THE CHALLENGES OF AI DECISION-MAKING IN GOVERNMENT AND ADMINISTRATIVE LAW: A PROPOSAL FOR REGULATORY DESIGN

BENEDICT SHEEHY & YEE-FUI NG*

ABSTRACT

Artificial Intelligence ("AI") decision-making is rapidly becoming a key technology in government. AI decision-making has a number of significant benefits. These include reduced costs, consistency in decision-making, and the potential for reduced bias. It is not, however, an unalloyed good. There are a number of risks associated with the technology. For example, it may have inherent biases or be prone to developing certain biases. These issues are most evident when AI is used to deal with vulnerable populations, populations unaware of or less able to deal with advanced technologies and sophisticated systems in general. With these populations, AI has the potential to do significant harm. Governments have struggled to understand and address AI decision-making appropriately. And nowhere is this clearer than in dealing with vulnerable populations. This Article addresses these issues using theories of effective regulation and two case studies of AI decision-making in Australia. It proposes two regulatory innovations: a default advocate for negative decisions "Golden Rule", and a monitoring body that, among other things, implements Professor Aziz Huq's proposed "right to a well-calibrated decision."[1]

INTRODUCTION

Significant technological advances in artificial intelligence ("AI") and machine learning over the last two decades have enabled the widespread automation of decision-making in business and government in Western liberal democracies and elsewhere. Automating high-volume decision-making, however, is clearly not an unqualified good.[2] It can have adverse effects on significant parts of populations, particularly the intended recipients of government social programs, who are the least able to address errors in government decision-making. A recent example can be drawn from Australia, where an AI system erroneously identified overpayments and calculated debts deemed to be owed by social security beneficiaries; errors of methodology led to incorrect or inflated

---

* Benedict Sheehy is a Professor of Law, University of Canberra. Yee Fui-Ng is a Senior Lecturer, Faculty of Law, and Deputy Director, Australian Centre for Justice Innovation, Monash University. The authors thank Professor Aziz Huq of the University of Chicago for many useful comments on earlier drafts. Of course, he holds no responsibility for any aspect of the final work, which rests with the authors.

1. Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 619 (2020).

2. Yee-Fui Ng et al., *Revitalising Public Law in a Technological Era: Rights, Transparency and Administrative Justice*, 43 U.N.S.W. L.J. 1041 (2020).

debt calculations for over 450,000 individuals.[3] These incorrect calculations have led to grave repercussions for vulnerable, low-socioeconomic debtors (a particularly vulnerable class in terms of dealing with AI, as will be discussed below), including individuals losing their housing and food, experiencing severe mental health issues, and even committing suicide.[4]

The debacle known nationally as "Robodebt" serves as an excellent example of the issues and concerns associated with the government's use of AI decision-making. It draws attention to the nature of issues that arise when technology design is deficient in terms of decision-making methodology.[5] It also illustrates the importance of government following the rule of law and other administrative law principles, regardless of political orientation.[6] Basic legal issues of legality, rule of law, procedural fairness, and accountability—in this case, of the administrative agency, "Centrelink"—came to the fore as matters of concern in addition to the novel transparency issues generated by the technology.[7] Among other things, users of the agency's services were unaware of the use of AI decision-making and enforcement.[8] Without that awareness, recipients mistakenly assumed that the attribution and calculation of overpayments had been conducted and checked by humans and, hence, were both appropriate and correct.[9]

This large-scale, deeply flawed implementation of a government AI system, Robodebt, has harmed vulnerable recipients, reduced public trust in AI-supported government decision-making in Australia, and drawn attention to the need to reconsider the limitations of AI decision-making.[10] In particular, there has been renewed interest in developing appropriate regulatory frameworks and systems for AI in government.[11]

Law traditionally has been conceived of as a set of rules granting authoritative decisions with respect to rights and duties.[12] In the context of administrative law, these decisions have been made within an institutional, normative framework where they are contestable and transparent (in terms of

---

3. COMMONWEALTH OF AUSTL., REPORT OF THE ROYAL COMMISSION INTO THE ROBODEBT SCHEME (2023), https://robodebt.royalcommission.gov.au/system/files/2023-09/rrc-accessible-full-report.PDF [https://perma.cc/3B7G-HUJJ] [hereinafter REPORT OF THE ROYAL COMMISSION INTO THE ROBODEBT SCHEME]; *see also* Terry Carney, *Robo-debt Illegality: The Seven Veils of Failed Guarantees of the Rule of Law?*, 44 ALT. L.J. 4 (2019).

4. REPORT OF THE ROYAL COMMISSION INTO THE ROBODEBT SCHEME, *supra* note 3, at 327.

5. *See generally* Asher Wright & Yee-Fui Ng, *Services Australia's Single Touch Payroll Program: The Enduring Legacy of Robodebt, or a Fundamentally Different System*, 33(2) PUB. L. REV. 127 (2022) (discussing Single Touch Payroll (STP), the successor to Robodebt).

6. Ng et al., *supra* note 2.

7. *Id.*

8. Wright & Ng, *supra* note 5.

9. *Id.*

10. Ng et al., *supra* note 2.

11. *See id.*

12. Paul Craig, *Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework, in* THE RULE OF LAW AND THE SEPARATION OF POWERS (Richard Bellamy eds., 2005).

rules, rulings, and procedures), and rules have been developed to ensure that decision-makers are limited with respect to bias, arbitrariness or capriciousness in decisions they may make.[13] Further, administrative law requires accountability, often in the form of an agency head or minister who can be called to account for a decision via the courts.[14] AI technology upends all of these institutional norms and so poses significant challenges for legal norms, administrative law, and regulatory design.[15] It is necessary, therefore, to consider how AI decision-making in government should be regulated.[16]

This Article addresses the question by arguing that careful and explicit attention to regulatory design using a systems perspective, as opposed to narrowly focused attention on the discrete failures in a particular benefits scheme, is required for AI decision-making to maximize its benefits of efficiency in terms of cost-effectiveness, objectivity, and consistency in terms of decision-making while minimizing the risk of harm to vulnerable populations.

This Article further draws attention to the particular applications of legal principles of administrative law within the broader legal system, ensuring that people receive their just entitlements as well as the right to procedural justice, the foundations of confidence and trust in government, and in the rule of law. This latter is a critical point because, as Chatila et al. observe, "[a] responsible approach to development and use of AI is needed to facilitate trust in AI and ensure that all can profit from the benefits of AI."[17]

This Article then applies a theory of effective regulation, a two-part theory based on systems thinking, composed of a normative theory and a positive theory, to the challenge of regulatory systems design.[18] Together the theories provide an analytical framework for regulatory systems of all types.[19] In terms of normative issues, the Article draws attention to the necessity of managing conflicting norms, such as efficiency and fairness. In terms of the positive regulatory structures, the Article proposes a new system function specifically for regulatory systems that include AI. This Article proposes a new structure that includes two functions: a complaints-driven adversarial function to advocate for benefit applicants whose claims have been denied, and an oversight function that continually monitors the AI decisions to ensure fairness, accountability, and appropriate AI functioning, as well as the calibration, remain intact. The first of these functions is supported

13. Wright & Ng, *supra* note 5.

14. Ng et al., *supra* note 2.

15. *Id.*

16. Yee-Fui Ng, *Institutional Adaptation and the Administrative State*, 44 MELBOURNE U. L. REV. 889 (2021).

17. Raja Chatila et al., *Trustworthy AI*, *in* REFLECTIONS ON ARTIFICIAL INTELLIGENCE FOR HUMANITY 13, 13-14 (Bertrand Braunschweig & Malik Ghallab eds., 2021).

18. *See* Benedict Sheehy & Donald P. Feaver, *Designing Effective Regulation: A Normative Theory*, 38 UNSW L.J. 392, 416 (2015) [hereinafter *Normative Theory*]; Donald P. Feaver & Benedict Sheehy, *Designing Effective Regulation: A Positive Theory*, 38 UNSW L.J. 961, 968 (2015) [hereinafter *Positive Theory*].

19. *Id.*

by a new AI-specific legal principle, the AI "Golden Rule" in decision-making: a rule that enables the implementation of AI decisions that administratively approve a rights claim granting access to a resource, while assigning negative determinations for review by a human administrator.[20] The second new function implements Professor Aziz Huq's newly proposed "right to a well-calibrated decision."[21] These measures will make it easier for beneficiaries to contest decisions, thus improving access to justice and decision-making ex-ante.

Accordingly, this Article contributes to the literature by analyzing the contours of AI decision-making, interlacing principles of administrative law, and applying principles of regulatory theory to the design and operation of regulation where AI decision-making is being employed by the government, with a particular focus on allocative decisions concerning benefit applicants.

To that end, this Article begins by defining AI and considering its uses in government decision-making, particularly, as noted, in the case of decisions about the granting of benefits to individuals. This first section includes consideration of the drivers and challenges of AI use (Part II). Next, the article turns to consider the term AI. It is a term used for a broad range of rapidly evolving technologies, and defining it properly focuses the article appropriately. It then turns to identify theories of regulatory design and analyze their applicability to AI (Part III). Following this, we will develop a normative framework of regulatory design for AI decision-making (Part IV). To illustrate the issues that arise from AI decision-making, we consider two Australian case studies to highlight necessary elements of AI regulatory design (Part V).

I. THE USE OF AI IN GOVERNMENT DECISION-MAKING

Artificial intelligence encompasses a wide-ranging constellation of technologies that include some form of automation and can make predictions or decisions using machine or human-based inputs, including machine learning.[22] AI includes digital systems that execute a "process."[23] In its most advanced forms AI is "autonomous" (by which it is meant that there may be limited need of human intervention after the setting of the goals), "adaptable" (meaning that the system is able to update its behavior in response to changes in the environment known as "machine learning"), and "interactive" (given that it acts in a physical or digital dimension where people and other systems co-exist).[24] Widely defined as such, AI as used for decision-making ranges from deterministic systems employing relatively simple binary logic all the way to "deep learning" machines, which make probabilistic predictions based on complex algorithms.[25]

---

20. Ng et al., *supra* note 2.

21. Huq, *supra* note 1, at 619.

22. *See, e.g.*, *id.* at 650 ("Machine learning denotes a large field of heterogenous and evolving computational forms.").

23. *Id.*

24. Chatila et al., *supra* note 17, at 15.

25. Monika Zalnieriute et al., *The Rule of Law and Automation of Government Decision-*

Deterministic systems are designed by using the derivation of rules from data, where a coded version of the rules (or law) can be understood and used by a computer.[26] This sort of expert system works well with objective criteria in decision-making, as a simple yes/no response will allow it to continue to work towards a decision.[27] An example of this type of system are systems that match data on welfare compliance, since they use pre-programmed rules to reach a decision, such as for example, that an applicant is eligible for a welfare benefit. In this type of binary deterministic system, there will be a predetermined output depending on the type of input the system receives.[28]

"Deep learning" or probabilistic systems derive rules from historic data to make inferences/predictions using machine learning.[29] With machine learning, there are models that are interpretable by humans and those that can generate models that are uninterpretable.[30] These reinforcement models are constantly updating their own models as new data comes through, as their learning base changes in step with their use.[31] An example of a predictive system would be one that determines whether a person is likely to be a recidivist, such as the U.S. sentencing tool Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)[32]—a tool that was foundational in the seminal case of *Wisconsin v. Loomis*.[33] Probabilistic predictive systems are more problematic from a rule-of-law point of view and lead to questions as to whether subjective or evaluative issues with dynamic elements for consideration can or should be assigned to automation in AI systems.

There is a further important distinction within the broader AI decision-making framework itself. That is, there are significant distinctions between *AI-assisted decisions*, where the system is used to inform a human decision-maker who relies on or utilizes certain information produced by the AI, and *AI-made decisions*, where the AI system itself makes decisions.[34] We take the position that these are not binary opposites but are best conceptualized as on a spectrum of sorts, which includes AI as advisor, AI as empirical input, and AI as co-pilot for

---

Making, 82 MOD. L. REV. 425, 443-45 (2019).

26. James Mohun & Alex Roberts, *Cracking the Code: Rulemaking for Humans and Machines* 16 (Org. for Econ. Coop. and Dev., Working Papers on Public Governance, Paper No. 42, 2020).

27. Per-Ola Olsson et al., *Automation of Building Permission by Integration of BIM and Geospatial Data*, 7 INT'L J. GEO-INFO. 307, 317 (2018).

28. Aziz Z. Huq, Constitutional Rights in the Machine Learning State, 105 CORNELL L. REV. 1875, 1915-17 (2020).

29. Zalnieriute et al., *supra* note 25, at 427.

30. Arun Rai, *Explainable AI: From Black Box to Glass Box*, 48 J. ACAD. MKTG. SCI. 137, 138 (2020).

31. Marion Oswald, *Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power*, 376 PHIL. TRANSACTIONS ROYAL SOC'Y A. 1-20 (2018).

32. Zalnieriute et al., *supra* note 25, at 437.

33. 881 N.W.2d 749 (Wis. 2016).

34. Zalnieriute et al., *supra* note 25, at 432.

ordinary/extraordinary decisions. As an example of one end of the spectrum, some forms of AI, such as advanced surveillance or facial recognition technology, are enhanced beyond mere recognition.[35] These forms of AI-assisted or "supportive" technology may be programmed to identify individuals regarded as suspects, or people fitting a profile considered likely suspect, which technology then supports the human "decision" to arrest, search, or charge.[36] At the other end of the spectrum, in terms of "replacement" technology or AI-made decisions, are AI systems that may identify relevant information based on predetermined criteria, such as the likelihood of compliance with a payment plan, and then "make a decision based upon that information without engaging a human decision-maker," such as whether an applicant qualifies for a welfare benefit.[37]

AI-assisted decision-making has become increasingly prevalent in the public sector of Western governments in a range of settings, including determining tax and social security benefits, preventing crime, and ensuring national security.[38] AI use in government spans across activities such as *enforcement activities*, using predictive tools as by the Internal Revenue Service to detect and punish fraud, *agency adjudication*, such as the determination of disability benefits, *regulatory analysis*, such as the Food and Drug Administration's utilization of machine learning to assess adverse medical events after drugs are approved, as well as *public engagement*, such as the use of chatbots.[39]

For the purposes of this article, we focus on a specific type of AI systems, namely those AI systems designed to make or assist in making decisions using the derivation of rules from data, and specifically in the area of allocation of individualized benefits.[40]

A. Drivers

There are obvious benefits of introducing automation: it allows greater speed, efficiency, and quality of decision-making. Achieving efficiency goes towards competency and adequacy of performance, towards attaining a desired effect.[41] These outcomes in turn increase the cost-effectiveness, accuracy, and precision of decision-making.[42]

---

35. Yee-Fui Ng & Stephen Gray, *Disadvantage and the Automated Decision*, 43 ADELAIDE L. REV. 641 (2022).

36. *Id.*

37. Zalnieriute et al., *supra* note 25, at 432.

38. *See, e.g.*, Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35, 42 (2018).

39. DAVID FREEMAN ENGSTROM ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES: REPORT SUBMITTED TO THE ADMINISTRATIVE CONFERENCE OF THE UNITED STATES 10 (2020).

40. *See* Huq, *supra* note 28.

41. *Aust. Sec. and Inv. Comm'n v RI Advice Grp.* [2022] FCA 496 (5 May 2022) (Austl.).

42. In an AI context, precision is defined as "the ratio of system-generated results that correctly predicted positive observations (True Positives) to the system's total predicted positive observations,

It is clear that automated government decision-making has advanced at a rapid pace in recent decades. The new public management ("NPM") movement, which has swept across Western democracies worldwide since the late 1970s, involves inserting purported private-sector management practices and principles into government, in turn leading to a focus on readily measured, narrowly construed performance measures and, consequently, an extensive effort being put into the measuring, monitoring, and auditing of agency outcomes.[43] It has provided a fertile ground for the introduction of AI technologies and a broader move in the public sector towards "digital era governance."[44]

This practice combined with AI has resulted in administrators being focused on case numbers and targets, with significantly less focus on the impacts of "digital governance" and AI on vulnerable populations.[45] As a result, new technologies in government have been deployed in ways that may negatively affect vulnerable populations—those most affected by allocative decision-making.[46] This can be seen particularly in the area of social security, where coercive interventions that have proved to be deeply detrimental to welfare recipients have been imposed in the name of protecting the vulnerable.[47] The use of AI in the welfare state has supported the "informatization" of organizations, including the surveillance of claimants through data-matching procedures to identify welfare fraud and over-payments, as well as the monitoring and measuring of departmental staff rates of processing cases.[48]

---

both correct (True Positives) and incorrect (False Positives)." *4 Things You Need to Know About AI: Accuracy, Precision, Recall and F1 Scores*, LAWTOMATED (Oct. 10, 2019), https://lawtomated.com/accuracy-precision-recall-and-f1-scores-for-lawyers/ [https://perma.cc/9PE6-6JKL] (emphasis omitted). Accuracy is defined as "a ratio of the correctly predicted classifications (both True Positives + True Negatives) to the total Test Dataset." *Id.*

43. *See* Paul Henman & Michael Adler, *Information Technology and the Governance of Social Security*, 23 CRITICAL SOC. POL'Y 139, 148 (2003); Christopher Hood & Guy Peters, *The Middle Aging of New Public Management: Into the Age of Paradox?*, 14 J. PUB. ADMIN. RSCH. & THEORY 267 (2004).

44. Michael Veale & Irina Brass, *Administration by Algorithm? Public Management meets Public Sector Machine Learning*, *in* ALGORITHMIC REGULATION 121, 122 (Karen Yeung & Martin Lodge eds., Oxford Univ. Press 2019); *see also* Christopher Hood, *A Public Management for All Seasons?*, 69 PUB. ADMIN. 3, 4-5 (1991).

45. Ng & Gray, *supra* note 35.

46. *See id.*

47. *See, e.g.*, ANDREW FORREST, THE FORREST REVIEW: CREATING PARITY 103 (2014) (arguing that there should be "a cashless welfare card system, not just for vulnerable first Australians, but for vulnerable people across Australia"); Explanatory Memorandum, Social Security Legislation Amendment (Debit Card Trial) Bill 2015 (Cth) 3 (Austl.) (stating that the restriction on welfare recipients using the cashless debit card on alcohol, gambling, and illegal drugs "is to ensure that vulnerable people are protected from abuse of these substances, and associated harm and violence.").

48. Henman & Adler, *supra* note 43, at 147-48; Huq, *supra* note 28, at 1892-99.

B. Challenges

When used as a decision-making tool, AI has some key differences from human decision-making that pose challenges to its implementation and adoption. These are referred to as "black box" problems.[49] In this AI context, they can be classified as the "technical black box" and the "legal black box."[50] The technical black box is simply a recognition of the complexities of the technology itself.[51] To the non-specialist, an AI decision-making system is akin to a magic box: some problem is put into one end and a wholly formed resolution, which takes account of all the elements, weighs them all, and places them against the criteria, comes out the other end.[52] This process is the result of highly technical system design, programming, and sophisticated algorithmic modeling—none of which is accessible to the non-specialist.[53] This technical black box leads to discomfort with AI decision-making in the first instance.[54] The discomfort is compounded by the legal black box.

The legal black box has two dimensions. First, the inability to explain AI decision-making is a significant problem when it comes to giving reasons.[55] From an administrative law perspective, one should always be able to discover the reasons why a system has made a decision,[56] and whether those reasons were legally relevant. If this is not possible, then affected persons are unable to bring an application for review, thus undermining the bedrock principle of government accountability through the contestability of government decisions in the courts. Such opaqueness of the AI decision-making systems and lack of reasons undermine its legitimacy, acceptability, and credibility—the democratic[57] and trust issues.[58] Although there is a parallel argument to be made for opaqueness in human decision-making,[59] this issue has given rise to the desire for "explainable AI" ("XAI")—a class of systems that provides visibility into how they make decisions and predictions.[60]

Breaking these down further, the first is the issue of transparency: due to the

---

49. Han-Wei Liu et al., *Beyond* State v Loomis*: Artificial Intelligence, Government Algorithmization, and Accountability*, 27 INT'L J. L. & INFO. TECH. 122, 135 (2019).

50. *Id.*

51. *Id.*

52. *See id.*

53. *Id.*

54. *Id.*

55. Jennifer Cobbe, *Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making*, 39 LEGAL STUD. 636, 648 (2018).

56. Dominique Hogan-Doran, *Computer Says "No": Automation, Algorithms and Artificial Intelligence in Government Decision-making*, 13 JUD. REV. 345, 381 (2017).

57. WENDELL WALLACH, A DANGEROUS MASTER: HOW TO KEEP TECHNOLOGY FROM SLIPPING BEYOND OUR CONTROL (2015).

58. Chatila et al., *supra* note 17, at 17.

59. Huq, *supra* note 1, at 643.

60. Rai, *supra* note 30, at 137.

nature of the system itself as a "black box" of algorithm written in computer code, as well as due to proprietary interests in the AI, which holds the inner workings to be trade secrets, leads to the outcome of the reasoning behind the decision not being readily discoverable.[61] There is, however, an even more fundamental issue in dealing with explainability in algorithms. Philosophically, it may not be possible to explain certain algorithms and indeed, even what may actually count as an explanation when dealing with algorithms due, in part, to their disputed nature.[62] The second consideration relates to issue of algorithmic bias, where the training of machine learning programs has the possibility of ingraining existing biases in the AI (or even creating new ones)—although progress is being made in this area.[63] Third, the issues of privacy and data protection must be considered due to the new challenges AI systems present.[64] These issues are at the socio-technological interface, where ethics meets the machine.[65]

For society generally, and for the social institution of government—including the legal system—to function, there needs to be confidence, a sense of trust, that the system is ethical, fair, and just—the embodiment of the principles underpinning the rule of law and doctrines animating administrative law.[66] There has been considerable dialogue on the issue as it pertains to AI applications.[67] For public confidence to be institutionalized, it must be founded on credible evidence about the AI systems, and the ethics of their interactions at the socio-technological interface, and that ultimately relies on their oversight as being within the hands of humans accountable for their operations including

61. Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1090-94 (2018) (discussing machine learning decision-making and problems machine learning pose).

62. *See* Robin K. Hill, *What an Algorithm Is*, 29 PHIL. & TECH. 35, 35-59 (2016) (discussing what defines an algorithm).

63. FINN LATTIMORE ET AL., AUSTL. HUM. RTS. COMM'N, USING ARTIFICIAL INTELLIGENCE TO MAKE DECISIONS: ADDRESSING THE PROBLEM OF ALGORITHMIC BIAS 5 (2020); *see also* CENTRE FOR DATA ETHICS & INNOVATION, REVIEW INTO BIAS IN ALGORITHMIC DECISION-MAKING 119 (2020); Jacob O. Arowosegbe, *Data Bias, Intelligent Systems and Criminal Justice Outcomes*, 31 INT'L J. L. INFO. TECH. 22, 22-45 (2023) (discussing the nature and impact of data bias in the development and deployment of AI on criminal justice outcomes).

64. Moira Paterson & Maeve McDonagh, *Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data*, 44 MONASH U. L. REV. 1, 1-31 (2018) (discussing new challenges posed by Big Personal Data).

65. TANIA SOURDIN, JUDGES, TECHNOLOGY AND ARTIFICIAL INTELLIGENCE: THE ARTIFICIAL JUDGE 237 (2021); *but see* Paul Henman, *Improving Public Services Using Artificial Intelligence: Possibilities, Pitfalls, Governance* 42 ASIA PAC. J. PUB. ADMIN. 209, 209-21 (2020) (In comparison, Paul Henman identifies issues of accuracy, bias and discrimination, legality, due process and administrative justice, responsibility, accountability, transparency and explainability; and power, compliance, and control).

66. *See* Chatila et al., *supra* note 17, at 13-34.

67. *Id.* at 14.

decisions—a "human in the loop."[68] These are all matters, principles, and insights, as argued below, that must inform regulatory design.

It is thus necessary to consider how to best regulate AI in government decision-making. We begin by turning to theorizing the interface among technology, AI, and administrative law.

II. Theorizing Technology, AI, and Administrative Law

The invention of AI and its application in law is but one instance of the general de-centering of humans in the world as knowledge and technologies extend their reach. Automating technologies, from printing presses to photocopiers to AI, all push aside humans from central positions to increasingly peripheral, supporting roles in terms of production.

As a society, we have a dualistic approach to new technologies—a mix of fascination and fear—and there is usually some cause for both. Among these technologies, however, AI uniquely de-centers not only human individuals, but also institutions such as law.[69] As Huq observes: "As a result [of AI], many people feel a loss of control over key life decisions. Machines, they fear, solve questions of critical importance on grounds that are beyond individuals' ken or control. [As a result,] individuals experience a loss of elementary human agency, and a corresponding vulnerability to inhuman and inhumane machine logic."[70]

With respect to AI in legal decision-making, it is clear that in certain areas it has both great potential along with significant risks. There are three major benefits or fascinations of AI, along with related risks or fears. As Brownsword and Honen put it: "technology intrudes on the domain of law and regulation potentially to exacerbate problems but also to offer solutions."[71] First, the benefits of AI include efficiency in decision-making. The contribution of efficiency in AI must first be considered from a legal perspective and secondarily from an economic perspective.[72] In terms of efficiency, AI should be able to make speedier decisions than humans.[73] Efficiency in decision-making by allowing cases to be processed in a more timely manner increases greater procedural

---

68. Carlo Casonato*, AI and Constitutionalism: The Challenges Ahead, in* Reflections on Artificial Intelligence For Humanity (Lecture Notes in Computer Science) 127, 143 (Bertrand Braunschweig & Malik Ghallab eds., 1st ed. 2021).

69. Roger Brownsword & Han Somsen, *Law, Innovation and Technology: Fast Forward to 2021*, 13 L. Innovation & Tech. 1, 1-28 (2021) (discussing recent changes in approaches to researching law, innovation, and technology).

70. Huq, *supra* note 1, at 613-14.

71. Brownsword & Somsen, *supra* note 69, at 27.

72. Giovanni Sartor, *Artificial Intelligence and Human Rights: Between Law and Ethics,* 27 Maastricht J. European & Compar. L. 705-19 (2020). Other values are important to users of the legal system. Natali Helberger, Theo Araujo, & Claes H. de Vreese, *Who is the Fairest of Them All? Public Attitudes and Expectations Regarding Automated Decision-making*, 39 Comput. L. & Sec. Rev. 11 (2020).

73. Huq, *supra* note 1, at 639.

fairness and improved justice as expressed in the maxim drawn from the Magna Carta "justice delayed is justice denied",[74] a principle that applies across all areas of law.[75] Efficiency in decision-making also reduces the potential injustices involved in appeal processes that are costly for the parties and delays achievement of a just or legal decision.[76] Secondarily, from an economic perspective, resources saved from inefficient decision-making can be applied to other worthwhile causes reducing the burden on government coffers and ultimately the taxpayer. The fear is that too much reliance and too many tasks are assigned to AI with insufficient human supervision to ensure justice is done in the administration of the law an issue identified by Chatila et al. as "the social component of the socio-technical system."[77]

Secondly, it is clear that AI can be more objective. Using an algorithm to sort and classify data without attention to extraneous factors facilitates objective decision-making. It is potentially less biased, with the critical caveat that the initial algorithms or training data are not biased against any particular population contrary to the legislative framework (data that reflect historic biases will continue to perpetuate those biases, and some legislation is intentionally biased against certain populations such as wealthy people over specific income thresholds, genders or race in affirmative action, etc.). The fear is that objectivity is not guaranteed as it depends on the initial programming of the algorithm—that is, the decision-making algorithm reflects intended discrimination or targeted regulation and avoids the inevitable unintended.

Further, despite the promises of efficiency and cost-effectiveness, algorithms can be trained on datasets that contain human bias and insufficiencies,[78] thereby causing the predictions to be tainted with unfair discrimination and inaccuracies.[79] For instance, a U.S. study has shown that facial recognition technologies generate a disproportionate number of false positives among non-white people, with an error rate of 40%, compared to only 5% for white people.[80] As academics at New York University noted, "[g]iven the deep and historical racial biases in the criminal justice system, most law enforcement databases are unlikely to be 'appropriately representative.'"[81] Given the nature of machine learning and

74. J. C. HOLT, MAGNA CARTA 179 (Cambridge Univ. Press, 2nd ed. 1992).

75. Robert M. O'Neil, *Of Justice Delayed and Justice Denied: The Welfare Prior Hearing Cases*, 1970 SUP. CT. REV. 161-214 (1970).

76. For an example in the criminal system, see Salma S. Safiedine & K. Jeannie Chung, *The Price for Justice: The Economic Barriers that Contribute to an Unfair and Unjust Criminal Justice System,* 32 CRIM. JUST. 40 (2017).

77. Chatila et al., *supra* note 17, at 15.

78. NICHOLAS DAVIS ET AL., AUSTL. HUM. RTS. COMM'N & WORLD ECON. F., ARTIFICIAL INTELLIGENCE: GOVERNANCE AND LEADERSHIP 9 (2019).

79. *Id.* at 2.

80. Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots*, ACLU (July 26, 2018), https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28 [https://perma.cc/B7LL-7SJJ].

81. MEREDITH WHITTAKER ET AL., THE AI NOW INST. AT NEW YORK UNIV., AI NOW REPORT

reinforcement learning in particular, AI decision-making carries a very significant risk where insufficiently monitored by appropriately trained humans[82]—a matter to which we return again in our theorizing of administrative law.

Finally, AI has the potential to be less arbitrary, more consistent. Studies have shown that human decision-makers are all too susceptible to the biological and circadian rhythms that govern all animals—bail applicants, for example, are less likely to make bail just before lunch than any other time of the day.[83] Research on asylum adjudication in the United States and Canada has shown its arbitrariness, where the outcome of refugee status determinations has largely depended upon the identity of the particular adjudicator that an application was randomly assigned to—a phenomenon that the researchers termed as "refugee roulette".[84] AI as a non-biological decision-making locale does not suffer these variations and would produce more consistent decisions—an important feature of administrative and judicial decision-making. If inaccuracies are programmed into AI design or decision-making, however, then errors are compounded at a scale with potentially hundreds of thousands of decisions or more. As seen in the Robodebt debacle, which involved the large-scale failed implementation of an automated debt system, significant mistakes in the translation of legal rules and policies into code led to incorrect or irrational determinations, with an error rate exceeding 30%.[85] These potential improvements resulting from the implementation of AI, however, are not the whole of the matter. There are other significant issues to be considered, particularly in the realm of administrative law.

Turning to consider the administrative law aspect, we adopt Stewart's approach in which administrative law "defines the structural position of administrative agencies within the governmental system, specifies the decisional procedures those agencies must follow, and determines the availability and scope of review of their actions by the independent judiciary."[86] As Stewart expounded: "[t]he traditional core of administrative law has focused on securing the rule of

---

2018, at 16 (2018).

82. Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 282 (2019) (discussing both the positive and negative effects of artificial intelligence).

83. Shai Danziger et al., *Extraneous Factors in Judicial Decisions*, 108 PROC. NAT'L. ACAD. SCI. 6889, 6889-92 (2011) (study discussing extraneous factors that affect judicial decisions made by judges); *see* Konstantin Chatziathanasiou, *Nudging after the Replication Crisis: On Uncertain Effects of Behavioral Governance and the Way Forward*, VERFASSUNGSBLOG (Aug. 30, 2022), https://verfassungsblog.de/nudging-after-the-replication-crisis/ [https://perma.cc/Q74Z-GWY8].

84. Jaya Ramji-Nogales et al., *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295, 295 (2010) (discussing research on asylum adjudication in the United States and outcome of refugee status determinations).

85. SENATE CMTY. AFFS. REFERENCES COMM., DESIGN, SCOPE, COST-BENEFIT ANALYSIS, CONTRACTS AWARDED AND IMPLEMENTATION ASSOCIATED WITH THE BETTER MANAGEMENT OF THE SOCIAL WELFARE SYSTEM INITIATIVE 32-34 (2017).

86. Richard B. Stewart, *Administrative Law in the Twenty-First Century*, 78 N.Y.U. L. REV. 437, 438 (2003) (discussing administrative law as it relates to governmental regulation).

law and protecting liberty by ensuring that agencies follow fair and impartial decisional procedures, act within the bounds of the statutory authority delegated by the legislature, and respect private rights."[87]

Administrative law is thus predicated upon the control of government action, in ensuring that government acts within legal confines, and are subject to the dictates of rationality, accountability, transparency, participation, and procedural fairness.[88] These safeguards should protect individuals against arbitrary and unlawful government decisions.

A related stream of literature on administrative justice focuses on the nature and quality of administrative decision-making by government agencies, particularly those that determine the legal entitlements of individuals, as well as the systems of redress by which people can challenge decisions of public bodies.[89] Administrative justice aims to enable accurate administrative decision-making through internal agency procedures and redress mechanisms; which is congruent with the rule of law, which seeks to restrain arbitrary power by law.[90] There are thus a broad set of public law norms and values that governments as instruments of democracy need to address: the demand for transparency, rationality, and accountability. AI decision-making poses challenges for each of these principles.

One way in which AI brings into focus and challenges these norms and values is in the means versus ends debate. This debate, which has been carried on in Western society for more than two millennia[91] from Aristotle to Machiavelli,[92] is foundational to society's conception of what is good governance—what is the good social organization is to pursue.[93] American legal scholar, Jonathan Wiener, puts it thus: "regulatory design should be about consequences—what works, how much, with what costs and side effects."[94] It is very much a live debate in the AI regulatory space: should regulation target the means—i.e., should AI technology

---

87. *Id.*

88. Ng et al., *supra* note 2, at 1041 (discussing legal practices used in governmental decision-making).

89. Simon Halliday, *Administrative Justice and Street-level Emotions: Cultures of Denial in Entitlement Decision-making*, PUB. L. 727, 727-46 (2021) (discussing nature and quality of administrative decision-making).

90. Yseult Marique, *The Rule of Law and Administrative Justice*, *in* THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE 1-12 (Marc Hertogh et al. eds., 2022).

91. W.F.R. Hardie, *The Final Good in Aristotle's Ethics*, 40 PHIL. 277, 277-95 (1965) (discussing Aristotle's regarding the means versus ends debate).

92. Basil S. Nnamdi & Tamunosiki V. Ogan, *Niccolo Machiavelli and the Morality of "The End Justifies the Means" in The Prince: A Philosophical Perspective*, 4 J. GOOD GOVERNANCE & SUSTAINABLE DEV. AFR. 33, 33-39 (2019) (discussing Machiavelli's position regarding the means versus ends debate).

93. BENEDICT SHEEHY, PARADIGMS OF LEGAL SCHOLARSHIP THAT CONNECT THEORIES, METHODS AND PHENOMENA: DOCTRINAL, REALIST, AND NON-LAW FOCUSED LEGAL RESEARCH 27 (2022).

94. Jonathan B. Wiener, *The Regulation of Technology, and the Technology of Regulation*, 26 TECH. SOC'Y 483, 491 (2004).

be prohibited from making certain decisions or activities?[95] Or, should regulation focus on the ends such as the intended policy outcomes that Wiener argues?

This debate is a foundational issue in law. Law as an institution is concerned with both procedural and substantive norms. Achieving substantive ends by illegal means is prohibited as the term "illegal" indicates. Procedural norms are foundational rights in and of themselves regardless of substantive legal outcomes. Such norms are of particular importance in dealing with state power. They form part of the legal bulwark that protects individuals from state overreach and are deeply embedded in administrative law. The introduction of AI into procedural matters challenges this legal foundation. Indeed, there is an emerging scholarship and regulatory response to the very issue of AI and procedural law.[96] AI may excel in achieving substantive outcomes but fail miserably on procedural aspects.

For example, an important principle of procedural administrative law is the principle of transparency. Transparency in government is a democratic ideal, based on the notion that an informed citizenry is better able to participate in government, thus providing an obligation on the government to provide public disclosure of information.[97] Increasing transparency in government also reduces the risk of corruption. As Louis Brandeis put it pithily: "sunlight is said to be the best of disinfectants."[98] In terms of government decision-making, it is desirable for persons affected by a decision to know *why* it was made, and understanding such requires access to the reasons for decisions and underpinning principles.[99]

The rise of AI decision-making in government, however, has created significant challenges in terms of maintaining the transparency of government decisions.[100] AI decision-making can be opaque in two ways. The first is its invisibility; citizens often do not realize that they are interacting with the technology, and generally know little about the programs that are used to make decisions about them.[101] The second challenge of AI is the "black box" problem discussed above, whereby it is possible to observe incoming data (input) and outgoing data (output) in algorithmic systems, but internal operations are poorly understood if even visible.[102] As noted by Professor Marion Oswald,

---

95. Lyria Bennett Moses, *How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target*, 5 L., INNOVATION & TECH. 5 (2013).

96. Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1973 (2021); Bart Custers, *AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law*, *in* LAW AND ARTIFICIAL INTELLIGENCE 205 (Bart Custers & Eduard Fosch-Villaronga eds., 2022).

97. Daniel J. Metcalfe, *The History of Government Transparency*, *in* RESEARCH HANDBOOK ON TRANSPARENCY 247, 249 (Padideh Ala'i & Robert G. Vaughn eds., 2014).

98. LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 92 (Frederick A. Stokes Co., 1914).

99. *Id.* at 429-30.

100. Ng et al., *supra* note 2, at 1042; Yee-Fui Ng & Maria O'Sullivan, *Deliberation and Automation: When Is a Decision a Decision?*, 26 AUSTL. J. ADMIN. L. 21, 27-31 (2019).

101. Ng et al., *supra* note 2, at 1052.

102. *See* discussion, *supra* Part I.

incorporating an algorithm into decision-making, "may come with the risk of creating 'substantial' or 'genuine' doubt as to why decisions were made and what conclusions were reached."[103] A basic requirement of AI systems used in decision-making is that they must produce decisions that are rational or in compliance with the general framework under which they are authorized.[104] Artificial Intelligence, like all apparatuses of the state whether human or non-human, must be able to account to the citizenry for decisions and actions, and make restitution for harms suffered. These principles form the foundations of accountability.

There are three facets of accountability that are particularly relevant to AI decision-making. The first is the question of responsibility: who is responsible for AI decisions? The second is the ability of individuals harmed by government actions to get legal redress. The third is independent monitoring and oversight of government AI decisions. In this administrative law context, the underlying assumption has been that an administrative decision will be made by a human, or a body comprised of humans, so that in turn there will be a responsible, liable party.[105] Where a responsible person can be identified, individuals harmed by a decision are able to seek legal redress for government decisions, which is the second aspect of accountability. Government decisions must be reviewable in the courts and tribunals, and restitution must be made to those aggrieved by erroneous decisions by public officials.

The third element of accountability is independent oversight of government decisions. AI decision-making in government, as we shall argue in detail below, should be subjected to monitoring and be audited not just internally within a department or organization, but by external bodies. This accountability function is supported by a plethora of oversight bodies or officeholders, such as ombudsmen, auditors, commissions and tribunals; or what Hood et al. colourfully called the "waste watchers, quality police and sleaze busters," who are tasked with monitoring the executive.[106] Another source of independent scrutiny is through Congress or parliamentary committees, which provide both periodic audit-like oversight and "fire alarm" responses to political problems.[107] These scrutineers can play a significant role in investigating and exposing issues relating to AI decision-making in government agencies.

The challenges to administrative law values and accountability posed by AI have not discouraged a focus on digitized services and administration by the executive branch of government. The executive branch has been focused on

---

103. Oswald, *supra* note 31, at 3.

104. Chief Justice Robert French AC, *Public Law – An Australian Perspective*, SCOTTISH PUB. L. GRP. 1, 16 (2012).

105. Oswald, *supra* note 31, at 3.

106. CHRISTOPHER HOOD ET AL., REGULATION INSIDE GOVERNMENT: WASTE-WATCHERS, QUALITY POLICE, AND SLEAZE-BUSTERS 11 (1999).

107. Barry R. Weingast, *Caught in the Middle: The President, Congress, and the Political-Bureaucratic System*, *in* THE EXECUTIVE BRANCH AND AMERICAN DEMOCRACY 312, 329-30 (Joel D. Aberbach & Mark A. Peterson eds., 2005).

installing AI systems as part of various "new political or ideological agendas, such as job shedding, replacement of skilled with un/semi-skilled staff, enhanced managerial control of workers, and increasing surveillance and control of citizens."[108] We argue that current executive action and related regulation of AI decision-making is insufficiently attentive to the larger, legal and social concerns of compliance with legal and political institutions aimed at ensuring that government behavior respects not only individual rights but also sustains the social fabric which forms the foundation for modern society. As Brownsword has it: "where in this institutional design do we find the responsibility for stewardship of the commons and for the community's distinctive values?"[109] Nowhere are these values and foundations more evident than in the allocative decisions regarding individual benefits which form the focus of this article.

This consideration of larger societal goals and coordination leads to theorizing about regulation. When considering the matter of regulatory design, the basic issues of who is to be regulated, by whom, with what powers and duties, and using what type of incentives are critical. While the notion of incomplete laws—laws that lack corresponding jural correlatives—is not new, and the expansion of thinking about law leads to a high level of clarity about the overall design of a regulatory system—a clarity that is critical to ensuring that the system is not working against itself—that the regulatory system is coherent—and works to sustain rather than undermine trust in the institutions of government.[110]

In the case at hand, it is about ensuring that the AI does not become a type of "Sorcerer's Apprentice"—a mechanism tasked with mundane human tasks that runs out of control. It is about ensuring that AI remains appropriately under the supervision of humans as it goes about achieving the desired policy ends. That is, ensuring that AI achieves the desired normative ends, and does so with minimal harms or unintended consequences, is a basic objective of regulatory design. In addition to the normative and structural elements, there needs to be attention to the procedural elements when considering regulatory design. Such issues as the preferred legal procedures, whether formal adversarial hearings, inquisitorial or mediative approaches, along with decisions about the onus of proof, standard of proof and acceptable evidence all need to be considered.

These issues are novel in the AI context because, in that setting, the traditional controls of legal decision-making, such as procedural fairness, attention to equity and substantive fairness, are no longer obviously in operation.[111] Understanding these new issues and revisiting them from the

---

108. Paul W. Fay Henman, *Administrative Justice in a Digital World: Challenges and Solutions*, *in* THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE (Marc Hertogh et al. eds., 2022).

109. Brownsword & Somsen, *supra* note 69, at 96.

110. *Normative Theory*, *supra* note 18, at 416; *Positive Theory*, *supra* note 18, at 968.

111. As Roger Brownsword and Han Somsen write: "[i]n practice, technological measures are employed for regulatory purposes by both public and private actors (e.g. by the police, the revenue, and financial regulators as much as by BigTech corporations, banks and insurance companies) without there being any prior public authorisation or debate." Brownsword & Somsen, *supra* note 69, at 1-28 (discussing recent changes in approaches to researching law, innovation, and technology).

perspective of regulatory design provides not only the potential to avoid breaching the norms of administrative law, but also reduce the risk of administrative errors. Together, these contribute to the broader and more important moral objective of avoiding visiting injustice on the populace and particularly, those vulnerable parties already in need of public support.

We turn next to consider in detail regulatory design considerations and recommendations for dealing with AI in decision-making, particularly in the instance of allocative benefits.

III. APPLYING THEORIES OF REGULATORY DESIGN TO AI IN PUBLIC DECISION-MAKING

Regulation and legal solutions more broadly are always predicated on prior theorizing about law, and hence, attention to legal theory is an important preliminary.[112] In addressing AI and legal issues, some scholars have taken an approach in which the law is given a new definition—an approach which could be said to side-step the problem.[113] In this vein, Brownsword helpfully distinguishes three versions of law. He refers to "Law 1.0" as the traditional positivist view of law in which law is conceptualized as no more than rights and duties.[114] Next, he offers an instrumentalist view of law which he denominates "Law 2.0" and which emphasizes policy outcome.[115] Finally, he arrives at "Law 3.0": a technological approach in which technologies are devised to preclude non-compliance or impose compliance.[116] Brownswords' Law 3.0 side-steps traditional positive law analysis of rights and duties and focuses nearly exclusively on state policy implementation and outcomes.[117] It is an approach that ignores procedural law with the predictable concerns of administrative law.[118] It requires a thorough reconceptualization not only of law but of the rule of law, coherence, and legal institutions more broadly—a matter for other scholars.[119]

The approach adopted in the analysis which follows adopts a more traditional approach, "Law 2.0" and conceptualizes law as a purposive normative social system. In designing effective regulation, we turn to Sheehy and Feaver's two theories of Effective Regulation—a normative theory[120] and a positive theory[121]—founded upon the philosophico-legal principle of coherence, rather

---

112. Benedict Sheehy, *Paradigms of Legal Scholarship that Connect Theories, Methods and Problems: When and How to use Doctrinal, Realist and Non-Law Focused Legal Research*, RUTGERS J. L. & PUB. POL'Y (forthcoming 2024).

113. Brownsword & Somsen, *supra* note 69.

114. *Id.* at 3-5.

115. *Id.*

116. *Id.*

117. *Id.*

118. *Id.*

119. *Id.*

120. *Normative Theory*, *supra* note 18.

121. *Id.*

than Law 3.0's ideas of technology as a rule-free solution to social problems.

Effective Regulation, it is argued, must take account of existing social institutions, formal and informal, and consider how law reform and legal actors interact with these institutions.[122] It further accounts for the political nature of problem identification and policy development, particularly the identification of norms, even potentially conflicting norms, and describes strategies for their dynamic ordering in a dynamic environment.[123] These theories provide a framework for the development, analysis and critique of legislation and its reform.

Law relies on legitimacy, understood as institutional acceptance by a wide swathe of society, because the fabric of society is dependent upon buy-in, built upon the foundational political philosophies and ideas of social contract as the basis of democratic legitimacy.[124] The legitimacy of any system of public control in a liberal democracy thus depends in the first instance on democratic support, some form of respect for law, legal institutions, and government.[125] Therefore, law, as a form of public control, relies on a sense of legitimacy.

A fundamental idea in law is that like any system, the legal system must be coherent within itself to expand and develop. Self-contradictory rules cause distortions in a system and may even lead it to self-destruct. This coherence principle is particularly important in the case of AI systems. As AI systems expand and different policy areas come into contact with one another, developing coherent approaches to problems that AI is being used to solve becomes critical. To address this problem, a commitment to coherence in law is required as a foundational commitment.

A. Normative Regulatory System Design

In terms of developing coherent regulation, following Sheehy and Feaver, there must be agreement on the problem to be solved—an agreement about an issue of concern. In Sheehy and Feaver's terminology, that issue of concern around which people and parties organize themselves and resources is referred to as the "organizing problem."[126] The organizing problem is always a social phenomenon—a group of people conclude from their deliberations that "there ought to be a law" of some type or another, regulating how people interact with one another, the social environment, and/or with the physical environment.[127] In the current instance, the use of AI technology in decision-making in the

122. *Id.*

123. *See id.*

124. *See* Benedict Sheehy, *Corporations and the Lateral Obligations of the Social Contract*, 8 NEWCASTLE L. REV. 48-71 (2004).

125. Brownsword & Somsen, *supra* note 69, at 10-16.

126. Brownsword and Somsen express it thus: the hope for law 3.0 is that "technology might be the solution to our regulatory [organizing] problems." *Id.* at 17; *see Normative Theory*, *supra* note 18.

127. *Normative Theory*, *supra* note 18, at 394-96.

administration of law is the organizing problem.

After settling on the organizing problem, the next step following the Effective Regulation framework is characterization of the problem. This step is critical in setting the boundaries of the problem. The problem posed by AI in public decision-making is clearly a public matter, a broad social problem, and does not belong in the private sector. As such, it is a matter to be dealt with by government action rather than being left to private individuals and institutions such as markets.

Having characterized it as a governance problem, the next step in Effective Regulation's regulatory design process is framing the policy response. Framing is necessary as it sets the parameters of the solution. In this instance, the framing is either as an administrative law problem or as a programming problem. This framing provides guidance on potential pathways forward. In the case of AI in decision-making, as noted some parties frame it as a "human-versus-machine" problem—a problem that is likely to result in focusing on the nature of the decision-maker. Such characterizations are reminiscent of Samuel Butler's tale, *Erehwon*, in which all machines are banned for fear of the harm they could cause[128]—an unhelpful response, although as noted, one of the typical responses to new technology.

Rather than framing it as a machine-versus-human problem, it is better framed as a problem of governance, an issue of the institutions of the state's administration—an administrative law problem. In thinking about AI and public governance, there are at least two significant value aspects to the framing. These are first, the legal value of the rule of law as expressed in terms of accountability for and contestability of administrative decisions along with due process, rationality (reasonableness), consistency, and timeliness in decision-making.[129] Law, regardless of AI technology, as argued above, still demands procedural fairness, including transparency, throughout the process.[130] There is also a second set of values, however, namely, the economic values of efficiency and cost-effectiveness. These two distinct value constructs are in conflict and thus lead to a normative ordering problem.

Explicit political decisions will need to be made about normative ordering. The decision will be about when to prioritize legal values and when to prioritise economic values. This decision will be crucial for allocative decisions about benefits as a matter of law's fairness principles, whereas AI implementation is driven to a significant extent by economic priorities of efficiencies.

As a tool of government, AI must embody administrative law principles and doctrines, as well as policy objectives and institutional norms, as these are found

---

128. SAMUEL BUTLER, EREWHON, OR, OVER THE RANGE (London: David Bogue, 1880); *see* Brownsword & Somsen, *supra* note 69.

129. ADMIN. REV. COUNCIL, AUSTL. GOV'T: ATT'Y-GEN'S. DEP'T, REPORT 46 - AUTOMATED ASSISTANCE IN ADMINISTRATIVE DECISION-MAKING 2004 (2004).

130. Australian Academy of Science, *Artificial Intelligence: Future Directions in Technology and Law*, YOUTUBE (Oct. 7, 2021), https://www.youtube.com/watch?v=229xD-gKGS8 [https://perma.cc/9H34-EUU9].

encapsulated in the notions of administrative justice and the rule of law, combined with the institutional norm of being a model litigant.[131] Administrative justice theories focus on achieving a "just" decision through various institutional mechanisms for challenging government decisions, such as internal review, judicial or tribunal review, and the operation of oversight bodies such as Ombudsmen and Auditors-General.[132] At a broader level, administrative law derives its foundations from tenets of the rule of law, which incorporate the institutionalization of the social contract in the sense that parties surrender power to a higher authority who wields that power in a manner that treats people equally, is unbiased, without favor or fear of any person or group in society. This institutionalization is embodied in positive law, particularly in administrative law, the norms of which require the exercise of discretion, merit-based decision-making, and as noted, holds the procedural safeguards of fairness, transparency, contestability and accountability.[133] The institutionalization of these principles is realized in the ultimate goal of the administrative state which is to "bring competence to politics through the progressive submission of power to reason," towards achieving "legitimate, liberal democratic governance."[134]

The noted issue in the case of reliance on AI decision-making is achieving algorithmic transparency which would require "clarity in the procurement, implementation and technical mechanisms associated with automated decision-making systems," and this extends to disclosure of the fact, extent, and operation of AI in decision-making.[135] As the UK House of Lords noted in its major report on AI: "[e]ach individual should have access to the rationality behind a decision being made. The process needs to be transparent and easily understood by society."[136] Contestability suggests that AI-made or AI-assisted decisions must be able to be challenged by affected individuals in the tribunals or courts; for example, through judicial review procedures. Accountability suggests that AI-made decisions need to be explainable, transparent, and subject to periodic review by an independent authority.

---

131. Theodore Eisenberg & Henry S. Farber, *The Government as Litigant: Further Tests of the Case Selection Model*, 5 AM. L. & ECON. REV. 94, 94 (2003); Gabrielle J. Appleby, *The Government as Litigant*, 37 U.N.S.W. L. J. 94, 94 (2014).

132. T. T. Arvind et al., *Judicial Review and Administrative Justice*, *in* THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE (Marc Hertogh et al. eds., 2022).

133. Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1203 (2017); Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 50 (2019); Jeffrey L. Mashaw, *Small Things Like Reasons Are Put in a Jar: Reason and Legitimacy in the Administrative State*, 70 FORDHAM L. REV. 17, 26 (2001).

134. JERRY L. MASHAW, REASONED ADMINISTRATION AND DEMOCRATIC LEGITIMACY: HOW ADMINISTRATIVE LAW SUPPORTS DEMOCRATIC GOVERNMENT 11 (Cambridge Univ. Press 2018).

135. TOBY WALSH ET AL., CLOSER TO THE MACHINE: TECHNICAL, SOCIAL, AND LEGAL ASPECTS OF AI 50 (Cliff Bertram et al. eds., 2019).

136. ALL-PARTY PARLIAMENTARY GROUP ON ARTIFICIAL INTELLIGENCE, ETHICS AND LEGAL IN AI: DECISION MAKING AND MORAL ISSUES, BIG INNOVATION CENTRE (2017).

There are a variety of reasons why AI decisions may lack transparency. Expanding on the earlier discussion of the technical and legal black boxes, they can be described in summary as the issue of opacity. This opacity, as noted, may stem from three sources: "(1) opacity as intentional corporate or state secrecy (2) opacity as technical illiteracy, and (3) an opacity that arises from the characteristics of machine learning algorithms and the scale required to apply them usefully."[137] In the context of administrative decision-making, it is suggested that at least parties subject to decisions, if not human decision-makers too, suffer from reasons two and three. These limitations themselves could pose a serious challenge for administrative agencies that choose to rely on AI; however, before restricting AI on this basis alone, one ought to consider the opacity issues with human decision-makers. As Huq argues for a "[r]ough equality between human and machine transparency . . . . Other minds are just as much black boxes as are machine-learning instruments."[138] Accordingly, transparency arguments may not have significant traction when considering AI. A further concern arises with consideration of the principles of procedural fairness. These principles require a right to participate in the decision-making process, a right to advocate one's own behalf, and the right to have reasons.

Returning to the main discussion of framing and explicit attention to normative ordering—a discussion about the overall values government is seeking to implement, whether justice or economic—requires clear and explicit acknowledgement of and deliberation about the trade-offs, in this case, the distribution of costs and benefits. It is the task of legislators and policy makers to address these matters directly and explicitly. In the case of the of the implementation of AI, it is explicitly driven by a government imperative of saving costs—Wiener's economic argument.[139] In that process, however, the use of AI will certainly impose costs on others including, of particular concern, vulnerable populations who may have additional difficulties navigating government bureaucratic procedures, website-based forms and related technologies, and simply understanding their legal rights and remedies—Chatila's social interface concern.[140] Perhaps most significantly in cases of social services for underprivileged and vulnerable populations, due to their lack of knowledge of the legal, governmental, and technological interfaces, the procedural barriers to challenge decisions—whether done by AI or humans—are overwhelming, precluding even the lodgement of claims for benefits, objections when denied. Additionally, without a human interface to provide some assistance, empathy, and advice, these disadvantaged people are more likely to be wholly without options. Accordingly, as government increasingly relies on AI decision-making for decision-based entitlements on social security based on "rigid eligibility criteria

137. Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1-2 (2016).

138. Huq, *supra* note 1, at 643.

139. Wiener, *supra* note 94.

140. Chatila et al., *supra* note 17.

and tight arithmetic logic,"[141] the intended beneficiaries will face a numerical increase in the number of decisions which may go against them, and thus they will need additional support in contesting such decisions. As a result, explicit attention and debate regarding the normative ordering and related cost-benefit analysis—especially as they are imposed on vulnerable intended beneficiaries—is an imperative on policy makers.

In making decisions about the normative ordering, debate concerning the underlying values—whether justice, fairness, efficiency, equity, or some other value—are of a foundational nature. These values include the liberal democratic principles of the rule of law and egalitarianism. Egalitarianism in particular needs to address the effects of inequity when evaluating the distribution of the social costs of AI. Sustaining these foundational values requires three distinct regulatory design features, which we discuss next in terms of Effective Regulation's policy objectives, targets, and distributions.

First, in terms of policy objectives, to achieve the desired policy objective, the algorithms on which the AI is operating must include the politically determined normative ordering. That is, the algorithm must have a broad default setting that preferences equity over economics. Further, it must include a specific algorithm to implement the broad default at the individual case level, again preferencing equity over economics where a case is on the margin.

Second, in terms of policy objectives, it is necessary to accept that in certain instances it is preferable to have AI autonomy limited in some or other aspect. While AI may have the technical capacity to function autonomously, in these instances direct human oversight is preferable. Brownsword and Somsen offer an example in which the technology is capable of operating without such oversight arguing that "a community might attach particular value to preserving both human officials (rather than machines) and rules (rather than technological measures) in the core areas of the criminal justice system."[142] The preference for limiting AI autonomy comes from the normative aims.

In essence, the limitations are necessary for purposes of normative ordering—a critical matter in regulatory design. While economic values prioritize efficiency, the welfare state prioritizes equity and rule of law considerations.[143] Rule-of-law itself embodies prior democratic values, and these can become pitted against economic and political concerns in the debate about values in AI decision-making. As Wendell Wallach puts it: "Bowing to political and economic imperatives is not sufficient. Nor is it acceptable to defer to the mechanistic unfolding of technological possibilities. In a democratic society, we—the public—should give approval."[144] Certainly, Wallach is correct: that to surrender

141. Terry Carney, *Social Security Law: What Does the Politics of 'Conditional Welfare' Mean for Review and Client Representation?*, 46 AUSTL. J. SOC. ISSUES 233, 236 (2011).

142. Brownsword & Somsen, *supra* note 69, at 82.

143. Donald Feaver & Benedict Sheehy, *The Shifting Balance of Power in the Regulatory State: Structure, Strategy, and the Division of Labour*, 41 J. L. SOC'Y 203-226 (2014).

144. Brownsword & Somsen, *supra* note 69, at 97 (quoting WENDELL WALLACH, A DANGEROUS MASTER: HOW TO KEEP TECHNOLOGY FROM SLIPPING BEYOND OUR CONTROL 10

to unbridled political ambitions and technological innovation is antithetical to democracy. It is insufficiently precise, however, as a response to the problem of AI technology and administrative decision-making, and that pushes the normative ordering task to delve deeper. Accordingly, it should be ensured that, where these normative principles are in tension, equity and rule of law should take primacy such that all design incorporates those values into the algorithms aimed at improving efficiency. One way to ensure that this occurs is by requiring human intervention whenever a decision adverse to an individual applicant is made.

To a certain degree, the idea is that drawing in human cognition and decision-making has the potential to enliven empathy—something which AI is without and is a hallmark of law's ideas of equity.[145] To this end, there is emerging discussion of a right to human decision-maker. As Huq explains, it is "a right to a human decision rather than a decision reached by a purely automated process (a 'machine decision')."[146] While the EU has a directive granting that right,[147] it is still early in its development as a legal phenomenon.[148]

Third, addressing the social side of the socio-technology innovation, requires particular attention. As noted, government is to be a model litigant. In this case, where executive government is embodied in AI technology, interacting with a disembodied decision-maker puts an additional burden on the already disadvantaged, vulnerable benefit seeker. Accordingly, attention is required to ease the burden on the applicant to contest adverse decisions and so improve access to justice.[149] This issue could be addressed by consideration of the design of the regulatory infrastructure, a matter to which the article turns next.

B. Positive Regulatory System Design

A solution presents itself for structuring a regulatory system by turning from Sheehy and Feaver's *normative* theory to their *positive* theory of Effective Regulation.[150] The positive theory approaches governance problems by considering structural options in regulatory systems that would work in a coherent manner and align with the normative policy objectives—in this case, the delivery of social services in a fair, effective and efficient manner.[151] At a very basic level, two decisions need to be addressed: who ought to be regulated and

---

(2015)).

145. *See* Stephen Humphreys, *Equity before 'Equity'*, 86 MOD. L. REV. 85-121 (2023).

146. Huq, *supra* note 1, at 615.

147. Article 22 provides a natural individual with "the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Regulation 2016/679, 2016 O.J. (L 119).

148. Christina Tikkinen-Piri et al., *EU General Data Protection Regulation: Changes and Implications for Personal Data Collecting Companies*, 34 COMPUT. L. & SEC. REV. 134, 134-35 (2018).

149. A right proposed in Kaminski & Urban, *supra* note 96.

150. *Positive Theory*, *supra* note 18.

151. *See id.*

what structures ought to be in place.

In the first instance, the issue of which party will be subjected to regulation—the administrative agency, or the AI decision-maker—needs to be considered. Given that the AI decision-maker—a technology—is at the heart of the question, it would appear to be the appropriate party. While the inanimate object cannot be morally, and hence legally, responsible for its actions, nor is it consciously making decisions, neither of these issues of morality or consciousness are necessary conditions for it to be subjected to regulation. Quite simply, developing rules which limit what a technology can do or how it is used is no different from placing limits on what side of the road a car must travel. These basic rules place restrictions on how the technology can be used. Limiting AI decision-making to certain types of decisions—i.e., positive outcomes—and limiting its power to execute negative decisions, is an appropriate regulation in the first instance. Responsibility and liability for doing so and ensuring it continues to work in this manner would appropriately rest with the agency authorized by law to see to its operation.

Secondly, in designing a new regulatory system, novel components could be constructed to address the social policy and rule of law objectives being implemented. Such new design answers Brownsword's call for a redesigned institutional framework to address technological innovations in a Law 3.0 context.[152] We propose a new component for a regulatory system aimed at AI. It would be a body or agency dually tasked as follows: first, it would have some form of administrative review which would provide advocacy for negative outcomes and second, a system monitoring task which provides constant review of system operations, outcomes. As a review body, in the first instance, it could be a complaint-driven advocacy focused agency action on behalf of those denied benefits—taking the cases of those denied and ensuring that the decision-making has paid due attention to the broader policy objectives of relieving hardship. While traditionally this role has been taken on by community legal aid clinics, funded through legal aid programs or as pro bono work by lawyers,[153] given the savings generated by the automation of decision-making, some of the savings could be redirected to ensure benefits are not unfairly denied to society's more vulnerable populations. A limitation of this model is that it still requires negatively affected individuals to have the knowledge, determination, and resources to challenge the administration—characteristics scarce among that population by definition. One way to address this challenge, foreshadowing the Golden Rule discussed below, would require certain high stakes decisions, such as disability pension entitlements, to create an automatic review obligation on the review body to contest the decision.

As to the second task, the body could contain the functions of an oversight executive agency such as a commissioner or ombudsman who would be empowered to oversee and scrutinise government AI decision-making on an

---

152. Brownsword & Somsen, *supra* note 69, at 96-97.

153. April Faith-Slaker, *What we Know and Need to Know about Pro Bono Service Delivery*, 67 S.C. L. REV. 267 (2015).

ongoing basis, shifting the oversight from complaints-driven adversarial procedure to a continual monitoring function—implementing something like Huq's "right to a well-calibrated machine decision" ensuring that the AI is properly serviced, updated and functioning as intended.[154] For instance, the Australian Human Rights Commission recommended the establishment of an AI Safety Commissioner, which the function of supporting "regulators, policy makers, government and business" to "apply laws and other standards in respect of AI-informed decision making."[155] Another option, as implemented in the United Kingdom, is to ensure oversight in the procurement of AI technologies—government departments are required to obtain approval from the UK Government Digital Service before spending on AI technologies.[156] This approach would not suffer from the weakness of relying on the energy and resources of disempowered complainants and would be one way to implement the Golden Rule, discussed below.

We turn now to examine how these proposals would have worked in two specific cases.

IV. TWO AUSTRALIAN CASE STUDIES

Although there is growing interest in AI by regulators as demonstrated by recent reports by the Australian Human Rights Commission,[157] the Office of the Victorian Information Commissioner,[158] the Australian Information Commissioner,[159] and Data61's Ethical Framework for AI,[160] Australia remains in its relative infancy in conceptualising AI regulatory design. Accordingly, the following two case studies are used to offer insight into dramatically different approaches to AI decision-making and related outcomes for government, benefit claimants and Australian society as a whole. The case studies are the Robodebt case referred to earlier, and a case drawn from the Department of Defence and Veteran's Affairs.

A. Robodebt

The Robodebt debacle illustrates failures of design and implementation of AI decision-making on the back of very poor governmental initial decision-making.

---

154. Huq, *supra* note 1.

155. AUSTL. HUM. RTS. COMM'N, FINAL REPORT: HUMAN RIGHTS AND TECHNOLOGY (2021).

156. CENT. DIGIT. & DATA OFF. AND OFF. FOR A.I., ASSESSING IF ARTIFICIAL INTELLIGENCE IS THE RIGHT SOLUTION (2019).

157. AUSTL. HUM. RTS. COMM'N, *supra* note 155; AUSTL. HUM. RTS. COMM'N, HUMAN RIGHTS AND TECHNOLOGY: DISCUSSION PAPER (2019) [hereinafter DISCUSSION PAPER].

158. OFF. OF THE VICTORIAN INFO. COMM'R, CLOSER TO THE MACHINE: TECHNICAL, SOCIAL AND LEGAL ASPECTS OF AI (2019).

159. AUSTL. INFO. COMM'R, GUIDE TO DATA ANALYTICS AND AUSTRALIAN PRIVACY PRINCIPLES (2018).

160. DAVE DAWSON ET AL., ARTIFICIAL INTELLIGENCE AUSTRALIA'S ETHICS FRAMEWORK: A DISCUSSION PAPER (2019).

Robodebt was a system designed to identify benefit recipients who appeared to have funds in excess to threshold requirements.[161] It identified those recipients by averaging their income over a period of time.[162] The system then issued debt notices to those recipients and where the repayments were not received, commenced recovery actions against them.[163] This system was designed following a government decision to reverse the onus of proof of debt: while traditionally the creditor, in this case the government, is required to prove the debt, in this case, government decided that recipients would be required to disprove debt.[164] We turn next to examine the AI and related legal flaws of both the AI and the government in the system.

The system relied on a data-matching program between the tax office and the department charged with responsibility for payment of benefits. In the first instance, the design of the system of data matching was flawed and resulted in a discrepancy or error rate of 20%,[165] meaning that in many instances the government was pursuing incorrect, inflated or even non-existent debts. For instance, in the case of *Fletcher*, the Tribunal noted that despite fortnightly income being correctly recorded as $3,563.00, annual income had been recorded as $23.83.[166] In the judgement against the government, the Tribunal's Senior Member Puplick stated that "[t]he Tribunal cannot understand how such errors occur, nor why it appears that the Department expects its clients to understand this either."[167] Similarly, in the case of *Amato v Commonwealth* challenging the Robodebt scheme of incorrect debt calculation, while the debt was initially calculated at $2,700, a Freedom of Information request revealed that Ms. Amato only owed $1.48.[168]

These errors exemplify the importance of basic administrative law principles; namely, the importance of transparency, and contestability, and particularly for vulnerable populations. Given the high stakes involved for those people dependent on these payments, understanding the calculations which could deprive them of food and shelter, the basics of human existence, the necessity of prioritizing transparency and contestability are obvious.

The second aspect to the lack of transparency was that the benefit recipients were unaware of the automated nature of the decision-making.[169] As a result, they assumed that the debts had been checked by human operators and were

---

161. Ng et al., *supra* note 2.

162. *Id.*

163. *Id.*

164. *Id.*

165. SENATE CMTY. AFFS. REFERENCES COMM., *supra* note 85, at 33-34, 36-38, 79-84, 107.

166. *Fletcher; Secretary, Department of Social Services (Social Services Second Review)* [2021] AATA 577, 90 (Austl.) [hereinafter *Fletcher*].

167. *Id*.

168. National Legal Aid, Submission No. 26 to Senate Community Affairs References Committee, Parliament of Australia, *Inquiry into Centrelink's Compliance Program* 25-26 (Sept. 27, 2019).

169. Ng et al., *supra* note 2.

accurate.[170] They had much less confidence in contesting the debt assessed against them.[171] The lack of transparency on the issue further disempowered these beneficiary recipients, who felt further out of control of their lives due to the lack of transparency on the nature of the decision-maker.[172] Again, it was a government choice to use AI and to not disclose that use—and as such, it is not a necessary feature of the employment of AI. Rather, it was a government choice to obscure its decision-making procedure.

Third, the government decision to reverse the onus of proof forced vulnerable welfare claimants to disprove their alleged debts.[173] This reversal was unconscionable and wrong from an administrative law perspective: pragmatically recipients did not understand how the debts were calculated in the first place, making it nearly impossible to challenge them. Further, the reversal of the onus undermines a core principle of the rule of law and administrative law, namely, the principle of accountability.[174] Government is to be accountable for its decisions providing the evidence in the first instance. The use of an AI system may have served the government as it provided a type of a shield or smokescreen—a type of "the AI system did it" blame shifting. But it was the government's choice in the first instance, rather than a necessary feature of AI. Further, as noted, this decision is contrary to properly construed debtor-creditor law in which the onus clearly rests on the creditor-claimant not the debtor.[175] There is no reasonable excuse or lawful justification for the government's decision in this instance.

This reversed onus aspect of the debacle was exacerbated by further pragmatic difficulties imposed by government on vulnerable recipients when attempting to contest their supposed debts. Government decisions concerning complaint processes resulted in efforts to contact the responsible department, the Department of Human Services, to discuss the debt, being delayed and frustrated inexcusably—calls were unanswered for hours.[176] These difficulties in contacting the Department were clear procedural barriers that undermined recipients' ability to exercise the administrative law right to contest debt decisions. Again, it is important to note that such poor responsiveness is a matter of government choice rather than a necessary consequence of using an AI system.

From the perspective of regulatory design, the lack of an appropriate accountability body in the regulatory system allowed this system to be structured positively as it was and as a result undermined the government's ability to comply with administrative law obligations. Had a dually tasked agency (advocacy and monitoring) been inserted into the regulatory system, it could have picked up and addressed these calls, and quickly identified and remedied the issues, or terminated the program in a timely manner.

---

170. *Id.*
171. *Id.*
172. *Id.*
173. *Id.*
174. *Id.*
175. *Id.*
176. SENATE CMTY. AFFS. REFERENCES COMM., *supra* note 85, at ix, 33-34,107.

Fortunately, the calamity made its way through the courts, the Law 1.0 system, and the government was held liable in a test case challenging the Robodebt system.[177] The government ultimately responded to what became class action and which was settled for $1.8 billion.[178]

The extent of the Robodebt disaster was such that among the first acts of the new government, adopting the recommendation of a senate investigation established the Robodebt Royal Commission in 2022.[179] Chaired by a former judge, the Commission investigated the establishment, design and implementation of the Robodebt scheme, as well as issues and outcomes that arose from that scheme.[180] As Murphy J, when hearing the *Prygodicz* case noted, the Robodebt debacle has been "a very sorry chapter in Australian public administration"[181]—a view echoed by the Commission. It represented a monumental failure of government in its design and implementation of AI, resulting in incorrect debts being levied on hundreds of thousands of citizens, many of them among the most vulnerable members of the population.

The Commission's final report was scathing, lambasting the Robodebt scheme as an "extraordinary saga" of "venality, incompetence and cowardice".[182] The Commissioner found that former conservative Prime Minister, Scott Morrison, who was Social Services Minister at the time, had misled Cabinet and "failed to meet his ministerial responsibility to ensure that Cabinet was properly informed about what the [income averaging] proposal actually entailed, and to ensure that it was lawful."[183] Another conservative Minister who oversaw the Robodebt scheme, Alan Tudge, and his ministerial adviser, were criticized for a "mean-spirited" media strategy that sought to put forward media case studies about "welfare fraudsters," as well as leak negative personal information to the

177. Order of Davies J in *Amato v Commonwealth* (Federal Court of Australia, VID611/2019, 27 November 2019) 6 [9].

178. *Prygodicz v Commonwealth of Australia [No 2]* [2021] FCA 634 (11 June 2021) (Austl.).

179. *About the Royal Commission*, ROYAL COMM'N INTO THE ROBODEBT SCHEME, https://robodebt.royalcommission.gov.au/about#:~:text=The%20Robodebt%20Royal%20Commission%20was,the%20intended%20or%20actual%20outcomes [https://perma.cc/2URT-JZJP] (last visited Mar. 15, 2024); SENATE CMTY. AFFS. REFERENCES COMM., ACCOUNTABILITY AND JUSTICE: WHY WE NEED A ROYAL COMMISSION INTO ROBODEBT (2022); Luke Henriques-Gomes, *Senate Committee Calls for Royal Commission into Robodebt Scandal,* GUARDIAN (May 13, 2022), https://www.theguardian.com/australia-news/2022/may/13/senate-committee-calls-for-royal-commission-into-robodebt-scandal [https://perma.cc/2ZVF-2EG6].

180. *Terms of Reference*, ROYAL COMM'N INTO THE ROBODEBT SCHEME, https://robodebt.royalcommission.gov.au/about/terms-reference [https://perma.cc/GU3X-UMWH] (last visited Mar. 15, 2024).

181. Luke Henriques-Gomes, *Robodebt Responsible for $1.5bn Unlawful Debts in 'Very Sorry Chapter', Court Hears*, GUARDIAN (May 7, 2021), https://www.theguardian.com/australia-news/2021/may/07/robodebt-responsible-for-15bn-unlawful-debts-in-very-sorry-chapter-court-hears [https://perma.cc/ZE7F-Q8G4].

182. REPORT OF THE ROYAL COMMISSION INTO THE ROBODEBT SCHEME, *supra* note 3, at 702.

183. *Id.* at 106.

media about social security recipients who complained about the Robodebt scheme, which was intended to intimidate Robodebt victims from speaking up.[184]

Worryingly, the Commission's report showed that government ignored legal advice. Senior public servants had legal advice from 2014 that the scheme was unlawful, but under pressure from government, scrambled desperately to hide the illegality of the scheme by refusing to finalise legal advice and leaving it in draft form, deceiving the Commonwealth Ombudsman, wilfully ignoring tribunal decisions ruling the scheme to be unlawful, and deliberatively misleading Cabinet about the nature of the income averaging scheme.[185] Front line public servants who raised concerns about the scheme were silenced by managers and faced professional retribution, and the position of a Departmental Secretary who sought to stop the scheme was abolished.[186] Consequently, although the "unfairness, probable illegality and cruelty" of the scheme was apparent from the beginning of 2017, the scheme persisted until late 2019.[187]

These deliberate and illegal governmental choices in implementing AI decision-making led to great stress, anxiety, stigma, and even suicide within the vulnerable populations who were supposed to be the beneficiaries, the recipients of government support rather than alleged debtors.[188] Such perverse consequences are precisely the types of safeguards administrative law is designed to provide. And indeed, in its review of the Robodebt program, the Senate stated: that the scheme "relied extensively on online systems and data-driven processes . . . [u]se of technology by Government must be supported by appropriate safeguards, especially to protect vulnerable people, . . . [including] rights to an explanation of administrative decisions and to have those decisions reviewed."[189] The Robodebt saga was an unnecessary and serious failure on the part of government. These decisions and errors are precisely the types of errors that have been used by advocates in the USA to promote a right to a human decision-maker.[190] An impoverished woman was denied public housing as a result of an AI error only overcome when human intervention came to her aid.[191] Such errors undermine not only rule of law, but also public confidence in AI decision-making, particularly for the most vulnerable, the recipients of public benefits.

This leads to a consideration of the issue of norms addressed in Sheehy and

---

184. *Id*. at 177.

185. *Id.* at 217-19, 541, 641.

186. *Id.* at 390-91, 658.

187. *Id.* at xxvii.

188. Tom Stayner, *Mothers whose Sons Took Their Lives after Robodebts Detail Anguish in Heartbreaking Letters*, SBS NEWS (Aug. 18, 2020, 11:20 AM), https://www.sbs.com.au/news/mothers-whose-sons-took-their-lives-after-robodebts-detail-anguish-in-heartbreaking-letters [https://perma.cc/8TVB-RVWH].

189. SENATE CMTY. AFFS. REFERENCES COMM., *supra* note 179, at ¶ 2.46.

190. Bart Custers, *AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law*, *in* LAW AND ARTIFICIAL INTELLIGENCE (Bart Custers & Eduard Fosch-Villaronga eds., 2023).

191. *See* Huq, *supra* note 1, at 616.

Feaver's Effective Regulation.[192] They theorize that a regulatory system must be coherent to work and in particular, that conflicting normative foundations will cause the system to fail.[193] Where there are conflicting norms, they must be ordered and given different priorities. Clearly the case of Robodebt included a failure of normative ordering. As noted above, there was a conflict between the administrative law norms for fairness and transparency and the norms of efficiency underpinning economics.[194] The lack of effective handling of normative ordering was noted by the Royal Commission, "[a]n enthusiasm for savings would seem an anathema to the underlying policy and rationale for social security spending, of supporting those in need."[195] Failure to identify and address the normative policy objectives and prioritise them appropriately, was a fundamental flaw in the design of the regulatory scheme.

Although the Robodebt scheme was set up as a cost-saving measure, prioritising efficiency norms over legal, to claw back $1.7 billion in allegedly overpaid welfare from recipients, it ended up recouping just $406.2 million, while costing the government $971.4 million in implementation, administration and wind-up costs.[196] Ironically, the scheme that was supposed to save the government money ended up costing the government $565.12 million overall.[197] It was a scheme of a conservative government, seeking to undermine the social security safety net that forms a foundation for post-Great Depression developed economies world-wide. In other words, at the foundation of the scheme, the government aimed to advance its own normative agenda in priority in contradiction to the foundational normative purpose of the legislation.

Among the Commission's other recommendation was that the government consider legislative reform to introduce a consistent legal framework for government's AI decision-making, as well as that it establish an independent body to monitor and audit AI decision-making processes for their technical aspects and their impact in respect of fairness, the avoidance of bias, and client usability.[198] In other words, the proposed new regulatory system component tasked with monitoring identified above, be implemented.

The Robodebt case study showed a willingness by government to implement punitive AI debt recovery system on the Australian population, without regard for the legality of such actions or its proper normative foundations. The methodology of uncovering the debts was unlawful, yet legal advice was completely disregarded. This scenario illustrates not only poor regulatory design from the perspective of Effective Regulation, but also the poor design and implementation of an AI system all of which resulted from failed government policy and decision-making in the first place in laying the foundations of the scheme.

---

192. *See Normative Theory*, *supra* note 18.
193. *Id.*
194. *See* discussion, *supra* Part III.
195. REPORT OF THE ROYAL COMMISSION INTO THE ROBODEBT SCHEME, *supra* note 3, at 28.
196. *Id.*
197. *Id.* at 401-02.
198. *Id.* at xvi.

B. Veterans and the Golden Rule

By way of contrast, the Australian Department of Home Affairs and Department of Veterans' Affairs have developed an AI Golden Rule in AI decision-making[199] with significant success. Under the AI Golden Rule, decisions that have a beneficial outcome for claimants are automated.[200] AI decision-making is thus used as a 'triage' tool which makes the granting of positive decisions to applicants for benefits more efficient.[201]

From the perspective of Effective Regulation, the AI Golden Rule is a form of implementing and entrenching a specific normative ordering. It strikes the right balance between efficiency and fairness. As noted, decisions that have a beneficial outcome for citizens are automated, while negative decisions are subject to review. Further, as a triage tool, it is not an unsupervised AI decision-maker. Rather, it is a preliminary sorting tool that provides benefits to vulnerable populations where obvious and identified criteria are met and sends the complex cases to humans.

In addition, this use of AI as a decision-making mechanism maximises the efficiency potential of the technology. It by-passes the routine, but time-consuming processing average applications require, freeing up more costly human decision makers for the challenging cases. This efficiency enhances law's fairness mandate—justice delayed is justice denied, and in this case, timely justice is true justice. Thus, from an Effective Regulation perspective, the insertion of AI decision-making is a system enhancement.

Following from that same perspective, the accountability structure of the regulatory system, at least in this aspect, is built on an appropriate theory of law. It does not abandon a theory of law as being the posited law, or Law 1.0; rather, the AI Golden Rule allows Law 3.0 AI decision-makers to make decisions while Law 1.0 administrative lawyers review decisions that are averse to applicants.[202] This structure permits vulnerable applicants to deal with accountable people rather than algorithms when seeking to contest their denied benefits.

The AI Golden Rule further embodies an important rule of law principle: that is that law attends to the means, the procedure, as well as the ends. Where a procedure is obscure as AI decision-making is likely to be, such a decision can be made open and transparent by human involvement and contested in a way understandable to non-specialists. In doing so, it addresses Wendall Wallach's contention that technology cannot sustain Kant's moral imperative.[203] By defaulting to human intervention, the AI Golden Rule allows human decision-

199. TOBY WALSH ET AL., OFF. OF THE VICTORIAN INFO. COMM'R, CLOSER TO THE MACHINE: TECHNICAL SOCIAL AND LEGAL ASPECTS OF AI 48 (Cliff Bertram et al. eds., 2019).

200. Ng et al., *supra* note 2.

201. *Id.*

202. *See* discussion, *supra* Part II.

203. WENDELL WALLACH & COLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG 94-97 (2009).

makers to attend to that imperative. This rule could be implemented in practice by way of regulatory design, and in particular, by empowering the dually tasked oversight body discussed above, with the contestation obligations.

V. ADDITIONAL REGULATORY SAFEGUARDS FOR AI DECISION-MAKING

Beyond the assessment resulting from the application of the theory of Effective Regulation, there are recommendations specific to AI. To combat both the legal and technical black boxes, we support the recommendation of the Australian Human Rights Commission that reasons be provided in the form of (1) a non-technical explanation of the algorithmic decision, which can be understood by a layperson, and (2) a technical explanation of an AI decision, which can be assessed by a person with technical expertise.[204] As we have argued, requiring government to provide reasons for AI decisions will preserve fundamental administrative law values, in particular fairness and transparency,[205] and will enable affected persons to challenge decisions. Given the challenges of explaining algorithms,[206] however, this latter task will remain a serious challenge. Government agencies employing such AI will need to have staff suitably trained in AI to be able to develop explanations suitable to administrative law specialists.

There is also an additional element to the legal black box associated with legal ownership of AI technologies. Private companies will claim proprietary rights over the inner workings of such technology. There are a number of notable examples of this issue. For example consider the case of the Correctional Offender Management Profiling for Alternative Sentences (COMPAS) system in the United States mentioned above, which generates a risk score showing the likelihood of *recidivism* to help judges with sentencing.[207] The company owns the proprietary rights to the technology and, as a result, the methods employed by the system are not known by the judiciary (and presumably not by the legislature or the executive).[208] As such, persons subject to a COMPAS score, and indeed all other parties involved in the sentencing, have no means of challenging the score as they cannot point to any particular error in data or process that the system has used or made. Nevertheless, in *State of Wisconsin v. Loomis*,[209] the Supreme Court of Wisconsin approved the use of such tools, providing the judge did not fully delegate their decision-making function, and still considered the defendant's arguments on the question of future re-offending.[210]

Another case that illustrates the American Court's efforts to balance the protection of intellectual property against constitutional rights is that of *Houston Federation of Teachers, Local 2145 v. Houston Independent School District*,

---

204. DISCUSSION PAPER, *supra* note 157, at 85.
205. Ng et al., *supra* note 2, at 1046.
206. *See* Hill, *supra* note 62.
207. Liu et al., *supra* note 49, at 122, 126.
208. Chatila et al., *supra* note 17, at 436.
209. 881 N.W.2d. 749 (Wis. 2016).
210. Liu et al., *supra* note 49, at 122, 126.

where it held that public disclosure of general information about the Educational Value-Added Assessment System (EVAAS) methodology used to evaluate teacher effectiveness and to fire teachers with a low score was insufficient to alleviate American constitutional due process concerns.[211] This decision was made because the methodology of EVAAS could not establish the reason for a teacher's dismissal in sufficient detail for the teacher to establish whether an error existed.[212] The court was able to balance the company's IP rights because it did not require the disclosure of trade secrets that might put the company that developed the statistical model out of business.[213]

To avoid such legal black boxes of proprietary secrets, the government should own or have a licence to the AI sufficient to allow its processes to be within the view of humans affected by its decisions as well as the courts. This is particularly important with the trend in recent decades of outsourcing and privatisation of governmental functions, especially in the procurement of AI decision-making systems by agencies.[214] Where the government owns or has the licence to the AI, it will enable disclosure of AI processes to persons affected by its decisions and to the courts. It will also ensure that agencies are able to evolve the system with changes in industry practice, technological developments, government policy and the law.

Another significant concern is algorithmic bias. As noted above, machine learning can develop, carry, and perpetrate biases in its AI decision-making functions because the data that is provided and pre-labelled usually carry biases as a result of contexts—the nature of the data, its collection, the construction of its categories and parameters, quantity and quality. For example, crime data will reflect to a significant degree its collection—collected from over-policed minorities.[215] This algorithmic bias is a particular concern, as the bias is often not readily noticeable, and further undermines the basic legal principle of equality before the law.[216] In terms of regulatory design, therefore, there is a need to critically evaluate potential biases and continually audit systems so that systemic errors that lead to incorrect categorisation do not occur, or are at least discovered.[217] Again, this task would be appropriate for the dually tasked AI regulatory body proposed in the Effective Regulation analysis. The UK Framework for *Ethics, Transparency and Accountability Framework for Automated Decision-Making*, for example, encourages stringent testing of

---

211. Hous. Fed'n of Tchrs., Local 2145 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1178 (S.D. Tex. 2017).

212. *Id.* at 1176.

213. *Id.* at 1179.

214. Deirdre Mulligan & Kenneth Bamberger, *Procurement as Policy: Administrative Process for Machine Learning*, 34 BERKELEY TECH. L.J. 781(2019).

215. Zalnieriute et al., *supra* note 25, at 434.

216. Paul Schwartfeger, *AFR Technology: Automating Bias*, NEW L.J. 15 (Aug. 18. 2020); LATTIMORE ET AL., *supra* note 63; CENTRE FOR DATA ETHICS & INNOVATION, *supra* note 63, at 119.

217. Admin. Conf. of the U.S., Statement #20, Agency Use of Artificial Intelligence 86 Fed. Reg. 6616 app. 1 (2021).

automated systems to avoid unintended outcomes or consequences.[218] In addition, the UK Framework recommends that teams working on AI decision-making are diverse and multidisciplinary, and that testing of the system is representative.[219]

In arguing for the addition of a human element, it is important to acknowledge Huq's caveat: "[t]he flawed quality of a machine decision does not imply that a human decision maker would do better. Nor will a human decision maker better serve a putative non-instrumental interest in participation."[220] Rather, the role of the human decision-maker must be to deal with those individual errors for the purposes of correcting them and, more importantly, identify those that are the result of systemic biases or other types of errors in the AI system—all matters to be built into the regulatory system as a matter of Effective Regulation.

CONCLUSION

AI decision-making has much to offer government and society: it can deliver objective, consistent and timely decisions. Further, there are significant efficiencies available which enhance both justice and the finances of government. Simultaneously, there are significant risks to the basic tenets of the rule of law that allow liberal democracies to thrive.

We believe that best practice regulatory design for government AI decision-making should address both the benefits and the risks. It should incorporate safeguards protecting the rule of law and administrative law principles, doctrines, and institutions. Effective regulatory design will attend to the society-technology interface to ensure that the technology is not further entrenching unjust, undesirable outcomes and undermining trust in government as a result of flawed AI decision-making. In particular, it is an important part of regulatory design for AI decision-making that a human be directly involved—wherever decisions adverse to applicants are made, and wherever the system has potential for less than optimal calibrations. Further, we believe that at a system level, there should be an AI impact assessment for all decisions, the provision of reasons for decisions, the ownership of AI proprietary interests by government, and auditing to reduce the risk of algorithmic bias is an important consideration. Finally, we believe that active steps need to be taken to facilitate access to review of AI-generated or assisted administrative decisions for vulnerable disenfranchised groups—the dually tasked body proposed above as a matter of regulatory system design.

Rule of law and administrative law norms are particularly pertinent in AI decision-making, for as Brownsword notes:

---

218. *Guidance: Ethics, Transparency and Accountability Framework for Automated Decision-Making,* OFF. FOR A.I. (Nov. 29, 2023), https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making#ensure-that-you-are-compliant-with-the-law [https://perma.cc/R3G4-MDGL].

219. *See* CENTRE FOR DATA ETHICS & INNOVATION, *supra* note 63.

220. Huq, *supra* note 1, at 671.

> in practice, technological measures are employed for regulatory purposes by both public and private actors (e.g. by the police, the revenue, and financial regulators as much as by BigTech corporations, banks and insurance companies) without there being any prior public authorisation or debate. In this sense, law 3.0 is a conversation that is everywhere and yet, publicly, transparently, and officially, nowhere.[221]

We have argued that AI decision-making should not be neglected, narrowly implemented or otherwise rejected on the basis of fears or concerns about the de-centering of humans.

We believe that AI decision-making, however, needs to be subject to uniquely rigorous regulatory design and particularly with vulnerable populations. This necessity arises because of its potential broad harm to individuals, governments as well as to the underlying trust of people in democratic societies with respect to fairness, accessibility to government, accountability and contestability of government decisions—the principles and doctrines that make up the rule of law, form the substance and inform the procedures of administrative law.

221. Brownsword & Somsen, *supra* note 69.