Toward Algorithmic Checks and Balances: A Rejoinder

Eyal Benvenisti[*]

In their thoughtful responses to my 2018 Foreword, Lorenzo Casini and Lorna McGregor focus mainly on its third part that addresses the challenges posed by new technologies in governance. McGregor observes that 'the use of new and emerging technologies ... is central to the sustainability and adaptability of accountability principles',[1] and Casini emphasizes the challenge of fake news being 'probably one of the most harmful contemporary threats against democracy in our time'.[2] Indeed, the spread of new information and communication technologies has already transformed the way government functions. This move to automated decisions requires public lawyers to grasp the resulting seismic shifts in democratic governance and realize the need to readjust the traditional checks and balances by providing novel legal tools for ensuring voice and holding the algorithmic government to account.

The challenge of artificial intelligence (AI) in governance is significantly different from all previous challenges posed by technological innovations. The inventions of the printing press, the radio, the television and, even recently, the smartphone[3] have not altered the basic form of engagement with government since the Athenian agora; two-way communications between the government and the governed has always been perceived as both necessary and sufficient for maintaining a functioning and accountable administration. This understanding informed the law controlling public authority that sought to facilitate the bidirectional flow of information between the citizens and their representatives in government.

[*]   Whewell Professor of International Law, University of Cambridge; CC Ng Fellow of Jesus College, Cambridge; Global Professor of Law, New York University, USA; Professor, Faculty of Law, Tel Aviv University, Israel. Email: eb653@cam.ac.uk.
[1]   McGregor, 'Accountability for Governance Choices in Artificial Intelligence: Afterword to Eyal Benvenisti's Foreword', 29(4) *European Journal of International Law* (2018) 1079, at 1079.
[2]   Casini, 'Googling Democracy? New Technologies and the Law of Global Governance: Afterword to Eyal Benvenisti's Foreword', 29(4) *European Journal of International Law* (2018) 1071, at 1072.
[3]   On the democratic effects of the spread of mobile phones, see Srinivasan and Abreu Lopes, 'Africa's Voices Versus Big Data? The Value of Citizen Engagement through Interactive Radio', in O. Hemer and T. Tufte (eds), *Voice and Matter: Communication, Development and the Cultural Return* (2016) 155; M. Mutiga and Z. Flood, 'Africa Calling: Mobile Phone Revolution to Transform Democracies', *The Guardian* (8 August 2016).

With the turn to algorithms that process raw data rather than information, the foundational premise that 'the more communication the better' has lost its bite. In the age of automated decisions, communications are neither necessary nor sufficient to ensure citizens' voice or governmental accountability. Communications are no longer necessary for a government that relies on machine learning that has no interest in citizens' input and is premised on the assumption of being omniscient. At the same time, communications are no longer sufficient for the affected public: first, because these machines are incapable of explaining their decisions (an explanation might also be protected by the designer's proprietary rights) and, second and more importantly, because there is now a much better way for monitoring government. The same AI technology can now be enlisted to review the governments' policies. As much as AI pledges better-informed automated decisions, it also promises more comprehensive monitoring of those decisions, checking them for possible bias, for weighing irrelevant considerations and for unreasonable or disproportionate balancing among relevant considerations.[4] There is no reason why the public could not benefit from an 'equality of arms' with their representatives – employing 'an AI for an AI'.

This is why a person affected by an automated governmental decision should not be content with a simple right 'to express his or her point of view [and] obtain an explanation of the decision reached' and even with the right to a review, as required by the European Union's General Data Protection Regulation.[5] The limited capacity of the inscrutable algorithm to explain its decision, coupled with 'automation bias'[6] potentially affecting the human reviewer, limits the prospects of a rigorous examination of the decision's motives and consequences. Instead, a meaningful right to an explanation must encompass the opportunity to examine the decision in its wider context, including by assessing the design of the algorithm and the assumptions underlying it as well as the data that it is instructed to process (and the data that is left out). And, as McGregor rightly emphasizes, the right to an explanation must also encompass

---

[4]   Berman, 'A Government of Laws and not of Machines', 98 *Boston University Law Review* (2018) 1277, 1325–1326 (noting that 'a programmer must make dozens of decisions that, consciously or unconsciously, impact the outcome. Questions such as which features to employ and calibrating their relative weight, how to address issues of incomplete or incorrect data, what types of models to use, which type of algorithm from among several plausible choices to employ, how to interpret the outputs, and how to measure the model's performance and determine whether it is sufficiently reliable, all must be answered and all introduce a specific bias').

[5]   Council Regulation 2016/679, OJ 2016 L119, Art. 22, Recital 71. There is a debate raging about the scope of the right to an explanation that Art. 22 provides, and the discrepancy between this article and Recital 71. See Goodman and Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"', 38 *Artificial Intelligence Magazine* (2017), available at https://arxiv.org/abs/1606.08813; I. Mendoza and L.A. Bygrave, 'The Right Not to Be Subject to Automated Decisions Based on Profiling', University of Oslo Faculty of Law Legal Studies Research Paper Series no. 2017–20 (2017), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2964855; S. Wachter, B. Mittelstadt and L. Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', *International Data Privacy Law* (2017), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469.

[6]   On the automation bias, see K. Yalin-Mor, 'The Incorporation of Judicial Decision Support Systems in Judicial Decision-Making' (2017) (PhD dissertation on file at Tel-Aviv University), at 158–163 (on file with author).

preliminary and more systemic questions about whether, when, which and how it was decided to rely on an automated decision,[7] taking into account not only instrumental considerations but also the need to ensure human dignity and a vibrant deliberative community.[8]

Such an approach could inform human decision-makers, affected parties and the general public about flawed automated decisions both before and after they are made. *Ex-ante*, accountability machines could offer alternatives to policy choices given by the government's machines, for example, by assigning slightly different weights to relevant hiring criteria: less weight to the applicants' place of residence, more weight to their education or ethnic background.[9] *Ex-post*, 'automated monitoring systems' could sift through reams of data to assess the consequences of implemented governmental policies. Just like proving discrimination through the disparate impact of policies,[10] the examination of the consequences of policies could help assess their appropriateness and compatibility with demands of equality and compliance with other constitutional demands or statutory goals.

In other words, while there is a real cause for concern that the algorithmic government is inherently inscrutable, and that it might undermine human dignity and a sense of a political community, there is also a genuine hope. The availability of AI technologies and big data open up new possibilities for making government more accountable to the public than ever before. 'Monitoring machines' or 'accountability machines' can portend greater accountability, more robust democratic controls of government and more effective human agency by auditing government for functionality (as well as for neglect and corruption) and by providing alternative policies and new findings that could reinvigorate public deliberations about policies and their alternatives. Properly designed AI-based accountability tools could probably become the most effective strategy to rebalance the newly structured governance playing field, regain citizens' ownership of democratic decision-making and ensure a community of knowledge and commitment.

AI accountability tools could and should be promoted and implemented by the same governance bodies that employ these new technologies as part of their commitment to accountability. Courts and ombudspersons could sharpen their monitoring capacities by relying on such tools. And with increased demand, it can be expected that AI technology will become widely available and used by private actors such as media outlets seeking to offer reliable news as well as by civil society activists striving to expose corruption, promote human rights and secure environmental sustainability. Academic

---

[7]   McGregor, *supra* note 1 at 1082-1085.

[8]   On these, see Benvenisti, 'EJIL Foreword: Upholding Democracy amid the Challenges of New Technology: What Role for the Law of Global Governance?', 29(1) *European Journal of International Law* (2018) 9, at 62-63.

[9]   E. Israni, *Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis* (2017), available at https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1 (critically reviewing *State v. Loomis*, 881 N.W.2d 749, 767 (Wis. 2016)).

[10] On 'disparate impact', see *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971). On monitoring the disparate impact of automated decisions, see Feldman *et al.*, 'Certifying and Removing Disparate Impact', arXiv:1412.3756v3 [stat.ML], 259 (2015), available at https://arxiv.org/pdf/1412.3756.pdf; Barocas and Selbst, 'Big Data's Disparate Impact', 104 *California Law Review* (2016) 671.

research could also benefit from AI-based research tools that produce independent knowledge about governmental policies.

The key to the success of such algorithmic checks and balances is the requirement that public and private bodies that accumulate data maintain and manage their databases and allow access to them. Obviously, this aspect raises a myriad of concerns ranging from privacy and trade secrets protection to property rights and national security issues. But such considerations, and a number of countervailing arguments, have always informed the evolution of administrative law and the law of global governance, and they are likely to continue to shape the development of these bodies of law as lawyers and activists engage with politicians and bureaucrats while seeking to adjust to the advent of new technologies and rebalance considerations of functionality and democracy.