Trust Us, We Know What We're Doing:
Machine Learning and the Automation of the Public Sector

Minnowbrook 50th Anniversary Conference Concept Paper

Matthew Young
Maxwell School of Citizenship and Public Affairs
Syracuse University

The term 'revolution' cannot adequately describe the current state of public administration and the administrative state in the West. Rather, public administration is experiencing simultaneous and interrelated existential crises: a crisis of legitimacy spurred by laissez-faire and reactionary ideology that paints the public sector as either obstacle or enemy; and a crisis of capacity where the state seems ill-equipped to understand and address present-day challenges. These crises are the product of decades of contention about the proper role of government, increasing economic inequality and fiscal austerity, and social and technological change.

The present and future use of machine learning is an important case of such change, with implications for the capacity and legitimacy of democratic public administration. As the public sector is subjected to normative and material pressure to increase capacity while maintaining or reducing inputs, the substitution of civil servants with automation via machine learning systems is increasingly likely. What's more, this change is also likely to take place before the technology can perform with the consistency or transparency expected – whether tacitly or legally – of public institutions. Whether machine learning systems merely augment or fully automate public sector positions, we need to be aware of the implications and ready to assess their effects and train our students appropriately.

Motivations for Adopting Machine Learning in Public Administration

Bureaucracy's *raison d'être* is the systematic collection and processing of data to understand and reshape the state of the world. The exponential growth in data generation and computational capacity over the past three decades has given rise to the era of "Big Data" –digital data are now generated at volumes and dimensionalities that the human mind cannot process. Demand for the tools necessary to put Big Data to use in improving organizational performance has contributed to an increased emphasis on "data-driven decision-making" in the public and nonprofit sectors.

Machine learning is an ideal – indeed, purpose-built – tool for analyzing Big Data. While traditional software involves providing inputs to a human-designed program with outputs as a result, machine learning involves providing inputs and outputs into a recursive system with a program as the result. One consequence of this difference is that machine learning systems are theoretically superior to human-designed software for inductively identifying relationships between complex, multidimensional data and a given outcome of interest. Another is that machine learning systems are much more adaptable in the face of increasing data dimensionality; they require neither changes to their programming nor *a priori* theory to inform how they weigh different variables in their predictive algorithm.

One example of machine learning applications in the public sector is the National Security Agency's SKYNET program. Based on documents leaked by former Booz Allen Hamilton contractor Edward Snowden, at a minimum SKYNET feeds metadata collected from cellular phone networks in Pakistan into a machine learning system that produces a probabilistically weighted assessment of an individual's likelihood of being a terrorist. This assessment is then used to determine the targets of military strikes by drone or other delivery system (Grothoff and Porup, 2016). Another example is the development of 'risk assessment tools' in criminal justice. These tools use machine learning to produce estimates of an individual's likelihood of failing to appear in court or committing a new offense if paroled (Monahan and Skeem, 2016).

The Automation of the Administrative State

In these examples machine learning systems are used to inform and support decisions made by both street- and mid-level bureaucrats. But eventually machine learning may obviate the need for decision-making in these positions or, in conjunction with physical automation where necessary, replace the bureaucrats altogether. This should be expected for several reasons.

First, substituting labor with capital becomes more appealing as employee compensation costs (particularly medical benefits and pension obligations) increase while fiscal constraints remain unchanged or worsen. Machine learning was not a viable alternative to FTE positions during the 2008 recession and its aftermath; whether public organizations replace staff with these systems during the next economic contraction is an empirical question (but see Frey and Osborne 2017 for a model of skilled positions' vulnerability to replacement by machine learning). Second, the reliance on contracting out for internally-oriented information technology products and services creates supply-side pressure as private firms market machine learning to public organizations. Even as a nascent market, marketing of machine learning solutions to the public sector occurs not only from large technology companies but also from specialized firms (see www.datarobot.com/public-sector/).

Third, there are normative pressures in the form of (1) a growing distrust of (human) expertise in general and in the public sector/administrative state in particular; (2) a commonly-held belief in the objectivity and superiority of quantitative data to other forms of epistemology; and (3) technological determinism – the belief in technological progress and its capacity to solve problems (soft determinism) and/or its centrality in determining social relations (hard determinism). An example of the latter of these pressures is the argument that machine learning systems are equity-enhancing because they can eliminate the biases, whether conscious or implicit, present in human decision-makers. Finally, replacing employees with machine learning is a logical extension of elected and appointed officials and senior management's use of technology to limit bureaucratic discretion. Traditional principal-agent models of administrative delegation suggest that machine learning systems should be the optimal agent; shirking is eliminated completely, and moral hazard could be minimized or eliminated through system design. Empirically, Eubanks (2017) documents the use of algorithmic-based decision support systems in this way, particularly in healthcare and social service provision.

The Dangers of Machine Learning in Public Administration

Machine learning is subject to two significant limitations that have significant consequences for its use in public administration. The first is its dependence on the quality of the data used to 'train' the system. If these data are incomplete or biased the system will not only reproduce these effects, they are likely to be magnified due to the reinforcement effect of backwards propagation. Unfortunately, not only are public sector administrative data notoriously incomplete, but as with all data they contain the biases of the organizational processes and individual administrators who generated them. Indeed, the empirical study of representative bureaucracy is predicated on the existence of these biases (e.g., Nicholson-Crotty et al, 2016). And to the extent that such research finds equity-enhancing effects from representative bureaucracies, the mechanism for these effects is the same bureaucratic discretion that machine learning systems reduce or eliminate.

Machine learning's second major limitation is that the processes by which the system arrives at a decision cannot be reverse engineered. If one wants to know how or why a machine learning system made a decision, we can only say that it was a function of the data (both training and applied) and the system architecture. We cannot put a machine learning system on the stand in a courtroom and interrogate it to judge whether its decision process was flawed – it is a true 'black box.' It is hard to overstate the potential consequences this poses for the institutional legitimacy of the democratic administrative state. The inability to so much as develop an interpretable logic model for decision-making is fundamentally incompatible with the concept of due process.

Jointly these limitations are likely to create a negative feedback loop: machine learning's adoption is motivated in part by a distrust of bureaucratic discretion and dismissal of the value of human expertise, and its implementation will result in decisions that cannot be justified and are therefore likely to be considered illegitimate by any individual or group that disagrees with the outcome, further contributing to the erosion of shared trust in public institutions. There are active research agendas in the fields of computer and data science that are dedicated to minimizing and solving these limitations, but as of this writing they remain inescapable.

Considerations

The challenges that machine learning pose to our field mirror the broader, existential challenges of the current milieu. I choose to view this as an opportunity, because it allows for focused efforts on research and teaching with respect to machine learning to carry over to other debates. In research, the pressure to automate public administration demands a new assessment of the normative value of human agency and decision-making in democratic states, and further highlights the tradeoffs between equity, efficiency, and transparency. Empirical analysis of bias in public sector data also calls direct attention to the systemic and idiosyncratic sources of bias in our institutions. In teaching, we can help public managers understand the inherent limitations of machine learning, the epistemological implications of different data generative processes, and the ethical issues posed by data-driven decision-making. Rather than retreat to our respective intellectual silos, it is incumbent upon us to acknowledge and engage with the technological automation of public administration.

References

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

Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation?. *Technological forecasting and social change*, *114*, 254-280.

Grothoff, C., & Porup, J. (2016). The NSA's SKYNET program may be killing thousands of innocent people. *Ars Technica*.

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual review of clinical psychology*, *12*, 489-513.

Nicholson-Crotty, S., Grissom, J. A., Nicholson-Crotty, J., & Redding, C. (2016). Disentangling the causal mechanisms of representative bureaucracy: Evidence from assignment of students to gifted programs. *Journal of Public Administration Research and Theory*, *26*(4), 745-757.