *Public Administration Research & Theory*, 29(3), 491–510.

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016). Machine Bias. *ProPublica*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Appel, S. & Coglianese, C. (2021). Algorithmic Governance and Administrative Law, in Woodrow Barfield, ed., *Cambridge Handbook on the Law of Algorithms: Human Rights, Intellectual Property, Government Regulation*, 162-181.

Arrow, K.J. (1950). A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, 58, 328–346.

Baekgaard, M. et al. (2019). The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians. *British Journal of Political Science*, 49, 1117.

Barton, B.H. & Bibas, S. (2017). *Rebooting Justice: More Technology, Fewer Lawyers, and the Future of Law*.

Berk, R. & Kuchibhotla, A.K. (2021). Improving Fairness in Criminal Justice Algorithmic Risk Assessments Using Conformal Prediction Sets, https://arxiv.org/abs/2008.11664.

Bertrand, M. & Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, *American Economic Review*, 94, 991.

Bray, D.A. (2014). An Update on the Volume of Open Internet Comments Submitted to the FCC. *Federal Communications Commission*, https://www.fcc.gov/news-events/blog/2014/09/17/update-volume-open-internet-comments-submitted-fcc.

Brown, J. (2016). Chatbots Debut in North Carolina, Allow IT Personnel to Focus on Strategic Tasks. *Government Technology*.

Bureau Labor Statistics (BLS), Automated Coding of Injury and Illness Data, https://www.bls.gov/iif/autocoding.htm.

Busuioc, M. (2020). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, 1–12.

Chenok, D. & Yusti, C. The Future Has Begun: Using Artificial Intelligence to Transform Government (2018). *IBM Center for The Business of Government*, https://ourpublicservice.org/wp-content/uploads/2018/01/0c1b8914d59b94dc0a5115b739376c90-1515436519.pdf.

Citron, D.K. (2008). Technological Due Process, *Washington University Law Review*, 85, 1249.

Coglianese, C. (2021). Administrative Law in the Automated State. *Dædalus*, 150, 104-120.

Coglianese, C. (2009). The Transparency President? The Obama Administration and Open Government. *Government*, 22, 529.

Coglianese, C. & Ben Dor, L.M. (2021). AI in Adjudication and Administration, *Brooklyn Law Review*, forthcoming.

Coglianese, C. & Hefter, K. (2021). From Negative to Positive Algorithm Rights, *William & Mary Bill of Rights Journal*, forthcoming.

Coglianese, C. & Lampmann, E. (2021). Contracting for Algorithmic Accountability, *Administrative Law Review Accord*, 6, 175.

Coglianese, C. & Lehr, D. (2017). Regulating by Robot: Administrative Decision Making in the Machine-Learning Era. *Georgetown Law Journal*, 105, 1147.

Coglianese, C. & Lehr, D. (2019). Transparency and Algorithmic Governance. *Administrative Law Review*, 71, 1.

Cowan, N. (2001). The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences*, 24, 87.

Cummings, M.L. (2020). The Surprising Brittleness of AI. *Women Corporate Directors*.

Deeks, A. (2019). The Judicial Demand for Explainable Artificial Intelligence. *Columbia Law Review*, 119, 1829-1850.

Dekel, O. & Schurr, A. (2014). Cognitive Biases in Government Procurement – An Experimental Study. *Review of Law & Economics*, 10, 169.

Eberhardt, J.L. (2019). *Biased: Uncovering the Hidden Prejudice That Shapes What We See, Think, and Do*.

Engstrom, D.F., Ho, D.E., Sharkey, C.M. & Cuéllar, M.-F. (2020). Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. *Administrative Conference of the*

United States, https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.

Fischhoff, B. (1975). Hindsight is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty. *Journal of Experimental Psychology*, 1, 288.

Food & Drug Administration (FDA) (2011). FDA Commissioner's Fellowship Program, https://www.fda.gov/media/83569/download.

Fisher, M. (2020). Coronavirus 'Hits All the Hot Buttons' for How We Misjudge Risk. *New York Times*.

Forrest, J. (2005). The Space Shuttle Challenger Disaster: A Failure in Decision Support System and Human Factors Management. *Decision Support Systems Resources*, http://dssresources.com/cases/spaceshuttlechallenger/index.html.

Freund, C. & Özden, C. (2008). Trade Policy and Loss Aversion. *American Economic Review*, 98, 1675.

Gates, S.W. et al. (2010). Automated Underwriting in Mortgage Lending: Good News for The Underserved? *Housing Policy Debate*, 13, 369.

Goldsmith, S. & Crawford, S.P. (2014). *The Responsive City: Engaging Communities Through Data-Smart Governance*.

Halpern, S.D., Truog, R.D. & Miller, F.G. (2020). Cognitive Bias and Public Health Policy During the COVID-19 Pandemic. *Journal of the American Medical Association*, 324, 337-338.

Hansson, D.H. (2019). *Twitter*, https://twitter.com/dhh/status/1192540900393705474.

Hino, M., Benami, E. & Brooks, N. (2018). Machine learning for Environmental Monitoring, *Nature Sustainability*, 1, 583-588.

Hinsz, V.B., Tindale, R.S. & Vollrath, D.A. (1997). The emerging conceptualization of groups as information processors, *Psychological Bulletin*, 121, 43-64.

Janis, I. (1972). *Victims of groupthink*.

Kearns, M. & Roth, A. (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. & Mullainathan, S. (2018). Human Decisions and Machine Predictions, *The Quarterly Journal of Economics*, 133, 237-293.

Kogan, N., & Wallach, M. A. (1967). Risky-shift phenomenon in small decision-making groups: A test of the information-exchange hypothesis. *Journal of Experimental Social Psychology*, 3, 75–84.

Kremer, M., Moritz, B. & Siemsen, E. (2011). Demand Forecasting Behavior: System Neglect and Change Detection. *Management Science*, 57, 1827.

Lai, A. (2021). Artificial Intelligence, LLC: Corporate Personhood as Tort Reform. *Michigan State Law Review*, 2021, 599.

Latane, B., Williams, K. & Harkins, S. (1979). Many Hands Make Light the Work: The Causes and Consequences of Social Loafing. *Journal of Personality & Social Psychology*, 37, 822.

Lewis, M. (2021). *The Premonition*.

Light, P.C. (2014). A Cascade of Failures: Why Government Fails, and How to Stop It. *Center for Effective Public Management at Brookings*.

Lord, C.G., Ross, L. & Lepper, M.R. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality & Social Psychology*, 37, 2098.

Mathews v. Eldridge, 424 U.S. 319 (1976).

Mayson, S. (2019). Bias In, Bias Out. *Yale Law Journal*, 128(8), 2218.

McKinsey Global Institute (2018). Applying AI for Social Good, https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx.

Mullainathan, S. (2019). Biased Algorithms Are Easier to Fix Than Biased People. *New York Times*, Dec. 6.

National Institute of Standards and Technology (NIST), Artificial Intelligence, https://csrc.nist.gov/Topics/technologies/artificial-intelligence.

Neustadt, R.E. & May, E.R. (1986). *Thinking in Time: The Uses of History for Decision-Makers*.

Noveck, B.S. (2015). *Smart Citizens, Smarter State: The Technologies of Expertise and the Future of Governing.*

Nussle, J. & Orszag, P. (2015). *Moneyball for Government*.

Nutt, P.C. (1999). Surprising But True: Half the Decisions in Organizations Fail, *The Academy of Management Executive (1993-2005)*, 13, 75-90.

Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019). *Dissecting racial bais in an algorithm used to manage the health of populations*, *Science*, 366, 447.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*.

Panigrahi, B.K. et al. (2018). Nature Inspired Computing. *Advances in Intelligent Systems & Computing*, 652, 203.

Pascalis, O. et al. (2005). Plasticity of Face Processing in Infancy. *Proceedings of the National Academy of Sciences*, 102, 5297.

Peters, R. (2006). Aging and the Brain. *Postgraduate Medical Journal*. 82, 84.

Pretorius, A. & Parry, D. (2016). Human Decision Making and Artificial Intelligence: A Comparison in the Domain of Sports Prediction. *Annual Conference of the South African Institute of Computer Scientists and Information Technologists*.

Rehavi, M.M. & Starr, S.B. (2014). Racial Disparity in Federal Criminal Sentences, *Journal of Political Economy*, 122, 1320.

Report of the Presidential Commission on the Space Shuttle Challenger Accident (1986), https://history.nasa.gov/rogersrep/v2appg.htm.

Rigano, C. (2018). Using Artificial Intelligence to Address Criminal Justice Needs. *National Institute of Justice*.

Schuck, P.H. (2014). *Why Government Fails So Often: And How It Can Do Better*.

Selbst, A.D. & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines, *Fordham Law Review* 87, 1085.

Shrestha, Y. et al. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, 6, 66.

State v. Loomis, 371 Wis.2d 235 (2016).

Sunstein, C., Kahneman, D. & Sibony, O. (2021). *Noise*.

TRAC (2011). Social Security Awards Depend More on Judge than Facts: Disparities Within SSA Disability Hearing Offices Grow (2011), https://trac.syr.edu/tracreports/ssa/254/.

Tschandl, P. (2019). Comparison of The Accuracy of Human Readers Versus Machine-Learning Algorithms for Pigmented Skin Lesion Classification: An Open, Web-Based, International, Diagnostic Study. *The Lancet Oncology*, 20, 938.

Tversky, A. & Kahneman, D. (1974). Judgment Under Uncertainty. *Science*, 185, 1124.

Vaughan, D. (2016). *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA.*

Wyden, P. (1980). *The Bay of Pigs: The Untold Story*.

Yoo, C.S. & Lai, A. (2020). Regulation of Algorithmic Decision-making in the United States. *Journal of Law & Economic Regulation*, 13, 7.

Young, M.M., Bullock, J.B. & Lecy, J.D. (2019). Artificial Discretion as a Tool of Governance: A Framework for Understanding the Impact of Artificial Intelligence on Public Administration. *Perspectives on Public Management & Governance*, 2019, 301–313.