A Framework for Governmental Use of Machine Learning

Cary Coglianese[*]

Computerized algorithms increasingly make decisions that previously had been made by humans. These new types of algorithms—known as machine-learning algorithms—have recently found themselves in use in so many products and settings that they may even appear to portend the reshaping of many important aspects of human life.[1] With their distinctive ability to find complex patterns in large datasets, machine-learning algorithms are being used to help make forecasts about who to hire[2] or lend money,[3] how to trade stocks,[4] and what products consumers are likely purchase.[5] They also drive both Internet search and autono-

[*] This is a preliminary draft. The author gratefully acknowledges major contributions to the drafting of Parts I and II by Alicia Lai, as well as many helpful contributions both to this project and to related collaborations by Steven Appel, Lavi Ben Dor, and David Lehr. Emma Ronzetti and Roshie Xing provided valuable research assistance.

[1] *See, e.g.,* ERIK BRYNJOLFSSON & ANDREW MCAFEE, THE SECOND MACHINE AGE: WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES 251 (describing the current trend toward "machine intelligence" as creating a societal "inflection point" that will generate a "shift as profound as that brought on by the Industrial Revolution"); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 13 (2016) (decrying the "dark side" of machine-learning algorithms and the purported risk that people will be "increasingly controlled by secret models wielding arbitrary punishments"). Machine-learning algorithms can learn to identify patterns across the vast quantities of data that can now be stored and processed digitally, and they can do so autonomously—that is, without human specification of the form of a particular model or key variables, and subject mainly to overarching criteria or parameters to be optimized. As such, these algorithms are often discussed under the banner of "Big Data" or "artificial intelligence." For a discussion of machine learning and how it works, see Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1156-60 (2017); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 UC DAVIS L. REV. 653, 669-702 (2017).

[2] *See* Clare Cain Miller, *Can an Algorithm Hire Better Than a Human?*, N.Y. TIMES (June 25, 2015), https://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html.

[3] *See* Scott Zoldi, *How To Build Credit Risk Models Using AI and Machine Learning*, FICO BLOG (Apr. 6, 2017), http://www.fico.com/en/blogs/analytics-optimization/how-to-build-credit-risk-models-using-ai-and-machine-learning/.

[4] *See* Jigar Patel et al., *Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques*, 42 EXPERT SYSTEMS WITH APPLICATIONS 259 (2015).

[5] *See Using Machine Learning on Computer Engine to Make Product Recommendations*, GOOGLE CLOUD PLATFORM (Feb. 14, 2017), https://cloud.google.com/solutions/recommendations-using-machine-learning-on-compute-engine.

mous vehicles, and they provide the backbone for both advanced medical techniques as well as the everyday use of smartphones.[6] The superior speed and accuracy of machine-learning applications have made them valuable in helping humans make decisions and, increasingly, even in driving automated systems that effectively make decisions themselves.

Today, the reach of machine-learning algorithms extends beyond seemingly banal private-sector uses, such as video recommendatons on Netflix or results on Internet search engines. Militaries are investigating the possibility of automated, robotic warfare—both on the battlefield and in cyberspace.[7] Law enforcement agencies around the world are turning to machine learning to predict when and where crime will occur, as well as which individuals would be likely to commit crimes, say, if they were released on probation or parole.[8] Other governmental bodies are starting to use machine learning algorithms to enhance the administration of social services programs, adjudicate claims for government benefits, and support regulatory functions.[9]

Scholars and commentators have begun to scrutinize applications of machine learning in a variety of settings, but especially by governmental authorities. They postulate that learning algorithms may produce discriminatory effects for members of historically underrepresented groups by reproducing human biases baked into datasets,[10] or by using flawed input[11] or output variables.[12] They

---

[6] Thomas A. Peterson, Emily Doughty, and Maricel G. Kann, *Towards Precision Medicine: Advances in Computational Approaches for the Analysis of Human Variants*, 425 J. MOLECULAR BIO. 4047 (2013); Alexis C. Madrigal, *The Trick That Makes Google's Self-Driving Cars Work*, ATLANTIC (May 15, 2014), http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-thatmakes-googles-self-driving-cars-work/370871/; Nikhil Dandekar, *What are Some Uses of Machine Learning in Search Engines?*, MEDIUM (Apr. 7, 2016), https://medium.com/@nikhilbd/what-are-some-uses-of-machine-learning-in-search-engines-5770f534d46b; Steffen Herget, *Machine Learning and AI: How Smartphones Get Even Smarter*, ANDROIDPIT (Jan. 24, 2018), https://www.androidpit.com/machine-learning-and-ai-on-smartphones.

[7] *See* Andrew Tarantola, *The Pentagon Is Hunting ISIS Using Big Data and Machine Learning*, ENGADGET (May 15, 2017), https://www.engadget.com/2017/05/15/the-pentagon-is-hunting-isis-using-big-data-and-machine-learning/.

[8] *See* Richard Berk et al., *Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning*, 172 J. ROYAL. STAT. SOC'Y SERIES A 191 (2009).

[9] *See generally, e.g.,* Cary Coglianese & Lavi Ben Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. (forthcoming Mar. 2021); DAVID FREEMAN ENGSTROM ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES (2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.

[10] *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 680-87 (2014).

[11] *See id.* at 688-92.

[12] *See id.* at 677-80.

worry as well that these so-called black-box algorithms are too opaque and their results too inscrutable to provide adequate reasons to individuals who want to know why they were denied government benefits or were predicted to pose a crime risk if released on parole or probation.[13] Despite machine learning's reputation for accuracy, some critics even question whether learning algorithms are in fact sufficiently accurate to be used in certain governmental settings that can result in life-altering decisions.[14] Still more broadly, a few commentators express concern about a future where humanity becomes subjected to unaccountable robotic overlords.[15]

These critics express strong claims raising serious potential concerns about the expanded use of artificial intelligence by governmental entities. Yet, as I have indicated elsewhere, governments also face compelling reasons to take advantage of machine learning's potential to improve decision making.[16] The same benefits that machine learning has delivered in the private sector can justify their use by public sector organizations. And in fact, federal, state, and local governments have begun to use algorithms in a variety of administrative contexts.[17] After all, the adjudicatory and enforcement actions that governmental entities make on a daily basis are predicated on certain facts—whether someone has qualified for a benefit or violated a rule, for example—and machine learning is well suited to predict or estimate factual predicates.[18] If augmented with agent-based modeling systems, machine learning could even be applied to the making of new government policies and rules by selecting the "best" rule from among multiple human-specified rules.[19]

Yet, even in day-to-day administrative settings focused on the delivery of government services and the enforcement of government rules, many machine-learning applications could find themselves subjected to the very charges that skeptics have raised elsewhere about artificial intelligence's potential for reinforcing discrimination, its lack of transparency, and its incompatibility with human autonomy and governmental legitimacy. For reasons I have explained in

---

[13] *See, e.g.*, Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1, 1-2 (2016); *2018 Program*, FAT* CONFERENCE (2018), https://fatconference.org/2018/program.html (listing papers focused on topics like the right to explanation, interpretable machine learning, and auditing).

[14] *See, e.g.*, Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

[15] *See, e.g.*, Samuel Gibbs, *Elon Musk: Artificial Intelligence Is Our Biggest Existential Threat*, GUARDIAN (Oct. 27, 2014), https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat; Rory Cellan-Jones, *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC NEWS (Dec. 2, 2014), http://www.bbc.com/news/technology-30290540.

[16] *See generally* Coglianese & Lehr, *supra* note 1.

[17] Coglianese & Ben Dor, *supra* note 9; ENGSTROM ET AL., *supra* note 9.

[18] Coglianese & Lehr, *supra* note 1, at 1167-71.

[19] *Id.* at 1171-75.

other work, these concerns should serve as no intrinsic legal bar to most applications of machine learning in the administrative context, at least under prevailing concepts of administrative law.[20] Still, the critics of artificial intelligence raise concerns that merit consideration. Certainly a wholesale shift on the part of government agencies to the reliance on automated regulatory and administrative decisionmaking systems would mark an major change from the status quo. As with any change, proposals to automate major facets of public administration through artificial intelligence should be suitably analyzed. Any such analysis should begin, though, by recognizing that the status quo itself is far from perfect—and, on some dimensions, it may well be much worse than a world that involves greater reliance on what might be considered to be governing by robot.

At times, critics of machine learning seem to suggest that machine-learning algorithms would produce entirely new problems—that is, that these algorithms are distinctively complex, inscrutable, difficult to scrub of bias, and lacking in accountability. But any realistic assessment of the use of machine learning in public administration needs to acknowledge that government as it exists today is already based on "algorithms" of arguably still greater complexity and potential for abuse. These existing algorithms are those inherent in human decisionmaking, both on an individual and organizational level, and they also can be highly complex, inscrutable, prone to bias, and lacking in accountability.[21] Government already operates through a process involving many individuals, each with their own unique interests, cognitive biases, and limitations. Furthermore, when governmental institutions make collective decisions involving multiple individuals, they are prone to some of the same limitations of individual decision-making as well as to new ones—and it can often be difficult to understand exactly how any collective decision came to be made.

This report begins with the recognition that human decisions in government today can themselves be prone to producing many of the same kinds of harms that critics worry accompany the deployment of machine learning. In fact, in some cases human decisionmaking might arguably be more prone to these limitations or harms—and perhaps to others that do not afflict machine learning. Thus, when evaluating the use of machine learning in governmental settings, any anticipated shortcomings of machine learning must be placed in proper perspective. The choice will not be one of algorithms versus a Platonic ideal; rather, the choice will be one

[20] *See generally, e.g.*, Coglianese & Lehr, *supra* note 1; Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 4-5 (2019); Cary Coglianese & Steven M. Appel, Algorithmic Governance and Administrative Law, in WOODROW BARFIELD, ED., CAMBRIDGE HANDBOOK ON THE LAW OF ALGORITHMS: HUMAN RIGHTS, INTELLECTUAL PROPERTY, GOVERNMENT REGULATION (forthcoming).

[21] *Infra* Part II.

of digital algorithms versus human algorithms, each with their own potential advantages and disadvantages.

Part I of this report begins by detailing the well-documented physical limitations and cognitive biases that affect human decision-making, as well as the types of problems that can arise when humans make collective decisions. The purpose in pointing out these limitations with human decision-making is not to discredit existing governmental processes as much as to show that existing processes can often leave room for improvement. To the extent that automated systems that rely on machine learning can make an improvement over human judgment, then it is worth considering ways that they could be implemented throughout government.

Part II focuses on machine learning and its promise for improving governmental decision-making. I begin by explaining what machine learning is and why it presents an opportunity for improvement, at least in some circumstances. After all, machine-learning systems are already delivering improvements in private-sector applications in terms of accuracy, capacity, speed, and consistency. These private sector uses, combined with what we know about current applications of machine learning in the public sector, suggest that machine learning can deliver similar advantages in the public sector. Of course, just as human decision-making has its limitations, so too does machine learning. Part II concludes with a review of the concerns that have been raised about the use of machine learning by government.

Part III attends to the main legal issues that governmental use of machine learning could raise. These issues implicate legal and policy principles of delegation and accountability, procedural due process and reason-giving, transparency, privacy, and equal protection. None of these principles, as applied under current doctrine, should create any categorical bar to governmental use of machine learning. But ultimately the way that government deploys machine learning with respect to any given application will affect the legal issues. As with the use of other technologies, government will need to engage in responsible planning to ensure that machine-learning applications that substitute for human decision-making will address the core legal and policy principles, especially in high stakes contexts where litigation may ensue.

Finally, Part IV presents a framework for public officials to use in deciding when to develop and deploy automated decision tools that rely on machine learning. This part returns to the starting premise that human decision-making exhibits limitations and focuses government officials' attention on the extent to which machine learning may improve upon a status quo founded on human judgment. It emphasizes the need to consider first whether a new use case for digital algorithms would likely satisfy the preconditions for successful deployment of machine learning: such as, a well-defined objective for repeated tasks for which there exist

large quantities of data on outcomes and related correlates. In this final part, I emphasize the need to be able to validate that a machine-learning system would indeed make an improvement over the status quo, both in terms of the principal objective that the algorithm in question is designed to optimize while avoiding side effects, legal issues, or other new types of problems. In addition to ensuring that improvements from machine-learning systems can be validated, government officials will need to engage in adequate planning for their use, take care in procuring any private contractor services to create these systems, and ensure appropriate opportunities for public participation in the design, development, and ongoing oversight of such systems.

I. Limitations of Human Decision-Making

As I have already suggested, human judgment is algorithmic. Our individual minds operate through processes that can be understood in algorithmic terms to constitute human cognition. Indeed, one form of machine learning, neural networks, draws inspiration from the algorithms that drive human judgment. The algorithmic nature of human judgment is evident beyond individual neurology and psychology. It applies when humans interact socially as well. Any government that operates under the rule of law is a government of algorithms. The law outlines myriad procedures and other steps that are necessary to generate legal outcomes and other authoritative governmental decisions.

The central challenge in choosing to rely on algorithmic decision-making lies not necessarily in honing the accuracy of the results obtained by a machine-learning algorithm and making a digital system work effectively—as difficult as that task may be. Rather, the core challenge will be ultimately making a decision to substitute a digital algorithm for humans in a governmental decision-making process.

As social creatures, humans regularly trust in the integrity and rationality of other humans. Such trust undergirds social life.[22] Trust in the legitimacy of government helps governmental institutions to function.[23] But that trust is not equally shared with digital machines. Empirical research has documented what has

[22] Jillian J. Jordan et al., *Uncalculating Cooperation Is Used to Signal Trustworthiness*, 113 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES 8658 (2016).

[23] *See generally, e.g.*, DONALD F. KETTL, CAN GOVERNMENTS EARN OUR TRUST? (2017); OECD, TRUST AND PUBLIC POLICY: HOW BETTER GOVERNANCE CAN HELP REBUILD PUBLIC TRUST (2017), http://www.oecd.org/corruption-integrity/reports/trust-and-public-policy-9789264268920-en.html; WHY PEOPLE DON'T TRUST GOVERNMENT (Joseph S. Nye, Jr. et al. eds., 1997).]

come to be known as "algorithmic aversion."[24] They do not always trust machines as much as they do humans, even when the machines are shown to be more accurate in the forecasts and outcomes they produce. And as a corrollary, research indicates that people may also tend to be less forgiving when machines make mistakes than when humans do.[25] It is unsurprising that the "right to a human decision," as Aziz Huq has noted, is one of the fundamental assumptions of our legal system.[26]

However, despite people's tendency to view human decision-making more favorably, it is undeniable that human judgment exhibits a series of widely understood limitations and biases. Since the pathbreaking work of Daniel Kahneman and Amos Tversky, an extensive body of research in behavioral economics and cognitive psychology has demonstrated that humans quite regularly make decisions contrary to conventional welfare-maximizing rationality.[27] Instead, individuals are prone to all sorts of cognitive limitations that result in decisions that can be characterized either as irrational or at least puzzling.

Many of the consequences of the limitations and biases in human decision-making do not stem from outright malice, but from taking shortcuts, relying on heuristics, leaping to conclusions before gathering information, or even from the selective gathering or processing of information. Some may stem from expediency or self-interest. Machine-learning algorithms often have few of these limitations, leading to optimism that digital systems can be designed to compensate for and improve upon human limitations. Expressing such optimism, chess champion Garry Kasparov once said after being defeated by the IBM supercomputer Deep Blue, "Anything we can do, … machines will do it better."[28]

To understand whether machines might in fact do better—and, more importantly, when and under what conditions they might—it is necessary to understand the foibles or inefficiencies of the human mind. This Section summarizes a range of human limitations that affect decision-making, separating physical or biological capacities from cognitive biases (although recognizing that these intuitive categorizations are not airtight classifications). By shedding light on this flawed status quo, this Section seeks to make room for the possibility of expanding the role of machine-learning algorithms in governmental processes and decision-making.

---

[24] Berkeley J. Dietvorst et al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCHOL.: GEN. 114, 114 (2015).]

[25] CÉSAR A. HIDALGO ET AL., HOW HUMANS JUDGE MACHINES (forthcoming 2021) (manuscript at ix-x).

[26] Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 615-20 (2020).

[27] For recent syntheses of such research, see generally RICHARD THALER, MISBEHAVING: THE MAKING OF BEHAVIORAL ECONOMICS (2015); DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011).

[28] DAVID EPSTEIN, RANGE: WHY GENERALISTS TRIUMPH IN A SPECIALIZED WORLD 22 (2019).

A. Physical limitations

Physical limitations constitute biological ceilings of human performance. The human brain, after all, is composed of soft tissue, blood vessels, and fatty acids. Each of our one billion of neurons forms connections with other neurons, interweaving into a network of connections in the mind. As humans develop and age, aspects of this brain circuitry are strengthened with use but can also be weakened with neglect, injury, illness, or age. Overall, human decision-making is naturally limited by biological constraints, including physical ceilings, environmental influences, and the natural passage of time. I highlight here five physical capacities that can limit the quality of human decision-making.

1. Memory Capacity. Neuroscientists have estimated that humans have the memory capacity to remember 10^{8432} bits of information—without question making the human brain a highly efficient and high-capacity tool.[29] Nevertheless, practical decision-making often involves reliance less on long-term, aggregated memory, but more on short-term, working memory. For many professional tasks today, the volume and complexity of modern knowledge has exceeded the ability of individuals to deliver effectively.[30] By simply relying on the vastness and faith in human memory capacity, professionals too often fail to consistently, correctly, and safely treat their patients and clients across healthcare, government, law, and finance.

Neuroscientists estimate that human working memory is limited to about four variables (plus or minus one).[31] If a decision-maker exceeds the limit of four variables, decision quality is typically degraded. For instance, medical diagnostic errors occur in 12 million adult outpatients per year, based on medical records, insurance claims, malpractice claims.[32] These errors are largely due to limits on human memory: inadequate collection of patient information, inadequate knowledge amongst physicians, and incorrect interpretation and integration of information.[33] Recognizing this limitation, most clinical decisions are restrained to one to three input variables in order to promote rational decision-making.[34] Humans, in other words, often need to block out a lot of potentially relevant information if they are to make decisions.

---

[29] *Id.*

[30] ATUL GAWANDE, THE CHECKLIST MANIFESTO (2009).

[31] Nelson Cowan, *The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity*, 24 BEHAV. BRAIN SCI. 87 (2001).

[32] JOHN HALAMKA, REINVENTING CLINICAL DECISION SUPPORT.

[33] *Id.*

[34] Alan H. Morris, *Human Cognitive Limitations Broad, Consistent, Clinical Application of Physiological Principles Will Require Decision Support*, NOBEL PRIZE SYMPOSIUM (2018).

One way to overcome the limits on working memory is to rely on ordinary, non-digital algorithms called checklists.[35] The World Health Organization, for example, has developed a surgical safety checklist that reduces the surgical process to a single page of "yes/no" questions. Its use has led significant reductions in morbidity and mortality rates through medical errors.[36]

While the reliance on such simple decision aids have proven effective, it is important to remember that the initial need for these aids is merely a by-product of one aspect of the system: the humans.[37] Even healthy individuals have finite memory capacity and limited working memory. Efforts to increase memory capacity (such as institutionalizing knowledge by congregating a large number of people, or making digital or physical records) may only complicate the decision-making environment and increase demands on human memory.

2. Fatigue. Fatigue has tangible, negative effects on human decision-making. A fatigued individual will be less alert, have difficulty mentally processing information, have slower reaction times, experience memory lapses, and be less situational aware.[38] These factors lower productivity and increase the risk of work-related errors and accidents.[39]

Professions demanding long hours and acute attention, like clinical surgeons, are a breeding ground for examples of avoidable, fatigue-induced human errors.[40] One study of orthopedic surgical residents found that residents were fatigued during 48% of their time awake, increasing the risk of medical error by 22% as compared with well-rested control subjects.[41] Government agencies have

---

[35] *Id.*

[36] *WHO Surgical Safety Checklist*, WORLD HEALTH ORG. (2020), https://www.who.int/patientsafety/safesurgery/checklist/en/.

[37] PAUL CERRATO & JOHN HALAMKA, REINVENTING CLINICAL DECISION SUPPORT (2019) (proposing that optimistically, AI will allow each of us to "have the electronic equivalent of a personal physician who has access to the very latest research, the best medical facilities that specialize in each individual's health problems, access to cutting-edge data sets, predictive analytics, testing options, clinical trials currently enrolling new patients, and much more." For instance, a machine learning tool to distinguish between a normal mole and skin cancer, after analyzing more than 100,000 images, or precision medicine that is personalized to an individual's genetic predisposition).

[38] Paula Alhola et al., *Sleep Deprivation: Impact on Cognitive Performance*, 3 NEUROPSYCHIATRIC DISEASE AND TREATMENT 553 (2007).

[39] Katrin Uehli et al., *Sleep Problems and Work Injuries: A Systematic Review and Meta-Analysis*, 18 SLEEP MED. REV. 61 (2014).

[40] Sponges accidentally left in patients after surgery has become such a frequent occurrence that there is even a term for such items—"retained surgical bodies (RSB)." *See* Valon A. Zejnullahu et al., *Retained Surgical Foreign Bodies after Surgery*, 5 OPEN ACCESS MACED J MED SCI. 97 (2017).

[41] Frank McCormick et al., *Surgeon Fatigue: A Prospective Analysis of the Incidence, Risk, and Intervals of Predicted Fatigue-Related Impairment in Residents*, 147 ARCH SURG. 430 (2012).

recognized fatigue as a health concern: the Health and Safety Executive agency in the UK has listed "employee fatigue" as a top-10 human and organizational factors issue because it drastically increases the risk of human errors.[42]

In other contexts, people have been more reluctant to admit to the effects of fatigue. Corporations rarely proactively admit to exhausting workplace conditions.[43] Judges insist their rulings are shielded from external factors such as fatigue.[44] Yet one study tracked judicial rulings on parole decisions across three decision sessions, punctuated by food breaks. [45] At the start of each sessions, the well-rested judges averaged approximately 65% favorable decisions, a rate that dropped to zero as the judges fatigued.[46] After each food break, the rate reset at 65% and the cycle continued.[47] Justice for some individuals would appear to be affected by the idiosyncrasies of human fatigue.

3. Aging. Aging is a physical phenomenon at its core, affecting molecules, cells, vasculature, gross morphology, and finally, cognition.[48] As we age, the brain, particularly the frontal cortex, shrinks in volume. Our susceptibility to disease rises. Blood pressure, along with the risk of stroke, increases. Brain activation becomes more bilateral for memory tasks. Memory declines. Further, neurodegenerative disorders plague the population, such as Alzheimer's disease stripping its patients of their brain cells and cognition.[49]

---

[42] *Human Factors: Fatigue*, HEALTH & SAFETY EXECUTIVE, https://www.hse.gov.uk/humanfactors/topics/fatigue.htm.

[43] Harriet Agerholm, *Amazon Workers Working 55-Hour Weeks and so Exhausted By Targets They 'Fall Asleep Standing Up,'* INDEPENDENT (Nov. 27, 2017), https://www.independent.co.uk/news/uk/home-news/amazon-workers-working-hours-weeks-conditions-targets-online-shopping-delivery-a8079111.html.

[44] TARA SMITH, JUDICIAL REVIEW IN AN OBJECTIVE LEGAL SYSTEM (2015) (highlighting the importance of objectivity in judicial system, and proposing a theory on how to practically embrace objectivity's demands).

[45] Shai Danziger, Jonathan Levav, & Liora Avnaim-Pesso, *Extraneous Factors in Judicial Decisions*, 108 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES 6889 (2011).

[46] *Id.*

[47] *Id.*

[48] Ruth Peters, *Aging and the Brain*, 82 POSTGRAD MED J. 84 (2006).

[49] The dominant theory in Alzheimer's is the amyloid beta hypothesis—which, at the surface level, generally suggests that the amyloid beta protein accumulates in the brains of patients in the form of brain cell-killing fibers. The protein accumulates into oligomers, then into clusters, mats, and plaques, until it kills the brain cell. Cell-to-cell communication is disrupted (causing symptoms of neurodegenerative impairment) and brain cells are killed (causing movement impairment). As the hypothesis goes, one can then only clear a flooded apartment (the patient's brain) by both/either fixing the leaking pipe (prevent the accumulation of the protein) and opening the clogged drain (clear the already accumulated protein). *Beta-amyloid and the amyloid hypothesis*, Alzheimer's Association (2017), https://www.alz.org/national/documents/topicsheet_betaamyloid.pdf. *But see* Pam Belluck, *Why Didn't She Get Alzheimer's? The Answer Could Hold a Key to Fighting the*

Decision-makers in the legal profession are surely not immune to the effects of aging. With no mandatory retirement age, octogenarians and nonagenarians on the federal bench have doubled over twenty years. Vocal critics in a *ProPublica* report have alleged that dementia increasingly plagues federal judges.[50] The report alleges a variety of instances of forgetfulness and impulsiveness: citing judges who could not remember the route to walk out of their own courtroom, judges who had difficulty reading aloud, judges who seemed to lack memory of previous decisions, and judges who based their decision on non-existent evidence.[51] Interestingly, although judges' health might be scrutinized when they are younger (at the time of their appointment), apparently nothing dictates any medical evaluation for the rest of a judge's career.[52] Of course, age does not affect decision-making in all individuals the same: although information processing speeds tend to decline with age, there exists great variation between individuals in their ability to perform as they advance in years.[53]

4. Impulse Control. Impulsivity—defined as premature action without foresight—is thought to have genetic and neurobehavioral underpinnings, but is by definition difficult to anticipate in the midst of decision-making. Although there are evolutionary advantages to quick and risky responses, impulsivity may at times also be indicative of the psychiatric symptoms of a range of disorders.[54] According to data from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), about 10.5% of the general population is estimated to have an impulse control disorder.[55] Attorneys report higher levels of mental health issues such as depression and anxiety, maladies that are not infrequently self-medicated and exacerbated with alcohol or substance abuse. In a 2017 American Bar Association study, 36.4% of

Disease, N.Y. TIMES (Nov. 4, 2019), https://www.nytimes.com/2019/11/04/health/alzheimers-treatment-genetics.html?action=click&module=Well&pgtype=Homepage&section=Health.

[50] Joseph Goldstein, *Life Tenure for Federal Judges Raises Issues of Senility, Dementia*, PROPUBLICA (Jan. 18, 2011), https://www.propublica.org/article/life-tenure-for-federal-judges-raises-issues-of-senility-dementia.

[51] *Id.*

[52] Francis X. Shen, *Aging Judges*, 81 OHIO STATE L. J. 235, 238 (2020).

[53] *Id.*

[54] Such disorders include drug addiction, alcoholism, intermittent explosive disorder (impulsive and angry outbursts), oppositional defiant disorder (challenge authority figures, flout rules, bother others on purpose), conduct disorder (persistent behavior that violates social rules), kleptomania (theft), and pyromania (deliberately sets fires). T. W. Robbins & J. W. Dalley, *Impulsivity, Risky Choice, and Impulse Control Disorders: Animal Models*, DECISION NEUROSCIENCE 81 (2017).

[55] Harvard Medical School, *12-month prevalence of DSM-IV/WMH-CIDI disorders by sex and cohort*, https://www.hcp.med.harvard.edu/ncs/ftpdir/NCS-R_12-month_Prevalence_Estimates.pdf.

respondents had scores on the Alcohol Use Disorders Identification Test consistent with problematic drinking.[56]

5. Perceptual Inaccuracies. Human decisions are affected by mental models of the environment within which individuals act.[57] These perceptions are created from the interaction of different senses—taste, touch, smell, hearing, sight—and can be distorted through the lens of emotions, motivations, desires, and culture. In a noisy and chaotic world, individuals who lacked a mental model would be overwhelmed by the sheer volume of unfiltered information. Humans have thus developed perceptual filters to make sense of the flood of information: selective attention allows focus on some sensory experiences while tuning out others; sensory adaptation allows de-sensitization to unimportant changes in the environments; perceptual constancy allows a logically consistent perception of objects, despite actual changes in sensation.

Perceptual inaccuracies have been revealed in the laboratory setting as well as in real world tragedies. Approximately two-thirds of accidents on commercial airplane flights are caused by human error.[58] When a plane lands, there are no other distance cues visible, subjecting the pilot to a moon illusion: the city lights beyond the runway appear larger on the retina than they actually are, which can fool a pilot into landing prematurely.[59] Airlines have since instituted safety measures that align with human visual perception: during landing, copilots must call out the altitude progressively during the descent.[60] Not all perceptual limitations are as readily capable of being remedied.[61]

---

[56] *Addiction Recovery Poses Special Challenges for Legal Professionals*, BUTLER CTR. FOR RES. (Mar. 16, 2017), https://www.hazeldenbettyford.org/education/bcr/addiction-research/substance-abuse-legal-professionals-ru-317.

[57] To act, we generate a mental model by registering structural invariants, using assumptions to drive informational elaboration, and functionally organize stimuli. *See* Daniele Zavagno, Olga Daneyko, & Rossana Actis-Grosso, *Mishaps, Errors, and Cognitive Experiences: On the Conceptualization of Perceptual Illusions*, 9 FRONT. HUM. NEUROSCI. 1 (2015).

[58] Raymond S. Nickerson, *Applied Experimental Psychology*, 47 APPLIED PSYCHOL.: AN INT'L REV. 155-73 (1998).

[59] Conrad Kraft, *A Psychophysical Approach To Air Safety: Simulator Studies Of Visual Illusions In Night Approaches*, PSYCHOLOGY: FROM RESEARCH TO PRACTICE (1978).

[60] *Id.*

[61] For instance, civilian casualties through military target misidentifications. The U.S. has a track record of mistakenly killing civilians or allies due to human error. *See* Eric Schmitt & Anjali Singhvi, *Why American Airstrikes Go Wrong*, N.Y. TIMES (Apr. 14, 2017), https://www.nytimes.com/interactive/2017/04/14/world/middleeast/why-american-airstrikes-go-wrong.html.

B. Biases

It is of little surprise that humans do not always act in full accord with perfect rationality.[62] The sheer volume of information available in the world makes it necessary to use cognitive shortcuts. It may also be some shortcuts reflect traits that have given humans evolutionary advantages. But these shortcuts can be subject to systematic errors in information processing. Although the biological, neurochemical, or other physical mechanisms underlying cognitive biases in human judgement remain unclear, these phenomena have been replicated in many studies.[63]

This section details a series of widely documented biases that can predictably affect human decision-making and lead to errors in judgment. These cognitive biases are likely part of any human-based system of decision-making, although the precise decision task will affect how much of a risk any one of these biases will pose to effective decision-making. It may also be possible to countermand some of these tendencies through what is known as debiasing, but not always.[64]

1. Endowment Effect. An individual who owns a particular good tends to value it more than someone who does not. As one study found, participants demanded much more to give up a Cornell University mug than they would be willing to pay to acquire the same mug in the first place.[65] Furthermore, the longer owners possess an item, the more they will tend to value it.[66] Subsequent studies have explored plausible psychological mechanisms, such as the feeling of psychological ownership and possession. Owners—whether of a mug, a potential business deal, or a legislative proposal—are more likely to recall the positive attributes of the possession, focusing on reasons to keep what they already have. Non-owners are more likely to recall the negative attributes of the item in question, focusing on the reasons to keep their money and not to buy in. One result of the endowment effect may be to contribute to inertia toward the status quo, as it takes a great deal of effort to shift resources and perspectives away from those that

---

[62] Kahneman, *supra* note 27.

[63] Milton Friedman, *The Methodology of Positive Economics*, ESSAYS IN POSITIVE ECONOMICS 3, 14-16 (1953) ("Economics should not be judged on whether the assumptions are realistic or valid, but rather on the quality of its predictions.").

[64] Christine Jolls & Cass R. Sunstein, *Debiasing Through Law* 2-4 (Nat'l Bureau of Econ. Research, Working Paper No. 11738, 2005), https://www.nber.org/papers/w11738.pdf.

[65] Daniel Kahneman, Jack L. Knetsch & Richard H. Thaler, *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. OF ECON.PERSPECTIVES 193 (1991).

[66] Michael A. Strahilevitz & George Loewenstein, *The effect of ownership history on the valuation of objects*, 25 J. CONSUM. RES. 276 (1998).

currently possess them. The endowment effect may lead, for example, to friction and obstacles in reaching negotiated agreements.

2. Loss Aversion. Humans dislike losses far more than they like corresponding gains.[67] The effect is twice as large: we tend to take gambles only if the potential amount we could win is more than double the amount we stand to lose.[68] People tend to disregard the potential gains and focus on the losses associated with the activity, because the latter are cognitively "available" regardless of whether the statistical risk is high. Practically speaking, all else equal, this may help explain why "preventing losses looms large in the government's objective function."[69] Nation states are less likely to behave aggressively when doing so would produce gains than when the same behavior might prevent losses.[70] Policymakers and government officials may prefer cautious, preventative measures over aggressive efforts to take a gamble.

3. System Neglect. The system-neglect hypothesis posits that individuals overweigh signals relative to the underlying system which generates the signals.[71] Decisions are effectively made in isolation, as humans tend to neglect the systemic, rippling effects of a single decision. If the environment is stable, humans tend to overreact to forecast errors; if the environment is unstable, people underreact.[72] With 90% of U.S. corporations accessing forecasting software, the bias holds true with financial decisions at stake. This finding suggests that "managerial judgment in forecasting is better suited to unstable environments than to stable ones, so particular emphasis should be placed on automating decision making in stable environments." [73]

4. Hindsight Bias. Past events are easy to chalk up as predictable. It is common to play "Monday morning quarterback," to feel like you "knew it all along," or to view "hindsight as 20/20 vision."[74] People have a tendency to change

---

[67] Daniel Kahneman & Amos Tversky, *An Analysis of Decision Under Risk*, 47 ECONOMETRICA (1979).

[68] *Id.*

[69] Caroline Freund & Çağlar Özden, *Trade Policy and Loss Aversion*, 98 AM. ECON. REV. 1675 (2008).

[70] Robert Jervis, *Political Implications of Loss Aversion*, 13 POL. PSYCHOL. 187 (1992).

[71] Cade Massey & George Wu, *Detecting Regime Shifts: The Causes of Under- and Overreaction*, 51 MGMT. SCI. 932 (2005).

[72] Mirko Kremer, Brent Moritz, & Enno Siemsen, *Demand Forecasting Behavior: System Neglect and Change Detection*, 57 MGMT. SCI. 1827 (2011).

[73] *Id.*

[74] Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571 (1998).

their estimates of probabilities of past events when considering an event retroactively.[75] Faced with the results—of a scientific study, a football game, or a political election—spectators are prone to believe the result was obvious all along.[76] Experts are equally fallible to this bias: when asked to assess the probabilities of alternative diagnoses, given a set of symptoms, professional physicians offer significantly different estimates depending on what they are told the actual diagnosis turned out to be.[77]

Overoptimism is closely linked to hindsight bias. Overoptimism describes the tendency to think that bad events are far less likely to happen to oneself than to others. For instance, overoptimism can manifest when juries evaluate the risk of harm or when companies predict the success of certain internal and external strategies. One study found a negative correlation between startup founders' level of optimism and the performance of their new ventures.[78] Debiasing remedies appear to have little or no effect on reducing these biases.[79]

5. Availability bias. The availability heuristic or bias describes the assumption that the examples which come to mind easily are also the most important or prevalent things.[80] This bias is exacerbated by increased availability and volume of news or other information.[81] When a hazard is particularly salient or frequently observed, the hazard is more cognitively "available" and drives collective decision-making. In the context of legislation and agency decision-making, policy decisions will inevitably become anecdote-driven if preferences are shaped by a set of probability judgments that are themselves riddled with the spotlight effect. For instance, support for environmental legislation can be driven by recent and memorable instances of harm, such as explosions or fires.

---

[75] Baruch Fischhoff, *Hindsight is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXP. PSYCH. 288 (1975).

[76] Benedict Carey, *That Guy Won? Why We Knew It All Along*, N.Y. TIMES (Oct. 29, 2012), https://www.nytimes.com/2012/10/30/health/he-won-the-election-i-knew-it-all-along.html.

[77] Hal R. Arkes, David Faust, Thomas J. Guilmette & Kathleen Hart, *Eliminating the Hindsight Bias*, 73 J. OF APPLIED PSYCHOLOGY 305, 306 tbl.1 (1988).

[78] Keith Hmieleski & Robert Baron, *Entrepreneurs' Optimism and New Venture Performance: A Social Cognitive Perspective*, 52 ACAD. MGMT. J. 473 (2009).

[79] Martin F. Davies, *Reduction of Hindsight Bias by Restoration of Foresight Perspective: Effectiveness of Foresight-Encoding and Hindsight-Retrieval Strategies*, 40 ORGANIZATIONAL BEHAV. & HUMAN DECISION PROCESSES 61 (1987); Baruch Fischhoff, *Perceived Informativeness of Facts*, 3 J. EXPERIMENTAL PSYCHOL. 349 (1977).

[80] Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974).

[81] Thomas Gilovich, Victoria H. Medvec, & Kenneth Savitsky, *The Spotlight Effect In Social Judgment: An Egocentric Bias in Estimates of the Salience of One's Own Actions and Appearance*, 78 J. OF PERSONALITY & SOC. PSYCHOL. 211 (2000).

Available stories also tend disproportionately to be success stories—news outlets report largely on the startups darlings that have succeeded; military consultants only analyze war planes that have made it home.[82] If only successes are observed, then the factors crucial to tipping a failure into a success, or vice versa, will be overlooked.

6. Confirmation Bias. Confirmation bias is the tendency to search for and favor information that confirms existing beliefs, while simultaneously ignoring or devaluing information that contradicts them.[83] Sometimes this bias is referred to as "motivated reasoning."

In one study, the participant pool consisted half of individuals who supported capital punishment and half who did not.[84] Both groups were given the same two fictional studies—one supporting their views, one rejecting them—both backed with substantive evidence. Yet the participants merely ignored the inconvenient information and focused on the confirming their initial positions. Neither group's initial views changed.[85]

Civil servants are not immune to motivated reasoning. In a recent Danish study, elected politicians were showed the characteristics of two schools, and asked to choose the best-performing one.[86] When the options were labelled anonymously (e.g. "School A" and "School B"), they answered correctly; when the options revealed privatization, a contentious issue in Danish politics (e.g. "Private School" and "Public School"), the results changed dramatically. Counterintuitively, politicians performed worse when they were given more information, which allowed them to cherry-pick evidence that supported their pre-existing beliefs and entrenched values. Once decision-makers establish an initial position—perhaps by making a public commitment through a speech or other statement—they are less likely to make use of new evidence that might depart from their staked-out views.

Research also suggests that as humans acquire domain expertise, they can lose flexibility with regard to problem solving, adaptation, and creative idea and generation.[87] In other words, experts can become cognitively entrenched.

---

[82] DAVID EPSTEIN, RANGE: WHY GENERALISTS TRIUMPH IN A SPECIALIZED WORLD (2019).

[83] Charles G. Lord, Lee Ross & Mark R. Lepper, *Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence*, 37 J. OF PERSONALITY & SOC. PSYCHOL. 2098 (1979).

[84] *Id.*

[85] *Id.*

[86] Martin Baekgaard et al., *The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians*, 49 BRITISH J. POL. SCI. 1117 (2019).

[87] *See generally* Erik Dane, *Reconsidering the Trade-off Between Expertise and Flexibility: A Cognitive Entrenchment Perspective*, 35 ACAD. MGMT. REV. 579 (2010).

7. Framing. Psychologists have used prospect theory to propose that framing changes the evaluation of risks. For example, if a health policy is framed in terms of number of lives saved, people are more conservative and risk-averse; if the same policy is framed in terms of number of lives lost, people are much more willing to take risks for the opportunity to reduce that number.[88] The choice of framing affects decision-making toward risk, even when the situations are quantitatively identical. On the one hand, gain frames are more effective for disease prevention and treatment interventions, because these are perceived as "non-risky" in the sense they are preventing disease or returning health to normal. On the other hand, detection behaviors, such as cancer screening, are perceived as "risky," at least in the short-term, because they may discover early signs of a health problem and loss frames should be more persuasive for these types of risk-seeking behaviors.

8. Anchoring. Decisions are also shaped by anchored values.[89] People make estimates of unknowns by modifying an initial value—whether explicitly given or implicitly in the subconscious—to yield the final answer. Although anchoring typically affects decisions in negotiation, it also plays a part in how voters evaluate the costs of local government programs. One study found that when asked how much they believed a typical referendum question would raise their property taxes, the majority of participants anchored their estimate according to the number embedded in the question itself (higher estimates in response to "$50,000,000" of financing, as opposed to "$130 per capita" of financing).[90]

9. Susceptibility to Overpersuasion. Studies reveal cognitive biases in the legal system to be inherent in language, gruesome text, rhetorical devices, and PowerPoint presentations, among others.[91] In one study, cognitive psychologist Elizabeth Loftus found that extreme word choice influences recall of a car accident.[92] Research participants were shown a video of a single car accident and

---

[88] Alexander J. Rothman & Peter Salovey, *Shaping Perceptions To Motivate Healthy Behavior: The Role of Message Framing*, 121 PSYCHOLOGICAL BULLETIN 3 (1997).

[89] Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974).

[90] Kenneth A. Kriz, *Anchoring and Adjustment Biases and Local Government Referenda Language* (working paper), https://www.ntanet.org/wp-content/uploads/proceedings/2014/078-kriz-anchoring-adjustment-biases-local-government.pdf.

[91] Alicia Lai, *Brain Bait: Effects of Cognitive Biases on Scientific Evidence in Legal Decision-Making* (2018) (unpublished A.B. thesis, Princeton University) (on file with the Princeton University Library) (discussing the over-persuasiveness of scientific jargon and images).

[92] Elizabeth F. Loftus & John C. Palmer, *Reconstruction of Automobile Destruction: An example of the Interaction Between language and memory*, 13 J. OF VERBAL LEARNING AND VERBAL BEHAVIOR 585 (1974).

questioned, "How fast the cars were going when they ____ each other?" with one of the interchangeable verbs "smashed", "collided", "bumped", "hit", and "contacted." Results revealed that higher intensity of the chosen verb correlated with a higher estimate of speed.[93] Likewise, jurors' perceptions of events may be shaped simply with word choice. Even subtler grammatical choices may influence recall and testimony.[94] Loftus had participants respond to a video of a car accident. Some participants had questionnaires with indefinite articles ("Did you see *a* broken headlight?") and the others with definite articles ("Did you see *the* broken headlight?"). The latter responded with fewer unsure answers and increased recognition of events that did not actually occur. While the question asks for simple recall, the definite article implies that a broken headlight exists in the first place.

Another study found that gruesome text influences conviction rates.[95] 34% of the subjects who viewed gruesome textual evidence chose to convict, whereas 14% of those who did not view gruesome textual evidence did. The researchers hypothesize that gruesome text evokes visceral disgust, triggering withdrawal and feelings of moral and social unfairness, and leading to higher rates of conviction. Even in studies where disgust was delivered through unsavory smells in the room, subjects misattributed the disgusting smell to their conviction decision.[96]

Visual evidence may be even more influential: diagrams, photographs, and animations can evoke emotional states, hold juror attention, and easily elucidate complex concepts. Courts have begun to recognize the potentially prejudicial nature of visual advocacy. The use of PowerPoint slides in opening statements have been found to directly correspond to an increase in decisions for liability if the presentation is made by the prosecution, and vice versa.[97]

10. Implicit Racial and Gender Biases. Like the various cognitive biases noted above, race and gender biases can be implicit in that they imperceptibly affect human judgment. Calling attention to the existence of these widespread human

---

[93] *Id.* (resulting in average estimates of 40.5 mph for "smashed", 39.3 mph for "collided", 38.1 mph for "bumped", 34.0 mph for "hit", and 31.8 mph for "contacted").

[94] Elizabeth F. Loftus & Guido Zanni, *Eyewitness testimony: The influence of the Wording of a Question*, 5 BULLETIN PSYCHONOMIC SOC'Y 86 (1975).

[95] Gruesome text influences conviction rates. *See* David A. Bright & Jane Goodman-Delahunty, *Gruesome evidence and emotion: anger, blame, and jury decision-making*, 30 LAW & HUMAN BEHAV. 183 (2004); Beatrice H. Capestany & Lasana T. Harris, *Disgust and Biological Descriptions Bias Logical Reasoning During Legal Decision-Making*, 9 SOCIAL NEUROSCIENCE 265 (2014).

[96] Nicolao Bonini et al., *Pecunia olet: The Role Of Incidental Disgust In The Ultimatum Game*, 11 EMOTION 965 (2011).

[97] Jaihyun Park & Neal Feigenson, *Effects of a Visual Technology on Mock Juror Decision Making*, 27 APPLIED COGNITIVE PSYCHOL. 235 (2012). See, e.g., *In re Pers. Restraint of Glasmann*, 286 P.3d 673 (Wash. 2012); *State v. Robinson*, No. 47398-1-I, 2002 WL 258038, at *2 (Wash. Ct. App. Feb. 25, 2002).

biases "is not a new way of calling someone a racist."[98] Rather, it is to call attention to a "distorting lens that's a product of both the architecture of our brain and the disparities in our society."[99] The architecture of the mind in interaction with socialization may contribute to such biases. One study showed that human infants—exposed to very few faces and voices—are able to interpret foreign languages and differentiate individual monkey faces.[100] However, as the infants are socialized within their families and other social environments, they lose this ability.[101] Their perceptions become affected by their surroundings—and the subtle cues about race that appear in their environment. Another study exposed adult subjects to a series of flashes of light that contained letters that could not be consciously perceived. One randomly assigned group of subjects was exposed to flashes that made up words related to crime, such as "arrest" and "shoot," while the other group was exposed to jumbled letters. These flashes occurred at speeds of 75 milliseconds—too rapid for anyone even to know that they were being shown letters. Immediately after being exposed to these flashes, subjects were shown two human faces simultaneously—one a black face, one a white face. The subjects exposed to the crime-related words spent more time staring at the black face.[102]

Quantitative studies of racial biases in the legal system have gained significant attention over the last decade. Studies show evidence of racial bias in the influence of prosecutors over convictions[103] and federal sentences,[104] as well as the influence of defense attorneys,[105] police officers,[106] judges,[107] and juries.[108]

---

[98] JENNIFER L. EBERHARDT, BIASED: UNCOVERING THE HIDDEN PREJUDICE THAT SHAPES WHAT WE SEE, THINK, AND DO 6 (2019).

[99] *Id*.

[100] Olivier Pascalis et al., *Plasticity of Face Processing In Infancy*, 102 PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES 5297 (2005).

[101] *Id.*

[102] EBERHARDT, supra note 98, at 58-60.

[103] Carly W. Sloan, *Racial Bias by Prosecutors: Evidence from Random Assignment* (2019).

[104] Marit Rehavi & Sonja B. Starr, *Racial Disparity in Federal Criminal Sentences*, 122 J. POL. ECON. 1320 (2014).

[105] David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145 (2010).

[106] Kate Antonovics & Brian G. Knight, *A New Look at Racial Profiling: Evidence from the Boston Police Department*, 91 REV. OF ECON.AND STATISTICS 163 (2009).

[107] Briggs Depew et al., *Judges, Juveniles, and In-Group Bias*, 60 J. OF LAW AND ECONOMICS 209 (2017).

[108] Shamena Anwar, Patrick Bayer, & Randi Hjalmarsson, *The Impact of Jury Race in Criminal Trials*, 127 Q.J. ECON. 1017 (2012).

C. Problems with Group Decision-Making

Decisions made by groups are also plagued by flawed pathologies. At the forefront of the list is groupthink, a psychological drive for consensus at any cost that suppresses dissent and appraisal of alternatives in cohesive groups.[109] The striving for unanimity overrides their motivation to realistically appraise alternative courses of action. Groupthink contributed to NASA's decision to launch the starcrossed Challenger, Truman's invasion of North Korea, Kennedy's Bay of Pigs fiasco, Johnson's escalation of the Vietnam War, Nixon's Watergate break-in, and Reagan's Iran-Contra scandal coverups.[110]

Another group decision-making pathology is the lowest common denominator effect, meaning that the most restricting commonality controls the decision. As a marketing problem, firms use impersonal, generalized advertisements that are tested to appeal to the greatest number of people. As a management problem, organizations are forced to set the bar based on the ability of the worst of the organization, and devalue the most productive of the workforce. As a political problem, local governments should assume responsibility over problems where it is the lowest unit of government with jurisdiction over the majority of people with that problem.[111]

There are also limitations in how the group structure aggregates individual inputs. Organizational behaviorists have characterized group decision-making as a "garbage can," one in which participants often identify solutions first and then go in search of problems which might justify the preferred solutions.[112] Whether useful choice opportunities are generated within a group depends upon the mixture, the collection speed, the removal speed, and the availability of other problems and solutions generated.[113] According to Arrow's Impossibility Theorem, the aggregation of preferences within a group can be intrinsically difficult if individual

---

[109] IRVING JANIS, VICTIMS OF GROUPTHINK (1972).

[110] Irving Janis, *Groupthink*, A FIRST LOOK AT COMMUNICATION THEORY 235-46 (E. Griffin ed., 1991).

[111] For discussion of ways that the lowest common denominator effect (and other pathologies of group decision-making) can lead to problems in administrative and regulatory decision-making, see Cary Coglianese, *Is Consensus an Appropriate Basis for Regulatory Policy?*, *in* ENVIRONMENTAL CONTRACTS: COMPARATIVE APPROACHES TO REGULATORY INNOVATION IN THE UNITED STATES AND EUROPE 93 (Eric Orts & Kurt Deketelaere eds., 2001), and also see Cary Coglianese, *Is Satisfaction Success? Evaluating Public Participation in Regulatory Policymaking*, *in* THE PROMISE AND PERFORMANCE OF ENVIRONMENTAL CONFLICT RESOLUTION 69 (Rosemary O'Leary & Lisa Bingham eds., 2003).

[112] Michael D. Cohen et al., *A garbage can model of organizational choice*, 17 ADMIN. SCI. Q. 1 (1972).

[113] *Id.*

preferences are arrayed across more than a single dimension. Economist Kenneth Arrow has shown that in such circumstances there can be no clear decision rule, even majority voting, that rationally aggregates individual preferences.[114]

D. Implications for Decision-Making in Government

The various limitations associated with human decision-making manifest themselves in common complaints about governmental performance. Human-based governmental processes are frequently criticized for delays, inconsistencies, and disparities, and concerns about racial, gender, and other biases predominate discussions of the fairness of governmental decision-making.[115]

The faultiness of human decision-making perhaps may be most salient in the context of national security and military warfare. Public policy scholars have vehemently expressed wariness and opposition to any use of AI in matters where risks can be international, irrevocable, and fatal.[116] But the argument against lethal autonomous systems presupposes that lethal human decision systems are a superior method of conducting warfare. Unfortunately, current weaponry, military measures, and human biases are far from infallible. In Kunduz, Afghanistan, for example, senior officials reportedly approved an American Special Forces gunship to open fire upon a Doctors Without Borders hospital, continuing even after the doctors notified the American military, killing 42 people.[117] In Belgrade, Serbia, CIA analysts apparently mistook the wrong address and bombed the Chinese Embassy, killing 3 people.[118] Mistakes like these are rooted in human error—often in situations with great uncertainty, stress, and time pressures. Even in domestic policy circumstances with less taxing or pressured conditions for making decisions, the physical limitations, cognitive biases, and group pathologies highlighted above

---

[114] Kenneth J. Arrow, *A Difficulty in the Concept of Social Welfare*, 58 J. POL. ECON. 328 (1950).

[115] *Cf. generally* Lucie E. White, *Subordination, Rhetorical Survival Skills, and Sunday Shoes: Notes on the Hearing of Mrs. G.*, 38 BUFFALO L. REV. 1 (1990).

[116] *See, e.g.*, David Nield, *This Horrifying 'Slaughterbot' Video Is The Best Warning Against Autonomous Weapons*, SCI. ALERT (Nov. 22, 2017), https://www.sciencealert.com/chilling-drone-video-shows-a-disturbing-vision-of-an-ai-controlled-future (publishing a video campaign by the Campaign to Stop Killer Robots where autonomous drones break free of human control to independently identify, pursue, and assassinate human targets).

[117] Gregor Aisch et al., *How a Cascade of Errors Led to the U.S. Airstrike on an Afghan Hospital*, N.Y. TIMES (Apr. 29, 2016), https://www.nytimes.com/interactive/2015/11/25/world/asia/errors-us-airstrike-afghan-kunduz-msf-hospital.html?_r=0.

[118] Steven L. Myers, *Chinese Embassy Bombing: A Wide Net of Blame*, N.Y. TIMES (Apr. 17, 2000), https://www.nytimes.com/2000/04/17/world/chinese-embassy-bombing-a-wide-net-of-blame.html.

–either individually or in combination—can lead to poor governmental performance.[119]

II. Machine Learning's Promise for Improving Governmental Decision-Making

Recognizing the limitations of human decision-making does not automatically mean that digital systems based on machine learning will always perform better than current human-based systems. It also cannot be said that machine learning tools will be infallible either. But recognition of human foibles does indicate that room for improvement exists in current tasks performed by humans. Government administrators and system designers ought to be open to the possibility that machine learning tools could help overcome some of the limitations of human decision-making.

Government performs important responsibilities in domestic administration by helping provide social services and enforce rules in the service of promoting social welfare. It is now increasingly realistic to imagine a future where, seeking to fulfill these responsibilities, government agencies can develop sophisticated systems to help it identify those applicants who qualify for support. But it is also possible to imagine further that, in the end, such a system could turns out to award benefits arbitrarily, or to prefer white applicants over black applicants. Such a machine-based system would be properly condemned as unfair. It is exactly such outcomes that worry those who oppose the use of artificial intelligence in administering social programs.

Yet we need not merely imagine such a system developing that would have such inconsistent and unfair outcomes. That system actually was adopted decades ago in the United States and other countries—and remains in use to this day. The "technology" underlying that current system is not digital, but human. The U.S. Social Security Administration's (SSA) disability system, for example, relies on more than a thousand human adjudicators. Although most of these officials are no doubt well-trained and dedicated, they also work under heavy caseloads. And they are human.

Any system that relies on thousands of human decision-makers working at high capacity will surely yield variable outcomes. A 2011 report issued by independent researchers offers a stark illustration of the potential for variability across humans: among the fifteen most active administrative judges in a Dallas SSA

[119] For recent discussions of governmental performance, see FRANCIS FUKUYAMA, POLITICAL ORDER AND POLITICAL DECAY: FROM THE INDUSTRIAL REVOLUTION TO THE GLOBALIZATION OF DEMOCRACY 548 (2014), BO ROTHSTEIN, THE QUALITY OF GOVERNMENT (2011), and PETER H. SCHUCK, WHY GOVERNMENT FAILS SO OFTEN—AND HOW IT CAN DO BETTER 30 (2014).

office, "the judge grant rates in this single location ranged … from less than 10 percent being granted to over 90 percent."[120] The researchers reported, for example, that three judges in this office awarded benefits to no more than 30 percent of their applicants, while three other judges awarded to more than 70 percent.[121] Other studies have suggested that racial disparities exist in SSA disability awards, with certain black applicants tending to receive less favorable outcomes compared with white applicants.[122]

In light of reasonable concerns about arbitrariness and bias in human decisions, the relevant question to ask about artificial intelligence is not whether it will be free of any bias or unexplainable variation. Rather, the question should be whether artificial intelligence can perform better than the current human-based system. Anyone concerned about fairness in government decision-making should entertain the possibility that digital algorithms might sometimes prove to be fairer and more consistent than humans. At the very least, it might turn out to be easier to remedy biased algorithms than to remove deeply ingrained implicit or cognitive biases from human decision-making.[123]

A. What Makes Machine Learning Different?

An algorithm is simply a set of clear, instructional steps designed to solve a problem. A cookbook recipe is an algorithm. The process of efficiently folding a shirt or a piece of origami is an algorithm. Algorithms are not unique to the computer age; they have been part of human societies for millennia.

But today, modern computing power allows businesses and governments to take advantage of a distinctive type of algorithm known as a machine-learning algorithm. Machine learning is a subspecies of artificial intelligence involving algorithms that learn autonomously by deciphering patterns and generating inferences in large datasets that contain images, numbers, dense text, and natural languages.[124] Machine-learning algorithms can assume multiple forms. In "supervised learning," the algorithm is provided with numerous labeled examples—e.g. images categorized as "dog" or "cat"—and must then develop an algorithmic model to accurately identify unlabeled images of dogs and cats. In "unsupervised learning," the algorithm must learn how to differentiate between

---

[120] *Social Security Awards Depend More on Judge than Facts*, TRAC (July 4, 2011), https://trac.syr.edu/tracreports/ssa/254/. The Social Security sharply disputed aspects of this study.

[121] *Id*.

[122] *See, e.g.*, U.S. GEN. ACCOUNTING OFFICE, B-247327, RACIAL DIFFERENCE IN DISABILITY DECISIONS WARRANTS FURTHER INVESTIGATION (1992), https://www.gao.gov/assets/160/151781.pdf; Erin M. Godtland et al., *Racial Disparities in Federal Disability Benefits*, 25 CONTEMP. ECON. POL'Y 27 (2007).

[123] *See generally* MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM (2019).

[124] *See* Coglianese & Lehr, *supra* note 1; Lehr & Ohm, *supra* note 1.

images of dogs and cats, for example, without the benefit of learning from labeled data. As the algorithm is fed an increasing number of images of dogs and cats, the algorithm builds predictive models for how to reliably distinguish between the two.[125]

Unlike traditional statistical analysis techniques, machine learning does not require humans to specify at the outset which variables to use.[126] While humans are not completely out of the machine-learning loop—humans must still select the machine-learning's meta-algorithm, feed the algorithm its data, and tweak the algorithm's optimization process for analyzing "test data"—machine-learning algorithms largely design their own predictive models through autonomous learning. This process of self-education happens in a black box. From the simplest "random forests"[127] machine-learning algorithms to the most intricate "deep learning" neural networks,[128] it is virtually impossible for humans to decipher how the algorithm developed its predictive model. In fact, these models are often so complex that the algorithm's original designers cannot predict how they will behave.[129]

As the amount of data generated on a daily basis has exponentially increased in the last decade, a subfield of machine learning—deep learning—has grown in popularity. Deep learning algorithms are made of layers of computational neural networks loosely inspired by the human brain. These neural networks, which thrive on massive amounts of unstructured data (Facebook deep learning algorithms, for example, process more than one billion photos) and can be very expensive to run, are comprised of an input layer where data is fed and an output layer where final predictions are generated. Numerous hidden layers exist between these input and output layers. Neurons in each hidden layer communicate with each other just as they do in the human brain. As data is passed between layers, subsequent layers grow more intelligent and are able to process more sophisticated features of the data. After numerous training cycles, the model gradually learns which links between neurons are critical in generating accurate predictions and prioritizes those pathways by altering the values or weights ascribed to each neuron. Deep learning is very effective at determining how abstract forms like human faces are generated from combinations of discreet characteristics embedded in the data.[130]

Deep learning is behind a number of recent technological achievements including predictive maintenance warnings from Internet of Things (IoT) sensor

---

[125] Coglianese & Lehr, *supra* note 1.

[126] Lehr & Ohm, *supra* note 1.

[127] Leo Breiman, *Random Forests*, 45 MACHINE LEARNING 5, 5–6 (2001).

[128] Lehr & Ohm, *supra* note 1.

[129] Coglianese & Lehr, *supra* note 1.

[130] Nick Heath, *What is Deep Learning? Everything You Need to Know*, ZDNET (Aug. 7, 2018), https://www.zdnet.com/article/what-is-deep-learning-everything-you-need-to-know/.

data, virtual assistants like Siri and Alexa, automatic translation between languages, machine vision for autonomous cars and drones, chatbots and servicebots, image colorization, facial recognition, disease and tumor identification in x-rays and other scans, personalized medicine based on one's genome, and personalized shopping recommendations on websites like Amazon.[131] Natural language processing has also greatly benefitted from advances in deep learning.[132] Federal agencies are increasingly interested in harnessing the enormous power of deep learning algorithms.[133]

Machine-learning algorithms have the potential to vastly improve upon human decision-making. Machine-learning algorithms used by retail businesses, for example, can make exceedingly accurate judgments about what new products a particular consumer might want to purchase based simply on an analysis of their past purchases. Machine-learning algorithms used by government can be fed a variety of input variables—satellite photos, municipal complaint calls, and sensor readings—and generate confident forecasts about outputs—say, an oil spill, the location of a local disturbance, or a water pipeline leak. The downside of such analytics is that it is not easy to explain exactly how and why the algorithm reached its particular prediction. These algorithms do not generally support causal claims, nor do they allow a government official to provide a compelling reason for the precise output obtained.[134]

B. Machine Learning's Advantages in Private-Sector and Medical Decision-Making

The private sector has long begun to reap the benefits of algorithms, delivering promising results across industries such as medicine, finance, online search, marketing, and autonomous vehicles. These applications have revealed several key benefits, including increased *accuracy*, greater *capacity* to analyze volumes of data, faster computation *speeds*, and more *consistent* outcomes across cases. These benefits might even be said to be inherent to digital algorithms in

---

[131] Bernard Marr, *What is Deep Learning AI? A Simple Guide with 8 Practical Examples*, FORBES (Oct. 1, 2018), https://www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/#7deb288f8d4b.

[132] Yoav Goldberg, *A Primer on Neural Network Models for Natural Language Processing* 57 J. OF ARTIFICIAL INTELLIGENCE RESEARCH (2016).

[133] Phil Goldstein, *What Is Deep Learning? A Look at Machine Learning in Federal IT Environments*, FEDTECH (Jul. 31, 2019), https://fedtechmagazine.com/article/2019/07/what-deep-learning-look-machine-learning-federal-it-environments-perfcon.

[134] Sometimes machine-learning analysis can be incorporated into, and assist with, broader analysis of causal connections. For related discussion, see Sendhil Mullainathan & Jann Spiess, *Machine Learning: An Applied Econometric Approach*, 31 J. ECON. PERSP. 87 (2017).

certain respects. They can, in some applications, offer a marked improvement over the corresponding characteristics of human-based systems.

1. Accuracy. Algorithms produce accurate results based on clear directives. By definition, algorithms consist of logical steps and equations; mathematical equations dutifully carry out rules created for them and produce outputs that are within those bounds. Accuracy will be assessed by several metrics, depending on the type of algorithm and type of task:

- Classification accuracy is the ratio of number of correct predictions to the total number of input samples.
- Logarithmic loss is calculated by penalizing false classifications.
- Confusion matrixes describes the complete performance of the model in a matrix of true positive, true negatives, false positives, and false negatives.
- Area under the curve compares true positive rate and false positive rate.
- F1 score measures precision and robustness.
- Mean absolute error measures how far the predictions were from the actual output.[135]

Each of these metrics are expressed clearly and numerically. These metrics are most useful for assessing the accuracy of non-causal[136] predictions by algorithms, such as identifying clusters, outliers, or associations within a population.[137]

---

[135] Aditya Mishra, *Metrics to Evaluate your Machine Learning Algorithm*, MEDIUM (Feb. 24, 2018), https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234.

[136] Kalev Leetaru, *Deep Learning And The Limits Of Learning By Correlation Rather Than Causation*, FORBES (Apr. 18, 2019), https://www.forbes.com/sites/kalevleetaru/2019/04/18/deep-learning-and-the-limits-of-learning-by-correlation-rather-than-causation/amp/?utm_campaign=A.I.%20%26%20Blockchain%20-%20Better%20Together%21&utm_medium=email&utm_source=Revue%20newsletter.

[137] *See, e.g.*, *Toxicity Forecasting*, U.S. ENVTL. PROTECTION AGENCY (2019), https://www.epa.gov/sites/production/files/2019-01/documents/toxcast_factsheet_dec2018.pdf (describing the EPA's use of automated screen assays to detect toxic chemicals as a categorical identification problem); *Grid Modernization and the Smart Grid*, DEP'T ENERGY, https://www.energy.gov/oe/activities/technology-development/grid-modernization-and-smart-grid (describing the Smart Grid Advisory Committee and Federal Smart Grid Task Force's task to optimize energy usage at the Department of Energy as a problem of optimizing flow, which has high accuracy). *See also* Deborah Gage, *Big Data Uncovers Some Weird Correlations*, WALL ST. J. (Mar. 23, 2014), https://www.wsj.com/articles/SB10001424052702303369904579423132072969654 ("Correlation … is different than causality. Finding surprising correlations has never been easier, thanks to the flood of data that's now available.").

Even the simplest algorithm can outperform supposed experts at predicting important outcomes.[138] For clinical diagnoses, state-of-the-art machine-learning classifiers are more accurate than human experts—including board-certified dermatologists—in diagnosing pigmented skin lesions from random dermatoscopic images.[139] In mortgage lending, automated underwriting algorithms "more accurately predict default" than human underwriters do, and "this increased accuracy results in higher borrower approval rates, especially for underserved applicants."[140] For each of these complex systems, human experts have tended to produce less accurate decisions.

Moreover, as various research groups seek to develop algorithms for similar tasks, the clear error rates of each serve as a useful indicator of accuracy, as algorithms compete with one another to be used in business and government. For instance, a survey of the 1,879 machine learning algorithms for breast cancer risk prediction clearly reveals one algorithm as the most accurate of the group, earning the endorsement of the research group.[141] Algorithmic accuracy may be further enhanced with technological improvements to ensure system performs as intended: without failure and within specified performance limits.[142]

2. Capacity. The high data capacity of algorithms affords the ability to comprehensively analyze volumes of data and to give each variable its proper consideration. Algorithms have the capacity to handle as many variables as their processing power allows. Algorithms typically face two main constraints: memory and processing speed. A modern computer typically has 16 gigabytes RAM—

---

[138] R. M. Dawes, *The Robust Beauty Of Improper Linear Models In Decision Making*, 34 AM. PSYCHOLOGIST 571 (1979) (describing statistical techniques that can test the accuracy and reliability of a machine learning algorithm, such as "leave-one-subject-out cross-validation").

[139] P. Tschandl, *Comparison Of The Accuracy Of Human Readers Versus Machine-Learning Algorithms For Pigmented Skin Lesion Classification: An Open, Web-Based, International, Diagnostic Study*, 20 LANCET ONCOL. 938 (2019), https://www.ncbi.nlm.nih.gov/pubmed/31201137.

[140] Susan W. Gates et al., *Automated Underwriting In Mortgage Lending: Good News For The Underserved?*, 13 HOUSING POL'Y DEBATE 369 (2010), https://www.tandfonline.com/doi/abs/10.1080/10511482.2002.9521447.

[141] Ricvan D. Nindrea et al., *Diagnostic Accuracy of Different Machine Learning Algorithms for Breast Cancer Risk Calculation: a Meta-Analysis*, 19 ASIAN PAC. J. CANCER PREV. 1747 (2018) (finding that the Super Vector Machine was superior, with an Area Under Curve (AUC) of over 90%).

[142] Other potential fixes include autoencoders, a class of deep learning frameworks, which are a powerful tool in extracting hidden representations and producing a robust reconstruction for further predicting tasks. Results showed that autoencoders can not only classify the states at a good accuracy, but also help to discover the failure mechanism. Peng Jiang, *Unsupervised Deep Learning for Data-Driven Reliability and Risk Analysis of Engineered Systems*, HANDBOOK NEURAL COMPUTATION 417 (2017), https://www.sciencedirect.com/science/article/pii/B9780128113189000235.

allowing for datasets of millions, possibly billions, of data points—more than enough for many algorithmic tasks.[143] Beyond on-site capacity, additional options include stochastic gradient descent, updating through mini-batches, or streaming data over the network.[144]

Machine learning algorithms ultimately feed upon large amounts of data for training and inference. The greater the volume and variety of training data, the more faithful and comprehensive the algorithm's performance. Unlike humans, who are vulnerable to memory limitations when faced with greater than four variables, algorithms have practically unlimited capacity except for the robust storage architecture necessitated by their heavy workloads. Some computer scientists note that even large datasets often lack unobserved variables and data.[145] For instance, in medicine, because we do not deny reliable treatments to subjects, there are no records on those classes of persistently unobservable events. But this complaint is not unique to algorithms: human speculation is also built off unobservable data points, and further subject to the availability heuristic upon recent, salient observations.

3. Speed. Digital algorithms are simply faster than humans at raw computations. Computers return speedier decisions for reasoned, non-reflexive decisions. The runtime analysis denotes how fast an algorithm takes to run, shaping whether it is practically feasible to use.[146]

Humans are better suited to reflexive, reactionary decisions, particularly for sensory inputs. But this is due to its efficiency, not its fundamental speed: the 140 million neurons on each lobe of the visual cortex can handle information processing, but the equivalent informational processing would require over 30 billion transistors on the latest processors, which require significant space and electricity.[147] Thus human decision-making presents the strongest challenge to consumer-facing algorithms (smart home devices, smart cars, etc.), not governmental tasks. A human driver may be able to reflexively respond faster than

---

[143] Håkon Hapnes Strand, *How Do Machine Learning Algorithms Handle Such Large Amounts Of Data?*, FORBES (Apr. 10, 2018), https://www.forbes.com/sites/quora/2018/04/10/how-do-machine-learning-algorithms-handle-such-large-amounts-of-data/#1442c680730d.

[144] *Id.*

[145] Joshua Loftus et al., *Causal Reasoning for Algorithmic Fairness* (May 15, 2018), https://arxiv.org/pdf/1805.05859.pdf.

[146] Runtime is relative to the size of the input. Could be exponential: For example, if the input consists of N integers, an algorithm might have a runtime proportional to N2. Large data sets require efficient algorithms, much less manual human review.

[147] But this may no longer be true. Development of 'neuromorphic' hardware that mimics the human brain to run brain-like software more efficiently. Sara Reardon, *Artificial Neurons Compute Faster than the Human Brain*, Nature (Jan. 26, 2018), https://www.nature.com/articles/d41586-018-01290-0.

an algorithm to swerve in the nick of time, but a human analyst will not be able to thoroughly comb through thousands of pages of documents as fast as an algorithm to reach a relatively well-informed decision.

4. Consistency. Uniform algorithmic standards and performance across the system can help achieve consistent results. By definition, digital algorithms approach tasks in a systematic manner—in theory, individual cases are subject to the same standard, although recent commercial usages may foster skepticism about this claim.[148]

C. Current Applications in the Public Sector

Governments in the U.S. are using machine-learning algorithms in a variety of contexts to support administrative decision-making.[149] The federal government relies on machine-learning algorithms to automate tedious, voluminous tasks and to parse through data to extract patterns that even experts could miss. Meteorologists at the National Oceanic and Atmospheric Administration use machine learning to improve forecasts of severe weather events.[150] Chemists at the Environmental Protection Agency use the program ToxCast to help the agency predict toxicities of chemical compounds to further analyze.[151] These new tools stem from efforts in the 1990s to "re-invent government through data-based performance management and oversight."[152] Already, this "new wave of AI technology is exhibiting early signs of transforming how government works."[153]

In a report published by the Administrative Conference of the United States,[154] researchers from Stanford University and New York University expand upon particularly promising use cases of federal agency deployment of AI. AI usage is primarily concentrated among only a handful of the hundreds of agencies, bureaus, and offices at the federal level: these include the Office of Justice Programs, Securities and Exchange Commission, National Aeronautics and Space Administration, Food and Drug Administration, United States Geological Survey, United States Postal Service, Social Security Administration, United States Patent

---

[148] For instance, ApplePay gave different credit limits to a husband-wife pair who filed joint tax returns, live in a community-property state, and have been married for a long time. David Heinemeier Hansson, (@DHH), TWITTER (Nov. 7, 2019, 3:34 PM), https://twitter.com/dhh/status/1192540900393705474.

[149] Coglianese & Ben Dor, *supra* note 9; ENGSTROM ET AL., *supra* note 9.

[150] *See* Coglianese & Lehr, *supra* note 1, at 1162.

[151] *Id.* at 1162-63.

[152] David F. Engstrom & Daniel E. Ho, *Artificially Intelligent Government: A Review and Agenda*, BIG DATA L., at *3 (2020).

[153] *Id.* at *1.

[154] *See* ENGSTROM ET AL., *supra* note 9.

and Trademark Office, Bureau of Labor Statistics, and Customs and Border Protection.[155] The report continues on to describe several use cases in detail.

First, for *civil enforcement*, algorithms may enforce agency regulations by "shrinking the haystack" of potential violators to better allocate scarce resources and assist the agency in balancing prosecutorial discretion with accountability. At the SEC, algorithmic enforcement targets fraud in accounting and financial reporting, trading-based market misconduct, insider trading, and unlawful investment advisors; these results are handed off to human enforcement staff who continue to work the cases.[156]

Second, for *law enforcement*, algorithms have been widely adopted, replacing traditional surveillance cameras with AI-powered gunshot detection technology, employing AI-driven automatic license plate readers, and deploying AI-powered predictive policing strategies to identify gang-related crimes. Customs and Border Protection, which straddles the civil and criminal divide, deploys two of the most controversial AI tools to expedite processing at airports and borders: facial recognition and risk prediction.[157] The agency is already using machine-learning algorithms to identify faces at airports when processing arrivals from international flights.[158] These tools can securely process private information, reduce wait times, and allocate scarce resources.

Third, for *formal adjudication*, federal agencies have turned to AI to guarantee that decisions are accurate and consistent in mass adjudication. Some agencies—such as the Social Security Administration, the Office of Medicare Hearings and Appeals, the Department of Labor, and the National Labor Relations Board—employ the vast bulk of formal adjudicatory procedures mandated under the Adminsitrative Procedure Act. Other agencies—such as the Board of Veterans Appeals and the Executive Office of Immigration Review—employ evidentiary hearings under administrative judges (ALJs). Given the enormous caseload of federal agencies, agencies experience a significant backlog of claims: the SSA received more than 2.5 million disability claims, with almost 700,000 appealed to the hearings level. The SSA has explored the use of clustering algorithms for micro-specialization, where adjudicators develop expertise in one area. The office has also used algorithms to identify claimants most likely to be awarded disability benefits to accelerate appeals with predicted likelihoods of success; they have even used natural language processing algorithms to improve the quality of decision

---

[155] Engstrom & Ho, *supra* note 152.

[156] *Id.* at *25.

[157] Engstrom & Ho, *supra* note 152, at *30.

[158] Pam Karlan & Joe Bankman, *Artificial Intelligence and the Administrative State with Guests David Engstrom and Cristina Ceballos*, STAN. LEGAL (Apr. 27, 2019), https://law.stanford.edu/stanford-legal-on-siriusxm/artificial-intelligence-and-the-administrative-state-with-guests-david-engstrom-and-cristina-ceballos/ (interview with David Engstrom).

writing.[159] Algorithms have been crucial to the effort to resolve cases in a reasonable processing time and with consistent grant rates.

Fourth, for *informal adjudication*, agencies also face the challenge of information management when making decisions upon government grants, permits, licenses, and inspections. The Patent and Trademark Office faces the quantitative challenge of tackling a considerable backlog of patent applications, and the qualitative challenge of processing patents that will not later be invalidated or wrongly denied.[160] The PTO has now incorporated AI into the processes of patent classification, patent prior art search, mark classification, and prior mark search. These AI-assisted systems ensure that examiners can focus their scarce time and expertise on other important parts of the informal adjudication process.

Fifth, for *regulatory analysis*, federal agencies must engage in rulemaking to establish legally binding regulations, such as producing standard-setting and guidance documents. The Food and Drug Administration has piloted natural language processing engines for postmarket surveillance of drugs and medical devices based on adverse event reports that contain substantial freeform text in its Federal Adverse Event Reporting System (FAERS) project. The results are used to update rulemaking and guidance, and occasionally to reevaluate an approval decision.[161]

Sixth, for *public engagement*, federal agencies seek to enhance "customer service" interactions such as applying for a passport, and "civic tech" applications such as open data portals and chatbots. For each of the thousands of final rules that federal agencies publish each year, they also seek to track, review, and integrate comments from interested parties. The Consumer Financial Protection Bureau faces a regulatory and administrative challenge of an unprecedented volume of consumer complaints relative to the CFPB's resources and personnel capacity; they have now deployed natural language processing algorithms to automatically analyze text to categorize narratives, identify trends, and predict consumer harm.[162]

Seventh, the U.S. Postal Service has embraced the physical, mobile manifestation of AI by proposing the use of *autonomous vehicles* in mail delivery. The agency has struggled with declining revenues and climbing operating costs. Its autonomous delivery vehicles and autonomous long-haul trucks are anticipated to

---

[159] Engstrom & Ho, *supra* note 152, at *37. *See also* Gerald Ray & Glenn Sklar, *An Operational Approach to Eliminating Backlogs in the Social Security Disability Program*, SSDI SOLUTIONS INITIATIVE 31–34 (June 2009), http://www.crfb.org/sites/default/files/An_Operational_Approach_to_Eliminating_Backlogs_in_the_Social_Security_Disability_Program.pdf.

[160] Engstrom & Ho, *supra* note 152, at *46; *see also* Arti Kaur Rai, *Machine Learning at the Patent Office: Lessons for Patents and Administrative Law*, 104 IOWA L. REV. 2617 (2019).

[161] *Id.* at *53.

[162] *Id.* at *59.

improve productivity and save money on overtime, fuel, and costs associated with collisions.[163]

State and local governments are also using these algorithms.[164] Police departments across the U.S. have take a systematic approach to allocating law enforcement resources through performance metrics and data analysis, including "place-based" and "person-based" predictive policing tools. Officials in Flint, Michigan now benefit from machine-learning predictions to identify priorities for replacing pipes contributing to lead contamination of the city's water supply.[165] New York City uses machine learning to identify potentially unsafe buildings and dispatch building inspectors.[166] Chicago deploys machine learning to forecast outbreaks of vermin and strategically place bait throughout the city.[167] And Los Angeles has implemented a machine-learning system that analyzes traffic patterns and automatically turns its street signals red or green according to a pattern the system determines will most efficiently minimize traffic congestion.[168] Local, state, and federal governments have made noticeable efforts to adopt, or at least study, the potential deployment of algorithmic decision-making.[169]

D. Advantages of Machine Learning in Governmental Decision-Making

As illustrated in the previous section, federal, state, and local governmental entities have already begun to implement algorithmic decision-making in various ways in assisting with domestic public administration. Not only do existing uses reap the benefits of algorithms, but these benefits presumably could continue to accrue with greater expereince in using algorithms in government. Not only must the government keep pace with technological development to learn to become smarter and more efficient than its private counterparts,[170] but machine learning

---

[163] *Id.* at *65.

[164] *See* Coglianese & Ben Dor, *supra* note 9 (manuscript at 21).

[165] *Id.* at 23.

[166] *Id.* at 22; *see also* Brian Heaton, *New York City Fights Fire with Data*, GOV'T TECH. (May 15, 2015), http://www.govtech.com/public-safety/New-York-City-Fights-Fire-with-Data.html.

[167] Linda Poon, *Will Cities Ever Outsmart Rats?*, CITYLAB (Aug. 9, 2017), https://www.citylab.com/solutions/2017/08/smart-cities-fight-rat-infestations-big-data/535407/; Ash Center Mayors Challenge Research Team, *Chicago's SmartData Platform: Pioneering Open Source Municipal Analytics*, DATA-SMART CITY SOLUTIONS (Jan. 8, 2014), http://datasmart.ash.harvard.edu/news/article/chicago-mayors-challenge-367.

[168] Ian Lovett, *To Fight Gridlock, Los Angeles Synchronizes Every Red Light*, N.Y. TIMES (Apr. 1, 2013), http://www.nytimes.com/2013/04/02/us/to-fight-gridlock-los-angeles-synchronizes-every-red-light.html.

[169] Coglianese & Ben Dor, *supra* note 9, at 24.

[170] Cary Coglianese, *Optimizing Regulation for an Optimizing Economy*, 4 U. PA. J.L. & PUB. AFF. 1 (2018).

may also help overcome some of the problems or limitations associated with current human-based systems. The same types of benefits that algorithms offer in the private sector may be extended to the governmental setting.

1. Accuracy. Although many new applications of AI in the public sector have yet to support long-term conclusions about their accuracy, several studies reveal promising results. In the criminal justice system, algorithms can achieve more equitable decisions by more accurately predicting recidivism, reducing jailing rates by 41.9% without increase in crime rates.[171] In the commercial workplace, the U.S. Bureau of Labor Statistics collects data on workplace injuries from 200,000 businesses, and must read and assign each incident a code—for occupation, event, injury, injury location, and injury source—to help the Bureau analyze preventions.[172] Implemented in 2014, the BLS's AI system is able to assign 81% of all codes, proving to be more accurate, on average, than a trained human coder.[173]

2. Capacity. The government sector is particularly vulnerable to struggles with capacity, saddled with personnel shortages, financial limits, and time constraints. Conversely, algorithms are particularly effective in balancing various factors in decision-making in ways that would tax individual human decision-makers. The government has begun to implement algorithmic tools to tackle volumes of data. The IRS uses data mining algorithms to predict fraud and abuse.[174] The FDA plans to develop automated tool for efficient adverse event labeling of medical device safety.[175] The General Service Administration (GSA) has automated "administrative 'cutting and pasting tasks," saving time and effort for the staff.[176]

During patent prosecution and trademark registration, Patent and Trademark Office commissioners comb through over 2 million trademark applications and over 1 million patent applications annually, not to mention the

---

[171] Jon Kleinberg et al., *Human Decisions and Machine Predictions* (Nat'l Bureau of Econ. Research, Working Paper No. 23180, 2017), https://www.nber.org/papers/w23180.

[172] *The Future Has Begun: Using Artificial Intelligence to Transform Government*, IBM CTR. BUS. GOV'T (2018), https://ourpublicservice.org/wp-content/uploads/2018/01/0c1b8914d59b94dc0a5115b739376c90-1515436519.pdf.

[173] *Automated Coding of Injury and Illness Data*, U.S. BUREAU LABOR STAT., https://www.bls.gov/iif/autocoding.htm (last modified Oct. 11, 2019).

[174] David DeBarr & Maury Harwood, *Relational Mining for Compliance Risk* (2004), https://www.irs.gov/pub/irs-soi/04debarr.pdf.

[175] FOOD & DRUG ADMIN., FDA COMMISSIONER'S FELLOWSHIP PROGRAM (2011), https://www.fda.gov/media/83569/download.

[176] Jory Heckman, *How GSA Turned An Automation Project Into An Acquisition Time-Saver*, FED. NEWS NETWORK (Mar. 29, 2018), https://federalnewsnetwork.com/technology-main/2018/03/how-gsa-turned-an-automation-project-into-a-acquisition-time-saver/.

unbounded realm of prior art.[177] Given the sheer volume of patent and trademark applications, the USPTO has pledged to implement core electronic examination tools for document management and searching, noting that its IT system was a "mission-critical enabler for every aspect of its operation."[178] In addition, PTO officers currently rely on simple keyword searches: using the Espacenet tool from the European Patent Office, the TotalPatent software developed by LexisNexis, or the PatSnap patent search.[179] Current search approaches require a significant amount of manual work and time, as they provide limited semantic search options and often fail to return relevant documents. Natural language processing techniques can automatically detect inventions that are similar to the one described in the submitted document.[180] These algorithms permit consideration of the extensive archive of prior art and bolster the quality of agency service.

3. Speed. Human decision-making can result in backlogs and unfair delays. Human personnel need time to work through requisite procedures, manual review, and at times heavy paperwork, often causing a bottleneck to justice. For example, the slow speed of judge-made bail decisions results in a backlog of defendants, who are subjected to pretrial detainment before any substantive conclusion is reached.[181]

Algorithmic processing speed can offer the prospect of improving analysis and decision-making speed, especially in domains where time is of the essence. They can be useful with real-time tracking and reporting, such as in the FDA's use of microbial sources tracking to assess foodborne outbreaks in real-time.[182]

During informal agency rulemaking, government employees receive millions of individual comments from the public in the course of notice and comment procedures. The strengths of AI lie in its capacity to comprehensively analyze volumes of data and variables, and its speed in returning computations for complex decisions. The process of analyzing public comments could be significantly more efficient under the purview of an algorithm: the notice of

---

[177] U.S. PATENT & TRADEMARK OFFICE, FY 2019 UNITED STATE PATENT AND TRADEMARK OFFICE PERFORMANCE 168, 181 (2020), https://www.uspto.gov/sites/default/files/documents/USPTOFY19PAR.pdf.

[178] *Id.* at 19 (pledging to implement "core electronic examination tools for document management and searching").

[179] Lea Helmers et al., *Automating The Search For A Patent's Prior Art With A Full Text Similarity Search*, arXiv:1901.03136v2 (Mar. 4, 2019).

[180] *Id.*

[181] Patrick Liu, Ryan Nunn, & Jay Shambaugh, *The Economics of Bail and Pretrial Detention*, HAMILTON PROJECT (2018), https://www.hamiltonproject.org/assets/files/BailFineReform_EA_121818_6PM.pdf ("The share of jail inmates who are unconvicted is high and has also increased, rising from about half the total jail population in 1990 to 65 percent in 2016.").

[182] FOOD & DRUG ADMIN, FDA COMMISSIONER'S FELLOWSHIP PROGRAM (2011), https://www.fda.gov/media/83569/download.

proposed rulemaking for the 2018 Open Internet Order elicited over 22 million comments.[183] Furthermore, the submission process was disaggregated: comments came from the FCC's Electronic Comment Filing System (ECFS), the "openinternet@fcc.gov" email address, and the recently launched CSV file option for large comment uploads.[184] Even for an efficient agency, any expectation for its human personnel to read and comprehend all 22 million statements would be unrealistic. But a natural language processing algorithm can breeze through the task, functioning as a screener to sort out the hundreds of thousands of fake, identical comments submitted by spambots.[185]

Some agencies have admitted in the past to having insufficient human personnel to handle the volume of data it must analyze and the decisions it must make.[186] Unfortunately, even when it comes to the government hiring humans to perform key decision-making roles, the government can lose top-tier job candidates to the private sector due to the slower pace of the hiring process.[187]

4. Consistency. A concern with consistency underlies the conception of any fair system, from the federal government's desire for uniformity in policy across the states, to the adoption of explicit multi-factor tests and standards in U.S. jurisprudence.[188] A system that uses a consistent approach may also be easier to modify and fix should errors or biases arise.

A single algorithmic system that can replace many different human decision-makers could allow for greater consistency. It is important to keep in mind that public administration consists of many different human actors. In 2011, the U.S. government employed over 2 million individual civil servants across all its functions—and many more government contractors.[189] Naturally, discrepancies

---

[183] David A. Bray, *An Update on the Volume of Open Internet Comments Submitted to the FCC*, FED. COMM. COMMISSION (Sep. 17, 2014), https://www.fcc.gov/news-events/blog/2014/09/17/update-volume-open-internet-comments-submitted-fcc.

[184] *Id.*

[185] Lauren Gambino & Dominic Rushe, *FCC Flooded With Comments Before Critical Net Neutrality Vote*, GUARDIAN (Aug. 20, 2017), https://www.theguardian.com/technology/2017/aug/30/fcc-net-neutrality-vote-open-internet.

[186] U.S. FOOD & DRUG ADMINISTRATION, THE COMMISSIONER'S FELLOWSHIP PROGRAM (Dec. 6, 2017), https://www.fda.gov/media/105686/download.

[187] Eric Katz, *The Federal Government Has Gotten Slower at Hiring New Employees for 5 Consecutive Years*, GOVERNMENT EXECUTIVE (Mar. 1, 2018), https://www.govexec.com/management/2018/03/federal-government-has-gotten-slower-hiring-new-employees-five-consecutive-years/146348/.

[188] Amanda Frost, *Overvaluing Uniformity*, 94 VIRGINIA L. REV. 1568 (2008).

[189] *Data, Analysis & Documentation: Federal Employment Reports*, U.S. OFF.PERSONNEL MGMT. (2011), https://www.opm.gov/policy-data-oversight/data-analysis-documentation/federal-employment-reports/employment-trends-data/2011/december/graphic-presentation-of-federal-civilian-employment/.

can arise between judgments of individual human actors. Unless deliberately modified, algorithms that accept the same inputs and training will be much more likely to produce more consistent outputs.

E. Concerns About the Use of Machine Learning in Government

These advantages are not necessarily inherent to all forms of algorithms or for all applications. An algorithm might not always be more accurate, capable, faster, or consistent than its human counterpart. A recent study meta-analyzes the many studies that claim that diagnostic deep learning algorithms for medical imaging perform better than their human clinician counterparts.[190] The hype of AI includes overpromising headlines such as "Google says its AI can spot lung cancer a year before doctors" and "AI is better at diagnosing skin cancer than your doctor, study finds."[191] However, many studies purporting to support these claims suffer their own methodological limitations, have not always been tested in real-world conditions, and involve at time suboptimal reporting and high risks of bias. Caution and further study will always be prudent antidotes to hype.

As already noted, machine-learning algorithms will not be infallible. They will make their own mistakes, and these may even be mistakes that humans would not make. But the promise of machine learning is fewer mistakes overall. Still, it is important to recognize that shifting from human-based systems to machine-learning ones can raise a variety of concerns when used in governmental setttings, including the *lack of human expertise*, *data storage and processing capacity*, *cybersecurity*, *privacy*, *transparency*, *bias*, and *abuse of power*. These concerns may be much more visible with artificial intelligence—but this may also mean that they can be more easily identified and, subsequently, systematically addressed. With sufficient resources and planning, these concerns will often be able to be successfully confronted. Indeed, researchers and policymakers are already developing solutions to many of the concerns raised by the use of machine-learning tools.

1. Adequate expertise. Perhaps ironically, effective design and deployment of algorithms requires sufficient human capital. This includes the need for analysts and data scientists with technical skills, but also such experts who can appreciate the specific challenges associated with the use of artificial intelligence in government.

---

[190] Myura Nagendran et al., *Artificial intelligence Versus Clinicians: Systematic Review Of Design, Reporting Standards, and Claims of Deep Learning Studies*, 368 BMJ 1 (2020).

[191] *Id.* at 2.

Analysts' and data scientists' expertise and time are needed to tailor and train algorithms to each specific task.[192] Currently, few advances have been made in general AI—the popular notion of general creativity that allows algorithms to autonomously tackle a range of tasks with little human interference.[193] Instead, human data scientists must form and test our foundation of *narrow AI*—that is, algorithms made for specific applications, such as self-driving cars, image recognition, facial recognition. This process of customizing each algorithm to each task can be labor-intensive.

Algorithms must be context-dependent. They must satisfy internal validity—the extent to which the observed results represent the truth in the population studied and, thus, are not due to methodological errors.[194] But then algorithms must also satisfy external validity—whether the study results apply to similar subjects in different settings.[195] Algorithms require time and data to train, to iterate through the data, and to learn and extrapolate patterns.

Even taking into account the likely role of private contractors in designing algorithmic systems, government needs to have sufficient personnel who are technologically sophisticated to deploy AI systems. Yet, there may be reason to wonder whether the government can both "stem the tide of out-flow from the ranks of governmental service" as well as cultivate "a new type of talent in-flow as well, one that brings even greater analytic capacities to the oversight of the optimizing economy."[196] Advancements in AI are expected not just to displace current jobs, but to create new ones for which human workers will need new skills: "more than 120 million workers globally will need retraining in the next three years due to artificial intelligence's impact on jobs."[197] An IBM report states that 67% of executives expect that "advancements in automation technology will require roles and skills that don't even exist today."[198] Private organizations have recognized the need to retrain human workers as a consequence of AI's impact on human jobs.[199]

---

[192] Coglianese, *supra* note 170, at 10.

[193] NATIONAL SCIENCE AND TECHNOLOGY COUNCIL, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016), https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

[194] MALCOLM GLADWELL, TALKING WITH STRANGERS (2019).

[195] *Id.*

[196] Coglianese, *supra* note 170, at 10.

[197] Shelley Hagan, *More Robots Mean 120 Million Workers Need to be Retrained*, Bloomberg (Sep. 6, 2019), https://www.bloomberg.com/news/articles/2019-09-06/robots-displacing-jobs-means-120-million-workers-need-retraining.

[198] Annette La Prade et al., *The Enterprise Guide to Closing the Skills Gap*, IBM (2019), https://www.ibm.com/downloads/cas/EPYMNBJA.

[199] *See, e.g.*, Daniel Newman, *The Digitally Transformed Workforce: How To Upskill And Retrain To Retain Talent*, FORBES (Mar. 11, 2018), https://www.forbes.com/sites/danielnewman/2018/03/11/the-digitally-transformed-workforce-

Some governments have also taken efforts to retrain their workforces: the United States has implemented the AI Initiative to train displaced workers;[200] Singapore adopted the SkillsFuture Initiative to pay for work-skills related courses in 23 industries;[201] the United Kingdom launched the National Retraining Scheme to assist low wage-earners learn about new opportunities and skills needed in the digital age.[202]

Not only has the government struggled to attract top talent away from private industry,[203] candidates with the experience and aptitudes needed for governmental use of AI appear to be in particularly short supply altogether. According to one report, there is a global AI skills shortage based on an analysis of PhD-qualified authors publishing academic papers at world-leading AI conferences.[204] "[T]he skills essential to AI and data science cannot be distilled into a single individual or role within the organization," an IBM Vice President explains. An effective design team "needs statistics experience; it needs science experience; it needs storytelling experience; it needs good visualization experience; it needs a lot of domain experience." Of the shallow pool of AI talent, so few enter government—as opposed to academia or industry—that it is rarely even mentioned in reported statistics on employment in this field.[205] This shortage of AI skills severely limits the pace of AI adoption and systemically affects the ability of governments to realize the full potential of algorithms.[206]

---

how-to-upskill-and-retrain-to-retain-talent/#1d031e551d6f (reporting that AT&T has reduced its product-development life cycle by 40% and accelerated time to revenue by 32% by focusing on retraining).

[200] Stephen Shankland & Sean Keane, *Trump Creates American AI Initiative To Boost Research, Train Displaced Workers*, CNET (Feb. 11, 2019), https://www.cnet.com/news/trump-to-create-american-ai-initiative-with-executive-order/.

[201] *SkillsFuture*, Government of Singapore (2020), https://www.skillsfuture.sg.

[202] Nicholas Fearn, *U.K. Government Invests £100 Million To Retrain Workers Replaced By Artificial Intelligence*, FORBES (July 18, 2019), https://www.forbes.com/sites/nicholasfearn/2019/07/18/uk-government-will-retrain-workers-replaced-by-artificial-intelligence/#bf66e0c45cd6.

[203] *Id.* (noting the government's tendency to fall behind the private sector in innovation and efficiency).

[204] Jean-Francois Gagne, *Global AI Talent Report 2019* (2019), https://jfgagne.ai/talent-2019/.

[205] Terry Brown, *The AI Skills Shortage*, IT Chronicles (Nov. 2019), https://itchronicles.com/artificial-intelligence/the-ai-skills-shortage/ (reporting that 77% of AI talent worked in academia and 23% in industry).

[206] Ciarán Daly, *How You Can Bridge the AI Skills Gap*, AI BUS. (Jan. 2018), https://aibusiness.com/bridging-ai-skills-gap-2018-long-read/ (reporting that 56% of senior AI professionals believed that a lack of additional, qualified AI workers was the single biggest hurdle to be overcome in terms of achieving the necessary level of AI implementation across business operations); Jeremy Kahn, *Just How Shallow is the Artificial Intelligence Talent Pool?*, BLOOMBERG (Feb. 7, 2018), https://www.bloomberg.com/news/articles/2018-02-07/just-how-

There are indications that this trend is shifting, as the government works to build a data-centric culture among the federal workforce in preparation for AI adoption. Under the Foundations for Evidence-Based Policymaking Act, signed into law in 2019, agencies must appoint "Chief Data Officers"[207] and "Evaluation Officers"[208] to understand and promote data, laying the stage for AI. It can be hoped that this attention will help foster greater recognition of the need to build adequate human capital to support the effective and responsible use of machine learning by governmental agencies.

2. Data storage and processing capacity. Algorithms are dependent upon an analytic infrastructure that includes a large volume of data as well as the hardware, software, and network resources needed to support machine-learning analysis of such data.[209] Some agencies have begun to realize the need for building this infrastructure. The U.S. Federal Aviation Administration (FAA), U.S. Federal Deposit Insurance Corporation (FDIC), and U.S. Federal Communications Commission (FCC) have released statements of their efforts to create large data sets to support agency function.[210] The Office of Financial Research within the U.S. Department of Treasury created the global Legal Entity Identifier (LEI) program in an effort to make big data more readily analyzable for regulators of financial markets.[211] The FDA, EPA, and SEC have begun to leverage cloud storage systems to store, consolidate, and enormous data sets.[212]

A frequent criticism leveraged against algorithms is that the training data are unrepresentative of intended outcomes—in other words, "rubbish in, rubbish out."[213] If the training data only consist of information from certain population groups, then the tool might work less well for members of missing communities.[214] For example, recognizing that white adult men are overrepresented in existing

shallow-is-the-artificial-intelligence-talent-pool (estimating that there are fewer than 10,000 people globally with the necessary skills to create fully-functional machine learning systems).

[207] Foundations for Evidence-Based Policymaking Act, H.R. 4174, 115th Cong. § 3520(c) (2017-18).

[208] *Id.* at § 313(d) (2017-18) (listing tasks including "(1) continually assess the coverage, quality, methods, consistency, effectiveness, independence, and balance of the portfolio of evaluations, policy research, and ongoing evaluation activities of the agency; (2) assess agency capacity to support the development and use of evaluation; (3) establish and implement an agency evaluation policy; and (4) coordinate, develop, and implement the plans required under section 312.").

[209] Coglianese & Lehr, *supra* note 1, at 1164-65.

[210] *Id.* at 1165.

[211] *Id.*

[212] *Id.* at 1166.

[213] Linda Nordling, *A Fairer Way Forward for AI in Health Care*, Nature (Sep. 25, 2019), https://www.nature.com/articles/d41586-019-02872-2.

[214] *Id.*

medical data sets at the expense of white women and people of all ages from other racial and ethnic groups, the U.S. National Institute of Health (NIH) has created the All of Us Initiative.[215] This $1.5 billion program has sought to capture a more diverse population to help generalize NIH findings to the general population. Criticisms of the implementation of this program aside,[216] the agency has recognized that any comprehensive algorithmic tool must rest on a comprehensive database.

However, other agencies are still funneling resources into maintaining legacy systems that are largely becoming obsolete.[217] Scholars have raised concerns about the maintenance of digital information, due to its physical deterioration and sensitivity to software obsolescence. One of the fathers of the internet, Vincent Cerf, has expressed concerns over "bit rot," where digital files are lost to progress and become unintelligible on new technology.[218] Users are "nonchalantly throwing all of [their] data into what could become an information black hole without realising it," Cerf states, going so far as to recommend users print out the information they want to preserve.[219] Digital preservation is not given much of a priority.[220] Storage should, ideally, have a "long life expectancy, a high degree of disaster resistance, sufficient durability to withstand regular use, and very large storage capacities."[221]

It should be noted, of course, that reliance on human decision-making raises similar challenges. Humans have limited memory capacity[222] and a limited life span.[223] Issues of knowledge transfer arise as employees transition between jobs and between organizations. As the Canadian and Dubai governments have fretted,

---

[215] *Id.*

[216] Michael Joyce, *NIH uses dodgy PR to enroll one million Americans in its 'All of Us' precision medicine program*, Health News Review (May 8, 2018), https://www.healthnewsreview.org/2018/05/nih-all-of-us-pr/.

[217] Coglianese, *supra* note 170, at 11.

[218] Ian Sample, *Google boss warns of 'forgotten century' with email and photos at risk*, The Guardian (Feb. 13, 2015), https://www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf.

[219] *Id.*

[220] Samuel Gibbs, *What is 'bit rot' and is Vint Cerf right to be worried?*, The Guardian (Feb. 13, 2015), https://www.theguardian.com/technology/2015/feb/13/what-is-bit-rot-and-is-vint-cerf-right-to-be-worried (stating that "commercial software developers, such as Microsoft and its Office suite, have no incentive to use open formats").

[221] Margaret Hedstrom, *Digital Preservation: A Time Bomb for Digital Libraries*, 31 COMPUTERS & HUMANITIES 189, 193 (1998); MEG LETA JONES, CTRL + Z: THE RIGHT TO BE FORGOTTEN (2016).

[222] *See supra* Part I.

[223] Grace Donnelly, *Here's Why Life Expectancy in the U.S. Dropped Again This Year*, Fortune (Feb. 9, 2018), https://fortune.com/2018/02/09/us-life-expectancy-dropped-again/ (finding human life span to be on average 78.7 years in 2018).

governments are particularly susceptible to knowledge transfer further exacerbated by an aging workforce.[224] Typical employment training is insufficient to smooth the transition between employees.[225] Demographic changes in the government workforce can lead to the potential for gaps in knowledge and expertise relevant to important functions.

3. Cybersecurity. Algorithmic decision-making must be secured against its cybersecurity vulnerabilities. Municipal networks are particularly vulnerable due to limited budgets and cybersecurity expertise, and simultaneously incredibly valuable due to the aggregation of personally identifiable information. Local networks may also act as compromised stepping stones to larger government networks, as they are trusted and connected to state and federal government activity. An algorithm with inadvertent cybersecurity flaws can be sabotaged by bad actors. In 2016, hackers bombarded Dyn DNS, an internet traffic handler, with information that overloaded its circuits, causing an internet slowdown swept the East Coast of the United States.[226] In 2019, hackers targeted government networks in a coordinated attack on local city and county government offices across Texas,[227] Florida,[228] Atlanta,[229] and New Orleans.[230] If a government chooses to host its data on networks—as many do—then it is prudent to ensure there are also solid backups of critical applications and data, as well as an effective data breach scanning strategy.[231]

---

[224] *Recognizing the value of knowledge and knowing how to transfer it*, Canadian Government Executive (May 26, 2015), https://canadiangovernmentexecutive.ca/recognizing-the-value-of-knowledge-and-knowing-how-to-transfer-it/; Mohammad H. Rahman, *Influence of organizational culture on knowledge transfer: Evidence from the Government of Dubai*, J. of Public Affairs (2018), https://onlinelibrary.wiley.com/doi/abs/10.1002/pa.1696.

[225] *Id.*

[226] David E. Sanger & Nicole Perlroth, *A New Era of Internet Attacks Powered by Everyday Devices*, N.Y TIMES (Oct. 22, 2016), https://www.nytimes.com/2016/10/23/us/politics/a-new-era-of-internet-attacks-powered-by-everyday-devices.html.

[227] Kate Fazzini, *Alarm in Texas as 23 towns hit by 'coordinated' ransomware attack*, CNBC (Aug. 19, 2019), https://www.cnbc.com/2019/08/19/alarm-in-texas-as-23-towns-hit-by-coordinated-ransomware-attack.html.

[228] Allison Ross & Ben Leonard, *Ransomware attacks put Florida governments on alert*, Tampa Bay Times (June 28, 2019), https://www.tampabay.com/florida-politics/buzz/2019/06/28/ransomware-attacks-put-florida-governments-on-alert/.

[229] Sarah Hammond, *Houston County Board of Education website hit with ransomware attack*, 13WMAZ (Sep. 24, 2019), https://www.13wmaz.com/article/news/local/houston-county-board-of-education-website-hit-with-ransomware-attack/93-dece14ea-9fef-4c3b-a913-ea972c5b46fc.

[230] Kirsten Korosec, *New Orleans declares state of emergency following ransomware attack*, TechCrunch (Dec. 14, 2019), https://techcrunch.com/2019/12/14/new-orleans-declares-state-of-emergency-following-ransomware-attack/.

[231] *See, e.g.*, Office of the Inspector General, *Semiannual Report to Congress*, U.S. Office of Personnel Management, 1, 8 (2019), https://www.opm.gov/news/reports-publications/semi-annual-

Even when the algorithm has no technical flaws, the responsive, reiterative learning nature of algorithms may cause it to be susceptible to malicious users. When Microsoft released a TwitterBot to learn to converse from its users, the automated account began spewing racist messages after learning and mimicking users.[232]

A major source of security vulnerability for digital systems are human beings. Scams often use social engineering to "hack human nature."[233] Senior officials in national security frequently intone that humans are the "weakest links in the information technology security chain of defense."[234] Not only are human intelligence agents susceptible to conversion, they are also susceptible to innocent mistakes: leaving laptops in unlocked cars, clicking on phishing emails, using weak passwords, and losing credentials. Not only are humans the source of security risks, but enforcers are also generally ineffective at detecting fraud and dishonesty by human actors.[235]

4. Privacy. Privacy concerns accompany most uses of big data. Even when personal information is not directly contained in such data, it may be inferred with considerable accuracy through the analysis of accumulated data. Given their strength and accuracy, algorithms can make accurate inferences about private matters, such political party affiliation or sexual orientation. A major retailer's predictive analytics system may only have access to a customer's Guest ID, purchases items, and other miscellaneous shopping habits, but the aggregation and algorithmic analysis upon this data is often enough to reveal private information useful for targeted advertising.[236] In one instance, the retail store Target deployed an algorithmic marketing software that sent coupons for maternity clothes to a list

---

reports/sar61.pdf (describing "the implementation and maintenance of mature cybersecurity programs [as] a critical need for OPM and its contractors").

[232] Daniel Victor, *Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk*, N.Y. TIMES (Mar. 24, 2016), https://www.nytimes.com/2016/03/25/technology/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html.

[233] Martin Kaste, C*ybercrime Booms As Scammers Hack Human Nature To Steal Billions*, NPR (Nov. 18, 2019), https://www.npr.org/2019/11/18/778894491/cybercrime-booms-as-scammers-hack-human-nature-to-steal-billions.

[234] Earl D. Matthews, *Incoming: The Weakest Link in Security Chain Is People, Not Technology*, Signal (Apr. 1, 2017), https://www.afcea.org/content/Article-incoming-weakest-link-security-chain-people-not-technology.

[235] Dan Ariely, THE HONEST TRUTH ABOUT DISHONESTY: HOW WE LIE TO EVERYONE—ESPECIALLY OURSELVES (2012); Malcolm Gladwell, TALKING TO STRANGERS (2019).

[236] Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

of women who were "likely pregnant," accidentally revealing to one father that his daughter was soon due.[237]

Another concern is related to security: users' data or personal information may be sold to third parties without user consent. A division of the U.K.'s National Health Service provided Alphabet's DeepMind with data on 1.6 million patients without their consent.[238] Google's partnership with health care system Ascension to collect medical data on millions of Americans drew sharp criticism from politicians as "[b]latant disregard for privacy, public well-being, & basic norms."[239] And of course, users' data can be leaked when platforms are hacked.[240]

Such privacy concerns are not insurmountable. One solution is to advance the technology, instead of forgoing flawed, personalized algorithmic services altogether. Platforms may develop privacy-preserving practices, such as cryptographic and randomization techniques,[241] or federated learning, differential privacy, and homomorphic encryption.[242] Platforms may adopt verification techniques to ensure that "influence of a user data point has been removed from a machine learning classifier."[243]

Future changes to the paradigm of privacy law may increasingly address these concerns. Intel,[244] the Center for Democracy and Technology,[245] and Senator Ron Wyden[246] have all drafted privacy legislation proposing a model focusing on business conduct, such as regulating how business collect, use, and share personal

---

[237] *Id.*

[238] Alex Hern, *Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind*, The Guardian (July 3, 2017), https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act.

[239] Richard Nieva, *Google reportedly collects health data on millions of Americans without informing patients*, CNET (Nov. 15, 2019), https://www.cnet.com/news/google-reportedly-collecting-health-data-on-millions-of-americans-without-informing-patients/.

[240] *See* Section II.E.3.

[241] Cong Wang et al., *Toward Privacy-Preserving Personalized Recommendation Services*, 4 ENGINEERING 21 (2018).

[242] Kyle Wiggers, *AI has a privacy problem, but these techniques could fix it*, Venture Beat (Dec. 21, 2019), https://venturebeat.com/2019/12/21/ai-has-a-privacy-problem-but-these-techniques-could-fix-it/; *see also* Kearns & Roth, *supra* note 123.

[243] U.S. Patent No. 10225277, Verifying that the influence of a user data point has been removed from a machine learning classifier, Symantec Corporation (filed June 8, 2018) (granted Mar. 5, 2019).

[244] Intel, https://usprivacybill.intel.com/wp-content/uploads/IntelPrivacyBill-05-25-19.pdf.

[245] *Baseline Privacy Legislation Discussion Draft*, Center for Democracy and Technology (Dec. 5, 2018), https://cdt.org/wp-content/uploads/2018/12/2018-12-12-CDT-Privacy-Discussion-Draft-Final.pdf.

[246] Ron Wyden, *Wyden Releases Discussion Draft of Legislation to Provide Real Protections for Americans' Privacy*, Ron Wyden (Nov. 1, 2018), https://www.wyden.senate.gov/news/press-releases/wyden-releases-discussion-draft-of-legislation-to-provide-real-protections-for-americans-privacy.

data. California has already set the bar high for data privacy regulation by enacting the California Consumer Privacy Act (CCPA) which offers California consumers new statutory rights to learn what personal information businesses have collected, sold and disclosed, and mandates businesses fulfil disclosure obligations and compliance procedures. The European Union enacted the General Data Protection Regulation, which included the "right to be forgotten."[247]

5. Transparency and explainability. Machine-learning algorithms are frequently described as "black box" algorithms because it can be difficult to explain intuitively how they reach the results they do. According to some scholars, algorithms fail to be transparent due to several characteristics: *complexity*, which can render the number of interdependent input factors involved too high for ready comprehension, even by experts; *non-intuitiveness*, where decision rules used by an algorithm, even if observable, do not make sense to experts; and *secrecy*, where details of algorithmic development are deliberately kept secret.[248] Machine-learning forecasts are typically not based on causation, so it is typically not possible to say that they generate explanations in causal terms.

Technological advances are increasing the practical ability to interpret machine learning models.[249] Data scientists have already begun to build transparency tools to "coax explanatory information out of ostensibly black-box algorithms." [250] Interrogation methods help in interpreting any machine learning models: using a "feature importance analysis" to determine that a feature is of high importance if shuffling the values causes a large change in loss; using a "partial dependence plot" to determine the relation between factors while controlling for all other features of the model; using a "individual row interpreter" to determine the contribution of every feature to a specific prediction.[251]

Whether these techiques for understanding the results of machine learning will be satisfactory to all those affected by them may be "much more a function of (evolving) cultural norms and comfort with technology than some intrinsic limitation."[252] Sometimes the assumed goal with explanation seems to be able to

---

[247] MEG LETA JONES, CTRL + Z: THE RIGHT TO BE FORGOTTEN (2016).

[248] David F. Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, YALE J. ON REG. at *22 (2020).

[249] Coglianese & Lehr, *supra* note 20.

[250] Michael Veale et al., *Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making*, PROCEEDINGS OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 1 (2018), https://dl.acm.org/doi/abs/10.1145/3173574.3174014.

[251] Zach Monge, *Machine Learning Algorithms are Not Black Boxes*, Medium (Aug. 26, 2019), https://towardsdatascience.com/machine-learning-algorithms-are-not-black-boxes-541ddaf760c3.

[252] Myura Nagendran et al., *Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies*, BMJ (2020).

make sense of the logic behind a decision in a "human-readable way."[253] But it should be remembered that human decision-making can also be considered as a black box.[254] Humans do not always exhibit transparency.[255] Seldom if ever are they "expected to furnish low-level explanations for their decisions—descending to physical, biochemical or psychological explanations for their motivations and prejudices."[256] As legal realists noted decades ago, the reasons that judges provide for their decisions may prove to be little more than rationalizations for decisions based on intuition, ideology, or stereotypes. Certainly no one should conclude that merely because machine learning requires specific techniques to interrogate that it is unacceptably opaque.

6. Bias. Numerous concerns have been raised that machine-learning algorithms will systemically exacerbate human biases.[257] Bias could creep into the outcomes of machine-learning algorithms at one or more steps:the framing of the objective function; the underlying collected data; or the variables in the selected to be used. As a White House report notes, "[i]f the data [are] incomplete or biased, AI can exacerbate problems of bias."[258] Data may be biased because they reflect existing prejudices—for instance, if historically, human decisions favored hiring men over women, then data on historic hiring patterns will reflect this bias.

The use of biased data has led to racial and gender biases in deciding whether to hire "Jamal or Brendan"[259] and whether to give more medical care to a black or white patient,[260] among others. A gender classification system sold by IBM and Microsoft was found to have an error rate as much as 34.4 percent higher for darker-skinned females than lighter-skinned males.[261] The credit assessment tool used by ApplePay and Goldman Sachs was found to allow up to 20 times more credit to a male applicant than a female applicant with the same financial history.[262] In perhaps one of the most well-known critiques to date, the authors of a 2016

---

[253] *Id.*

[254] Huq, *supra* note 26, at 21.

[255] Gladwell, *supra* note 194ag, at 42.

[256] Gavaghan et al., *supra* note 249.

[257] *See, e.g.*, Sonia K. Katyal, Private Accountability in the Age of Artificial Intelligence, 66 UCLA L. Rev. 54 (2019); Virginia Eubanks, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018).

[258] *Preparing for the Future of Artificial Intelligence*, *supra* note 193.

[259] Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 AMERICAN ECONOMIC REV. 991 (2004), https://www.aeaweb.org/articles?id=10.1257/0002828042002561.

[260] Ziad Obermeyer et al., *Dissecting racial bias in an algorithm used to manage the health of populations*, 366 SCIENCE 447 (2019), https://science.sciencemag.org/content/366/6464/447.

[261] Joy Buolamwini, Gender Shades, gendershades.org/overview.html.

[262] David Heinemeier Hansson (@DHH), Twitter (Nov. 7, 2019, 3:34 PM), https://twitter.com/dhh/status/1192540900393705474.

ProPublica article found racial bias in the (non-learning) algorithms in a system known as COMPAS that seeks to provide an objective measure of the likelihood a defendant will commit further crimes.[263] After collecting the COMPAS scores for more than 10,000 individuals arrested for crimes in Florida's Broward's County, *ProPublica* researchers found that black defendants were twice as likely to be incorrectly labeled as higher risk than white defendants. The risk assessment algorithm COMPAS has also been accused of discriminating by gender.[264]

Merely excluding data or variables related to suspect categories does not guarantee that machine learning will still not pick up on biases in the underlying data. A Goldman Sachs spokesman has stated that the bank's "credit decisions are based on a customer's creditworthiness and not on factors like gender, race, age, sexual orientation or any other basis prohibited by law,"[265] but exclusion of protected characteristics from training data does not guarantee that outcomes are not discriminatory or unfair.[266]

On the other hand, including data on suspect categories may well lead to outcomes that are are more fair. Some algorithms may result in outcomes with significantly less bias against underrepresented users: a job-screening algorithm at a software company actually favored "nontraditional" candidates, such as those without personal referrals or degrees from prestigious universities, much more than human screeners did.[267] Researchers tasked a hiring algorithm to select the best board members to a given company, and found the algorithm-directed companies would perform better; moreover, companies without algorithms "tend to choose directors who are much more likely to be male, have a large network, have a lot of board experience, currently serve on more boards, and have a finance background" than the algorithm-directed companies did.[268]

---

[263] Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[264] *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).

[265] Neil Vigdor, *Apple Card Investigated After Gender Discrimination Complaints*, New York Times (Nov. 10, 2019), https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

[266] Colin Gavaghan et al., *Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project* (2019), https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf.

[267] B. Cowgill, *Automating judgement and decisionmaking: Theory and evidence from résumé screening* (2020), conference.iza.org/conference_files/MacroEcon_2017/cowgill_b8981.pdf.

[268] Isil Erel, *Selecting Directors Using Machine Learning*, European Corporate Governance Institute (ECGI) - Finance Working Paper No. 605/2019 (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3144080; Isil Erel et aIsil Erel, *Research: Could Machine Learning Help Companies Select Better Board Directors?*, Harv. Business Rev. (2018), https://hbr.org/2018/04/research-could-machine-learning-help-companies-select-better-board-directors.

Moreover, even when they do lead to biased outcomes, algorithms are almost certainly easier to fix than biased people.[269] Organizations can leverage algorithms to identify potential biases and address them. Detecting bias in human decisionmakers can be incredible difficult,[270] as people dissemble, obfuscate, and lie. Others may rationalize post-hoc and fail to understand why and how they made their choices. One research group had to assemble a "complex covert operation" with fictitious resumes and ads to lure in prospective employers for months before they had "even one data point to analyze."[271]

By contrast, uncovering algorithmic discrimination merely requires feeding it data and observing its behavior—algorithmic auditing.[272] In an act of introspection, for example, Amazon checked the company's hiring process with algorithmic tools. Using 500 models to identify which cue predicted success at the company, Amazon also discovered a source of bias in hiring: certain keyword in applicants' resumes were correlated with being hired, including terms like "executed" and "capture"—terms predominantly used by males.[273] Consequently, Amazon was able to "pinpoint bias" in the overreliance of confidence as an indicator of competence in their past hiring decisions, and inoculate hiring managers from being overinfluenced by resume terminology.[274]

It is surely easier to retrain one algorithm than it is to retrain two million U.S. government employees. Algorithms provide a comprehensive, systemic fix: when racial bias was uncovered in an algorithm used to allocate healthcare systems, researchers quickly produced a prototype that fixed the algorithmic bias they found.[275] Humans are difficult to retrain: implicit bias training has a modest impact at best, and is ineffective and resource-draining at worst.[276] After a highly publicized incident of racial bias, the global coffee cafes, Starbucks, shut down its stores to conduct "bias training" for its employees. The one-day session will lose

---

[269] Sendhil Mullainathan, *Biased Algorithms Are Easier to Fix Than Biased People*, New York Times (Dec. 6, 2019), https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

[270] David A. Strauss, *Discriminatory Intent and the Taming of* Brown, 56 U. CHICAGO L. REV. 935 (1989).

[271] Mullainathan, *supra* note 276.

[272] *Id.*

[273] James Guszcza et al., *Why We Need to Audit Algorithms*, HARV. BUS. REV. (Nov. 28, 2018), https://hbr.org/2018/11/why-we-need-to-audit-algorithms.

[274] *Id.*

[275] Mullainathan, *supra* note 276.

[276] Edward H. Chang et al., *The mixed effects of online diversity training*, 116 PNAS 7778 (2019), https://www.pnas.org/content/116/16/7778.

the company millions in lost profits when the stores are shut down and a mountain of training resources.[277] Experts say it may not even be effective at all.[278]

7. Abuse of power. It may go without saying, but if algorithmic tools can help responsible governments improve their ability to deliver imporant public value, they could also make more effective the ability governments to pursue unjustified or illegitimate ends. For example, China has rolled out mandatory facial recognition program, awakening opposition and calls for the ban of the use of facial recognition altogether.[279] Algorithmic tools easily make authoritarian governments more effective at enforcing authoritarian rules than would have been achievable decades ago.

Of course, all abuses of power should be opposed, no matter what tools are used to facilitate them. But the possibility that someone might use machine-learning algorithms improperly or for illegitimate ends does not necessarily provide a reason for responsible governments not to use them in good faith and with reasonable care. Just as advances in medical care would presumably not be opposed simply because they could also make oppressive nations' armies or police forces healthier or stronger, it is unclear why otherwise beneficial algorithms would be opposed simply because they could facilitate unjust uses by oppressive regimes. Like other tools, algorithms can be used in good faith by responsible governments to improve public welfare. Indeed, when the use of these tools make meaningful improvements in public value over the status quo, then presumably their use should be applauded.

III. Legal Issues with Governmental Use of Machine Learning

Most of the concerns that have been expressed against governmental use of machine learning have their analogues in legal principles. Whenever a government agency considers using machine learning to assist or replace human decision-making, these related legal principles might be implicated by the proposed use. Any

---

[277] *Starbucks' Training Shutdown Could Cost It $16.7 Million in Lost Sales*, AdAge (Apr. 18, 2018), https://adage.com/article/cmo-strategy/starbucks-training-shutdown-cost-16-7-million/313191.

[278] Jennifer Calfas, *Starbucks Is Closing All Its U.S. Stores for Diversity Training Day. Experts Say That's Not Enough*, Time (May 28, 2018), https://time.com/5287082/corporate-diversity-training-starbucks-results/ ("Unfortunately, it doesn't really seem to do much good on average for companies to offer diversity training because they say you can't really change people's inherent biases with a training session . . . any positive results from a one-time session will fade within a day or two.").

[279] Paul Mozur, *One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority*, N.Y. TIMES (Apr. 14, 2019), https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

complete analysis of the legality of a particular governmental use of machine-learning algorithms will need to be conducted in the context of that use. Still, it is possible to conclude that nothing intrinsic about machine learning should lead government agencies to eschew consideration of its use to perform tasks previously undertaken by humans.

This Part of the report addresses legal questions that might be raised in connection with governmental use of machine learning. It proceeds with reference to several major principles of administrative law: delegation and accountability; procedural justice; transparency and reason-giving; privacy; and equal protection. I have already explored the implications of each of these principles for algorithmic administration in depth elsewhere, where I have reached the conclusion that the use of machine learning by federal administrative agencies should encounter no inherent legal barriers.[280] In this Part, I provide a summary of this general legal analysis and reaffirm that, when used for proper purposes and with due care, machine-learning algorithms can be implemented by federal agencies to replace human decision-making without offending traditional principles of administrative justice. Indeed, in some cases the responsible use of algorithmic tools might even enhance the cause of administrative justice. Given that machine learning faces no intrinsic legal bar, the challenge for federal agencies will be to decide when to rely on machine-learning tools to aid or even replace human decision-making in the administrative state—the issue to be addressed in the final Part of this report.

A. Delegation and Accountability

Perhaps the most fundamental question related to accountability is whether an automated, machine-learning system used by a federal agency to make administrative decisions would deprive individuals of their right to a human decision-maker. When humans remain in the loop, concerns about human accountability will be muted. As a result, the artificial intelligence systems that use forecasts to allocate human auditors or inspectors—which are among the more common uses today—would raise no concern regarding what California Supreme Court Justice Mariano-Florentino Cuéllar has referred to as "cyberdelegation."[281] Why? Because such algorithmic systems do not themselves make the case for a

[280] Coglianese & Lehr, *supra* note 1; Coglianese & Lehr, *supra* note 20; Coglianese & Appel, *supra* note 20; Cary Coglianese & Steven M. Appel, *Algorithmic Administrative Justice*, in MARC HERTOGH, RICHARD KIRKHAM, ROBERT THOMAS AND JOE TOMLINSON, EDS., THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE (forthcoming). Some of this work forms a basis for the discussion contained in this Part.

[281] Mariano-Florentino Cuéllar, *Cyberdelegation and the Administrative State, in* ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 134 (Nicholas R. Parrillo ed., 2017).

penalty. These systems only point humans to possible regulatory or tax violations, and any penalties would be justified only on the basis of what the human auditor or inspector finds.

But it is also not difficult to imagine systems in which the results of algorithmic analyses fully replace human judgment. Would such human-out-of-the-loop systems run afoul of legal accountability principles? Under current law, a longstanding constitutional principle known as the nondelegation doctrine limits, at least theoretically, the extent to which Congress can lawfully authorize an entity other than itself to make rules. Might the nondelegation doctrine serve as an accountability constraint on the use of algorithms as part of automated rule-generating systems—or what has been dubbed "regulating by robot"?[282]

If such robotic rulemaking systems are considered a part of the government agencies which develop and deploy them, then delegating legislative decision-making to these systems would likely require Congress to provide some guidance in the form of an "intelligible principle" as to the basis for these systems' regulatory decisions.[283] But the Supreme Court has determined that legislative principles as vague (and seemingly unintelligible) as "public interest, convenience, and necessity" satisfy the intelligible principle test. No machine-learning rulemaking system, however, could ever operate on the basis of such a broad principle; they must be precisely specified mathematically to work.

The nondelegation doctrine does, however, contain another version which might be relevant to machine-learning algorithms. If these algorithms are not considered part of any federal agency but are analogized instead to separate, private entities, then perhaps they could run afoul of what is sometimes called the private nondelegation doctrine. Under this version of the doctrine, Congress is prohibited from authorizing private actors or entities from making governmental decisions—a practice the Supreme Court has described as the most "obnoxious" form of delegation.[284]

Despite the intuitive appeal of analogizing algorithms to private entities—perhaps especially when government agencies contract with private consulting firms to create them—the principal rationale underlying the private nondelegation doctrine does not fit at all in the context of machine-learning algorithms. Private delegation is obnoxious because private actors or entities are more likely to make

[282] *See* Coglianese & Lehr, *supra* note 1. Although machine-learning algorithms by themselves cannot generate rules, they could be built into systems in which the forecasts from these algorithms are essentially determinative of the selection of predetermined regulatory options. Although somewhat banal, the city of Los Angeles' automated traffic signaling system is effectively a robotic rulemaking system. *See* Lovett, *supra* note 168.

[283] For a discussion of the canonical centrality of the intelligible principle to the nondelegation doctrine, see generally Cary Coglianese, *Dimensions of Delegation*, 167 U. PA. L. REV. 1849 (2019).

[284] Carter v. Carter Coal Co., 298 U.S. 238, 311 (1936).

decisions based on their own narrow interests instead of those of the broader public. Algorithms, however, are programmed to optimize the objectives defined by the operators. As long as those operators are accountable to the public, and the objectives are defined in non-self-interested ways, then the algorithms themselves pose no risk of corruption. Arguably they will be more accountable in their execution than even human officials might be.[285]

B. Procedural Due Process and Reason-Giving

Is a human decision, though, a required element of procedural due process? Procedural due process is a central legal precondition for governmental systems based on human decision-making that affect individual property and other protected interests. After all, when government agencies approve licenses or permits, grant benefits, or enforce regulations, they are making decisions with significant consequences for individuals and business. In these circumstances, affected parties have come to expect procedural fairness from their government officials perhaps even as much as achieving the substantive outcomes they desire.[286] Procedural due process calls for decision-makers who will listen, serve as neutral arbiters, and render reasoned judgments.

Under prevailing legal principles, the procedural fairness of a given administrative decision-making process is judged according to 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.[287] Algorithmic administrative decision-making would seem to pass muster quite easily under this balancing test—indeed, in some instances, due process might eventually even require reliance on algorithmic tools.[288] The private interests at stake are external to machine learning, but machine-learning systems are demonstrating great promise in reducing error and lowering administrative costs. Whenever machine-learning algorithms can be shown to achieve that promise, they will fully satisfy the conventional legal test for procedural due process.

---

[285] Overall, when machine-learning algorithms are recognized for what they are—merely digital machines—the accountability that the law traditionally demands will be accountability to the human governmental officials who create algorithmic systems or who commission their creation and oversee their operation. *Cf.* Prometheus Radio Project v. Fed. Commc'ns Comm'n, 373 F.3d 372, 387 (3d Cir. 2004).

[286] *See* ALLAN E. LIND & TOM R. TYLER, THE SOCIAL PSYCHOLOGY OF PROCEDURAL JUSTICE (1988).

[287] Mathews v. Eldridge, 424 U.S. 319, 335 (1976).

[288] In his probing consideration of the moral and legal principles surrounding a right to human decision-making, Aziz Huq suggests that "under certain circumstances a right to a well-calibrated machine decision might be the better option." *See* Huq, *supra* note 26.

Even if machine-learning systems can be designed so that the systems themselves operate in a manner that satisfies requirements of procedural justice, the process by which these systems are themselves developed might be considered an important, albeit indirect, governmental decision that should meet procedural fairness requirements. Affected interests, in other words, might need be afforded the opportunity to interrogate the design choices that the architects of these algorithms make. The key design choices could presumably be vetted through open processes, with input from advisory committees, peer reviewers, as well as public comments and hearings. Of course, this is how the federal government already operates when generating rules and regulations that will determine future adjudicatory decisions, so procedural justice in the sense of some process analogous to notice-and-comment rulemaking should prove no distinctive or insuperable barrier for machine learning. It will just call for applying well-established protocols for public input to the new world of algorithmically-driven administrative decision-making.

C. Transparency

Transparency is a hallmark of contemporary administrative process. The law of governmental transparency takes two forms.[289] One form—which has been called "fishbowl transparency"—calls upon administrators to make information available to the public and to hold key decision-making meetings open to the public. The second form—"reasoned transparency"—demands that administrators provide reasons for their decisions. The use of machine-learning algorithms in administrative decision-making implicates both forms of transparency.[290]

From the standpoint of fishbowl transparency, the key issue centers on what information government administrators must disclose about the design and operation of an algorithmic system. Such transparency, after all, is not absolute. For example, when algorithmic systems are used to direct scarce inspection or auditing resources, it would probably be counterproductive for government to let the public know too much about how these algorithms work for regulatory targets could simply learn how to game the algorithm. Fishbowl transparency might also become a salient issue when government contracts with private firms to develop their algorithmic systems, as the private contractors may claim proprietary protections for information related to their algorithms.

These examples demonstrate how algorithmic tools can give rise to questions about fishbowl transparency. But fishbowl transparency laws, such as the

[289] Cary Coglianese, *The Transparency President? The Obama Administration and Open Government*, 22 GOV. 529 (2009).

[290] For further elaboration of both types of transparency and their applicability to machine-learning algorithms, see Coglianese & Lehr, *supra* note 20, at 20.

Freedom of Information Act, contain exceptions to the required disclosure of public information. These exceptions include preventing the release of important law enforcement strategies or protecting trade secrets and other proprietary information.[291] In *Wisconsin v. Loomis*, the Wisconsin Supreme Court rejected fishbowl transparency arguments made under the 14th Amendment due process clause by a criminal defendant who contended he should have been provided with more details about a (non-machine-learning) algorithm that a trial court had relied upon in imposing the defendant's length of jail time.[292] The state supreme court held that the private company that had created the algorithm had a right to protect its proprietary information. It was enough to meet due process standards that the defendant had the opportunity to ensure the accuracy of his own personal information the algorithm processed; he was not entitled, however, to know any and all details concerning the proprietary algorithm's design.

With respect to the second form of transparency—reasoned transparency—the black-box nature of machine-learning algorithms raises a distinctive challenge because outputs cannot be intuitively understood or easily explained. A learning algorithm does not generate inferences in the form of "X causes Y," such that an administrator can use the algorithm's output to show that it is justified to regulate X to reduce Y. This inability to generate causal explanations would seem a particular challenge for administrators in meeting their legal obligation to provide adequate reasons for their decisions—an obligation that derives not only from due process but also from the Administrative Procedure Act's arbitrary and capricious standard.[293]

Despite the apparent tension between machine learning's black-box character and the law's reason-giving requirements, administrators are likely to be able to satisfy their reason-giving obligations by explaining in general terms how the algorithm was designed to work and demonstrating that it worked as designed. For example, by describing the type of algorithm used, disclosing the objective the algorithm was established to meet, and showing how the algorithm processed a certain type of data. Obtaining this kind of information should not come close to violating any protected trade secrets or disclosing other protected information. It should be enough to demonstrate to a reviewing court that the algorithm was designed to generate a certain type of relevant information and that it does so in a validated manner.

In this way, machine learning is not unlike administrators' reliance on other types of machines. For example, in justifying the imposition of an administrative penalty on a food processor for failing to store perishable food at a cool

---

[291] Coglianese & Lehr, *supra* note 1, at 1210.

[292] State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

[293] 5 U.S.C. § 706(2)(A) (2018); *see also* Motor Vehicle Mfrs. Ass'n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983).

temperature, an administrator need not understand exactly how a thermometer works—just that it reports temperatures accurately. Courts already defer to administrators' expertise in cases in which government officials have relied on complex machinery or mathematical analyses.[294] They will likely assume the same deferential approach in evaluating an agency's reasons for machine-learning decisions.

D. Privacy

Machine-learning algorithms require—indeed, thrive on—large quantities of data (so-called big data). Yet such data will often include sensitive or personal information related to businesses or individuals. Of course, privacy concerns raised by agency use of personal, sensitive, or confidential business information are hardly unique to machine learning. For that reason alone, there seems little reason to believe that, given adequate safeguards and cybersecurity protections, privacy law would seriously prevent federal agencies from greater use of machine-learning tools.

In fact, for years public management techniques have relied on personal data to deliver public services more efficiently—and the rise of big data and increased access to computing power (even without machine learning) has only accelerated a trend toward the "moneyballing" of government.[295] In carrying out administrative duties, agencies routinely handle an array of personal information from "names, addresses, dates of birth, and places of employment, to identity documents, Social Security numbers or other government-issued identifiers, precise location information, medical history, and biometrics."[296] Protecting such data while concomitantly advancing agency goals is a task many agencies already know how to handle.

In fact, it does not appear that existing legal requirements related to privacy have proven to be an insuperable barrier to federal agency use of big data. The Department of Health and Human Services (HHS), for example, collects personal health data to reduce bureaucratic costs and improve patient health.[297] The Departments of Education (DOE) and Defense (DOD) are harnessing big data for a variety of administrative uses, including: "human resources management; service improvement; fraud, waste, and abuse control; and detection of terrorist

---

[294] *See* Balt. Gas & Elec. Co. v. Nat'l Res. Def. Council, Inc., 462 U.S. 87, 103 (1983); *see generally* Coglianese & Lehr, *supra* note 20, at 43.

[295] MONEYBALL FOR GOVERNMENT (Jim Nussle & Peter R. Orszag eds., 2d ed. 2015).

[296] *See* OFFICE OF MANAGEMENT AND BUDGET, MEMORANDUM M-17-12, PREPARING FOR AND RESPONDING TO A BREACH OF PERSONALLY IDENTIFIABLE INFORMATION (Jan. 3, 2017).

[297] Kenneth A. Bamberger & Deirde K. Mulligan, *Privacy Decision-making in Administrative Agencies*, 75 U. CHI. L. REV. 75 (2008).

activity."[298] In several states, it is now legal for the Federal Bureau of Investigation (FBI) to use facial recognition software to scan Department of Motor Vehicle (DMV) databases that contain driver license photos.[299] The Customs and Border Protection's (CBP) facial recognition kiosks at U.S. airports are scanning the visages of passengers traveling internationally.[300] And the Department of Homeland Security (DHS) and the Department of Justice (DOJ) reportedly have developed "fusion centers" to mine personal data held by the military, CIA, and FBI to identify individuals worth pursuing based on agency criteria for investigation.[301]

The most applicable laws for agencies that deal with the collection, use, and storage of private information are the Privacy Act of 1974 and the E-Government Act of 2002.[302] The Privacy Act limits how agencies can collect, disclose, and maintain personal information in their records.[303] Changes to agency record systems must be disclosed to the public through the *Federal Register* so that the public is made aware of the existence of the types of records and information collected by agencies, the categories of individuals for whom records are kept, the purpose for which the information is used, and how the public can exercise their rights under the Act. The E-Government Act of 2002 requires agencies to conduct privacy impact assessments (PIAs) when developing or procuring technology that implicates privacy concerns.[304] The Office of Management and Budget (OMB) has provided guidance calling on agencies to ensure that their PIAs assess individual

---

[298] *Id.*

[299] Shirin Ghaffary and Rani Molla, *Here's where the U.S. government is Using Facial Recognition Technology To Surveil Americans*, VOX (July 18, 2019), https://www.vox.com/recode/2019/7/18/20698307/facial-recognition-technology-us-government-fight-for-the-future.

[300] *Id.*

[301] Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*," 55 B.C. L. REV. 104 (2014).

[302] In addition, federal agencies confront what is sometimes called a "reverse FOIA" situation, so named for the Freedom of Information Act (FOIA), which compels agencies, with some exceptions, to make information in its possession available to the public upon request. The reverse FOIA situation is one in which, as the U.S. Department of Justice has described it, "the 'submitter of information—usually a corporation or other business entity' that has supplied an agency with 'data on its policies, operations or products—seeks to prevent the agency that collected the information from revealing it to a third party in response to the latter's FOIA request.'" *See* U.S. DEPARTMENT OF JUSTICE, GUIDE TO THE FREEDOM OF INFORMATION ACT 863–880 (2009), https://www.justice.gov/sites/default/files/oip/legacy/2014/07/23/reverse-foia-2009.pdf.

[303] Pub. L. No. 93-579, 88 Stat. 1896 (1974), as amended; 5 U.S.C. § 552a; *see* also OMB, Privacy Act Implementation: Guidelines and Responsibilities, 40 Fed. Reg. 28,948, 28,962 (July 9, 1975).

[304] Pub. L. No. 107-347, § 208, 116 Stat. 2899, 2921 (2002); 44 U.S.C. § 3501 note.

privacy concerns, explore alternatives to the technology, and survey risk mitigation options, as well as articulate a rational for using the technology of choice.[305]

In addition to the Privacy Act and E-Government Act, other statutes also address privacy concerns—but only related to specific types of data. These other laws include: the Family Educational Rights and Privacy Act of 1974, which governs how education-related information can be shared and stored; the Driver's Privacy Protection Act of 1994, which protects personal information gathered by state motor vehicle departments; the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which protects individually identifiable health records; and the Children's Online Privacy Protection Act of 1998, which protects online information gathered from children under the age of 13.

These specialized laws governing data privacy—which tend to apply to private actors as much as government ones—are not matched by any single overarching data protection law or data protection enforcement agency. Instead, the Federal Trade Commission (FTC) is currently tasked with providing data protection in the commercial context under older, more general laws designed to protect consumers from fraudulent or otherwise abusive business behavior. Consequently, there are growing calls for the United States to adopt a vigorous, general data privacy regime. Some advocates look to the European Union's General Data Protection Regulation (GDPR) and its consumer-centric focus and heavy corporate fines as a potential model, while others have argued that GDPR-like legislation could never win sufficient congressional support in the U.S. anytime in the near future.[306]

A series of internal federal policies have sought to promote a privacy culture within administrative agencies. In 2015, President Obama established by executive order a Federal Privacy Council and revised Circular A-130, the government's policy guidance on managing federal information resources. Some of A-130's new provisions include the requirement that every federal agency hire a senior agency official for privacy (SAOP). A-130's new rules coupled with a serious June 2015 data breach at the Office of Personnel Management (OPM) provided a privacy wake-up call for many agencies.[307] Indeed, agencies appear to increasingly take privacy concerns more seriously. For example, in anticipation of the 2020 Census—and recognizing the small risk that public census data could be mined to re-identify

---

[305] *See* OFFICE OF MANAGEMENT AND BUDGET, *supra* note 296; Bamberger & Mulligan, *supra* note 297, at 75–79.

[306] Derek Hawkins, *The Cybersecurity 202: Why a Privacy Law Like GDPR Would be a Tough Sell in the U.S.*, WASH. POST (May 25, 2018), https://www.washingtonpost.com/news/powerpost/paloma/the-cybersecurity-202/2018/05/25/the-cybersecurity-202-why-a-privacy-law-like-gdpr-would-be-a-tough-sell-in-the-u-s/5b07038b1b326b492dd07e83/.

[307] Angelique Carson, *U.S. Government is Changing How it Does Privacy*, PRIVACY ADVISOR BLOG (Sept. 27, 2016), https://iapp.org/news/a/u-s-govt-is-changing-how-it-does-privacy-x/.

individuals—the Census Bureau announced that the agency would use cutting-edge "differential privacy" technology in its data systems. Differential privacy is used by many private firms to preserve the confidentiality of data. The specific differential privacy techniques being deployed by the Census Bureau introduce controlled noise into the data to add protection for individual information.[308]

But beyond machine learning's practical dependence on large amounts of data, it should not present any truly distinctive privacy issues under current laws or policies. All statistical and data systems implicate privacy concerns. However, one worst-case privacy concern can be said to be truly distinctive for machine learning: namely, the ability of such algorithms to combine seemingly disparate, non-sensitive data to yield predictions about personal information, such as the sexual orientation or political ideology of specific individuals. For this reason, machine learning might be said to undermine the ways that data anonymization can protect privacy, because an algorithm can be designed to put together seemingly unrelated data to make accurate predictions about individuals. The possibility of using machine learning to "unlock" private information has led some experts to surmise that anonymizing data is no longer a very useful means of protecting privacy. Legal scholar Paul Ohm has asserted that "data can be either useful or perfectly anonymous but never both."[309] Virtually any attribute of a person might now be traced back to that individual given sufficient publicly available information and machine-learning tools.[310]

But simply noting the possibility that machine learning could be used by government officials to reverse-engineer data and discover private details about individuals for nefarious purposes is far from an argument that there exist inherent legal limits on the responsible use of machine learning.[311] Quite the contrary, such a possibility only points to legal limits on the *irresponsible* use of machine learning. Nefarious uses, such as these worst-case scenarios, are what the law prohibits. On equal protection grounds, such uses would either be struck down as grounded in

---

[308] Ron Jarmin, *Census Bureau Adopts Cutting Edge Privacy Protections for 2020 Census*, CENSUS BLOGS (Feb. 15, 2019), https://www.census.gov/newsroom/blogs/random-samplings/2019/02/census_bureau_adopts.html.

[309] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1703–04 (2010).

[310] *See* Ira Rubinstein, *Big Data: A Pretty Good Privacy Solution*, FUTURE PRIVACY F. (July 30, 2013), https://fpf.org/wp-content/uploads/2013/07/Rubinstein-Big-Data-A-Pretty-Good-Privacy-Solution1.pdf

[311] In addition, such a possibility is also negated as a technical matter by the use of differentially private algorithms, such as those now reportedly being deployed by the Census Bureau. *See supra* note 308. Algorithms that deliver on differential privacy actually can protect against the possibility of reverse engineering. In this way, differentially private algorithms provide a counterexample to the most alarmist claims about the dangers of machine learning, as they reveal that at least sometimes mathematical tools can be designed to respond to and fix some of the problems that have been attributed to machine learning. *See* KEARNS & ROTH, *supra* note 123,at 22–56.

impermissible animus or found to be lacking in a rational basis. Barring that, efforts by federal agencies to target individuals' private details would surely offend administrative law's general prohibitions on "arbitrary and capricious" administrative action and the "abuse of discretion" by government officials.[312]

Without a doubt, privacy concerns are real any time government uses data. But no privacy-related legal strictures would appear to serve as any intrinsic or even serious impediment to the responsible governmental use of machine learning, any more than any other activity that would involve the collection or analysis of large quantities of data.

E. Equal Protection

Turning to equal protection, it is clear, as already noted, that algorithmic systems have the potential to exacerbate, or at least perpetuate, biases and prejudices.[313] Bias obviously exists with human decision-making, but it is also a real concern with machine-learning algorithms, especially when the underlying data used to train machine-learning algorithms is already biased.

Whenever a bias is intentional—in human decision-making, or with machine-learning systems—it will clearly offend constitutional equality protections. But absent an independent showing of such animus, with machine learning it will be difficult, if not impossible, to show any intentional discrimination of the type that would lead a court to rule against federal government officials for violating their constitutional equality responsibilities.[314] The algorithms, after all, are black box in nature, so even if an administrator elects to use data which includes variables on race, gender, or other protected classifications, it will not be readily apparent how the algorithm utilized such data. Even if certain individuals in protected classes are subjected to adverse outcomes due to an algorithm, it is possible that the algorithm might lead to better outcomes for that class overall—or that variables related to the relevant class did not factor dispositively into the algorithm's output.

---

[312] 5 U.S.C. §706(2)(A) (2018).

[313] For popular accounts of equality-related concerns with algorithmic tools, see, for example, CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016), and VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018). For an excellent, accessible technical treatment of algorithmic bias and its implications for the design of machine-learning tools, see KEARNS & ROTH, *supra* note 123.

[314] The Supreme Court has held that equal protection under the Fifth Amendment safeguards against intentional discrimination. *See* Washington v. Davis, 426 U.S. 229, 239 (1976). State administrators, however, may violate equal protection under the Fourteenth Amendment without a showing of intentionality.

Given the black-box nature of machine-learning algorithms, equal protection concerns will thus be unlikely to pose much of a legal barrier to the federal government's use of such algorithms. But questions will arise because, in developing algorithmic decision-making tools, federal administrators will often face the question of whether to include certain kinds of demographic data—e.g., race, gender, religion—in the datasets these tools analyze. That question will arise because machine-learning algorithms will generally increase in accuracy as the variables and data points available for processing increase. Moreover, even when race and other demographic details are excluded, machine-learning algorithms will still be training on existing datasets that themselves will often include biases.

Individuals who claim to have suffered an algorithmic injustice will need to show that the government's decision involves reliance on a suspect classification. Such a claim will face an uphill climb for at least two reasons. First, the Supreme Court has not clearly defined what constitutes a suspect classification. The Court has generally eschewed finding against the government in situations where factors beyond membership in a protected class have factored into administrative decisions—which will inherently be the case with big data systems. Second, although the Supreme Court will indeed impose heightened scrutiny on certain agency decisions which rely even in part on a protected class as among multiple decision-making variables. Heightened scrutiny applies when the government is classifying individuals *on the basis of* their group membership—something that will likely never be provable with algorithms. Administrative algorithms will have objectives defined in terms of the programmatic goal—e.g., identifying fraud, determining eligibility—and machine-learning classifications will be made on the basis of that objective.

Another obstacle facing claimants of algorithmic injustice under the equal protection clause would be the lack of "categorical treatment" in any adverse decisions produced through machine learning. The Supreme Court has disapproved of administrative decisions on equal protection grounds when the government has consistently treated members of a protected class either favorably or adversely.[315] Such categorical treatment is unlikely ever to arise with algorithmic administration because an algorithm's objective function will be defined in terms of some class-neutral outcome—not in terms of affording a categorial preference or disadvantage to certain classes. The algorithms will optimize for predictive accuracy in terms of its objective function and it would surely be unlikely to have a protected class be categorically harmed or advantaged from optimizing on a neutral, non-class objective function, even if the underlying data contains some taint of human bias.

[315] *See* Gratz v. Bollinger, 539 U.S. 244, 251–57 (2003); Grutter v. Bollinger, 539 U.S. 306, 312–16 (2003); Fullilove v. Klutznick, 448 U.S. 448, 491 (1980) (plurality opinion); Wygant v. Jackson Board of Education, 476 U.S. 267 (1985); and Fisher v. Univ. of Tex. At Austin, 133 S. Ct. 2411 (2013).

Moreover, emerging research suggests that biases already found within existing datasets may be reduced, if not eliminated altogether, even without always creating much loss of accuracy in algorithmic forecasts.[316]

Despite concerns about biased data or biased algorithms, it should be clear that the application of traditional equal protection doctrine to algorithmic administration will not constitute an intrinsic bar for the use of such algorithms. Given the black-box nature of these tools, the typical concerns animating equal protection doctrine will be confounded as notions such as "intent" and "categorical treatment" do not fit well with machine learning. As a result of this poor fit, we might reasonably expect few equal protection challenges to algorithmic administration to trigger heightened scrutiny. But even if we were to assume for sake of analysis that a court did apply heightened scrutiny to a machine-learning system, this would not necessarily preordain a court finding of an equal protection violation. After all, when administrators build and rely on machine-learning systems, they will often be doing so to advance important policy objectives and will thus likely have strong arguments that these systems serve compelling state interests, which will be sufficient to justify their use even under a heightened scrutiny standard.

F. Overall Legal Assessment

The legal issues that could be implicated by governmental use of machine learning are not ones that do not already arise with other types of analytic tools. Nor are they ones that should pose an insurmountable obstacle to governmental use of machine learning, at least with adequate planning and care in design. As I have concluded elsewhere, "when federal agencies use artificial intelligence to automate regulatory and adjudicatory decisions, they will likely face little difficulty in making machine-learning practices fit within existing administrative and constitutional constraints."[317]

IV. Deciding When to Deploy Machine Learning

The choice between the status quo of a human-run task for a particular governmental task, whether sorting mail or issuing grants or permits, and a future in which that same task is performed partly or even exclusively by a machine-learning system is a choice that government agencies will increasingly confront. Notwithstanding the futuristic overtones surrounding the use of artificial

[316] *See* James E. Johndrow and Kristian Lum, *An Algorithm for Removing Sensitive Information: Application to Race-Independent Recidivism Prediction*, 13 ANN. APPL. STAT. 1, 189 (2019).

[317] Coglianese & Lehr, *supra* note 1, at 1213.

intellience, the choice confronting agencies will not be much different than the choices agencies have long made between different ways of designing adjudicatory or rulemaking processes.

When contemplating a shift from human decision-making to reliance on machine learning to make decisions, the choice will essentially just be one between a human algorithm and a machine-learning algorithm. Choosing between a human or a digital processes always will itself require a process of some kind—or perhaps what might be called a "meta-process" to distinguish it from the processes under consideration to perform a specific governmental task. The choosing between humans or machines to perform task will come about through a *decision-making* meta-process. Later, the wisdom of the choice made will be assessed or validated through an *evaluation* meta-process. Both decision-making and evaluation can (and should) be assisted and informed by statistical analysis, but both meta-processes will ultimately be grounded in human thought, as humans will be making the judgment about whether to choose a human or digital process to perform the designated governmental task. In the end, there is no escaping the need for humans to think through the choice of having a task performed by or with a digital algorithm versus one having it driven by a human-based "algorithm."

This final Part of the report thus focuses on how humans—government officials—should approach making the choice between algorithm versus algorithm—that is, between maintaining a human process or shifting to a machine learning driven process to perform a given task. Given that governmental tasks are presently performed almost exclusively by humans, the most stark choice today, and for some time to come, will be one of deciding whether to replace exclusively human systems for performing certain tasks with ones that are partly or entirely replaced by a machine-learning systems. Other choices by governments to adopt machine-learning tools will be important too but they will also surely be less stark and presumably less controversial. It may be helpful, for instance, in many instances to use machine learning to supplement or inform what will remain firmly human-based decision processes—but deciding to adopt machine learning merely as an additional input into existing systems will not present as much of a challenge for public administrators.

Guidance and careful decision-making will be needed when administrators are confronted with the choice of whether to replace human processes with digital ones. In these cases, the decision to replace humans with a machine-learning system will implicate important substantive and procedural values. Such a decision can and should be approached in a systematic, analytic fashion.[318] By no means should governmental decision-makers rush unthinkingly into the adoption of and reliance

[318] See Cary Coglianese, *Process Choice*, 5 REG. & GOVERNANCE 250 (2011); CARY COGLIANESE, MEASURING REGULATORY PERFORMANCE: EVALUATING THE IMPACT OF REGULATION AND REGULATORY POLICY (2012), https://www.oecd.org/gov/regulatory-policy/1_coglianese%20web.pdf.

on machine-learning algorithms to perform key governmental tasks—no more than they should unthinkingly rush to shift from one type of human-driven process to another human-driven process.

The purpose in this final Part, then, is to offer guidance to government officials and members of the public as they contemplate replacing or even just complementing a human-driven process designed to perform a specific governmental task. That task could be one as vital but banal as sorting mail or as consequential as approving applications for commercial aircraft pilot licenses—or any number of other adjudicatory or regulatory tasks that have in the past been performed by human officials. With respect to each of these tasks, the core question will be whether a shift to reliance on a machine-learning algorithm would be better than the status quo that relies on human decision-making.

A. Multi-Factor Analysis of When to Use Machine Learning

As with much decision-making about public policy or the design of administrative procedures, what constitutes a "better" process will not always be easy, straightforward, or uncontroversial. Moreover, a judgment that machine learning will (or will not be) better than human decision-making will not be one that can be made in the abstract or across-the-board. Decision-making about machine learning will need to be made within specific contexts and with respect to particular tasks and problems. In some cases, machine learning will prove better than human decision-making, while other times it will not.

Even when machine learning is better, this will not necessarily mean it will be better in each and every relevant respect. It may be, for example, that with respect to a some tasks, machine learning will make demonstrable improvements over human decision-making in terms of speed and accuracy, but it may do so at some loss in the intuitive explainability of decisions. That may still lead to the judgment that, all things considered, machine learning is still better than human decision-making for that particular task, and in a particular context. But this will hardly mean that machine learning is perfect. Machine learning systems will still make mistakes and present downsides. A decision to whether to automate a task using a machine-learning tool will be justified when the gains from machine learning outweigh the downsides.

The meta-process of deciding whether to rely on machine learning for decision-making in performing a governmental task will necessitate balancing of different, and perhaps often competing, values. The kind of balancing could take one of at least three forms. As I will explain, the last of these—multi-factor policy analysis—is likely to be the best approach for administrators to use when facing the meta question of whether and when to use machine-learning tools to automate tasks previously handled by humans.

The first kind of balancing approach would be one reflected in the prevailing law of procedural due process, as articulated by the Supreme Court in its decision in *Mathews v. Eldridge*.[319] As already noted in Part III, the *Mathews v. Eldridge* test seeks to balance the government's interests afftected by a particular procedure (such as the costs of administering the procedure) with the degree of improved accuracy the procedure would deliver and the private interests at stake. Although the *Mathews* formula is often used by courts to assess the justifiability of a single process under challenge, it could easily be adapted by administrators to be used as a framework for choosing between a status quo of a human-based process and a proposed shift to a digitally algorithmic process. The question would be which system delivers the most value on net, taking decisional accuracy and private stakes into account and then "deducting" the costs to the government.

As already noted, well-designed machine learning systems would seem almost inherently to be superior to human systems under an *Mathews* calculus: they are likely to be less costly than systems that must rely on hundreds (if not thousands) of human decision-makers, and their main appeal is that they can be more accurate than humans. The private interests at stake are essentially exogenous and will presumably be unaffected by the choice of whether to use a human or digital algorithm. The *Mathews* calculus, in other words, almost seems hard-wired to support the digital algorithm, provided that the specific machine-learning application in question can be show to result in more accurate decisions than reliance on human decision-makers. For that reason, reliance on the *Mathews* calculus would often collapse the choice between human systems and digital ones into a single question: Which will produce more accurate decisions? Yet, even though improvements in accuracy are vital, just as surely the decision to shift to a machine learning algorithm will entail other, perhaps more fine-grained, considerations beyond accuracy.

A second balancing approach would thus sweep more broadly and take into account of both accuracy and all the other consequences that a shift to machine learning might entail. It would call for administrators to make an all-things-considered judgment about the use of machine learning: in other words, to conduct a benefit-cost analysis. Machine learning would be justified under this approach when it can deliver net benefits (i.e., benefits minus costs) that are greater than those under the status quo. An advantage of this approach is that it takes into account more factors than the *Mathews* calculus. The *Mathews* factors are clearly important, but they may not be always complete. By contrast, benefit-cost analysis is, in principle, complete, as it calls for a quantification and monetization of all consequences. But benefit-cost analysis will also have its limits in this setting—at least if it is to be approached in a "hard" fashion that seeks to place every

[319] 424 U.S. 319.

consequence into a common monetary equivalent that yields an estimate of net benefits. It may not be practical for administrators to achieve that level of precision because some of the consequences of the adoption of machine learning might not always be capable of being placed in a common unit. For example, if a particular machine-learning application would be more accurate and efficient but would result in a greater (and more disproportionate) number of adverse errors for individuals in a historically discriminated racial group, it may be neither meaningful nor justifiable to put the efficiency gains and the equity losses in the same units.[320]

A third balancing approach—a variation on the first two—will more feasibly accommodate a range of values and consequences: a qualitative (or "soft") benefit-cost analysis. This approach is also called multi-factor policy analysis.[321] Basically, it calls for the public administrator to run through a checklist of criteria against which both the human-based status quo and the digital alternative should be judged. These criteria will be more extensive than the three *Mathews* factors but they need not be placed in the same precise common units as in a hard form of benefit-cost analysis. The administrator compares how well each alternative will fare against each criteria, without converting them into a common unit. To aid in making a choice between alternatives, each alternative could be qualitatively rated on each criterion using a rough metric, such as a three-point scale (e.g., "+," "+/-" or "-"), with the ratings for each placed in a summary table. The decision-maker can then see the advantages and disadvantages that each alternative would have on each criterion. For purposes of choosing whether and when to deploy machine-learning tools in government, this third approach is likely to be the most practical and best approach for administrators to follow. [322] The main question will be what criteria or factors should such a multi-factor analysis take into account.

B. Key Factors for Deciding Whether to Use Machine Learning

The starting point for any multi-factor analysis of proposals to adopt machine-learning processes will be to identify the range of relevant factors or criteria. These criteria will vary from use to use, depending on the tasks which a machine-learning system would take over from humans. The precise criteria for a system used to read the hand-writing on U.S. postal mail (one of the first non-

[320] ARTHUR OKUN, EQUALITY AND EFFICIENCY: THE BIG TRADEOFF (1975).
[321] DAVID L. WEIMER & AIDAN R. VINING, POLICY ANALYSIS: CONCEPTS AND PRACTICE (2017).
[322] With respect to choosing whether to use machine learning, a multi-factor framework can be used at different stages of the development process, when different information is available. That is, it can be used at the outset in deciding whether an agency first should even invest in the development of a machine-learning based system, as well as later, whenever such system has been developed, in deciding whether to deploy the system. It can provide a basis for subsequent evaluation of the system in operation as well as making decisions about future modification of the system.

military uses of machine learning by the federal government), for example, will differ from those that might be appropriate for deciding whether to use a machine-learning system to automate decisions about whether to grant license applications for commercial airline pilots.[323] But in general, the criteria for deciding whether to shift to a process based on machine learning will fall into the following three categories: preconditions for success, improved outcomes, and legal and political considerations.

1. Preconditions for Use. As discussed above in Part II.E, to use machine effectively, agencies will need access to adequate human expertise and they will also need adequate data storage and processing capacities. In addition to these tangible human and technology resources which agencies will either need to have in place or secure through government contracts, there exist other, even more fundmental preconditions for government to rely on machine-learning tools. Currently these tools produce what it sometimes called "narrow" artificial intelligence, as they are focused on specific, human-specified goals working on well-defined problem. That is to be contrasted with "general" artificial intelligence which, like humans, can exhibit creativity, flexibility, and learning beyond the domains of a well-defined task. Where the preconditions for narrow artificial intelligence are very poorly met, machine learning is not likely even to be feasible for an agency to consider. Meeting at least the following three preconditions can be thought of as a necessary even if not sufficient condition for a potential shift from a human- to machine-based process:

A. Goal clarity and precision. Machine-learning algorithms need to be programmed to optimize a given objective, and the establishment of the objective function for an algorithm must be mathematically defined. What this means is that machine-learning tools will only be appropriate for an operating task where the objective can be well-defined. For example, if the goal is simply to make the most accurate decisions about claimants' eligibility for benefits, the goal of the algorithm can be specified in terms of reducing forecasting error.

   But if the goal is understood both to make accurate forecasts about who will be eligible while also minimizing unfairness to applicants from a racial minority group, then the clarity may not be sufficient for two reasons. First, it may be unclear what fairness entails. Must the benefits awarded be proportionate to the distribution of each racial group in society overall or in

[323] The latter use is a hypothetical discussed at some length in Coglianese & Lehr, *supra* note 1, and Coglianese & Lehr, *supra* note 20.

the applicant pool? Or perhaps what must be proportionate is the degree of false negative errors? Second, even if fairness is defined with sufficient clarity, given how machine learning works there will almost surely be a tradeoff between maximizing accuracy (the minimization of forecasting error) and addressing fairness. But in making such tradeoffs, agencies may have insufficient statutory direction or social consensus around how to define such a tradeoff in the precise mathematical terms.[324] Exactly how much unfairness should be tolerated to avoid how much diminution in accuracy?

In their need for goal clarity, machine-learning algorithms bear many affinities with performance-based regulation—sometimes called regulation by objectives. But as has been noted elsewhere, it may not always be clear what the full social objective is.[325] For example, federal regulators for years relied on a performance-based approach to standards for child-resistance of packages containing drugs and household chemicals, in an effort reduce childhood poisonings. Only after discovering how these standards designed to optimize for child resistance failed to allow adults to open such containers easily—and thus induced many adults to leave containers opened once they did manage to open them—did regulators redefine their objectives.[326] One of the most vexing preconditions for the use of machine learning may be to define a goal that is both acceptable on policy grounds as well as capable of mathematical definition.

B. 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 not be available for a variety of reasons. They may only have been recorded and stored by an agency in paper form, rather than in digital form.[327] Or they may not be available because disparate digital datasets that need to be combined lack a common identifier that would allow data for each business or individual in each to be linked to each other.

---

[324] Elsewhere I have how in human decision-making systems the existend of such tradeoffs may be obscured and their resolution made through what Cass Sunstein calls "incompletely theorized agreements." But machine-learning algorithms demand more than such incomplete agreements, such as about what may be "reasonable." They need the value choices reflected in the algorithm's objective to be stated with mathematical precision.

[325] Cary Coglianese, *The Limits of Performance-Based Regulation*, 50 UNIVERSITY OF MICHIGAN JOURNAL OF LAW REFORM 525-563 (2017)

[326] *Id*.

[327] Coglianese, *Optimizing Regulation*, *supra* note 170.

More conceptually, data may be unavailable because there just lack a sufficient number of narrow, repeated events around which data exist. It may be easier to find data to support machine-learning analysis of x-rays to determine if a miner qualifies for black lung benefits, but harder to find common data that could be used to determine whether asylum applicants satisfy the test of having a "well-founded fear of future persecution." The standard for the latter requires "both a subjectively genuine and an objectively reasonable fear,"[328] which may afford many unique circumstances to qualify. Similarly, data may be available to show the probability that a particular defendant's DNA could be contained within a mixed DNA sample from a crime scene,[329] but not for whether, in the absence of any DNA sample, that defendant was driving a red Corvette that passed through the intersection of Sixth and Main Streets at 12:35 a.m. on July 23rd. The point is simply that it will obvously be difficult to find a large set of data for cases that are truly *sui generis*.[330]

C. External validity. Related to the availability of data is a question of the likely representativeness of the data available for training a machine-learning algorithm with the population to which the algorithm will be applied. The world is ever-changing, so at a minimum there will need to be a steady stream of available data to keep updating an algorithm and retraining it as the world—and the data about the world—keep changing. If the relevant parts of the world change more quickly than an algorithm can be replenished with current data and retrained, then the algorithm will be "brittle"—that is, it will suffer from what statisticians call an external validity problem.[331] A machine-learning algorithm used for economic forecasting, for example, might not produce much accuracy in forecasting business or employment conditions during unprecedented pandemic-induced recession.

Now, any kind of forecasting and decision-making tool—even human judgment—will be limited in unprecedented times or periods of rapid dynamism. These circumstances of what Robin Hogarth calls "coconut

[328] INS v. Cardoza-Fonseca, 480 U.S. 421, 430-1 (1987).
[329] https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs.
[330] *Cf*. Gary Marcus & Ernest Davis, *A.I. is Harder Than You Think*, N.Y. TIMES (May 18, 2018).
[331] M.L. Cummings, *The Surprising Brittleness of AI* (2020), https://www.womencorporatedirectors.org/WCD/News/JAN-Feb2020/Reality%20Light.pdf.

> uncertainty"[332]—or others might call unknown unknowns—present inherent levels of uncertainty. The key question is whether, under such cirumstances machine-learning algorithms will more or less "brittle" than other types of analysis. It is certainly conceivable that, with the right kind of data acquisition and feedback process a machine-learning system could be designed so that it actually fares better than alternatives in periods of disruption. The high level of uncertainty endemic to such periods, though, will make it hard to be sure that machine learning—or anything else—fares better than alternatives.

Taking these criteria together, machine-learning systems will realistically only be plausible substitutes for human judgment for tasks where the objective can be defined with precision, for tasks of a kind that are repeated over a large number of instances (such that large quantities of data can be compiled), and for tasks where data collection and algorithm training and re-training can keep in sync with relevant changing patterns in the world. This is not to say that these preconditions are absolute or must be perfectly satisfied. But if they are not even minimally satisfied with respect to a given use case, it will make little sense to contemplate the use of machine learning. On the other hand, where these preconditions are sufficiently satisfied, there can be some reason for an administrator to think that machine learning could improve on the status quo and that it will be worth considering taking the next steps to start the development and design of an algorithmic system.

2. Improved Outcomes. The ultimate test for machine learning will be how it operates compared to the status quo. As Part I of this report has made clear, the status quo that is based on human decision-making need not be viewed as lacking in room for improvement. Whether a machine-learning system in fact can be exepected to fare better—and in fact does so—will make up a centerpiece in any multi-factor analysis aimed at deciding whether to adopt machine learning. The precise definition of better will need to be informed by each specific task, whether weather forecasts, identifying instances of tax fraud, determining eligibility for licenses or benefits, or any number of tasks. Although the specific specification of the relevant criteria for success will vary across these different uses, it is possible to identify three general types of impacts that should be considered in determining whether machine-learning leads to improved outcomes:

A. Goal performance. Current human-based systems have goals or tasks that they are supposed to perform, so the first set of

[332] Robin Hogarth, On Coconuts in Foggy Mine-Fields: An Approach to Studying Future-Choice Decisions (2006), https://www.researchgate.net/publication/228499901_On_Coconuts_in_Foggy_Mine-Fields_An_approach_to_studying_future-choice_decisions.

performance factors should be guided by those goals. The relevant factors can be captured by a series of straightforward questions: Would machine learning achieve those administrative goals more accurately? Would it operate more quickly? Would it be less costly, needing fewer FTEs and other scarce resources? Would machine learning yield a greater degree of consistency? Each of these and related questions can be asked from the standpoint of the current statutory purpose or operational goal of any human-driven system in the federal government. The overarching idea would be to determine, in effect, how well machine learning helps an administrative agency do its job.

B. Impacts on directly affected users. The ways that machine learning might help an agency do its job are only one way to consider machine learning's impacts. Unless already fully captured in the agency's own performance goals, it is also important to ask what the effects of machine learning will be on the applicants, beneficiaries, or other individuals and businesses who are (or would be) directly affected by a specific machine-learning system. How does the system treat them? Do some portion of users suffer disproportionate adverse effects? Do they feel like the system has sufficiently served them well? Keep in mind, again, that machine-systems do not need to be perfect nor completely problem free—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 not be ideal in the abstract but much better in relative terms. It is instructive that eBay deploys a fully automated dispute resolution software system that resolves disputes so satisfactorily that customers who have disputes are actually more inclined to return to eBay than those who never have a dispute.[333]

C. Impacts on broader public. And unless already factored into the agency's own performance goals (item "A" above), an administrator should include broader societal effects in any multi-factor analysis of machine learning. How would machine learning affect those who might not be directly interacting with or affected by the system? Will the errors that remain with machine learning, for example,

[333] *See* BENJAMIN H. BARTON & STEPHANOS BIBAS, REBOOTING JUSTICE: MORE TECHNOLOGY, FEWER LAWYERS, AND THE FUTURE OF LAW (2017).

> prove have broader consequences? They might not for a system used to sort the mail, but they could for a system to determine who should receive a commercial pilot's license—in the latter instance, the impact on members of the flying public surely would need to be considered. Again, though, the key question would be whether any broader societal effects would be more adversely consequential than similar effects under the status quo.

It is conceivable that a machine-learning system could deliver improved outcomes across all the outcome factors. Yet it might also be the case that few processes—digital or otherwise—will perform better than the status quo on each and every possible criterion. As a result, some effort will need to be made to characterize the *degree of* improvements and performance losses resulting from a shift to machine learning. Administrators, in other words, should ask not only whether machine learning improves accuracy, but by how much.

Some effort will then need to be made to determine a priority among different outcomes. If the use of machine learning for a particular task were to prove to be much less costly but slightly less accurate than the status quo, how important is accuracy for the use case at hand? Are the errors that remain with machine learning all that consequential? It would be one thing for the U.S. Postal Service to tolerate mistakes in letter sorting to lower the costs of mail sorting dramatically. But it would be another altogether to make that kind of tradeoff with respect to identifying safety risks at offshore oil rigs.

Finally, validation of machine-learning systems will be vital to assessing whether machine learning leads to improved outcomes.[334] That validation should be conducted in advance. In effect, it necessarily is when training and then testing an algorithm on historic data. But agencies might also consider setting up pilot efforts to run the algorithm in tandem with human decision-makers for a period of time to see how it will operate in practice. Validation efforts should occur later as well. Once a digital system has replaced a human-driven system, it could be evaluated relatively early in its use before the loss of human capital is irreversible. 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 or other problems.

[334] *Cf.* Admin. Conf. of the U.S., Recommendation 2017-6, *Learning from Regulatory Experience*, 82 Fed. Reg. 61,738 (Dec. 29, 2017).

3. Legal and Public Acceptance Risks.

Although I concluded Part III by noting that machine learning poses no intrinsic nor insurmountable legal barriers to adoption by government agencies, this does not mean that those agencies that make a switch to digital algorithmic systems will face no risk of getting sued nor of generating public controversy. On the contrary, it should be clear that any governmental body's decision to choose to use algorithmic tools could become the source of considerable controversy and even litigation.[335] Agency officials will thus want to take into consideration both litigation risks as well as risks of public or political controversy. These risks will likely be affected by (1) the degree to which machine learning determines agency action, and (2) the stakes—financial and otherwise—associated with the use in question.

When it comes to the degree to which the machine-learning algorithm determines the agency action, we can distinguish three ways that the results of a machine learning algorithm could play a role:

- *Input*: The result produced by a machine-learning algorithm could provide information to the (human) agency decision-maker, making the algorithm but one factor among others in the agency's decision.

- *Default*: A machine-learning algorithm could be part of a digital system that generates a default decision that can be overridden by a human (a so-called human-in-the-loop system).

- *Decision*: An algorithm could make a final decision—subject only to judicial review (a human-out-of-the-loop system).

All things being equal, agencies can expect that new uses of machine learning that only provide inputs into agency actions will be easier to defend in court (and thus less likely to be challenged in the first place) than those that create defaults or make decisions. Likewise with respect to public resistance or political controversy surrounding the implementation of a digital system.

[335] For a review of litigation that has arisen to date over administrative agencies' use of (non-learning) algorithms, see Coglianese & Ben Dor, *supra* note 9. For an especially insightful account of the political controversy stirred up by a city's use of machine learning, see Ellen P. Goodman, *The Challenge of Equitable Algorithmic Change*, REGUL. REV. (Feb. 12, 2019), https://www.theregreview.org/wp-content/uploads/2019/02/Goodman-The-Challenge-of-Equitable-Algorithmic-Change.pdf.

All other things being equal, the lower the stakes of the action to which machine learning is directly connected, the lower the risk of litigation or controversy. Surely among the lowest conceivable stakes would be uses that support purely internal staff functions at an agency. For example, consider an IT department within a government agency that chooses to deploy a machine-learning algorithm as part of a chat bot that answers calls from staff for technology assistance. That chatbot could even be designed to work autonomously to process password reset requests entirely on its own, without any human intervention, but given that the stakes to any member of the public could hardly be lower, litigation will not be a risk for the agency.[336] (In addition, of course, such internal matters will typically not be subject to judicial review, and it would be hard in any case to imagine how anyone would meet the requirements of standing to challeng such an internal chatbot.) On the other hand, digital 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 some degree of litigation risk, all other things being equal.

Putting the two factors together, it is possible to visualize the risks of litigation and public criticism arising from different uses of machine learning as shown in Table 1. The traffic signal colors indicating the degree of caution: red posing the greatest litigation risk, green the least. The U.S. Postal Service's use of machine learning to help read handwriting when sorting letters and packages would fall within the *low-stakes* row and the *default* column—because a postal worker can always intercede to redirect mistakenly sorted piece of mail. Presumably this use of machine learning gives rise to no meaningful litigation risk.

On the other hand, the use of machine learning as part of a digital system to make criminal sentencing recommendations would clearly fall into the *high-stakes*

Table 1: Risks to Agency from Government Use of Machine Learning

	Input	Default	Decision
Low Stakes			
High Stakes			

[336] Justine Brown, Chatbots Debut in North Carolina, Allow IT Personnel to Focus on Strategic Tasks (Oct. 12, 2016), https://www.govtech.com/Chatbots-Debut-in-North-Carolina-Allow-IT-Personnel-to-Focus-on-Strategic-Tasks.html

row. But the risk of such a system being struck down might not be very high if the system only provides judges with an *input* to be used as one of many factors in a judge's sentencing decision. In *State v. Loomis*, the Wisconsin state supreme court upheld against a due process challenge the state's use of a risk assessment algorithm in the sentencing process because it was merely an input into the sentencing decision.[337] The court emphasized that the sentencing decision in Loomis's case "was supported by other independent factors" and that the algorithm's "use was not determinative."[338]

Even when machine learning would be determinative of relatively high stakes matters, this does not mean that it should be avoided. The high stakes may make it all the more imperative for an agency to use a machine-learning system if it makes a significant improvement in accuracy, consistency, speed, or other performance goals. After all, when the stakes are high the government should do all it can to maximize its decision-making performance—and sometimes that need will weigh in favor of machine learning. But even in those contexts, it will be possible for agencies to manage the potential risks of litigation through careful planning, validation efforts (as described above), and public engagement when appropriate in the development and design of an algorithmic system. Finally, when contracting out for technical support and services in developing a machine-learning system, agencies should take into account the need to have access to and be able to disclose sufficient information about the algorithm, the underlying data, and the validation results to satisfy transparency norms.[339]

Conclusion

A wholesale shift by administrative agencies to reliance on automated decision making systems would mark an major change in how the federal government operates and how it interacts with beneficiaries, regulated entities, and the public overall. Yet the increasing use of machine learning to fuel automation in business, medicine, transportation, and other facets of society make more widespread use of machine-learning tools by government part of the foreseeable future. Indeed, already government agencies have been developing artificial intelligency tools to assist with enforcement, benefits administration, and other important government tasks.

The move to a future of algorithmic governance naturally gives rise to concerns about how new digital tools will affect the efficacy, fairness, and transparency of governmental processes. The aim of this report has been to show that whatever the validity of the concerns about machine-learning systems, they

[337] 881 N.W.2d 749.
[338] *Id*. at 753.
[339] Coglianese & Lehr, *supra* note 20.

should be kept in perspective. The status quo that relies on human "algorithms" is far from perfect. If the responsible use of machine learning can usher in a federal government that, at least for specific uses, achieves better results at constant or even fewer resources, then both government administrators and the public would do well to support such use. The challenge for agencies will be to decide when to use artificial intelligence. This report suggests looking at whether a particular candidate use will satisfy the general preconditions for the deployment of machine-learning algorithms and then to ask whether such algorithms will indeed deliver improved outcomes. Proper planning and risk management can help ensure that the federal government is able to make the most of what advanced digital algorithms may be able to deliver over the status quo.