McGill Law Journal — Revue de droit de McGill

RESPONSIBLE AI: BINARIES THAT BIND

*Jennifer Raso**

This article examines responsible AI as a public law-like movement that seeks to (self)regulate the design and use of AI systems. Using socio-legal methods, and the *Montréal Declaration for a Responsible Development of Artificial Intelligence* as an illustrate example, it explores responsible AI's upshots for digital government. Responsible AI initiatives, this article argues, rely on two binary distinctions: (1) between artificial and natural intelligence, and (2) between the future and present/past effects of AI systems. These conceptual binaries "bind" such initiatives to an impoverished understanding of what AI systems are, how they operate, and how they might be governed. To realize justice and fairness, especially in digital government, responsible AI projects must reconceive of AI systems and their regulation infrastructurally and agonistically.

Cet article examine l'IA responsable comme un mouvement de type droit public qui cherche à (auto)réglementer la conception et l'utilisation des systèmes d'IA. En utilisant des méthodes socio-juridiques et la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle* comme exemple illustratif, il explore les retombées de l'IA responsable pour le gouvernement numérique. Selon cet article, les initiatives en matière d'IA responsable reposent sur deux distinctions binaires : (1) entre l'intelligence artificielle et l'intelligence naturelle, et (2) entre les effets futurs et présents/passés des systèmes d'IA. Ces binaires conceptuels « lient » de telles initiatives à une compréhension appauvrie de ce que sont les systèmes d'IA, de leur fonctionnement et de la manière dont ils pourraient être gouvernés. Pour réaliser la justice et l'équité, en particulier dans le gouvernement numérique, les projets d'IA responsables doivent reconcevoir les systèmes d'IA et leur réglementation de manière infrastructurelle et agonistique.

* Assistant Professor, McGill University Faculty of Law. Thanks to Gajanan Velupillai and Marguerite Rolland for superb research assistance, participants at the McGill Law Journal's *Reimagining Justice: AI's Power for Redress and Division* symposium (9 February 2024), and anonymous peer reviewers for their feedback. This work is funded by the Social Sciences and Humanities Research Council.

Citation: (2024) 69:4 McGill LJ 417 — Référence : (2024) 69:4 RD McGill 417

Introduction

From crossing a border to voting, receiving social benefits to obtaining a medical diagnosis, today algorithmically-driven tools affect intimate aspects of our lives.[1] With attention focused on generative AI, digital government initiatives are minimally scrutinized despite their ability to do maximum harm. In digital government, artificial intelligence (AI) systems comprised of technologies, state officials, administrative agencies, and members of the public generate crucial decisions: whether someone can cross the border, vote in an election, or access food, clothes, housing, and healthcare. Legal scholars increasingly recognize that such decisions are (or should be) subject to basic public law protections, including the rule of law.[2] Yet, how to regulate a decision-making system remains an ongoing challenge.[3]

AI systems are not lawless, however. These systems are already governed by the conventions that technologists rely upon, and which technologists help to craft, as they develop and maintain AI networks.[4] Some of these features are evident in the "responsible AI" movement. While responsible AI proponents may describe their initiatives as "ethical," socio-legal scholars would recognize them as regulatory, because they aim to govern how AI systems are designed and used. Yet they remain underexplored by legal scholars, who have instead examined state-based regulatory efforts including Canada's *Directive on Automated Decision-Making* and its proposed *Artificial Intelligence and Data Act* (*AIDA*).[5]

---

1 Petra Molnar, *The Walls Have Eyes* (New York: The New Press, 2024); Louise Amoore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Durham, NC: Duke University Press, 2020); Terry Carney, "Robo-debt Illegality: The Seven Veils of Failed Guarantees of the Rule of Law?" (2019) 44:1 Alternative LJ 4.

2 Jennifer Cobbe, "Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making" (2019) 39:4 J Leg Stud 636; Jennifer Raso, "AI and Administrative Law" in Florian Martin-Bariteau & Teresa Scassa, eds, *Artificial Intelligence and the Law in Canada* (Toronto: LexisNexis, 2021) 181 [Raso, "AI & Admin Law"]; Karen Yeung, "The New Public Analytics as an Emerging Paradigm in Public Sector Administration" (2023) 27:2 Tilburg L Rev 1 at 27, 32.

3 Paul Daly, Jennifer Raso & Joe Tomlinson, "Researching Administrative Law in the Digital World" in Carol Harlow, ed, *Research Agenda for Administrative Law* (London: Edward Elgar, 2022) 255.

4 Louise Amoore, *The Politics of Possibility: Risk and Security Beyond Probability* (Durham, NC: Duke University Press, 2013); Gavin Sullivan, "Law, Technology, and Data-Driven Security: *Infra*-Legalities as Method Assemblage" (2022) 49:1 JL & Soc S31.

5 Canada, Treasury Board, *Directive on Automated Decision-Making*, online: <tbs-sct.canada.ca> [perma.cc/GN9A-US6X]; *AIDA* forms part of Canada's *Digital Charter Implementation Act, 2022* (see Bill C-27, *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial*

This article analyzes responsible AI as a regulatory movement guiding the design and use of AI systems through self-government initiatives. Drawing on documentary evidence, including the *Montréal Declaration for a Responsible Development of Artificial Intelligence*,[6] ethnographic fieldwork, and multidisciplinary scholarship on AI, it explores the upshots of this movement for digital government initiatives.[7] The *Declaration* is a useful example, because it remains a touchstone for responsible AI advocates in Montréal, Canada's "new Silicon Valley," and beyond.[8] Exploring how responsible AI advocates conceptualize AI and its harms, and how they propose to redress them, reveals obstacles to crafting effectively responsible AI systems.

The responsible AI movement, I argue, rests on binary thinking that ultimately structures and limits its own regulatory potential. Two binaries, in particular, are my focus: the distinction between *artificial and natural intelligence*, and the distinction between *future and present/past effects of AI systems*. Below, I show how these binaries are *individualizing*, as they segment the components that make up an AI system, and *futurizing*, as they conceptualize AI system risks as futuristic, disconnected from present and past algorithmic systems. These binaries, evident in the *Declaration*, "bind" responsible AI initiatives to an impoverished understanding of what AI systems are, how they operate, and how they might be governed.

This article first sketches how widespread AI systems have become in digital government. Next, it details the public law-like features of the responsible AI movement. The article then uses the *Declaration* to show how distinctions between artificial versus natural intelligence and future versus present/past individualize and futurize how AI systems function.

Intelligence and Data Act and to make consequential and related amendments to other Acts, 1st Sess, 44th Parl, 2022 [*Digital Charter Implementation Act*]).

6 "Montréal Declaration for a Responsible Development of Artificial Intelligence" (2018), online: <montrealdeclaration-responsibleai.com> [perma.cc/E4JC-9UJ3] [*Declaration*]. A literature review of all published articles on the *Declaration* yielded only five articles from the humanities and social sciences, one article from computer science, and none written by legal scholars.

7 This ongoing study includes a systematic review of major reports and policy documents on responsible AI from Canada, the United States, Europe, and the United Kingdom, attendance at responsible AI workshops, seminars, and events in Montreal and Toronto, a review of online talks on responsible AI, and interviews with key actors, including technologists, ethicists, and others.

8 Ana Brandusescu, *Artificial Intelligence Policy and Funding in Canada: Public Investments, Private Interests* (Montreal: McGill Centre for Interdisciplinary Research on Montreal, 2021) at 33; Fenwick McKelvey, Sophie Toupin & Jonathan Roberge, eds, *Northern Lights and Silicon Dreams: AI Governance in Canada (2011-2022)* (Montreal: Shaping AI, 2024) at 16, 18.

The paper concludes by arguing for an approach to responsibility that is *infrastructural*, distributing responsibility beyond tech developers, and *agonistic*, drawing on a wider range of disciplines to learn from the past and present effects of algorithmic systems.

I. AI Systems and Digital Government

In Canada and elsewhere, AI systems are increasingly integral to digital government, yet both remain under scrutinized. Governments may announce digitalization efforts as neutral developments that will "optimize" administrative processes.[9] While this goal seems universally beneficial, optimization's ends are skewed. As Karen Yeung notes, digitalization provides administrative agencies wide latitude to pursue their own organizational goals (such as reduced costs and increased efficiency) that may misalign with or contradict the public's interests.[10] Indeed, many such projects can further marginalize already marginalized communities, and anyone outside of the "norm," by wrongly denying them benefits, undermining their credibility, imposing heavy administrative and evidentiary burdens, and generating debts and other punishments.[11]

Legal scholars and lawyers are increasingly aware of how algorithmic tools (AI or otherwise) already affect administrative decisions. For instance, there is growing concern that immigration authorities use facial recognition tools to undermine the refugee status of migrants from east Africa.[12] Likewise, batch processing techniques that assess multiple visa applications at once are gaining attention.[13] Law reform bodies, most notably the Law Commission of Ontario, are studying how AI tools intersect with administrative and human rights laws.[14] At the federal level, the Treasury Board has developed and regularly revises its *Directive on Au-*

---

9 See Immigration, Refugees and Citizenship Canada, Standing Committee on Government Operations and Estimates, "OGGO – Digital Platform Modernization" (May 2023), online: <canada.ca> [perma.cc/KYB6-UA3W].

10 Yeung, *supra* note 2 at 18–22.

11 *Ibid*; see also Petra Molnar, "Territorial and Digital Borders and Migrant Vulnerability Under a Pandemic Crisis" in Anna Triandafyllidou, ed, *Migration and Pandemics: Spaces of Solidarity and Spaces of Exception* (Cham, Switzerland: Springer, 2022) 45 at 48–53; Philip Alston, *Report of the Special Rapporteur on Extreme Poverty and Human Rights*, UNGA, 74th Sess, UN Doc A/74/493 (2019) at 5.

12 *Barre v Canada (Citizenship and Immigration)*, 2022 FC 1078 at paras 25, 56.

13 See Zynab Ziaie, "Chinook and Canadian Immigration: An Efficiency-Enhancing Tool or Cause for Harm?" (6 December 2021), online (blog): <cila.co> [perma.cc/Q578-C2FD].

14 Law Commission of Ontario, *Accountable AI* (Toronto: LCO, June 2022); Law Commission of Ontario, *Regulating AI: Critical Issues and Choices* (Toronto: LCO, April 2021).

tomated Decision-Making, and federal departments are slowly but steadily completing algorithmic impact assessments.

Meanwhile, across all levels of government, administrative actors from police to tax authorities continue to integrate algorithmic tools into their everyday decisions. These efforts are variably attentive to how such tools interact with public officials, institutions, and "users" to generate decisions. Sometimes, it is unclear to what degree AI is involved in a specific decision. Instead, some level of "automation" may be at play, with AI capabilities presumably imminent.[15] What is clear, however, is that the records and knowledge generated by past and present digitalization efforts will inform how any new algorithmic or AI-enhanced system functions, especially as the system's components draw on legacy data to generate results.[16]

Administrative agencies and AI are decision-making *systems*. Each is often described as "making" decisions on its own. Yet, both administrative agencies and AI are constituted by algorithmic tools (including software, hardware, databases, etc.), humans (developers, civil servants, members of the public), and the institutions (administrative and otherwise) in whose name decisions are rendered. In digital government, these actors together create the algorithmic or AI systems delivering programs from border security to social assistance. Any attempt to conceptualize how AI systems operate, and how they might be made more responsible or lawful particularly in digital government settings, must therefore attend to their system-based qualities.

While governments have long used algorithmic tools to facilitate decision-making, they have only recently begun to "regulate AI" formally. Some of these measures, like Canada's draft *Artificial Intelligence and Data Act*, tackle a wide range of AI systems.[17] Other efforts, propelled by recent advances in generative AI, include new safety and security standards for AI developers.[18] In many documents, including the European Union's *AI Act*, legislators use regulatory techniques to ensure that AI is

---

15 Edana Robitaille, "Minister Fraser Clarifies How IRCC Uses AI in Application Processing", *CIC News* (31 May 2023), online: <cicnews.com> [perma.cc/Z3PN-TRX5].

16 Raso, "AI & Admin Law", *supra* note 2 at 5–6, 9.

17 For the *AIDA*, see *Digital Charter Implementation Act*, *supra* note 5.

18 See e.g. United States, Executive Order 14110, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (30 October 2023), online: <whitehouse.gov> [perma.cc/L9QH-V3TE]; see also the amendments to the European Union's forthcoming *AI Act* (Directorate General for Communication, *EU AI Act: First Regulation on Artificial Intelligence* (18 June 2023), online (pdf): <europarl.europa.eu> [perma.cc/V8KS-XQXR]).

"trustworthy."[19] Local governments, too, have undertaken their own efforts, from public consultations to outright bans on tools known to harm marginalized communities, such as facial recognition technology.[20]

Beyond government initiatives, technologists have also pursued self-regulation. Like technologies, these regulatory efforts have politics.[21] They "afford" some possibilities and frustrate others.[22] The responsible AI movement, for example, purports to "govern" AI, including the algorithmic tools developed for and eventually used in digital government projects. In doing so, the responsible AI movement sets the terms of the debate about what "responsibility" might require. In this way, responsible AI is a self-legitimizing project akin to corporate social responsibility.[23] Unlike corporate social responsibility, however, responsible AI remains understudied. It is nonetheless useful to examine how AI systems, particularly those underlying digital government initiatives, are (or might be) governed through self-regulatory efforts, and with what effects. These issues are well illustrated by an example central to the responsible AI movement: the *Montréal Declaration*.

II. Responsible AI and the *Montréal Declaration*: A Public Law-like Project

As governments regulate AI, a parallel "cottage industry" has sprung up among technologists to self-govern AI development and use.[24] Its mechanisms include statements of principles, development roadmaps, and even declarations not of statehood or jurisdiction, but of ethical principles and values. Behind this movement are many actors: big tech firms like Microsoft and Google, as well as consultancy firms like McKinsey.

---

19 Johann Laux, Sandra Wachter & Brent Mittelstadt, "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk" (2024) 18:1 Regulation & Governance 3 at 3.

20 Somerville, (Massachusetts), San Francisco, and Oakland, (California) have banned police use of facial recognition technology (see American Civil Liberties Union, "Oakland Approves Facial Recognition Technology Ban as Congress Moves to Require Government Transparency" (17 July 2019), online: <aclu.org> [perma.cc/9LQE-ET35]).

21 Langdon Winner, "Do Artifacts Have Politics?" (1980) 109:1 Daedalus 121 at 123.

22 Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Cheltenham, UK: Edward Elgar, 2015) at 176–7.

23 Ronen Shamir, "Capitalism, Governance and Authority: The Case of Corporate Social Responsibility" (2010) 6 Annual Rev L & **Soc** Science 531 at 532–33.

24 Jonathan Roberge, Marius Senneville & Kevin Morin, "How to Translate Artificial Intelligence? Myths and Justifications in Public Discourse" (2020) 7:1 Big Data & Society 1 at 5.

Each has adopted some version of a responsible AI policy.[25] Other actors include Canadian-based research hubs, such as Toronto's Vector Institute and Montréal's MILA, which partner closely with tech developers and academics and receive federal funding for their work.[26] These hubs have generated their own statements of principles articulating what responsible AI means and what it might require. Governments look to these sources to guide their own regulatory efforts. In some cases, they have even developed statements that structurally and rhetorically resemble those produced by other responsible AI movement actors.[27]

As a movement, responsible AI is a large tent housing similar but distinct initiatives. "Trustworthy" AI, "ethical" AI, and even "fairness, accountability, and transparency" projects all fit within this tent.[28] These initiatives commit to specific principles, such as transparency or accountability, with the overarching goal of ensuring that AI systems are "responsible" so that the public will "trust" system-generated results.[29] The Vector Institute's *AI Trust and Safety Principles*, for example, assume that the more trustworthy AI systems are, the more likely it will be that they are widely adopted.[30] Responsible AI advocates also aim to mitigate or avoid the potential harms of AI systems, such as their ability to create and entrench polarized outcomes, again to increase system uptake.

Although I describe it as "self-regulatory," the responsible AI movement is not quite "private" and not quite "public." Rather, it blends elements of both, being a mix of actors, concepts, institutions, and instru-

---

25 See e.g. McKinsey & Company, "Responsible AI (RAI) Principles" (last visited 12 September 2024), online: <mckinsey.com> [perma.cc/FT3N-CC6S].

26 Canada, Innovation, Science and Economic Development Canada, *Evaluation of Innovation, Science and Economic Development (ISED) Canada Funding to CIFAR*, Report (Ottawa: ISED, 2022) at 7; Brandusescu, *supra* note 8.

27 Canada, "Responsible Use of Artificial Intelligence in Government" (last modified 1 August 2024), online: <canada.ca> [perma.cc/3XFJ-GYWF].

28 See Ada Lovelace Institute, AI Now Institute & Open Government Partnership, *Algorithmic Accountability for the Public Sector* (New York & London: Open Government Partnership, 2021) at 13; David Leslie, *Understanding Artificial Intelligence Ethics and Safety* (London: Alan Turing Institute, 2019) at 7; Robyn Caplan et al, *Algorithmic Accountability: A Primer* (New York: Data & Society, 2018).

29 Laux, Wachter & Mittelstadt, *supra* note 19; McKinsey & Company, *supra* note 25.

30 Vector Institute, "AI Trust and Safety Principles" (14 June 2023), online: <vectorinstitute.ai> [perma.cc/8VH5-MERF]; for more on the link between trustworthiness and market adoption of AI, see Jessica Fjeld et al, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (Cambridge, Mass: Berkman Klein Centre for Internet & Society, 2020).

ments.[31] Many of its features seem private. Powerful tech firms and (in)famous AI developers, such as Geoffrey Hinton, Yann LeCun, and Yoshua Bengio, are key players. But so too are public funding agencies, such as the Fonds de Recherche du Québec, which supported events leading to the creation of the *Montréal Declaration*.[32] The AI industry has other public elements. It drives local economies and is also heavily subsidized by government grants, tax breaks, and land deals.[33] Big tech firms also shape how societies function as much as governments. Their most public personalities regularly capture attention with bold statements about how AI might impact society and how technologists ought to respond, including their recent call for a (not yet materialized) moratorium on AI development.[34] It is thus misleading to conceptualize responsible AI and its actors as merely private.

Responsible AI regulatory strategies also have distinct public law-like elements. For example, their instruments use broad terms that allude to but are distinct from public law concepts. Responsible AI proponents may support transparency initiatives rather than public law mechanisms like "notice" or "access to information." They may also favour explainable decision-making instead of requiring justifiable "reasons" for a decision.[35] In some instances, responsible AI actors use democratic-like consultative processes to produce declarations of broad principles to guide everyone who designs and uses AI systems.[36] Some actors, including the *Declaration*'s designers, may even hope that their efforts will guide lawmakers and public officials. To that end, they have been partly successful. The

---

31 Daniel Schiff et al, "AI Ethics in the Public, Private, and NGO Sectors: A Review of Global Document Collection" (2021) 2:1 IEEE Transactions on Tech & Society 31.

32 Armand Ngaketcha, "Une lecture technoprogressiste de la Déclaration de Montréal sur l'IA : quels enjeux pour l'éthique de demain" (2021) 3:3 Droit, Santé et Société 8 at 12.

33 See statement from Québec's then-Minister of Economic Development in Karl Rettino-Parazelli, "L'intelligence artificielle, moteur économique", *Le Devoir* (3 June 2017), online: <ledevoir.com> [perma.cc/7ZDD-J553]; Shannon Mattern, *A City is Not a Computer* (Princeton: Princeton University Press, 2021).

34 See "The 100 Most Influential People in AI", *Time Magazine* (September 2023), online: <time.com> [perma.cc/SM7T-LA3C]; Cade Metz & Gregory Schmidt, "Elon Musk and Others Call for Pause on AI, Citing 'Profound Risks to Society'", *New York Times* (29 March 2023), online: <nytimes.com> [perma.cc/DC2X-KH6J].

35 Ada Lovelace Institute, AI Now Institute & Open Government Partnership, *supra* note 28 at 18, 58.

36 As Bengio writes, "Its goal is to establish a certain number of principles that would form the basis of the adoption of new rules and laws to ensure AI is developed in a socially responsible manner. Current laws are not always well adapted to these new situations" (see Yoshua Bengio, "The Montréal Declaration: Why we must develop AI responsibly", *The Conversation* (5 December 2018), online: <theconversation.com> [perma.cc/S9TS- X543]).

Declaration has been promoted and even relied upon by the Office of Québec's Chief Scientist.[37] It also remains a touchstone in Canadian conversations about "ethical AI."[38] Such efforts reconfigure what responsibility means so as to sustain the business model underlying AI development. They also "avoid cognitive dissonance" between a declaration's principles and mainstream thinking within computer or data science on issues intersecting with "gender, race, class, history, and capitalism."[39] These self-regulation efforts are not simply performative; they create a platform or a jurisdictional sphere for self-governance.[40]

The *Montréal Declaration* illustrates how the responsible AI movement constitutes itself through a public law-like process that co-creates a series of normative principles. The *Declaration* was released in December 2018, following two years of consultations with actors across public and private sectors, including technologists, policy makers, and "citizens." According to Yoshua Bengio, himself a driving force behind the *Declaration*,

> It was forged on the basis of vast consensus. We consulted people on the internet and in bookstores and gathered opinion in all kinds of disciplines. Philosophers, sociologists, jurists and AI researchers took part in the process of creation, so all forms of expertise were included.[41]

These "co-construction" sessions were directed by researchers, primarily from the University of Montréal, with expertise in computer science, philosophy, and law.[42] Consultations echoed elements of government rule-making processes,[43] yet they were also distinctly opaque. For example, the publicly-accessible record of the *Declaration*'s creation lacks key details: who exactly participated? what sorts of "dialogues" unfolded? which information was presented to consultation leaders? and, to what extent did

---

37 Roberge, Senneville & Morin, *supra* note 24 at 5.

38 See e.g. discussions at the *World Summit AI Americas* (April 2024), online: <americas.worldsummit.ai> [perma.cc/EVZ6-YT39].

39 Adrian Daub, *What Tech Calls Thinking: An Inquiry into the Intellectual Bedrock of Silicon Valley* (New York: FSG Originals, 2020) at 9.

40 Nofar Sheffi, "We Accept: The Constitution of AirBnb" (2020) 11:4 Transnational Leg Theory 484 at 498–500; see also Thao Phan, Jake Goldenfein, Declan Kuch & Monique Mann, *Economies of Virtue – The Circulation of "Ethics" in AI* (Amsterdam: Institute of Network Cultures, 2022).

41 Bengio, *supra* note 36.

42 *Declaration*, *supra* note 6 at Credits, I.

43 Andrew Green, "Delegation and Consultation: How the Administrative State Functions and the Importance of Rules" in Colleen Flood & Paul Daly, eds, *Administrative Law in Context*, 4th ed (Toronto: Emond, 2022).

the process allow participants to dissent?[44] Unlike an agonistic public regulation-setting process, where legislators and representatives of both government ministries and civil society share divergent views about regulatory options over many stages of drafting, the team behind the *Declaration* seems to have avoided conflict through its very composition.[45] While it was multidisciplinary, the team lacked members from fields long critical of technological systems. Experts from race and disability studies, media studies, and science and technology studies, for instance, were notably absent. These absences may suggest why and how some of the binaries identified below arose and hint at their possible effects.

Like other responsible AI initiatives, the *Montréal Declaration* articulates ethical principles to guide designers and users of AI systems in a law-like way. The *Declaration* aims to "[d]evelop an ethical framework for the development and deployment of AI," to "[g]uide the digital transition so everyone benefits from this technological revolution," and to "[o]pen a national and international forum for discussion" to achieve responsible AI.[46] It then sets out a series of ten principles—*well-being, respect for autonomy, protection of privacy and intimacy, solidarity, democratic participation, equity, diversity and inclusion, prudence, responsibility,* and *sustainable development*—to govern the future development and use of AI systems.

These principles are unsurprisingly broad. Like written constitutions and statutes, their meaning depends on how they are interpreted and applied. The *Declaration* cautions that these principles "must be interpreted consistently to prevent any conflict that could prevent them from being applied," and that "the limits of one principle's application are defined by another principle's field of application."[47] To aid technologists in their interpretation, each principle is fleshed out by a series of sub-principles so broad that "they do not even specifically address AI."[48] For example, the *well-being principle* includes the sub-principle that AI systems must "help

---

44 While the *Declaration*'s website lists funders and the members of the committee behind it, no information is easily available as to who participated in the consultations that created it.

45 This challenge of integrating agonism extends into algorithmic tools themselves: see Kate Crawford, "Can an Algorithm Be Agonistic? Ten Scenes from Life in Calculated Publics" (2016) 41:1 Science, Tech & Human Values 77.

46 *Declaration*, *supra* note 6 at 5.

47 This statement resembles the "watertight compartments" concept within Canadian constitutional law (see *ibid*).

48 Roberge, Senneville & Morin, *supra* note 24 at 6; for more on how the principles and sub-principles relate to one another, see generally Thierry Ménissier, "Un 'moment machiavélien' pour l'intelligence artificielle? La Déclaration de Montréal pour un développement responsable de l'IA" (2020) 77:1 Raisons Politiques 67.

individuals improve their living conditions, their health, and their working conditions."[49] Elsewhere, under the *solidarity principle*, the *Declaration* states that AI systems "must be developed with the goal of collaborating with humans on complex tasks and should foster collaborative work between humans."[50]

Many of the *Montréal Declaration*'s principles echo public law concepts. For example, under the *democratic participation principle*, the *Declaration* explains that AI-generated decisions must be justifiable rather than merely explainable. Justifiability, here, resembles the principle of "deference as respect" in Canadian administrative law.[51] Decisions generated by an AI system that affect a person's "life, quality of life, or reputation" (which parallels administrative law's concern with bureaucratic decisions that impact rights, privileges, and interests)[52] must be "justifiable" using language that is accessible to those who use or who are subject to an AI-generated decision.[53] "Justification," the *Declaration* goes on to state, "consists in making transparent the most important factors and parameters shaping the decision, and should take the same form as the justification we would demand of a human making the same kind of decision."[54]

Responsible AI initiatives, including the *Montréal Declaration*, are often critiqued as being vague and co-optable. Their broad guidance on how technologists might act "ethically" resembles statements from other public law tools, such as the principles governing regulated professions, including engineers and lawyers. Yet, for regulated professions, these broad statements are backed up by licencing regimes, complaints processes, and officials who scrutinize complaints and impose penalties.[55] Responsible AI, meanwhile, makes broad commitments without "teeth." Like state-based regulatory initiatives that can be captured by corporate interests, even the responsible AI initiatives initiated in-house at a firm like Google,

---

49 *Declaration*, *supra* note 6 at 8.

50 *Ibid* at 11.

51 *Baker v Canada (Minister of Citizenship and Immigration)*, 1999 CanLII 699 at para 65 (SCC); David Dyzenhaus, "The Politics of Deference: Judicial Review and Democracy" in Michael Taggart, ed, *The Province of Administrative Law* (Oxford: Hart Publishing, 1997) 279.

52 *Cardinal v Director of Kent Institution*, [1985] 2 SCR 643 at 653, 1985 CanLII 23 (SCC).

53 *Declaration*, *supra* note 6 at 12.

54 *Ibid*.

55 For a famous administrative law example, see *Doré v Barreau du Québec*, 2012 SCC 12.

for example, are easily co-opted.[56] In some cases, the teams tasked with realizing responsible AI have been fired when they pursue their mandate too effectively.[57]

Social scientists have also critiqued the *Montréal Declaration*'s selective normative agenda. These critiques offer insights into the *Declaration*'s regulatory techniques, though their authors do not use regulatory terminology to do so. Jonathan Roberge et al, for example, show how the *Declaration* predefines the issues that it purports to resolve, thereby establishing limited roles for each actor who might govern those issues.[58] This "problematization" technique, the authors argue, is a type of techno-solutionism because it barely acknowledges the substantial risks raised by AI systems.[59] When such dangers are identified, the *Declaration* does so performatively to "signify a vague sense of awareness."[60] Through this process, "criticism is more or less neutralized, if not recycled, by justificatory discourses."[61]

Responsible AI documents do more than define the terms and conditions for AI's acceptability, however. Instruments like the *Montréal Declaration* also establish binaries that constrain how we conceptualize the issues likely to arise when governments use AI systems and the possible regulatory responses to those issues. The next section examines two such binaries: the division between artificial versus natural intelligence, and the distinction between the past or present and the future.

III. Responsible AI: Binaries that Bind

Responsible AI initiatives conceptualize AI systems and decision-making in ways that bar us from recognizing and tackling some of digital government's biggest problems. Many such binaries exist. This section fo-

---

56 Jacob Metcalf, Emanuel Moss & danah boyd, "Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics" (2019) 82:2 Soc Research: An Intl Quarterly 449 at 455.

57 Perhaps most notoriously, in December 2020, Timnit Gebru, the then-co-lead of Google's Ethical AI research team, was fired after refusing to withdraw a then-unpublished article about the dangers of Large Language Models. Margaret Mitchell, another top ethics expert, was fired weeks after (see Dieuwertje Luitse & Weibke Denkena, "The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI" (2021) 8:2 Big Data & Soc 1 at 2).

58 Michel Callon, "Some Elements of a Sociology of Translation: Domestication of the Scallops and Fishermen of St Brieuc Bay" (1984) 32:1 Sociological Rev 196.

59 On techno-solutionism, see Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (New York: Public Affairs, 2013).

60 Roberge, Senneville & Morin, *supra* note 24 at 4.

61 *Ibid*.

cuses on two pivotal ones: (1) the distinction between artificial and natural intelligence, which I argue individualizes responsibility; and (2) the distinction between AI system effects now versus in the future, which I argue futurizes the effects of digital government.

A. Individualizing AI Systems

First, the *Montréal Declaration* bifurcates artificial and natural intelligence in ways that individualize or segment different elements of decision-making processes. This individualizing makes it difficult to appreciate the system-based nature of not only AI but of decision-making, which is crucial for regulating digital government and AI systems more broadly.

This bifurcation appears throughout the *Declaration*, as the *Declaration* contrasts artificial (i.e., computer or algorithmically-based) intelligence with that of individual human beings. Of course, a text designed to govern the development and use of AI systems must define what "AI" is, and the act of defining AI may distinguish "artificial" from other types of intelligence. But this distinction simplistically separates machines from humans. For example, the *Declaration* immediately announces in its preamble:

> For the first time in human history, it is possible to create autonomous systems capable of performing complex tasks of which natural intelligence alone was thought capable: processing large quantities of information, calculating and predicting, learning and adapting responses to changing situations, and recognizing and classifying objects.[62]

This bifurcating tendency runs deep within the *Declaration* where "intelligent" machines are contrasted with humans. In these areas, machine or AI tools are conceptualized as either outperforming humans, as acting independently from them, or as caring for them (alluding to social robots).[63]

This account of what AI systems can do "autonomously" and what humans used to do through "natural intelligence" ignores how the boundaries between "artificial" or "machine" and "human" blur in practice. Even the tasks listed in the preamble, which humans supposedly once performed on their own using natural intelligence, would have to be completed by many humans, tools, and even institutions working *together*, harmoniously or adversarially. For example, how would humans "process" large quantities of information without other humans, or without devices that record, calculate or process, and store data? Where would these large

---

[62] *Declaration*, *supra* note 6 at 7.

[63] See e.g. *ibid.*

quantities of information come from (an individual human would likely need a variety of tools, institutions, and other humans to collect them)? Similarly, how would "artificial" devices (computers, for instance), process information without first being designed by people to do so? Would this information not also be curated, input, labeled (again, often by humans), and refined to achieve a result (itself the product of humans, tools, and institutions)?[64]

While some disciplines may commonly distinguish "artificial" from "human" intelligence to advance theoretical arguments, this bifurcation is itself highly artificial. Its presence in the *Declaration* reflects the expertise of the *Declaration*'s architects: computer scientists, moral philosophers, and legal scholars (who sometimes abstract away important elements of law's relationality).[65] Social scientists and socio-legally inclined lawyers would bristle at the *Declaration*'s account of intelligence and decision-making. Social scientists, for example, have long demonstrated that we cannot fully understand *how* complex tasks like the ones described in the *Declaration*'s preamble are performed without understanding how networks or webs of actors (humans, machines, institutions, etc.) co-produce such results.[66] The same is true in digital government processes, which require public officials, algorithmic tools, ministerial offices, and administrative agencies to co-generate results.[67]

This individualizing tendency limits the *Declaration*'s ability to robustly conceptualize and tackle AI systems and their effects in digital government settings. For example, as noted above, under the *democratic participation* principle, the *Declaration* proposes that AI system-generated decisions must be justifiable. This principle requires that the "most important factors and parameters" that shaped a decision ought to be communicated to the person(s) affected. Likewise, the *Declaration* states that "the code for decision-making algorithms used by public authorities must be accessible to all" (unless the algorithmic tool had a "high risk of serious danger if misused"). It also notes that, for AI systems that

---

64 See Mary L Gray & Siddharth Suri, *Ghost Work* (Boston: Houghton Mifflin Harcourt, 2019) at ix–x; Kate Crawford, *Atlas of AI* (New Haven: Yale University Press, 2021) at 8; Salomé Viljoen, "A Relational Theory of Data Governance" (2021) 131:2 Yale LJ 573 at 580.

65 See Jennifer Nedelsky, *Law's Relations: A Relational Theory of Self, Autonomy, and Law* (Oxford: Oxford University Press, 2011) at 3–4.

66 Callon, *supra* note 58; Amoore, *supra* note 4; see also Geoffrey Bowker & Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, Mass: MIT Press, 1999) at 33–34; Nick Seaver, "What Should an Anthropology of Algorithms Do?" (2018) 33:3 Cultural Anthropology 375 at 375, 378.

67 Sullivan, *supra* note 4; Fleur Johns, *#HELP: Digital Humanitarianism and the Remaking of International Order* (Oxford: Oxford University Press, 2023) at 183–84.

significantly impact people, those people should have access to the skills and opportunities needed to "deliberate on the social parameters of these AI [systems], their objectives, and the limits of their use."[68]

These descriptions exemplify the *Declaration*'s binary distinction between artificial and natural intelligence, while also suggesting the artificiality of that distinction. Individual humans are imagined as operating separate and apart from AI systems and as individually affected by those systems. Yet, these passages also suggest that AI systems involve *many* actors, including the people who are impacted by the system itself (positively, negatively, or otherwise). A tension is thus built into the *Declaration*. Even as it divides artificial from human intelligence to explain how AI systems operate and to compartmentalize the system's components, its very terms reaffirm that AI systems are just that: *systems* shaped by a rich infrastructure made up of many actors. Any attempt to effectively guide or regulate developers and users of AI to achieve "responsibility" *must* tackle the systemic nature of AI head on.[69] Failing to do so frustrates the possibility of ever realizing robustly responsible AI systems.

B. Futurizing Risks and Harms

Second, the *Montréal Declaration* conceptualizes AI system effects as future risks, disconnecting that future from challenges algorithmic systems have raised today and in the past. This distinction between future and present/past prevents the *Declaration* from identifying and addressing ongoing problems raised by algorithmic tools, thus degrading the quality of responsibility to which the *Declaration* might contribute.

The fact that the *Declaration* looks to the future is to be expected from a visionary document using public law-like practices to articulate its principles. This technique, however, frames technological progress as inevitable and harms as theoretical when *both* exist today. The *Declaration*'s title—*The Montréal Declaration for a Responsible Development of Artificial Intelligence*—itself suggests that AI systems are in the process of being developed rather than already here. The body of the *Declaration* similarly futurizes AI development, use, and harms as potentialities rather than persistent realities. This discursive technique defers governance and responsibility into the future. AI system development is a near-future issue, the use of AI will occur even farther in the future, and, presumably, AI harms will materialize even further into the distance, some point after AI systems are used.

---

68 To find all quotes, see *Declaration*, *supra* note 6 at 12.

69 Josh Dzieza, "Inside the AI Factory", *New York Magazine* (December 2023), online: <nymag.com> [perma.cc/5ASP-YWZL]; Gray & Suri, *supra* note 64.

This technique delinks the near and distant future from algorithmic systems' past and present effects. The *Declaration*'s principles avoid placing effective boundaries around present-day AI development and use, ignoring situations where such limits have been required to tackle real, well-documented harms.[70] The *Declaration* thus avoids asking whether AI technologies are "safe, should be developed, or whether certain surveillance technologies should be made illegal."[71] Instead, it artificially severs the development and use of AI systems from their predecessors' present and past effects.

Perhaps most troubling, the *Declaration*'s drafters seem unaware of how algorithmic tools easily divide populations and facilitate the punishment and annihilation of some "categories" of people.[72] It is painfully ironic when a drafting team, the members of which support building AI systems to iteratively learn from the past (in the form of data), seems to avoid learning from the past itself.

This feature of the *Declaration* is clearest when its commitments to *equity, diversity, and inclusion* are read together with its *well-being* principle. Under its *diversity and inclusion* principle, the *Declaration* states that AI systems "must not lead to the homogenization of society."[73] Its *equity* principle also clarifies that AI systems must "help eliminate relationships of domination between groups and people based on differences of power, wealth, or knowledge."[74] Language elsewhere hints at egregious potential and past uses of algorithmic and data-driven systems, though. The very first principle, on *well-being*, states, "The development and use of [AI systems] must permit the growth of the well-being of all sentient beings."[75] Related sub-principles elaborate, indicating that AI systems

---

70 See Virginia Eubanks, *Automating Inequality* (New York: St Martin's Press, 2017); Joy Buolamwini & Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (2018) 81 Proceedings Machine Learning Research 1 at 12; Joanna Redden, Lina Dencik & Harry Warne, "Datafied Child Welfare Services: Unpacking Politics, Economics and Power" (2020) 41:5 Pol'y Studies 507; Carney, *supra* note 1; Joanna Redden et al, *Cancelled Systems Report* (Cardiff: Carnegie UK Trust and Data Justice Lab, 2022) at 13–23.

71 Roberge, Senneville & Morin, *supra* note 24 at 6.

72 Wendy HK Chun, *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition* (Cambridge, Mass: MIT Press, 2021); Rocco Bellanova et al, "Towards a Critique of Algorithmic Violence" (2021) 15:1 Intl Political Sociology 121 at 128; Ali Alkhatib, "To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes" (Paper delivered at CHI '21: Conference on Human Factors in Computing Systems, Yokohama, 8–13 May 2021), (2021) 95 ACM 1.

73 *Declaration*, *supra* note 6 at 14.

74 *Ibid* at 13.

75 *Ibid* at 8.

must allow people to "exercise their mental and physical capacities," and must "allow individuals to pursue their preferences, so long as they do not cause harm to other sentient beings."[76] Finally, a sub-principle states that AI systems "must not become a source of ill-being, unless it allows us to achieve a superior well-being than what one could attain otherwise."[77] Who (or what) are "sentient beings"? And who decides which forms of "superior well-being" might outweigh the "ill-being" caused by an AI system?

These tensions, the *Declaration* states, must be resolved by reading these principles together.[78] Based on their wording, however, it seems entirely possible that, while responsible AI systems cannot "homogenize" society, they can certainly divide populations along lines of "sentience" (or non-sentience). Likewise, while AI systems must eliminate relationships of domination based on "power, wealth, or knowledge," other hierarchies based on race, gender, immigration status, and even "sentience" appear perfectly acceptable. "Ill-being," too, is acceptable so long as it achieves "a superior well-being" (for an unidentified group of people and perhaps even for AI tools themselves).

These seemingly bizarre elements of the *Declaration* reflect the history of algorithmic systems and the worldviews of at least some of their proponents. The pursuit of "superior well-being" and the distinction between sentient and non-sentient beings are at the heart of current debates on longtermism, effective altruisim, and transhumanism.[79] Leading figures in these movements argue for policies that seem indistinguishable from eugenics, including that the most "intelligent" people in society should produce as many genetically related offspring as possible.[80] They also support unrestrained capitalism, arguing that rich and powerful people should continue to accumulate capital and power because they know best how to donate it to philanthropic causes.[81] Some within the responsible AI movement propose that "intelligent" machines (i.e., technologies that can mimic things that humans do, like generating sentences

---

76 *Ibid.*

77 *Ibid*.

78 *Ibid* at 5.

79 Ethan Zuckerman, "Two Warring Visions of Artificial Intelligence", *Prospect Magazine* (2024), online: <prospectmagazine.co.uk> [perma.cc/DB2Q-ZCU5]; for more on the history of AI, correlation, and eugenics, see Katherine Bode & Lauren ME Goodlad, "Data Worlds: An Introduction" (2023) 1:1–2 Critical AI 1.

80 Timnit Gebru & Émile P Torres, "The TESCREAL Bundle: Eugenics and the Promise of Utopia Through Artificial General Intelligence" (2024) 29:4 First Monday.

81 Yeung, drawing on James Scott's observations, recalls how data-centric tools were central for mapping and rounding up Jews in Nazi-occupied Amsterdam (*supra* note 2 at 21–22).

that seem to make sense) might develop sentience. Taking the Cartesian maxim "I think, therefore I am" literally, they argue that AI systems ought to possess rights associated with living humans, such as the right to marry and to adopt human children.[82] These ideas have a history that goes unacknowledged in many corners of the responsible AI movement.[83] This history colours the *Declaration*'s proposal that AI development and use must benefit all sentient beings, as it suggests that AI systems might "responsibly" be developed and deployed *without* the well-being of non-sentient beings in mind.

The *Declaration*'s futurizing tendency may be influenced by the co-consultation process that created it. Consultations seem to have focused on "future" problems and abstract hypotheticals rather than examples of present-day algorithmic tools, such as those used in border security, facial recognition, or pre-emption and risk assessment. As noted above, the experts who interpreted consultation results came from disciplines that abstract problems (moral philosophy, for instance) and that draw "technical" boundaries around dilemmas to make their social and political drivers "irrelevant" (i.e., law).[84] Co-consultation seems to have avoided any deep grappling with the history of AI systems and the worldviews upon which such systems rely.

This futurizing technique allows the *Declaration* to bypass the effects of past and present algorithmic systems. It overlooks the growing literature that shows how technological systems create and sustain harms. It ignores the inconvenient truth that many algorithmic systems wrong already vulnerable communities, which social science, humanities, and critical computer and data science scholars convincingly show are a feature, not a bug, of algorithmic governance.[85] The *Declaration* tries to break free from this present/past, but the present/past is inescapable. Any AI system that learns from data generated by earlier algorithmic systems will replicate and even intensify the biases lodged within that data. By severing future from present/past, the *Declaration* commits to a naïve, perverse notion of responsibility.

---

[82] This proposal was made to me quite seriously by a leading AI developer during my fieldwork, and has been mentioned during interviews: fieldnotes, June 21 2023; also interview A.

[83] Or, as Daub states, "in a place that likes to pretend its ideas don't have any history" (*supra* note 39 at 3).

[84] Mariana Valverde, "Jurisdiction and Scale: Legal Technicalities as Resources for Theory" (2009) 18:2 Soc & Leg Studies 139.

[85] Chun, *supra* note 72; Crawford, *supra* note 64 at 117; Joy Buolamwini, *Unmasking AI* (New York: Penguin Random House, 2024); Yeung, *supra* note 2; see also Eubanks, *supra* note 70; Redden, Dencik & Warne, *supra* note 70; Alkhatib, *supra* note 72.

IV. Realizing Fairness and Justice: An Infrastructural and Agonistic Approach

What might realizing fairness and justice require of those who pursue responsible AI initiatives? In my view, any attempt to achieve responsible AI must approach AI systems and their regulation infrastructurally and agonistically.

In an era of digital government, responsible AI and administrative law communities must reconceptualize AI systems and the decision-making processes to which they contribute infrastructurally. By "infrastructurally," I mean as distributed systems that generate results based on the inputs and interactions of many actors: tools, databases, workers, members of the public, and institutions. This approach to conceptualizing what, exactly, is at issue will ensure that responsibility initiatives are scoped broadly enough to account for the webs of actors that co-generate decisions in public sector settings.

One may ask, why conceptualize these systems as infrastructures and not networks? Certainly, legal scholarship increasingly analyzes algorithmic decision-making as a networked phenomenon. The growing field of legal materialism and socio-legal-technical studies, for example, draws on actor network and systems theory to demonstrate how webs of tools, software, people, and institutions create results.[86] A networked approach may helpfully map these relations, but it may also underemphasize each actor's relative power within the network. This underappreciation can make strategizing for social, technical, and legal change difficult.[87]

Thinking infrastructurally, by contrast, highlights the differential agency and regulatory effects of various elements of a decision-making infrastructure.[88] Understanding AI systems as infrastructures allows us to design responsibility strategies that require discrete actions at specific sites within an infrastructure, and to appreciate these actions as regulatory in and of themselves. For instance, it might identify database construction, front-line worker training, and multi-disciplinary design teams as effective ways to ensure responsible AI system development and use.

---

[86] This proposal diverges, then, from that of actor network theory proponents (see e.g. Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory* (Oxford: Oxford University Press, 2005) or Callon, *supra* note 58).

[87] I expand on this argument in Jennifer Raso, "Digital Border Infrastructures and the Search for Agencies of the State" in Gavin Sullivan, Fleur Johns & Dimitri Van Den Meerssche, eds, *Global Governance by Data: Infrastructures of Algorithmic Rule* (Cambridge, UK: Cambridge University Press, forthcoming 2024).

[88] Benedict Kingsbury, "Infrastructure and InfraReg: On Rousing the International Law 'Wizard of Is'" (2019) 8:2 Cambridge Intl LJ 171.

This approach is evident in the strategies proposed by critical AI scholars, such as lifecycle audits of algorithmic systems attentive to a system's technical and regulatory effects,[89] or datasheets detailing the datasets on which digital government tools rely.[90] Statements of principles, too, might use an infrastructural approach to articulate obligations for specific actors within an AI system. These obligations could account for each actor's place and role within the infrastructure, thus ensuring that responsibilities correspond with one's infrastructural location.

An infrastructural approach also enables us to examine how AI systems modulate the process by which an administrative decision is produced. This focus directs attention to how algorithmic decision-making may channel processes in certain directions and create bottlenecks elsewhere.[91] These details are particularly salient in digital government contexts. The flows and bottlenecks afforded by AI systems can significantly impact the fairness and justice of a decision-making process. They might also help to distinguish those processes that provide individuals with contestation opportunities from those that efficiently streamline individuals into procedural dead ends. These features would enrich thinking about what responsible AI might mean for digital government.

Infrastructural thinking tends to be forward-looking,[92] however, and it should thus be paired with an agonistic approach to responsibility that learns from the present and past. By "agonism," I mean a commitment to norm-setting processes open to dissent and contestation. One way to achieve this goal is to envision responsible AI's potential by learning from how responsible AI conversations have previously been orchestrated. This critical reflection on the history of such initiatives would reveal fields of inquiry whose relevant knowledge has thus far been excluded or avoided, many of which I have identified above. Future responsible AI initiatives could seek to incorporate and substantively grapple with that knowledge

---

89 This proposal extends the auditing proposals technicians have been working on to include audits to ensure that algorithmic tools are consistent with statutory provisions and public law principles. For a critical account, see Abeba Birhane et al, "AI Auditing: The Broken Bus on the Road to AI Accountability" (Paper delivered at 2nd IEEE Conference on Secure and Trustworthy Machine Learning, Toronto, 9–11 April 2024), online: <arxiv.org> [perma.cc/XQ5N-YK4T].

90 Timnit Gebru et al, "Datasheets for Datasets" (2021) 64:12 Communications ACM 86.

91 To expand on the concept of bottleneck as a theoretical framework to examine governance, see Caroline Melly, *Bottleneck: Moving, Building and Belonging in an African City* (Chicago: University of Chicago Press, 2017).

92 Kingsbury, *supra* note 88 at 183.

precisely *because* it is agonistic.[93] This method is the only way that responsible AI proponents will come close to realizing the robust type of responsibility required of AI systems, especially in digital government applications.

This proposal is deceptively simple. It agitates against the values and methods used by those advancing mainstream responsible AI initiatives. Tech developers, for example, may perceive themselves as "disruptors." But my fieldwork reveals a consistent and obstinate resistance, even hostility, to engaging with knowledge that disrupts developers' own visions of an AI-driven future.[94] More broadly, the field of AI development resists and marginalizes those who question whether AI systems can easily promote well-being or prevent domination without first acknowledging and addressing their present and past effects.[95]

While it remains unclear to what degree algorithmic tools can ever be agonistic,[96] there is no convincing reason why the forums that produce responsible AI frameworks cannot invite dissent and contestation.[97] Here, the responsible AI movement would benefit from adopting the constitutional conventions. Drafters of principles should look beyond friendly disciplines and engage with scholars whose work *is* agonistic. Doing so would reflect evolving public law consultation practices. Understood generously, these practices include consulting communities with diverse expertise and interests, including those with vital knowledge of a problem's past and present features, grappling with that knowledge, and using it to

---

93 This argument extends some of Mireille Hildebrandt's arguments in "Algorithmic Regulation and the Rule of Law" (2018) 376:2128 Philosophical Transactions A: Mathematics, Physical & Engineering Sciences 1 at 6–9.

94 Fieldnotes, June 8, 2023; Fieldnotes, June 20, 2023; see also Jenna Burrell, "Automated Decision-Making as Domination" (2024) 29:4 First Monday.

95 This characteristic has led figures, like Timnit Gebru and Joy Buolamwi, to respectively create organizations such as the *Distributed AI Research Institute* (or *DAIR*) and Algorithmic Justice League: see Distributed AI Research Institute, "About Us" (last visited 12 September 2024), online: <dair-institute.org> [perma.cc/PV2L-PAY4]; Algorithmic Justice League, "About" (last visited 12 September 2024), online: <ajl.org> [perma.cc/B2BB-R3L5].

96 Andrew D Selbst et al, "Fairness and Abstraction in Sociotechnical Systems" (Paper Delivered at the Conference on Fairness, Accountability, and Transparency (FAT* '19), Atlanta, 29–31 January 2019). Suresh Venkatasubramanian, a co-author of this article, explores to what degree agonism can be baked into algorithmic tools in his innovative (but rare) computer science course on the subject at Brown University. It remains an open question whether algorithmic tools can robustly respond to polycentric issues, or whether binary approaches will persist given most technologists' disciplinary training.

97 See e.g. the approach taken by Jenna Burrell, Ranjit Singh & Patrick Davison, eds, *Keywords on the Datafied State* (New York: Data & Society, 2024).

inform the content of a statement of principles or other document. Such processes, too, provide groups with varied, conflictual expertise the opportunity to inform, challenge, and protest a norm-setting process. This contestation is itself important. It can be the source of grassroots education and advocacy campaigns, and may bolster further efforts to regulate AI systems in-house or externally.

Despite the above challenges, responsible AI matters. To what degree might it ensure that AI systems are developed and used to guarantee not only "trustworthy" but ultimately high quality, well informed, justifiable administrative decisions that can be challenged when necessary? This is the big question. Developing such tools will require deeper system redesigns, both technological and legal.[98]

---

[98] Bjorn Kleizen et al, "Do Citizens Trust Trustworthy Artificial Intelligence? Experimental Evidence on the Limits of Ethical AI Measures in Government" (2023) 40:4 Government Information Quarterly 101834.