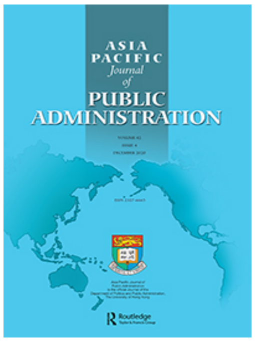

Asia Pacific Journal of Public Administration

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rapa20

Improving public services using artificial intelligence: possibilities, pitfalls, governance

Paul Henman

To cite this article: Paul Henman (2020) Improving public services using artificial intelligence: possibilities, pitfalls, governance, Asia Pacific Journal of Public Administration, 42:4, 209-221, DOI: 10.1080/23276665.2020.1816188

To link to this article: https://doi.org/10.1080/23276665.2020.1816188

Published online: 14 Sep 2020.

Asia Pacific Journal of Public Administration, 2020
Vol. 42, No. 4, 209–221, https://doi.org/10.1080/23276665.2020.1816188

Improving public services using artificial intelligence: possibilities, pitfalls, governance

Paul Henman *

School of Social Science, University of Queensland, QLD 4072, Australia

Artificial intelligence arising from the use of machine learning is rapidly being developed and deployed by governments to enhance operations, public services, and compliance and security activities. This article reviews how artificial intelligence is being used in public sector for automated decision making, for chatbots to provide information and advice, and for public safety and security. It then outlines four public administration challenges to deploying artificial intelligence in public administration: accuracy, bias and discrimination; legality, due process and administrative justice; responsibility, accountability, transparency and explainability; and power, compliance and control. The article outlines technological and governance innovations that are being developed to address these challenges.

Keywords: artificial intelligence (AI); chatbots; automated decision making; ethical AI; algorithms; digital government

Introduction

New digital technologies are rapidly changing the landscape for the delivery of public services. Mobile devices teamed with apps bring online public services to wherever the citizen is. Networked and wi-fi technologies enable the provision of information and collection of geo-coded data to be integrated with traditional administrative data, creating "big data" sets for building knowledge about populations and individuals. Automated administrative decision-making processes are being expanded, with artificial intelligence (via machine learning) providing more nuanced ways to make decisions within complex circumstances.

This article focuses on the potential of artificial intelligence (AI) for improving public services. It focuses on three areas: automated administrative decision-making; chatbots; and public governance. The article first provides examples of emerging uses of digital tools and processes and the potential benefits they provide. It then considers some of the technological, ethical, legal and political challenges and pitfalls to such technologies. Finally, it highlights ways in which regulation and governance of AI-based public administration can be enhanced in order to advance its potential, while mitigating its negative consequences to individuals and society.

Before proceeding it is important to clarify terminology. What is meant by "artificial intelligence"? The term artificial intelligence (AI) has been used and enlivened the imaginations of people for decades. Each new wondrous development of computer algorithms and systems – from Weizenbaum's (1984) ELIZA who engaged people in

*Corresponding author. Email: p.henman@uq.edu.au

conversation in the 1960s, and IBM's Deep Blue program that defeated the world chess master in 1996, to driverless cars of the present – have repeatedly displayed intelligent-like behaviour and challenged our collective thinking about what counts as intelligence. Today, the AI moniker is used as much as hype and marketing as to signal a real shift in how algorithms are being developed. AI is typically used to refer to (systems developed with) machine learning algorithms that self-organise their internal variables and values to achieve desired outcomes. For example, instead of human programmers trying to identify the facial structures that can be used to distinguish between and identify individuals, AI facial mapping systems learn by being trained on large sets of faces and matched faces with names. Current use of AI terminology is often about a specific mode of machine learning, namely artificial neural networks, which are inspired by brain structure and operation. Hardware (and software) developments have led to the recent explosion of AI as they have enabled what previously was not technologically feasible. Hereafter, this article uses the acronym AI to reflect computer systems built (at least partially) using machine learning.

The recent dramatic enthusiasm for AI (as machine learning) and associated plethora of policy discussion papers (Fjeld et al., 2020) signals a fundamental change in digital technologies. In terms of usage in government, it is important to keep in mind that AI is part of a trajectory in digital technologies, systems and algorithms, rather than a fundamental rupture. For example, many of the legal, ethical and social concerns about the use of AI in government are identical to those of non-machine learnt algorithms. The emergence of AI discourse has been beneficial in generating public discussions of how digital technologies are governed to achieve collective principles and aims.

Deploying Artificial Intelligence for improving public services

Automated decision-making

Governments have long been using computer algorithms to assist government officials make decisions and increasingly to automate those decisions without human involvement. To date, the use of AI in government decision making is still limited (Engstrom et al., 2020; De Sousa et al., 2019; Sun & Medaglia, 2019), though similar systems based on sophisticated statistical analyses are widely deployed. Indeed, is it sometimes hard to know if a system is a true AI. I outline three different forms of operation of AI in public sector decision-making: detecting patterns; sorting populations; and making predictions.

One, AI is very useful is detecting patterns. While the human brain also operates through pattern construction and recognition, due to their superior big data processing capacity, AI can detect patterns that evade humans. Cross matching human DNA with disease is an example whereby AI is used to identify multiple DNA segments that correlate in very complex ways to diseases. AI can also observe patterns without human guidance, such as the case in 2012 when Google's AI organically distinguished human faces and cats after processing ten million YouTube videos (Clark, 2012).

Two, a consequence of creating and recognising patterns, is AI's ability to subsequently delineate or sort individuals into different categories or groups according to different outcomes, scores or risks. For example, in 2017 the US Department of Homeland Security issued a competition, The Passenger Screening Algorithm Challenge, to improve their current non-AI algorithm using advanced machine learning. The renamed system Secure Flight,

> "is a risk-based passenger prescreening program that enhances security by identifying low and high-risk passengers before they arrive at the airport by matching their names against trusted traveler lists and watchlists" (https://www.tsa.gov/travel/security-screening).

Through more nuanced differentiation of people's risk profiles, the enhanced AI algorithm seeks to expedite the screening process and reduce passenger pat downs.[1]

In a similar manner, AI is increasingly replacing more traditional programed algorithms to enable government to identify the level of risk of child maltreatment in notifications of possible maltreatment. Statistically-based profiling, in the form of Structured Decision Making developed by the National Council on Crime and Delinquency in the USA,[2] has been widely used in child protection services. More recently, AI approaches have been developed (Cuccaro-Alamin et al., 2017; Vaithianathan et al., 2017) or are in development (Schwartz et al., 2017). An impact evaluation found the Allegheny system resulted in improved accuracy and consistency, and reduced staff workloads (Goldhaber-Fiebert & Prince, 2019).

AI used to segment and sort populations enables governments to enhance targeting of resources and services to better address particular challenges – be it long-term unemployment, child maltreatment, recidivism, tax/welfare fraud or food violation vendors – and, in doing so, generates government efficiencies and cost savings.

Three, AI is also used for predictive modelling of demand to enable administrators to respond and direct resources more nimbly. Australia's Commonwealth Scientific Industry and Research Organisation (CSIRO) has developed The Patient Admission and Prediction Tool (PAPT) using AI, which "provides an accurate prediction of the expected daily patient load [to hospital emergency departments] as well as patients' medical urgency and specialty, and how many will be admitted and discharged".[3] This tool thereby enables hospital administrators to better plan resourcing and staffing needs to reduce patient waiting times and enhance treatment. CSIRO has also developed a system called Spark, which uses AI to predict bushfire (forest fire) spread.[4] This system allows better allocation of firefighting resources and development of strategies to best combat fires' progression, thereby saving property and the lives of people and animals. AI is also being developed for automated traffic control (Abduljabbar et al., 2019), such as in Singapore whereby AI is coupled with real time sensors.[5] Policing is another area where predictive algorithms have been used to direct police resources. For example, China's Suzhou Municipality has been using the Suzhou Public Security Crime Prediction System since 2014 to predict the locations and times of theft, which in turn received greater police numbers.[6]

Chatbots

A chatbot (or virtual assistant) is an algorithm that conducts a textual or oral conversation. Chatbots require considerable sophistication in their design. Firstly, they require an ability to parse human language to identify the meaning of the statements given to it. Moreover, if the conversation is oral, chatbots require an ability to translate spoken into written word. Secondly, chatbots then need to identify how to respond to each human statement by drawing on its knowledge base, taking into account the context of the previous statements made by that person. While significant progress has been made over the years in developing algorithms to perform these various tasks, AI has enhanced their quality and accuracy. Apple's iphone virtual assistant, Siri, is perhaps the most famous AI chatbot, but their pre-AI history includes Microsoft Office's infamous virtual assistant, Clippy.

AI-based chatbots are increasingly being employed by governments to help manage large citizen contact volumes as well as to help navigate complex policies and legislation. From 2016 the Australian Taxation Office (ATO) has deployed a text only chatbot named Alex.[7] The chatbot interacts live with clients online to answer taxpayer queries, and when the questions go beyond Alex's level of knowledge the conversation progresses to a human ATO employee chat operator. Over time, Alex's knowledge base has been expanded, and analysis of client interactions are used to provide insight into where to expand its abilities, as well as alert management to trends in client questions and concerns. Alex is also available for ATO staff to help them find information to answer clients' questions. Alex resolved first contact questions in 75–90% of the time, and telephone call demand dropped by 7%. The ATO has assessed that Alex has reduced the time taxpayers take to find the information needed, a savings valued at approximately AUD$9.7 million per annum. Described as "the government helper", the Alex chatbot has also been repurposed (with a different colour shirt) by IP Australia (the patents office) as a buffer before citizens speak to human operators in the telephone call centre. IP Australia's Alex is also back facing, as well as front facing, providing

> "automated checks to run a comb over a trademark examiner's work, as well as for basic assessments before it gets to the patent examiner. It also performs: Application prioritisation when a new request is received; patent auto-classification, which previously saw an individual sift through all documentation to determine which class the particular patent falls under".[8]

Ongoing developments in automated decision-making and chatbots are expected to enhance governments' and citizens' abilities in areas of complex policy and administration areas.

Public governance and public security

Governments are also increasingly deploying AI to enhance public security and public governance. Greater use of online public services comes with greater risks of data protection and privacy breaches, as well as broader cyber security challenges. AI is being used to identify emerging real time patterns to enable governmental responses in assessing denial of service cyber-attacks or other malicious cyber warfare activities. Governments are also investigating how to use AI to detect and respond to disinformation campaigns, which can be used to disrupt political processes or public safety (as with the case of COVID-19 disinformation).

Regulatory enforcement is also an important area for government use of AI. The USA Securities and Exchange Commission uses AI to identify financial reporting fraud as well as detect suspicious share trading (Engstrom et al., 2020, pp. 25–29). In a similar way, Australia's tax office and social security systems have used risk-based assessments to algorithmically identify possible tax and social security fraud, though it is not known whether such algorithms are now AI. The USA Food and Drug Administration also deploys AI for oversight of new medications, by analysing patterns in reported adverse reactions in order to ensure public safety (Engstrom et al., 2020, pp. 53–56). Writing in the midst of the COVID-19 lockdown, AI is being quickly developed to diagnose COVID-19 on xrays, enhance treatment procedures for those in intensive care, predict infection progression and identify infection hotspots (Alimadadi et al., 2020).

While AI is offering new ways to enhance public governance, there remains the potential for AI to enhance accountability of government and public services. While not evident at this stage, such accountability can occur by deploying AI to test government algorithms for accuracy. This process involves the AI submitting a range of input variables to a government algorithm to determine how the latter makes its decisions. This enables an assessment of potential bias by protected social categories (such as ethnicity, sex and religion). Using social media data, AI can also be used to identify patterns and changes in public opinion, thus providing feedback loops for identifying emerging problems, and enhancing government responsiveness to the public.

Challenges and pitfalls of using AI in government

Paralleling the excitement of the opportunities for using AI in industry and government, a raft of ethical, social, political and legal/regulatory challenges are increasingly being examined. Many national governments and international governmental organisations have prepared discussion, issues and position papers, including Australia (Dawson et al., 2019; Human Rights Commissioner, 2018); China (China. Ministry of Science & Technology, 2019); Europe (EC. High-Level Expert Group on Artificial Intelligence, 2019); Singapore (Personal Data Protection Commission, 2018); UK: (Leslie, 2019); USA (Select Committee on Artificial Intelligence, 2019); and the OECD (2019). Broad ethical and human rights concerns have also been articulated by research and non-government groups, such as Access Now (2018) and Data & Society (Latonero, 2018), and companies such as Deloitte (Hashmi et al., 2019) and IBM (Desouza, 2019).

In addition to traditional digital rights concerns around privacy, surveillance and data protection, these reports often highlight challenges of: bias and discrimination; transparency, accountability and explainability; technical accuracy; and legality and due process (Fjeld et al., 2020; Mittelstadt et al., 2016). In what follows, I speak to these broad, interrelated concerns.

Accuracy, bias, discrimination and symbolic power

Since their origins in the mid-twentieth century, computers have been accorded an aura of accuracy and objectivity. By embedding laws and policies into code, computers have been framed as making decisions based on the facts, and not based on subjective human perceptions that could be biased or mistaken. Through automation, human error is also reduced. More recently these ideas have been further propelled in discourses surrounding "big data", which are cast as providing unadulterated accounts of reality and, when comprehensive enough, form the basis for accurate models of the "real world" (boyd & Crawford, 2012; Burrows & Savage, 2014; Mayer-Schönberger & Cukier, 2013).

Such is this symbolic power of digital technologies, that, despite the well circulated aphorism Garbage In Garbage Out, academic acceptance that "raw data is an oxymoron" (Gitelman, 2013) and an awareness that categories are socially constructed and historically transient (Bowker & Star, 2000), the allure of algorithmic objectivity and neutrality remains widespread (Gillespie, 2014).

In the context of the use of AI, empirically-based concerns about bias around gender/sex and race/ethnicity are well documented. For example, Amazon discovered that its AI to sort through and select job applicants was biased against women (Dastin, 2018). The use of the COMPAS system in American sentencing and parole decision making has also been well analysed for reproducing systemic biases against Afro-Americans (Allen, 2019; Benjamin, 2019). Facial recognition systems developed in the Western world have also been critiqued for having much higher error rates for non-White people (Bacchini & Lorusso, 2019). These problematic outcomes can result from the training data used to build the AI. If that data is incomplete, inaccurate or reflect historical structural inequalities, this gets learnt by the AI and informs its outputs. The case of Microsoft's 2016 twitterbot Tay is a dramatic illustrative case. Trained on the Twitter-verse, Tay quickly started posting inflammatory and offensive tweets, resulting in it being shut down after only 16 hours online (Horton, 2016).

Technical accuracy is also a challenge with the use of AI. This is particularly acute when used in a probabilistic decision-making or predictive manner, such as to predict child maltreatment or crime recidivism. In these cases, AI calculates future events based on probabilities, which may or may not occur. False negatives and false positives can have serious consequences, and the context of these decisions often requires carefully balancing between them (Henman, 2005). For example, a false negative that a child is unlikely to be maltreated, resulting in no child protection investigation, may mean the child dies. Alternatively, a false positive may result in a child unnecessarily being removed from the care of its parent/s. Accuracy is therefore especially pertinent in such high stakes decisions.

Legality, due process and administrative justice

AI decisions need to be lawful (Miller, 2016). This seems self-evident; however, non-AI algorithms have led to decisions–such as ceasing benefit payment–that have not had legal authority. Indeed, such was the case in Australia's robo-debt scandal whereby debts were unlawfully raised by algorithms that wrongly equated annual taxation income data with fortnightly social security income data (Carney, 2019; Henman, 2017). When perceived as objective, such algorithmic decisions are accepted without question, resulting in the algorithm becoming de facto policy/law.

Given the deference accorded to AI, decisions by AI must adhere to administrative law and procedural fairness principles (Henman, 2020; Surden, 2018, 2020). Key tenants of administrative justice are confounded, including: an ability to put one's case to the decision maker; understanding the basis for an administrative decision; and a realistic (not just formal) capacity to appeal, overturn a decision, and seek remedy. Three key dimensions of AI decision-making give rise to administrative justice challenges.

Firstly, using AI to make risk assessments (of individuals) is an exercise in calculating probabilities. These are not realities or certainties. Rather, such assessments are made by relating the current case (or individual) to cases (or individuals) with similar characteristics (or profiles) and deducing similar outcomes. The legal basis of acting based on a likelihood, rather than an actuality, can be problematic (Harcourt, 2008; Henman, 2005; Schauer, 2003). For example, criminal law is based on people being tried on an offence having been committed, not on what they may do in future.

Secondly, in many areas of administration, human administrators are required to exercise professional judgement and discretion to best determine how the rules apply in complex situations. As is well documented, automating administrative decision making has reduced human discretion and gives rise to concerns that an individual's situation may not have been appropriately considered in making an administrative decision (Adler & Henman, 2009; Evans & Hupe, 2020; Garson, 1989). While AI has the capacity to be more nuanced than standard algorithms, it is also less clear how decisions on specific cases are made.

Thirdly, the foregoing observation points to the "black boxed" (Pasquale, 2015) nature of algorithms in general, and AI in particular. While complex algorithms have always made it hard to understand the basis for decision-making (Weizenbaum, 1984), they have always been coded by humans to directly implement rules and procedures. Machine learning algorithms, in contrast, generate their own internal rules to determine input to output. In doing so they add another layer of opaqueness to their decision making processes.[9]

Responsibility, accountability, transparency, explainability

Complex legal and administrative systems have been developed over time to ensure government decision making by humans is accountable, transparent and appealable. As the foregoing observations highlight, translating such systems to AI-based decision making is not always straightforward. Sometimes new laws are required to ensure that an algorithmic decision is treated as equivalent to a human decision. However, given that an algorithm does not have human autonomy and agency, identifying responsibility when an AI decision is made is challenging. Who (or what) has responsibility or liability for the error: the machine, the creators of the machine, the coders, the managers who decided to deploy the machine? This situation is further exacerbated when AI tools are developed by external organisations or companies, and deployed by governments, a situation that is likely to increase as AI becomes more mainstream and used in an "off the shelf" manner.

A key part of ensuring accountability and administrative justice is understanding how a decision was made, particularly for those who were the subject of that decision. In principle, prior to AI, policies or the computer code created by humans provide an explanation of how data about a particular case is applied to the law/code to generate an outcome/decision. With AI, this is no longer possible, because the algorithm develops its own very complex approach to process the input data. Hence, a key issue being investigated is creating explainable AI (Adadi & Berrada, 2018).

Power, compliance and control

A final challenge in using AI in government is the opportunities it provides for greater and more differentiated control of people (Henman, 2010, ch. 12). For example, with facial recognition systems embedded in CCTV, the behaviours of people in public spaces can be readily logged, as in the case of China's social credit system (Creemers, 2018). Often this increased exercise in control is not readily obvious. This is particularly the case when AI is used to generate different levels of control or scrutiny for different types or groups of people. As Benjamin (2019) observes, when new technology is deployed to increase control of minorities (in her analysis, African-Americans), the majority population is often not aware of this. Such dynamics lead to an intensification in equalities of surveillance and suspicion (Henman & Marston, 2008; Lyon, 2003), thereby exacerbating social inequalities with the potential to fragment society. As a result, it is crucial that governments' use of AI has good oversight procedures to ensure that the use of AI is done in accordance with overall collective objectives and that government officials are accountable for how they use AI.

Regulation and governance of AI

Given that the use of digital algorithms in government is not new, governments have already developed a range of ways to manage their use in government. These include developing data protection and privacy laws, clarifying and strengthening administrative laws for digital decision-making, and providing oversight bodies for use of digital technology in policing and security. AI bolsters the case to review, develop and enhance these processes. There are both technological and governance innovations that can help in regulating, governing and shaping use of AI use by governments to generate desired social outcomes.

Technological innovations

Explainability is a considerable challenge for AI based decision-making. In short, how did an AI reach the decision it did with the input data it received? Computer and information scientists are working to develop algorithms that can independently provide an approximate explanation of how an AI generated a decision in a particular case (Samek et al., 2017). Such an approach involves giving the AI lots of separate sets of input data and assessing patterns in output data to identify the key input variables that appear to make a difference in AI decision-making. While such explanations are not precise, they are able to draw attention to key drivers emergent in the AI decision-making. Similarly "adversarial testing" is an approach whereby people try to "break" an AI or make it make very wrong decisions (Qiu et al., 2019).

A similar technological approach can also be to assess the level of "bias" (or differentiation by social characteristic) in algorithms. However, the question of bias is not a simple yes/no question: when does treating different people differently count as bias/discrimination and when does it ensure appropriate forms of personalisation? For example, public health initiatives purposefully targeting women over men, and the old over the young would typically not constitute discrimination, but be appropriate differentiation. Work by McNamara et al. (2019) on bias in domestic violence recidivism, demonstrates that the question of bias is actually a question of what level of differential treatment we want between groups, in their case between indigenous and non-indigenous Australians.

While not new, software development processes are increasingly using a "privacy by design" approach (Cavoukian, 2012) whereby legal considerations of data protection and privacy are not left to the end of the product development process, but built into the very architecture (and even software) of algorithmic decision making systems. Such approaches can be made more widespread and can also be extended to incorporate other ethical considerations in AI development (Morley et al., 2019). The International Standards Organisation (ISO) and its Australian counterpart – Standards Australia – are working on building technical standards for ethical AI (www.iso.org/committee/6794475.html).

Governance innovations

Revising and strengthening legal frameworks is one area being explored. The EU's General Data Protection Regulation is arguably one of the best frameworks encompassing AI decision making in government. For example, it provides a "right to know" when a government decision was made entirely by automation, a right to an explanation, and a right to have that decision made by a human operator (Kaminski, 2019). However, it is not without limitations, as these protections do not cover cases when there is some human decision-maker involvement alongside the AI process (Wachter et al., 2017). In addition, there needs to be legal clarification around responsibility, appeal and redress in AI based decision-making.

Another approach to governing AI in government is to provide an independent quality assurance mechanism to test AIs' compliance with ethical/legal considerations. Australia's Chief Scientist, Alan Finkel, has referred to such an idea as a "Turing Certificate" (2018). Ensuring that AIs used in government are available for independent assessment is essential, and also allows the opportunity to subject them to the bias and adversarial algorithmic testing mentioned above.

A further approach is to develop practical tools and processes by which AIs can be assessed. To date, much of the discussion has been abstract and at the level of principles. Currently little work has been done on developing these practical tools, though these are emerging. Australia's Department of Industry, Innovation and Science (Dawson et al., 2019) provides a "toolkit for ethical AI" that includes impact assessments, risk assessments, review, best practice guidelines, industry standards, collaboration, monitoring, improvement and recourse mechanisms, and consultation. The UK government issued a guide to using artificial intelligence in the public sector (Government Data Service & Office of Artificial Intelligence, 2019), which focuses on assessing, planning and managing AI, and using AI ethically and safely. A more comprehensive Ethics and Algorithms Toolkit led by John Hopkins University's Centre for Government Excellence provides a "risk management framework" (www.ethicstoolkit.ai). Technology organisation VDE has also developed a framework to operationalise AI ethics (Hallensleben & Hustedt, 2020), while the Ada Lovelace Institute (2020) has provided an overview of tools for assessing algorithmic systems.

Conlusion

Artificial intelligence is generating a lot of hype and excitement about the possibilities of it enabling governments to provide enhanced services and to engage with the public, particularly in complex policy and service domains. Simultaneously, the prospects of AI in government is also generating considerable concern about its possibilities for accountability, control and the impact on social relationships. It has ever been thus. New technologies always have both positive and negative possibilities. The challenge is shaping and using AI to enhance and protect social and economic objectives.

To date, much of the discussion has occurred at a highly abstract level. It has not been helped by the fact that the moniker "AI" is nebulous, multifaceted and more marketing lingo than technical description. AI, as machine learning, definitely represents a paradigmatic change in computer science and information systems possibilities. Yet, the many ways AI is being used, or envisaged for use, is not dissimilar to non-machine learned algorithms. Consequently, the ethical, legal and social challenges are not unique to AI, but exacerbated by AI. Where AI (and non-AI algorithms) are generating new ethical, legal and social challenges is in their deployment to automate current human activities (e.g., chatbots, automated vehicles), or to do things that were not previously possible at all (e.g., make decisions in human service delivery).

This continuity, rather than disrupture, is beneficial as it means we can draw upon the insights, experiences and responses of past deployment of algorithms in government, and can transfer learnings from use of AIs in one location to another. That said, there is still a lot of unmet or poorly managed challenges that AI is now highlighting. The challenge from here is to think carefully though the various efficiency, legal, social and ethical considerations when developing and deploying AI in government. For that, new technical and governmental innovations are being developed. It is an exciting time!

Notes

1. https://www.dhs.gov/science-and-technology/news/2018/07/09/news-release-st-announces-winners-15m-prize-competition.
2. https://www.nccdglobal.org/assessment/structured-decision-making-sdm-model.

3. https://www.csiro.au/en/Research/BF/Areas/Digital-health/Supporting-hospital-and-health-systems/Waiting-times.
4. https://www.data61.csiro.au/en/Our-Research/Our-Work/Safety-and-Security/Disaster-Management/Spark.
5. https://www.torque.com.sg/news/traffic-jams-in-singapore-may-lessen-with-use-of-a-i/.
6. http://www.globaltimes.cn/content/1070546.shtml.
7. https://www.businessinsider.com.au/the-ato-launched-a-siri-for-tax-and-has-called-it-alex-2016-12.
8. https://www.zdnet.com/article/ip-australias-alex-is-more-than-just-a-chatbot/.
9. The black boxed nature of algorithms can also occur as a result of the nature of data being used, whereby data collected for one purpose in particular circumstances is repurposed for other datasets and algorithms, but in doing so misunderstand the nature and meaning of the original data. For example, Eubanks (2018) observed how the variable "failure to cooperate" poorly represented what actually had occurred. Similarly, in criminal justice systems "breach parole" can result from deliberate non-compliant behaviour, or an inability to afford public transport to reporting locations.

Acknowledgements

This article was originally delivered at *The China-Australia Dialogue on Public Administration 2019 Workshop* 'Taking advantage of new technologies', held at Sun Yat sen University, Guangzhou, China on 5-7 November 2019. I gratefully acknowledge the financial support provided by the organisers to enable my attendance and participation, and for Workshop attendees for their comments on the earlier draft of this article. This paper was prepared while working in the UQ-CSIRO Responsible Innovation collaboration.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Paul Henman http://orcid.org/0000-0002-9923-6587

References

Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability*, *11*(1), 189. https://doi.org/10.3390/su11010189

Access Now. (2018). *Human rights in the age of artificial intelligence*. https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

Ada Lovelace Institute. (2020). *Examining the black box*. https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adler, M., & Henman, P. (2009). Justice beyond the courts: The implications of computerisation for procedural justice in social security. In A. C. Martínez & P. F. Abat (Eds.), *E-Justice* (pp. 65–86). IGI Global.

Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. (2020). Artificial intelligence and machine learning to fight COVID-19. https://doi.org/10.1152/physiolgenomics.00029.2020

Allen, J. A. (2019). The color of algorithms. *Fordham Urban Law Journal*, *46*, 219–270. https://heinonline.org/HOL/Page?handle=hein.journals/frdurb46&id=234&collection=journals&index=

Bacchini, F., & Lorusso, L. (2019). Race, again. How face recognition technology reinforces racial discrimination. *Journal of Information, Communication and Ethics in Society*, *17*(3), 321–335. https://doi.org/10.1108/JICES-05-2018-0050

Benjamin, R. (2019). *Race after technology.* John Wiley & Sons.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out*. MIT Press.

boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Burrows, R., & Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, *1*(1), 2053951714540280. https://doi.org/10.1177/2053951714540280

Carney, T. (2019). Robo-debt illegality. *Alternative Law Journal*, *44*(1), 4–10. https://doi.org/10.1177/1037969X18815913

Cavoukian, A. (2012). Privacy by design. In George O.M. Yee (Ed.), *Privacy protection measures and technologies in business organizations: Aspects and standards* (pp. 170–208). IGI Global.

China. Ministry of Science & Technology. (2019). *Development of responsible artificial intelligence*. http://most.gov.cn/kjbgz/201906/t20190617_147107.htm

Clark, L. (2012, June 26). Google's artificial brain learns to find cat videos. *Wired*. https://www.wired.com/2012/06/google-x-neural-network/

Creemers, R. (2018). *China's social credit system: An evolving practice of control*. SSRN 3175792.

Cuccaro-Alamin, S., Foust, R., Vaithianathan, R., & Putnam-Hornstein, E. (2017). Risk assessment and decision making in child protective services. *Children Youth Services Review*, *79*(August), 291–298. https://doi.org/10.1016/j.childyouth.2017.06.027

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

Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., Scowcroft, J., & Hajkowicz, S. (2019). *Artificial Intelligence: Australia's Ethics Framework*. Data61 CSIRO.

De Sousa, W. G., de Melo, E. R. P., Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? *Government Information Quarterly*, *36*(4), 101392. https://doi.org/10.1016/j.giq.2019.07.004

Desouza, K. C. (2019). *Delivering artificial intelligence in government*. IBM, Center for the Business of Government.

EC. High-Level Expert Group on Artificial Intelligence. (2019). *Ethical guidelines for trustworthy AI*. European Commission.

Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). *Government by algorithm: Artificial intelligence in federal administrative agencies*. SSRN 3551505.

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

Evans, T., & Hupe, P. (2020). Conceptualizing discretion. In *Discretion and the quest for controlled freedom* (pp. 1–13). Palgrave Macmillan.

Finkel, A. (2018). *Speech: Artificial Intelligence – A matter of trust*. Australia's Chief Scientist. https://www.chiefscientist.gov.au/2018/05/speech-artificial-intelligence-a-matter-of-trust

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center Research Publication No. 2020-1. https://ssrn.com/abstract=3518482

Garson, B. (1989). *The electronic sweatshop*. Penguin.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: essays on communication, materiality, and society* (pp. 167–194). MIT Press.

Gitelman, L. (2013). *Raw data is an oxymoron*. MIT press.

Goldhaber-Fiebert, J., & Prince, L. (2019). *Impact evaluation of a predictive risk modeling tool for allegheny county's child welfare office*. Allegheny County.

Government Data Service & Office of Artificial Intelligence. (2019). *A guide to using artificial intelligence in the public sector*. https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector

Hallensleben, S., & Hustedt, C. (2020). *From principles to practice: An interdisciplinary framework to operationalise AI ethics*. VDL.

Harcourt, B. E. (2008). *Against prediction: Profiling*. University of Chicago Press.

Hashmi, A., Lalwani, R., Senatore, A., Perricos, C., & Clemancon, V. (2019). *AI ethics: The next big thing in government – anticipating the impact of AI ethics within the public sector*. https://www.worldgovernmentsummit.org/docs/default-source/default-document-library/deloitte-wgs-report-en-lq.pdf?sfvrsn=1acfc90b_0
Henman, P. (2005). E-government, targeting and data profiling: Policy and ethical issues of differential treatment. *Journal of E-government*, *2*(1), 79–98. https://doi.org/10.1300/J399v02n01_05
Henman, P. (2010). *Governing electronically*. Palgrave.
Henman, P. (2017, September 6–7). *The computer says' DEBT': Towards a critical sociology of algorithms and algorithmic governance*. Data for Policy. https://zenodo.org/record/884117#.WcTlEsh97IU
Henman, P. (2020). Administrative justice in a digital age. In M. Hertogh, R. Kirkham, R. Thomas, & J. Tomlinson (Eds.), *The Oxford handbook of administrative justice*. Oxford University Press.
Henman, P., & Marston, G. (2008). The social division of welfare surveillance. *Journal of Social Policy*, *37*(2), 187–205. https://doi.org/10.1017/S0047279407001705
Horton, H. (2016, March 24). Microsoft deletes 'teen girl'AI after it became a Hitler-loving sex robot within 24 hours. *The Telegraph*. https://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/
Human Rights Commissioner. (2018). *New technology and human rights*. AHRC. www.tech.humanrights.gov.au/sites/default/files/2018-07/Human%20Rights%20and%20Technology%20Issues%20Paper%20FINAL.pdf
Kaminski, M. E. (2019). The right to explanation, explained. *Berkeley Technology Law Journal*, *34*, 189–218. https://heinonline.org/HOL/Page?handle=hein.journals/berktech34&div=9&g_sent=1&casa_token=&collection=journals#
Latonero, M. (2018). *Governing artificial intelligence: Upholding human rights & dignity*. https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf
Leslie, D. (2019). *Understanding artificial intelligence ethics and safety*. The Alan Turing Institute.
Lyon, D. (2003). Surveillance as social sorting. In D. Lyon (Ed.), *Surveillance as social sorting* (pp. 13–30). Routledge.
Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
McNamara, D., Graham, T., Broad, E., & Soon, C. (2019). Trade-offs in algorithmic risk assessment: An Australian domestic violence case study. In A. Daly, S. K. Devitt, & M. Mann (Eds.), *Good Data* (pp. 96–116). Institute of Network Cultures.
Miller, K. (2016). The application of administrative law principles to technology-assisted decision-making. *Australian Institute of Administrative Law Forum*, *86*, 20–34. https://search.informit.com.au/documentSummary;dn=354783163762849;res=IELAPA
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679
Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 1–28. https://doi.org/10.1007/s11948-019-00165-5
OECD. (2019). *What are the OECD Principles on AI?* Retrieved September 9, 2019, from https://www.oecd.org/going-digital/ai/principles/
Pasquale, F. (2015). *The black box society*. Harvard University Press.
Personal Data Protection Commission. (2018). *Discussion paper on artificial intelligence (AI) and personal data: Fostering responsible development and adoption of AI*. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD—050618.pdf
Qiu, S., Liu, Q., Zhou, S., & Wu, C. (2019). Review of artificial intelligence adversarial attack and defense technologies. *Applied Sciences*, *9*(5), 909. https://doi.org/10.3390/app9050909
Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence. *arXiv Preprint*, *arXiv*, 08296. https://arxiv.org/abs/1708.08296
Schauer, F. F. (2003). *Profiles, probabilities, and stereotypes*. Harvard University Press.
Schwartz, I. M., York, P., Nowakowski-Sims, E., & Ramos-Hernandez, A. (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward

County experience. *Children Youth Services Review*, *81*, 309–320. https://doi.org/10.1016/j.childyouth.2017.08.020

Select Committee on Artificial Intelligence. (2019). *The national artificial intelligence research and development strategic plan: 2019 update*. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector. *Government Information Quarterly*, *36*(2), 368–383. https://doi.org/10.1016/j.giq.2018.09.008

Surden, H. (2018). Artificial intelligence and law: An overview. *Georgia State University Law Review*, *35*(4), 1305–1337. https://readingroom.law.gsu.edu/gsulr/vol35/iss4/8

Surden, H. (2020). The ethics of artificial intelligence in law: Basic questions. In M. D. Dubber, S. Das, & F. Pasquale (Eds.), *Oxford handbook of ethics of AI*. University of Oxford. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3441303

Vaithianathan, R., Putnam-Hornstein, E., Jiang, N., Nand, P., & Maloney, T. (2017). *Developing predictive models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation*. Center for Social Data Analytics.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Weizenbaum, J. (1984). *Computer power and human reason*. Penguin.