SAFETY IN ARTIFICIAL INTELLIGENCE & ROBOTICS GOVERNANCE IN CANADA

Kristen Thomasen[1]

This paper attempts to reimagine our approach to "safety" in the context of artificial intelligence (AI) and robot governance in Canada. It begins with a discussion of varying ways of understanding of the goal of "safety." Drawing especially on abolitionist writing, the paper argues for broader, more comprehensive understandings of safety in AI and robot governance. It then examines and critiques how safety has been understood in recent state interventions involving AI/robot systems. Finally, a framework for working towards a comprehensive understanding of "safety" as a governance objective is provided.

L'auteure tente de repenser notre approche de la « sécurité » dans le contexte de l'intelligence artificielle (IA) et des mesures de gouvernance de la technologie robotique au Canada. Elle discute d'abord les différentes façons dont nous comprenons la définition de la « sécurité » comme objectif. S'inspirant notamment des écrits abolitionnistes, l'auteure plaide en faveur de compréhensions plus étendues et plus exhaustives de la sécurité dans le contexte de l'IA et de la gouvernance de la technologie robotique. Elle examine et critique ensuite la manière dont la sécurité a été interprétée dans les récentes interventions de l'État concernant les systèmes d'IA et de la technologie robotique. Enfin, l'auteure offre un cadre de travail pour parvenir à une compréhension exhaustive de la « sécurité » en tant qu'objectif de gouvernance.

[1] Assistant Professor, Peter A. Allard School of Law, University of British Columbia. I am deeply grateful for all of the assistance and support that made this paper possible. My enormous thanks to Sue Glueck for your enthusiastic support of my research, your recommended readings, and overall wonderfulness. Thank you to the Microsoft Corporation for the unrestricted funding support, which enabled the contributions of four phenomenal Allard Law research assistants over the course of this project: Mariah Friedrich, Jeremy Kellen, Jiwan Sangha, and Yuma Wu. A special thanks to Jeremy and Mariah for helping me develop my thinking around many of the ideas shared in this paper and for bringing my attention to some truly wonderful research. My deep gratitude also to Brenna Bhandar, Suzie Dunn, Woodrow Hartzog, Jon Khan, and Alexandra Mogyoros for their brilliant comments and feedback on an earlier draft, to Kim Bain, Cristie Ford, Yuvraj Joshi, Graham Reynolds, and Andrew Selbst for their suggestions and talking different ideas through, and to the incredible We Robot 2022 community and Allard Law Faculty Colloquium for all the feedback and comments that helped strengthen the paper. Thank you to the anonymous reviewers for their thoughtful suggestions and to everyone at the Canadian Bar Review for their support. Finally, my enormous thanks to John, Sterling, and Lyne always.

Contents

1. Introduction

This essay explores the idea of "safety" in artificial intelligence (AI) and robot governance in Canada.[2] Regulating robotic and AI-based systems through a lens of safety is a vital, but elusive, task. In Canada, much governance of robotic and AI systems occurs through public bodies and structures.[3] While various laws and policies aim to ensure that AI and robotic systems are used "safely," the meaning and scope of "safety"

[2] I use "governance" here to include but also look beyond formally enacted statutes and regulations, to also include written policies, agency decisions around the choice to adopt certain tools, or to fund certain programs, *etc*. On the limits of these and other forms of governance see: Luke Stark, Daniel Greene & Anna Lauren Hoffman, "Critical Perspectives on Governance Mechanisms for AI/ML Systems" in Jonathan Roberge & Michael Castelle, eds, *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies* (London, UK: Palgrave MacMillan, 2021). For an earlier consideration of Canadian regulatory responses to AI see: Aviv Gaon & Ian Stedman, "A Call to Action: Moving Forward with the Governance of Artificial Intelligence in Canada" (2019) 56 Alta LRev 1137 (identifying various regulatory considerations for federal Canadian law-makers, though not specifically considering safety); for a primer on US legal governance considerations see: Ryan Calo, "Artificial Intelligence Policy: A Primer and Roadmap" (2017) 51 UC Davis L Rev 399 (identifying a number of policy concerns related to AI, including safety, without normatively arguing for the direction policy should take).

[3] This paper is specifically focused on the notion of safety as a government/governance goal, and examines how government decision-makers understand safety and public safety in the scope of AI and robotics governance. The considerations in this essay intersect with (but are not exactly the same as) other pertinent "AI safety" discourses, like those engaging AI-safety from the design and development standpoint. For more on AI safety from this lens see e.g. Roel Dobbe, "System Safety and Artificial Intelligence" in Justin B Bullocket al, eds, *Oxford Handbook on AI Governance* (forthcoming, 2023); Inioluwa

are seldom, if ever, explicitly considered. Safety is not a neutral concept and determining what kinds of technologies and applications are "safe" requires normative choices that often go unexpressed in the law and policy-making process. Broad appeals to the policy goal of "safety" can bring conduct or regulation into conflict with the actual safety of individuals and communities.[4] Expanded thinking about "safety" and governance in relation to automated technologies is needed, along with greater precision in law and policy goals.

Scholars and activists, particularly those advocating for the abolition of state policing and the prison industrial complex,[5] have robustly critiqued and re-theorized the concept of "safety" in law and policy, particularly in ways that are cognizant of equitable and collectively beneficial outcomes. Black and Indigenous, queer, disabled, and abolitionist thinkers and writers have proposed some of the most comprehensive visions of public safety, in addition to critiquing the limits of some common deployments of the concept.[6] To imagine a society without policing and prisons, abolitionist thinkers must engage in a systemic critique of how society, communities, and the state understand and seek to attain "public safety." Thus, abolitionist writers participate in a deep rethinking of the concept of "safety" and methods for creating safety, providing a richness that can enhance current discussions about AI and robotics governance.

This paper explores some of this scholarship, particularly arising from the North American perspective (given the geographical focus of the paper), and relates it to how we might understand "safety" in AI and robotics governance in Canada.[7] The paper develops three central

---

Deborah Raji & Roel Dobbe, "Concrete Problems in AI Safety, Revisited" (Paper delivered at the ICLR 2020, April 26, 2020)(sites.google.com/nyu.edu/ml-irl-2020/accepted-papers).

4 The term could be understood as a floating signifier—an empty but significant term; a term that "points to no actual object and has no agreed upon meaning" yet can galvanize action in a particular direction. A seemingly objective straightforward term that is in fact "mired in hidden subjectivities." See e.g. Patricia Williams, *The Alchemy of Race and Rights: Diary of a Law Professor* (Cambridge, MA: Harvard University Press, 1991) at 11, 122-23.

5 A term used to describe the relationship between the various beneficiaries of incarceration. "The overlapping interests of government and industry that use surveillance, policing, and imprisonment as solutions to economic, social, and political problems." Angela Y. Davis et al, *Abolition. Feminism. Now.* (Chicago, IL: Haymarket Books, 2022) at 43-44; Ruth Wilson Gilmore, *Abolition Geography: Essays Toward Liberation*, ed by Brenna Bhandar & Alberto Toscano (New York, NY: Verso Books, 2022) at 272-73 [Gilmore, "Geography"](cautioning not to hollow-out the term so it focuses only on profit and erases the role of the state).

6 See for example, the citations throughout Section III.

7 The essay looks to many scholars who write from the Canadian perspective, but also draws upon many U.S. based scholars. While the political histories that ground

arguments. First, the people who design and evaluate the governance of AI and robotics in Canada ought to be cognizant of more comprehensive ways of understanding safety, as explored in abolitionist and other critical work, including and extending beyond the anti-carceral core of this writing.[8] Second, having a more comprehensive understanding of safety in-hand, law- and policy-makers should be more explicit in terms of how they will actually regulate, distinguishing between, for instance, narrower objectives like collision-mitigation and more comprehensive objectives like strengthening individual and community responsiveness to harm. The transparency that comes with accuracy can also help reveal limits and inequities within the safety framework so that simultaneously, regulators, agencies, and local and community organizations can contribute toward a more comprehensive experience of safety in relation to automated technologies. Finally, state agencies should stop appealing to the concept of "public safety" to justify using automated technology in ways that undermine community and individual safety; such uses ought to cease all-together.

In drawing upon the insights offered by abolitionist thinkers, I am mindful of the ways that abolitionist frameworks would encourage, at a minimum, approaching state governance with caution. Governance does not arise solely from the state, and in thinking about safety goals, community governance should not be overlooked.[9] Further, governance is not exclusively in the realm of the public—manufacturers, for instance, can have considerable power over governance through the design of automated systems, as well as influence in the public bodies that govern through partnerships and lobbying, so state regulation cannot be presumed to be driven solely by the public interest. Additionally, the Canadian state as it is currently structured might not be capable of enacting a vision of comprehensive and equitable safety discussed below, in which case, if this

---

abolitionist thinking in each country are different, writing from the U.S. perspective is nevertheless informative given the transnational nature of the analysis, and several key similarities in respective political histories, including structures of colonialism and racial capitalism.

[8] Abolition offers a "mode of analysis" or a framework for structural analysis of the world, as well as a strategy to transform it: see Davis et al, *supra* note 5 at xi; Mariame Kaba, *We Do This 'Til We Free Us: Abolitionist Organizing and Transforming Justice*, ed by Tamara K Nopper (Chicago, IL: Haymarket Books, 2021).

[9] See e.g. Barnard Center for Research on Women, "No Borders! No Prison! No Cops! No War! No State?" (15 November 2022), online (video): *YouTube* <youtube.com/watch?v=4ji7Z8mMe78> [perma.cc/9H5D-49R8] (discussing the question of whether abolition calls for a shift away from the state, asking a fundamental question of *who* should be (or will be) caring for our safety?); see also Hannah Bloch-Wehba, "Algorithmic Governance from the Bottom Up" (2022) 48:1 BYU L Rev 69 (on the governance of AI systems through social and labour movements in the U.S.).

vision is to be enacted, the structures making up the state would have to change.[10] However, the state is made up of many components, including "actors, agencies, rules, bodies" through which change can be generated, and it is to these components that this essay speaks.[11]

For now, this essay has a narrower focus. Overarchingly, it calls for a move away from broad appeals to "public safety" to justify the state use or restriction of AI and robotic systems, and encourages recognizing that governance alternatives can and already do exist.[12] This essay does not evaluate whether specific policies are themselves abolitionist. Rather, it points to the imperative nuance that can be brought to bear on automated technology governance from the range of works envisioning individual and public safety free from the prison industrial complex. Further consideration of the state itself, as well as specific policies and their modes of enforcement is imperative, though beyond the scope of this essay.

The first section of this essay explains how "safety" arises within the context of Canadian robot and AI governance. Next, the essay explores some of the clandestine politics in governance appeals to "safety" generally, and in AI and robot law and policy specifically. It highlights recent examples where state appeals to "public safety" have jeopardized safety efforts, in part by misconstruing the meaning of "safety" and in part by narrowing the scope of the beneficiary "public."[13] Finally, the essay concludes with a framework for expanded deliberation and critique of the meaning and method of regulating "safe" AI and robotic systems.

Notably, this essay and the proposed framework do not claim that every law and policy must on its own address a broad and comprehensive notion of safety. This would be understandably challenging and, in many

---

[10] As Ruth Wilson Gilmore notes in *Abolition Geography*, *supra* note 5 at 262: "a state is a territorially bounded set of relatively specialized institutions that develop and change over time in the gaps and fissures of social conflict, compromise and cooperation. Analytically states differ from governments, if states are ideological and instrumental capacities that derive their legitimacy and material wherewithal from residents, governments are animating forces, policies plus personnel, that put state capacities into motion and orchestrate or coerce people in their jurisdictions to conduct their lives according to centrally made and enforced rules."

[11] Gilmore, "Geography", *supra* note 5 at 275.

[12] It is my hope that, at least, these recommendations do not constitute "reformist reforms." This paper calls for a move away from state appeals to a "safety crisis" that justifies expanded policing, and police use of automated technology. See e.g. Davis et al, *supra* note 6—Appendix, "Reformist Reforms vs Abolitionist Steps to End Imprisonment" at 185.

[13] None of this would be surprising to abolitionist theorists, as discussed further below. See Gilmore, "Geography", *supra* note 5 at 269-272.

instances, inappropriate for the scope of a particular government agency. The framework proposed in this essay is intended to guide decision-makers to more explicitly place purported governance objectives within the context of a more comprehensive understanding of "safety," recognizing that risk-mitigation and other narrowly-focused rules may contribute to, but are not exhaustive of, safety in relation to automated technologies.

Also, while much of this discussion of safety is transferrable to other legal contexts, it is necessary to specifically consider how we identify "safe" automated systems. Automated systems already have, and will continue to have, significant impact on individual and collective rights, community interests, and physical and virtual spaces. Research and lived experience with these systems have already highlighted many ways in which discriminatory bias can be embedded in automated technologies.[14] Scholars have also warned about the ways in which harmful outcomes might be concealed through the perceived neutrality or superiority of technical systems over human actors.[15] Legal scholars have noted some of the emerging challenges for regulating automated systems, including the ways in which such systems can be predictably unpredictable—

---

[14] See e.g. Joy Buolamwini & Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (2018) 81 Proceedings in Machine Learning Research 1 (showing commercial FRT systems' inaccuracy worsens along lines of gender and race, with Black women's images returning highest rates of inaccuracies); Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Medford, MA: Polity, 2019) (demonstrating examples of how and why automated systems reinforce racism, even when designed to address racism); Safia Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, NY: New York University Press, 2018) (explaining how search algorithms reinforce racist discrimination, particularly prevalent against Black women and girls); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, NY: St. Martin's Press, 2017); Morgan Mouton & Ryan Burns, "(digital) neo-colonialism in the smart city" (2021) 55:12 Regional Studies 1890; Jason C Young, "The new knowledge politics of Digital Colonialism" (2019) 51:7 Environment and Planning A: Economy and Space 1424; Christopher L Dancy and P Khalil Saucier, "AI and Blackness: Toward Moving Beyond Bias and Representation" (2022) 3 IEEE Transactions on Technology and Society 31 (drawing on Sylvia Wynter in discussing the embeddedness of antiblackness in AI).

[15] Kate Crawford, *Atlas of AI* (New Haven, CT: Yale University Press, 2021) (exploring the many harmful material impacts hidden behind the concept of "artificial intelligence"); *Race After Technology*, *supra* note 14 (emphasizing how the perceived neutrality of automated tools can conceal the ways in which automation deepens discrimination); Eubanks, *supra* note 14 (many automated tools adopted to enhance administrative efficiency harm those who rely on the state for support); Jason Millar & Ian Kerr, "Delegation, Relinquishment and Responsibility: The Prospect of Expert Robots" in Ryan Calo, Michael Froomkin & Ian Kerr, eds, *Robot Law* (Northhampton MA: Edward Elgar Publishing, 2016) (examining the considerations at stake when delegating human decision-making to automated systems).

generating potentially harmful outcomes and decisions that a human might not.[16] Current legal frameworks meant to ensure safe operations of new technologies, like negligence law, are not always easily applied to automated systems.[17] By virtue of the pervasive nature of some automated technologies, these issues might generate quantitatively and qualitatively more harm than previously possible.[18] Furthermore, these very systems are already contributing to the expansion of the carceral state.[19] At least in the near-term, laws and policies that oversee the "safety" of emerging automated technologies will be important for mitigating potential harms.

Conversely, automated systems can also permit access to safety, help to maintain safety, or be used to secure rights, or protect against infringement. These potentially novel uses of automated systems also fall within the scope of safety governance, and require specific consideration. Developing a more robust understanding of "safety," as well as expanding AI and robotics governance, can contribute to a stronger foundation for more socially just development and use of automated technologies—an increasingly urgent task given the pace of technological change, adoption, and deployment.[20] Safety is an imperative goal, as "a positive precondition

---

16 See e.g. Ryan Calo's discussion of emergence in "Robotics and the Lessons of Cyberlaw" (2015) 103 Cal LRev 513. In the process of using a complex tool to solve a problem, that very complexity might raise other problems or harms. The complexity makes the tool useful for its identified task, but can also generate behaviour that is unintended or unexpected (and may initially go unnoticed).

17 See e.g. Kristen Thomasen, "AI and Tort Law" in Florian Martin-Bariteau & Teresa Scassa, eds, *Artificial Intelligence and the Law in Canada* (Toronto: LexisNexis, 2021); Cynthia Khoo, "Missing the Unintended Forest Despite the Deliberately Planted Trees: Reasonable Foreseeability and Legal Recognition of Platform Algorithm-Facilitated Emergent Systemic Harm to Marginalized Communities" delivered at the University of Ottawa, (April 2-4, 2020), *We Robot*, online: <techlaw.uottawa.ca/werobot/papers> [perma.cc/EV8V-YYGF].

18 As Cathy O'Neil helpfully explains, mathematical tools like machine learning systems are of particular concern in regards to potential harm because they are opaque (it is difficult to know when one is in use, the reasons for decisions can difficult if not impossible to understand), unregulated, difficult to contest (both practically because of the complexity of the system, and because of a popular perception that these tools are superior to human analysis), and importantly—they are scalable. Machine learning systems, for example, can exert quantifiably more harm than previous data analysis tools/processes when scaled. See Cathy O'Neil, *Weapons of Math Destruction* (New York, NY: Crown Publishers, 2016).

19 See e.g. Benjamin, *supra* note 14.

20 I am hesitant to engage the "technology outpacing law" debate here, as there are many ways in which law can be amenable to protecting rights and norms in spite of technological change (e.g., Lyria Bennet Moses, "Agents of Change: How Law 'Copes' with Technological Change" (2011) 20 Griffith L Rev 763). That said, where rights and norms

for individual and collective freedom, perhaps especially for those to whom it has been long denied."[21]

2. "Safety" in AI & Robotics State Governance in Canada

"Safety" is a common objective or guiding value in the governance of automated technologies. The concept arises as a law or policy objective in Canadian technology governance in different ways. Sometimes, concern for safety justifies restraints on the use of automated systems, limiting how, where, or by whom a system can be used, in order to protect the safety of other people or property. For example, drone regulations in Canada seek to protect aviation safety (e.g. preventing drones from interfering with commercial and crewed flights), and ensure drone operators know how to fly 'safely' (without endangering people or property on the ground below), through regulations that stipulate how, where and when a drone can be permissibly operated.[22] Similarly, in the regulation of automated vehicles, regulators seek to ensure 'safe' operations and conditions of the vehicle through requirements and restrictions on use, and assign responsibility for such safe operations to a human operator.[23] 'Safe' government and private uses of AI systems are also an objective in various public strategy and guiding documents.[24]

---

are already underdeveloped, a technology that threatens those rights could do considerably more harm if gaps in law are not promptly addressed.

[21] Monica C Bell, "Safety, Friendship, and Dreams" (2019) 54 Harv CR -CLL Rev 703 at 720. "Safety, friendship, and dreams are oft-ignored and misunderstood aspects of inclusion and solidarity, and thus, justice does not exist in their absence": at 708.

[22] See "Transport Canada's Drone Strategy to 2025" (2021), online (pdf): *Transport Canada* <tc.canada.ca/sites/default/files/2021-03/TC223-Drone-Strategy-ENG-ACC.pdf> [perma.cc/KG7H-GNN2]; "Transport Canada's drone safety initiatives" (last modified 21 December 2016), online: *Transport Canada* <canada.ca/en/transport-canada/news/2016/12/transport-canada-drone-safety-initiatives.html> [perma.cc/LR78-BNU9]; see also *Regulations Amending the Canadian Aviation Regulations (Remotely Piloted Aircraft Systems)*, SOR/2019-11.

[23] Meanwhile, potential increases in highway safety associated with automation simultaneously justify governance encouraging the use of automated vehicles. See "Automated Vehicle Pilot Program" (last modified 6 April 2022), online: *Ontario Ministry of Transportation* <ontario.ca/page/automated-vehicle-pilot-program> [perma.cc/M2JE-RQXV].

[24] See e.g. Canadian Institute for Advanced Research, "CIFAR Pan-Canadian AI Strategy Impact Report" (2020), online (pdf): *CIFAR* <cifar.ca/wp-content/uploads/2020/11/AICan-2020-CIFAR-Pan-Canadian-AI-Strategy-Impact-Report.pdf> [perma.cc/U6ZY-BJBY]; Ignacio Cofone, "Policy Proposals for PIPEDA Reform to Address Artificial Intelligence Report" (last modified 12 November 2020), online: *Office of the Privacy Commissioner of Canada* <priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/pol-ai_202011/> [perma.cc/9CFK-VNA9]. See also "Accountable AI" (June 2022) at 21, online (pdf): *Law Commission of*

Sometimes safety governance can take the form of a precautionary prohibition. In a recent example, Toronto City Council temporarily banned the use of sidewalk robots out of concern for the accessibility and safety of public sidewalks and other shared infrastructure in the city.[25] The ban is framed as temporary, permitting time to better assess accessibility concerns for persons with limited mobility or vision, elderly people, and children, as emphasized by the Toronto Accessibility Advisory Committee and others.[26]

In addition to regulating AI & robotic systems in the pursuit of safety, laws and policies may also justify the *use* of automated technologies with the purported goal of protecting safety. This has been the case with the gradual integration of automated vehicles onto public roads, which are expected to be safer (better at avoiding collisions) compared to human driven vehicles.[27] Policing agencies like law enforcement and border

---

Ontario <lco-cdo.org/wp-content/uploads/2022/06/LCO-Accountable_AI_Final_Report.pdf> [perma.cc/5NEN-XMHL] (one component of trustworthy AI is that it is "safe"). Important additional developments may be imminent. The recently introduced Bill C-27, *Digital Charter Implementation Act, 2022,*1st Sess, 44th Parl, 2022 includes as Part III the *Artificial Intelligence and Data Act*, which aims to "mitigate risks of harm and biased output related to high-impact artificial intelligence systems." The Act is meant to include and look beyond physical harm, noting the psychological and discrimination-based harm that can also arise from AI systems. The Bill has not yet been passed, and leaves much of the details to future regulations. From the structure set out in the Bill, it is focused on harm-mitigation (but with a relatively broad understanding of potential kinds of harms), and penalties for breaches.

25 Under Ontario's automated vehicle regulation, the provincial Ministry of Transport proposed a pilot program for "micro-utility devices" (MUD), colloquially called sidewalk robots: "Proposed Amendments to Ontario Regulation 306/15: Pilot Project—Automated Vehicles and Revised Regulations of Ontario 1990, Regulation 628: Vehicle Permits—Summary" (2021), online (pdf): *Ontario Ministry of Transportation* <ontariocanada.com/registry/showAttachment.do?postingId=39087&attachmentId=50976> [perma.cc/ZX3U-6JNH]; Bylaw: City of Toronto, by-law 1075-2021, *To amend City of Toronto Municipal Code Chapter 886, Footpaths, Pedestrian Ways, Bicycle Paths, Bicycle Lanes and Cycle Tracks, City of Toronto Municipal Code Chapter 950, Traffic and Parking, and City of Toronto Municipal, Code Chapter 610, Penalties, Administration of, in respect of regulating micro-utility devices* (17 December 2021) [Bylaw].

26 See Canadian Press, "Toronto company behind pink delivery robots to temporarily pull devices off city's sidewalks", *The Globe and Mail* (9 December 2021) online: <theglobeandmail.com/business/article-toronto-company-behind-pink-delivery-robots-to-temporarily-pull> [perma.cc/AW6C-WTWN]. The same article reports that following the proposed motion (which later passed), sidewalk robot company Tiny Mile announced its interest in working with the disability community to help "make [their] robots safer for the community, [and] greatly benefit people with disabilities."

27 Automated vehicles are seen as a harbinger of greater highway "safety" (through collision mitigation), compared to human driven cars, while also bringing special concerns about safety, including cybersecurity and algorithmic/sensor fallibility. Transport Canada,

services, as well as other government administrative bodies, have also engaged this justification for turning to facial recognition technology (FRT) and other automated surveillance technologies.[28] However, the extent to which these technologies are "safe" is widely contested.

For example, when investigative journalism revealed that various police forces in Canada had been using the Clearview AI facial recognition system, some agencies responded that the systems were used to protect "public safety."[29] Substantial public outcry followed this revelation. Clearview AI illegally collected photographs of people in Canada in order to develop the system.[30] But the use of the system itself also raised

---

Canada's Vehicle Cyber Security Guidance, Catalogue No T46-61/2020E-PDF (Ottawa: Transport Canada, 2020) at 5-8 [Transport Canada, Cyber Security].

28 See Office of the Privacy Commissioner of Canada, *Police use of Facial Recognition Technology in Canada and the way forward*, Catalogue No IP54-110/2021E-PDF (Ottawa: Office of the Privacy Commissioner of Canada. 2021); Kate Robertson, Cynthia Khoo, & Yolanda Song, "To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada" (1 September 2020), online: *Citizen Lab* <citizenlab.ca/2020/09/to-surveil-and-predict-a-human-rights-analysis-of-algorithmic-policing-in-canada/> [perma.cc/PK9Z-84WR]. See also Petra Molnar & Lex Gill, "Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System" (September 2018), online (pdf): *Citizen Lab* <citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf> [perma.cc/PK9Z-84WR]. With regard to other administrative contexts, see for instance: Bill 80, *An Act Respecting Highways and Traffic Safety*, 2nd Sess, 63rd Gen Ass, Nova Scotia, 2018 s 319 (assented to October 2018) [*Bill 80*], proposing the use of facial recognition for identification with respect to licences.

29 See e.g. Royal Canadian Mounted Police, News Release, "RCMP use of Facial Recognition Technology" (27 February 2020), online: *Royal Canadian Mounted Police* <rcmp-grc.gc.ca/en/news/2020/rcmp-use-facial-recognition-technology> [perma.cc/58SW-5C25] ("While we [the RCMP] recognize that privacy is paramount and a reasonable expectation for Canadians, this must be balanced with the ability of law enforcement to conduct investigations and protect the safety and security of Canadians, including our most vulnerable [referring to victims of online sexual abuse])" (at para 10). See also Office of the Privacy Commissioner of Canada, News Release, "News release: RCMP's use of Clearview AI's Facial Recognition Technology Violated Privacy Act, investigation concludes" (10 June 2021), online: *Office of the Privacy Commissioner of Canada* <priv.gc.ca/en/opc-news/news-and-announcements/2021/nr-c_210610/>[perma.cc/VJ3B-56T5] ("The RCMP will continue to strive for improvement in its ability to identify and use new technologies to support our mandate of protecting and policing our communities" at para 17).

30 See the report & findings of the Privacy Commissioners: "Joint investigation of Clearview AI, Inc by the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta" (2 February 2021), online: *Office of the Privacy Commissioner of Canada* <priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/> [Privacy Commissioner Joint Report].

concerns about harm and safety for those upon whom it would be used.[31] Facial recognition systems have been widely cited for inaccuracies, especially along lines of race and gender.[32] The harms associated with false identification and false arrest are obvious.[33] However, experts warn that even if an FRT system could work with high levels of accuracy, it would feed into an extension of state surveillance and activities that also have documented harm.[34] This is not to say that every single use of automated identification software should be decried.[35] Rather, the various ways in which FRT has been documented and anticipated to *cause* harm are not captured within the general appeal to "public safety" justifying its use.

---

31 The House of Commons Standing Committee on Access to Information, Privacy and Ethics (ETHI) report discusses RCMP use of Clearview AI beginning at page 20. The RCMP reports that the system was actually applied in three cases, one of which involved suspect identification abroad and the other two involved providing safeguards to and protecting victims of crime: at 22. House of Commons, Standing Committee on Access to Information, Privacy and Ethics, Facial Recognition Technology and the Growing Power of Artificial Intelligence (October 2022), (Chair: Pat Kelly), online: https://www.ourcommons.ca/DocumentViewer/en/44-1/ETHI/report-6/ [ETHI Report].

32 See e.g. Buolamwini & Gebru, *supra* note 14.

33 See e.g. in the US context: Thomas Germain, "Innocent Black Man Jailed After Facial Recognition Technology Got It Wrong, His Lawyer Says", *Gizmodo* (3 January 2023), online: <gizmodo.com/facial-recognition-randall-reid-black-man-error-jail-1849944231> [perma.cc/EY4D-YHEF]; Khari Johnson, "How Wrongful Arrests Based on AI Derailed 3 Men's Lives" *Wired* (7 March 2022), online: <wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/> [perma.cc/RS5S-6K2B]; Miriam Marini, "Farmington Hills Man Sues Detroit Police after Facial Recognition Wrongly Identifies Him" *Detroit Free Press* (13 April 2021), online: <freep.com/story/news/local/michigan/2021/04/13/detroit-police-wrongful-arrest-faulty-facial-recognition/7207135002/> [perma.cc/V3D9-N4U6].

34 See Evan Selinger & Woodrow Hartzog, "The Inconsentability of Facial Surveillance" (2019) 66 Loy L Rev 101; Benjamin, *supra* note 14; Evan Selinger & Brenda Leong, "The Ethics of Facial Recognition Technology" in Carissa Véliz, ed, *The Oxford Handbook of Digital Ethics* (forthcoming). This can also include dignitary harms in relation to the capture and unwanted use of one's image, see e.g. the proposed class action suit in Canada which names copyright and moral rights in the proposed claims: "Facial Recognition Lawsuit—Clearview AI" (3 February 2021), online: *Public Safety Canada* <publicsafety.gc.ca/cnt/trnsprnc/brfng-mtrls/prlmntry-bndrs/20201119/023/index-en.aspx> [perma.cc/2X2U-29FU].

35 Image matching software that uses a hashing system, for instance, has been used to identify and remove child sexual abuse material from the Internet. Notably, this software is not FRT, it is an image identification system. See for example: "What is Project Arachnid" (2023), online: *Project Arachnid* <projectarachnid.ca/en/#what-is-project-arachnid> [perma.cc/9UTE-2ZFS].

There is relatively little direct governance of the use of algorithmic policing technologies in Canada at present.[36] However, within the limited existing governance, there is a suggestion of a foundational expectation that use of these tools will enhance public safety.[37] For instance, the Toronto Police Services Board (TPSB) policy for Toronto Police use of AI-systems in policing, the first algorithmic policing policy of its kind in Canada, states: "Novel technologies making use of artificial intelligence (AI) applications hold the promise of improving the effectiveness of policing services and *increasing public safety* in Toronto."[38] The policy anticipates that algorithmic tools can increase safety, so long as potential harmful impacts can be mitigated or balanced against perceived "safety" benefits.

Generally speaking, increasing oversight over law enforcement use of algorithmic systems is constructive. Nevertheless, it is not self-evident that all automated tools promise to enhance "public safety," particularly given the harms already associated with some of them.[39] The Canadian House of Commons Standing Committee on Access to Information,

---

[36] See e.g. Bill C-27 proposing the *Artificial Intelligence and Data Act* (though much substance of this *Act* is remains to be laid out in future regulations); Treasury Board, "Directive on Automated Decision-Making" (1 April 2021), online: *Government of Canada* <tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> [perma.cc/QD7R-DTYH]; Privacy Commissioner Joint Report, *supra* note 30; "Use of Artificial Intelligence Technology" (28 February 2022), online: *Toronto Police Services Board* <tpsb.ca/policies-by-laws/board-policies/195-use-of-artificial-intelligence-technology>[perma.cc/Q8MY-95ZZ]. Additionally, as noted above, the House of Commons ETHI Committee issued a report and numerous recommendations for the Federal government in relation to facial recognition regulation, see note 19.

[37] See e.g. Letter from Kristen Thomasen, Suzie Dunn, Kate Robertson, et al to Dr. Dubi Kanengisser, Senior Advisor, Strategic Analysis and Governance, Toronto Police Services Board, "Submission to the Toronto Police Services Board's Use of New Artificial Intelligence Technologies Policy—LEAF and The Citizen Lab",online: <papers.ssrn.com/sol3/papers.cfm?abstract_id=3989271> [perma.cc/3E4M-24BK].

[38] Toronto Police Services Board, *supra* note 36. Emphasis added.

[39] See the examples of lawsuits following police encounters generated by FRT: *supra* note 33. Notably, the claimants in these suits are predominantly Black men. Numerous reports and investigations have demonstrated that policing in Toronto has been disproportionately violent and harmful for Black communities in the city, e.g.: "Paying the Price: The Human Cost of Racial Profiling" (21 October 2003), online (pdf):*Ontario Human Rights Commission* <www3.ohrc.on.ca/sites/default/files/attachments/Paying_the_price%3A_The_human_cost_of_racial_profiling.pdf> [perma.cc/99JF-HAHH]; "A Collective Impact: Interim report on the inquiry into racial profiling and racial discrimination of Black persons by the Toronto Police Service" (November 2018), online (pdf): *Ontario Human Rights Commission* <ohrc.on.ca/en/public-interest-inquiry-racial-profiling-and-discrimination-toronto-police-service/collective-impact-interim-report-inquiry-racial-profiling-and-racial-discrimination-black> [perma.cc/2YWE-T2SX];

Privacy, and Ethics (ETHI Committee) has recognized some of the socially harmful impacts of FRT specifically.[40] The Committee shared a number of governance recommendations with the federal government including a recommendation to adopt a temporary partial moratorium on law enforcement use of the technology, given its potential for harm (contradicting any assumptions that the technology already or inherently enhances "public" safety).[41] A deeper interrogation of what it means for a system to be "safe" is necessary, especially where "safety" is the justification for devoting public funds and state power toward the use of a system.

Despite the fact that "safety" often justifies or guides state governance of automated technologies, statutes and policies regulating AI and robotics in Canada have not explicitly or directly defined what "safety" means. For instance, in drone regulation, which centers around the concept of "aviation safety," and concerns for personal safety (e.g., of people on the ground), neither the *Aeronautics Act*, nor the *Canadian Aviation Regulations* define the terms "safety" or "safe."[42] The meaning of aviation and personal safety can be deduced from the legislation and regulations, which focus on issues relating to aircraft (including drones) operating in predictable ways such that they do not cause collisions, crashes, or personal or property damage. While these are all important objectives, this essay contends that the meaning of "safety" need not, and ideally will not, be limited to notions of physical and property harm resulting from collision.[43] A safe airspace requires more than collision-mitigation. Any conception of safety that is limited to specific risk-mitigation is insufficiently narrow, particularly given the ways in which "aviation" and "public safety" goals

---

Shanifa Nasser, "Toronto Police Chief to Apologize to Black Community as Force Unveils Race-Based Data: Sources", *CBC* (14 June 2022), online: <cbc.ca/news/canada/toronto/toronto-police-apology-black-community-race-based-use-of-force-1.6488225> [perma.cc/35W3-27YH]

40 ETHI Report, *supra* note 31 at 13-35.

41 ETHI Report, *supra* note 31. See in particular Recommendation 18, and the conclusion at 64, *supra* note 31: "Without an appropriate [governance] framework, FRT and other AI tools could cause irreparable harm to some individuals." There appears to be an ever-growing groundwork emerging within law- and policy-making bodies in Canada encouraging a cautionary assessment of when a system is "safe" for use in public and shared spaces, including the ETHI report recommendations and temporary sidewalk robot ban in Toronto. This essay builds on that groundwork.

42 *Aeronautics Act*, RSC 1985, c A-2; *Canadian Aviation Regulations*, SOR/96-433

43 Notably, Transport Canada has been increasingly attuned to the privacy implications of drone use, and has shared non-enforceable but helpful privacy guidelines for drone operators as well. See "Privacy Guidelines for Drone Users" (2 December 2021), online: *Transport Canada* <tc.canada.ca/en/aviation/drone-safety/learn-rules-you-fly-your-drone/privacy-guidelines-drone-users> [perma.cc/EX7P-ZZD8] [Transport Canada, "Privacy Guidelines"].

can be employed to override individual and community interests that might be engaged under the regulatory framework.[44]

Similarly, automated vehicles are being introduced and tested on Canadian roads through the lens of "traffic" and "highway safety."[45] The legal framework for automated vehicles is not as extensively developed yet as that which exists for drones. But similarly, current laws do not directly define the meaning and content of highway- or motor vehicle-"safety."[46] However, it is evident that the regulations are predominantly focused on collision mitigation. Some policy documents also include considerations around cybersecurity, privacy, and data management, reflecting an understanding of safety that includes and goes beyond

---

44 See e.g. *Gill v Canada (Minister of Transport)*, 2015 BCCA 344 at paras 27-31 (holding that the Minister's broad powers to protect aviation safety and ensure public safety operate against a finding that the Minister could be liable in negligence to an operator, even where a ministerial decision has severe economic consequences for the individual); *Brown v Canada (Attorney General)*, 2014 FC 1081 at para 71 (regulatory scheme gives the Minister wide discretion to refuse to grant a Transportation Security Clearance, even if that means loss of employment for the applicant, if the Minister feels there may be a risk to public/aviation safety; the Minister is entitled to prioritize public safety over the interests of the individual); *Rivet v Canada (Attorney General)*, 2007 FC 1175 at para 15 (interests of public safety take precedence over individual interests); *Canada (Minister of Transport, Infrastructure and Communities) v Farwaha*, 2014 FCA 56 at para 94 (it is appropriate for the Minister to prioritize public safety over individual interests, Security Clearance decisions do not ascribe blame as in criminal law, rather they are "forward-looking and predictive" of risks to public safety); *Randhawa v Canada (Minister of Transport)*, 2017 FC 556 at para 18 (concludes the same reasoning as noted in the preceding aviation safety cases also applies with regard to marine safety).

45 See *Pilot Project - Automated Vehicles*, O Reg 306/2015; *Bill 80*, *supra* note 28; Bill 23, *The Vehicle Technology Testing Act (Various Acts Amended)*, 2nd Sess, 42nd Leg, Manitoba, 2020 (not carried forward).

46 As Transport Canada explains with respect to federal regulations: "The objective of these regulations is to reduce the risk of death, injury, and damage to property and the environment." Transport Canada, Cyber Security, *supra* note 27 at 11. The Ontario Court of Appeal considered the safety objective in *R v Michaud*, 2015 ONCA 585 at paras 74, 106, 115, 127, and 148, where an accused brought a *Charter* challenge against the requirement for speed limiters set at 105km/hr in trucks, adopted for the purpose of increasing highway safety (implicitly, preventing accidents and reducing severity when they occur). The claimant contended that the requirement undermined his safety when he needed to drive faster than the limit in order to remain safe on the highway. The court recognized the requirement was part of a "complex regulatory response to" a complicated and highly technical social problem (road safety). The Court held that where a challenge like this arises, courts should be more deferential at the section 1 analysis where the law has been found to violate s. 7. Erring on the side of public safety is a legitimate exercise of public authority. See also *Di Cienzo v Attorney General of Ontario*, 2020 ONSC 4347, also finding exercise of precaution deserves deference under *Charter* section 1.

physical and property harm.[47] At the federal level, there has also been explicit recognition of the need to consider "safety" (seemingly physical/ property damage from collisions) for both drivers and more vulnerable users of the roadways, like pedestrians and cyclists.[48]

It is also worth noting that robotics and AI governance is not unique in its omission of a definition of "safety," or for using "safety" as a broad term where a policy goal might in fact be stated more narrowly (e.g., "collision mitigation" or "individual identification"). Across several statutes in Canada where "safety" is a legislative goal, the term is either vaguely or not at all defined.[49] Even statutes that are entirely premised on the notion of achieving a goal of "safety" do not provide a definition, including some cases where being deemed to have jeopardized "safety" can lead to severe consequences.[50]

---

47 See: Transport Canada, Cyber Security, *supra* note 27; Transport Canada, *Safety Assessment for Automated Driving Systems in Canada*, Catalogue No T86-52/2018E-PDF (Ottawa: Transport Canada, January 2019) [Transport Canada, Safety Assessment].

48 Transport Canada, Safety Assessment, *supra* note 47. See also Quebec's *Highway Safety Code*, CQLR c C-24.2, s 3.1.

49 A rare exception includes a policy document—the Public Safety Management Directive for Parks Canada, which defined its Guiding Principle 7 of "public safety" as "a coordinated effort to ensure that visitors to parks, canals and sites have a positive experience while minimizing the potential for suffering or loss. Public safety deals with the measures employed to reduce the risk of an incident occurring or to protect visitors from a hazard; and measures to be implemented in the event that an incident develops requiring emergency response capabilities." See e.g., *Public Service Alliance of Canada v Parks Canada Agency* (24 November 2008), 2008 PSLRB 97, online: PSLRB <canlii.ca/t/221m6> [perma.cc/W5VY-XRBT] at para 24 citing evidence from the Directive; Performance, Audit and Review Group, "Evaluation of Parks Canada's Public Safety Program" (February 2005) at 6, online (pdf): *Parks Canada Agency* <parkscanadahistory.com/publications/evaluations/public-safety-eval-e-2005.pdf> [perma.cc/72HP-LKKF]. The Parks Canada Agency approach to public safety is specifically a risk-based approach (at 3).

50 For example, the *Child, Family, and Community Service Act* permits removal of a child from their family if their health or safety is in immediate danger: RSBC 1996 c 46, s 30(1). "Safety" is not defined in the *Act*. Section 2 sets out guiding principles for the interpretation of the *Act*, which should apply to an interpretation of whether a child's safety is in danger. Still, this severity of consequence would seem to necessitate the clearest definition possible, but "safety" is not included. Section 77 of the *Criminal Code of Canada* bears the possibility of up to life imprisonment for Offences Against Air or Maritime Safety, without a direct definition of safety or specifically how one might endanger the safety of a craft. The subsections include examples like: "render an aircraft incapable of flight *or* … endanger the safety of the aircraft" (s. 77(c) emphasis added). While there are examples of specific unlawful conduct, these are distinguished from the notion of "endangering safety."

One explanation for this phenomenon could be that the meaning of safety is self-evident and needs no definition. Alternatively, one might say that the intended goal can simply be inferred from reading the legislation or policy document and noting which kinds of risks or harms are being mitigated.[51] By contrast, in the next section of this essay I contend that, while risk-mitigation is a crucial goal in the process of introducing automated technologies into society, the overarching governance of "safe" vehicles, highways, airspace, and technologies should also consider more broadly what it means to *create* safe public spaces and safe technologies, beyond just eliminating named risks.[52] The concept of safety, used without precision, can become a floating signifier—a nebulous term imbued with whatever meaning its user finds helpful to galvanize power in a given direction.[53] Canada already has a tumultuous history with broad governance appeals to "safety" to justify a range of harmful policies and state conduct. Safety is not a neutral or straightforward concept, and decision-makers would be well-served by devoting more attention to providing clarity to this protean objective. This is an issue that is not exclusive to AI and robotics regulation, but is pertinent in this context, given the significant social and legal impacts that might arise from different understandings of what it means for an AI or robotic system to operate safely within social contexts.

3. Toward a More Comprehensive Understanding of Safety

Laws and policies that fail to define safety, or its intended beneficiaries (sometimes simply couched as the "public"), leave a range of potential governance objectives unspoken. Broad appeals to an undefined notion

---

[51] A handful of judicial decisions have implicitly taken this approach, or have briefly relied on an Oxford Dictionary definition of safety to interpret a *Criminal Code* requirement, which I note below is a narrow understanding of the concept. See e.g., *R v Sinclair*, 2020 ONCA 61 at paras 23-25 (finding that jurors can understand the legal requirement of "safety" without more than an instruction that it includes not only physical but also psychological safety); *R v Shapira (H)*, [1997] AJ No 588 (QL), 203 AR 299 (Alta Prov Ct) at paras 42-43 (relying on the Oxford Dictionary definition to briefly interpret a *Criminal Code* requirement of causing someone to fear for their safety, meaning not just physical but also psychological and emotional).

[52] I have previously argued that the notion of aviation safety could consider the ways in which drones might engage privacy concerns and enhance surveillance, among other things: Kristen Thomasen, "Beyond Airspace Safety: A Feminist Perspective on Drone Privacy Regulation" (2018) 16:2 CJLT 307. [Thomasen, "Airspace"] At the time of writing, the drone regulatory framework did not include privacy considerations as part of "safety." I argued that excising privacy from the notion of "safety" leaves anyone who is particularly susceptible to privacy intrusion at greater risk of being unsafe due to surveillance and its possible consequences. See also: Michael Froomkin & Zac Colangelo, "Privacy as Safety" (2020) 95 Wash L Rev 141.

[53] *Supra* note 4.

of safety can be perilous. Legal regulations guided by the objective of safety/public safety have on many occasions improved the lives and lived experiences of members of the public.[54] However, at various times in Canadian legislative history, laws and policies aimed at attaining the purported goal of "safety" have compromised the safety of already politically oppressed communities, furthered the violent process of colonialism, prompted over-policing, and enabled dangers, undermining the safety of some under the misconstrued guise of attaining "safety" for all.[55]

More specifically, various writers have connected strategic or political appeals to "public safety" back to the early emergence of policing in Canada to protect the "safety" of one group (white-settlers) in the process of dispossessing Indigenous peoples from their lands, as well as in relation to the surveillance and control of enslaved or formerly enslaved Black people living in Canada.[56] Professor Robyn Maynard explains how more recently in Canada, policies like carding and drug programs emerged to address perceived threats to "public safety" in response to moral panic

---

[54] For instance, vehicle safety regulations gained traction in the 1960s, and reduced the number of deaths caused by motor vehicle accidents. It is important to note though, that even though these regulations had an overall positive effect in reducing driver mortality, the notion of "safe" motor vehicles was nevertheless still *not* neutral. Studies have shown numerous examples of gender bias in safety testing (testing for an assumed weight and height of an average man), resulting in injuries and deaths to smaller drivers, mostly women. See e.g., Ralph Nader, *Unsafe at Any Speed: The Designed In Dangers of the American Automobile* (New York, NY: Grossman Publishers, 1965); Caroline Criado Perez, *Invisible Women: Data Bias in a World Designed for Men* (New York, NY: Abrams Press, 2019); and see Sam Peltzman, "The Effects of Automobile Safety Regulation" (1975) 83:4 J Political Economy 677 (contending that while the rate of driver deaths had deceased following the adoption of safety regulation, this was offset by increased pedestrian deaths).

[55] As I will explain in more depth in sub-sections a and b below, these concerns arise both because of how "safety" is understood *and* how the "public" is understood (the "public" who is perceived as the beneficiary of safety can be implicitly or explicitly limited, such that the safety and wellbeing of some are compromised in the interests of a loose notion of "safety" for others).

[56] See e.g. Shiri Pasternak, "Canada is a Bad Company: Police as Colonial Mercenaries for State and Capital" in Shiri Pasternak, Kevin Walby, and Abby Stadnyk, eds, *Disarm, Defund, Dismantle: Police Abolition in Canada* (Toronto: Between the Lines, 2022) at 66; Rinaldo Walcott, *On Property* (Windsor, ON: Biblioasis, 2021); Robyn Maynard, *Policing Black Lives: State Violence in Canada from Slavery to the Present* (Winnipeg: Fernwood Publishing, 2018); Robyn Maynard & Leanne Betasamosake Simpson, *Rehearsals for Living* (Toronto: Knopf Canada, 2022) at 189; Simone Browne, *Dark Matters: On the Surveillance of Blackness* (Durham, NC: Duke University Press, 2015).

over drugs and false perceptions of the prevalence of gang activity.[57] She explains how these policies have directly harmed Black communities, rendering Black people and communities in Canada less safe.[58] Increased policing of public spaces has also been justified on the grounds of protecting safety, but with a focus on the safety of private property and its owners, to the detriment of safety for unhoused communities.[59] Many other examples of appeals to "public safety" have arisen through Canadian history wherein "safety" has justified direct or indirect violence against already marginalized communities—purportedly members of the very "public" meant to benefit from safety.[60] Technology has long facilitated the weaponization of "safety" against different communities.[61]

---

[57] See also Christina Sharpe's discussion of Mayor Bloomberg's claim that stop and frisk policies "save lives." In response to former New York City mayor Michael Bloomberg's statement to Black church congregants that stop-and-frisk policies 'save lives', Sharpe argues: "We must ask whose lives are being saved, who in fact is in possession of a life that can be saved, because it is clear that in at least one direction Black lives are being destroyed" discussed by Maynard, *supra* note 57 at 64.

[58] Maynard, *supra* note 57 at 88-102. This harm has been reiterated through empirical and statistical studies. See e.g. Jason Miller, "Peel Region Police Used Force on Black People 3.2 Times More Often Than Their Share of the Population in 2021, Police Data Shows", *The Toronto Star* (26 August 2022), online: <thestar.com/news/gta/2022/08/23/peel-regional-police-used-force-on-black-people-32-times-more-often-than-their-share-of-the-population-in-2021-police-data-shows.html?utm_source=share-bar&utm_medium=user&utm_campaign=user-share> [perma.cc/ECU6-D9T2]. I would be remiss not to recognize that there is not universal agreement, including from harmed communities, that the solution to these issues is the abolition of policing in all its forms. For the purposes of this paper though, the point that appeals to "public safety" can lead to harm to the public, including to the very communities meant to be protected, is important. See also, I Bennet Capers, "Afrofuturism, Critical Race Theory, and Policing in the Year 2044" (2019) 94 NYUL Rev 1 (imagining policing in a future informed by both critical race theory and Afrofuturism).

[59] Walcott, *supra* note 56 at 79-80. This has been stark in recent years in Canada with forceful removal of tent encampments in public spaces in major Canadian cities including Toronto and Vancouver, e.g. Canadian Press, "Toronto's Ombudsman to Investigate Homeless Encampment Clearings", *CBC News* (28 September 2021), online: <cbc.ca/news/canada/toronto/ont-toronto-encampments-1.6192498> [perma.cc/2JGX-PCC8]; Kwame Addo & Ciarán Buggle, "Interim Report: Investigation Into the City's Process for Clearing Encampments in 2021" (14 July 2022), online (pdf): *Ombudsman Toronto* <ombudsmantoronto.ca/Investigative-Work/Investigative-Reports/Investigation-Reports/Report-Folder/Early-Recommendations-to-Improve-City-Response-to.aspx?ext=.pdf> [perma.cc/HUF6-6PSS]. State removal of unhoused communities from public spaces undermines the safety of those who are dispossessed, especially when few if any alternatives are offered for housing and accommodation, and when the process of removal is itself violent. If the wellbeing of all were a genuine concern, housing would be offered as an alternative for people who share and rely on public space for private and personal needs like sleeping and survival. See e.g. the #StoptheSweeps Coalition in

These examples are obviously not comprehensive; they speak to the ways in which broad appeals to "public safety" have been used to attain political goals despite paradoxically compromising the safety of often already marginalized communities.[62] These examples might initially seem unrelated to the regulation of drones, automated vehicles, sidewalk robots, and AI-systems. But how the term "public safety" is understood and deployed, even in seemingly narrow contexts of technical regulations, can have significant impacts on the actual safety and lived experiences of members of the public. In Section IV below, I explore examples where broad appeals to "public safety" in AI and robotics governance have

---

Vancouver: "Stop the Sweeps: End daily displacement and dispossession in Vancouver's Downtown Eastside" (2022), online: *Stop the Sweeps* <stopthesweeps.ca/> [perma.cc/6XER-FJD3].

[60] Safety has been a justification for the interment of Jewish refugees, and members of the Japanese and Ukrainian communities in Canada, among others (see e.g. Denis Smith, "War Measures Act" in *The Canadian Encyclopedia* (25 July 2013) online: <thecanadianencyclopedia.ca/en/article/war-measures-act>) [perma.cc/JTU6-3FEK]; for terrorizing queer public service workers (see e.g. Gary Kinsman & Patrizia Gentile, *The Canadian War on Queers: National Security as Sexual Regulation* (Vancouver: University of British Columbia Press, 2010) at 122); demonizing drug users (see e.g. Johann Hari, *Chasing the Scream: The First and Last Days of the War on Drugs* (London, UK: Bloomsbury, 2016); Bigo Didier, "Security and Immigration: Toward a Critique of the Governmentality of Unease" (2002) 27:1 (Supplement) Alt J 63. See especially Maynard, *supra* note 57 at 95: "The only success of the War on Drugs was to demonize Black life, massively expand Black incarceration and cause irreparable harm to Black communities."); and endangering sex workers (see e.g. Ellie Ade Kur & Jenny Duffy on behalf of Maggie's Toronto Sex Workers Action Project, "Sex Worker Justice—by Us, for Us: Toronto Sex Workers Resisting Carceral Violence" in Pasternak, Walby, & Stadnyk, *supra* note 56).

[61] See Maynard, *supra* note 57; Browne, *supra* note 57; Molnar & Gill, *supra* note 28. See also Kinsman & Gentile, *supra* note 60 at 168: "The attempt to devise a machine or a battery of psychological tests that could scientifically detect homosexuals was part of the growing obsession with identifying queers as a national security threat."

[62] See e.g. Sylvia Wynter & Katherine McKittrick, "Unparalleled Catastrophe for our Species? Or, to give Humanness a Different Future" in Katherine McKittrick, ed, *Sylvia Wynter: On Being Human as Praxis*, (Durham, NC: Duke University Press, 2015) 9. See also Dancy & Saucier, *supra* note 14 (drawing on Wynter in discussing the embeddedness of anti-Blackness in AI). Maynard emphasizes how the "public" in public safety (also in "public health" in the context of COVID, "public space" in the context of the housing crisis) is not inclusive, and its exclusions are based in settler colonialism and racial capitalism. Maynard & Simpson, *supra* note 56 at 19, 58, 64-65, 72; Maynard, *supra* note 56 at 102, at 88: "Black existence in public spaces is itself seen as criminal and thus subject to scrutiny, surveillance, frequent interruption and police intervention." Limits on the scope of "public" entitled to safety are also reflected in the criminal law system's treatment of interpersonal colonial violence, see Gina Starblanket & Dallas Hunt, *Storying Violence: Unravelling Colonial Narratives in the Stanley Trial* (Winnipeg: ARP Books, 2020) (noting themes of "safety" favouring white property-owners over the lives of Indigenous "intruders").

already undermined the safety of members of the public—an outcome that is incompatible with governance goals of ensuring the safe integration of automated technologies into society. Before examining these specific governance examples and associated recommendations, though, the below sub-sections first draw upon critical and abolitionist writing to outline different ways that safety can be understood, to help ground the subsequent analysis.

A) Narrow Visions of Safety

Safety is commonly understood as a state of being free from dangers and threats of harm.[63] The typical method for attaining this vision of safety (freedom from danger) is to eliminate or mitigate known and anticipated risks and threats. While common, this approach to safety is descriptively and conceptually limited. It draws policy attention toward the removal or elimination of things, conduct, and people considered threatening, rather than building conditions of safety, strengthening capacity to handle harm, or addressing the root causes of risk/danger.[64] While some forms of risk-mitigation do contribute to the broader goal of safety, as a guiding principle for governance this view will not on its own create public safety in relation to AI and robotic systems. When the limits of this vision are left unrecognized, it can threaten to undermine safety in a range of contexts.

Scholar Meaghan McDowell has named this policy objective "carceral safety," referring to the method of attaining the goal of safety (actually, risk-mitigation) through which the state seeks to control and punish those deemed to be threatening or frightening.[65] The state achieves control through policing, surveillance, punishment, and the removal of people

---

[63] The Oxford English Dictionary defines "safety" as "the state of being safe," elaborated as "the state of being protected from or guarded against hurt or injury; freedom from danger." *Oxford English Dictionary* (Oxford, UK: Oxford University Press) sub verbo "safety, n.".

[64] While this might be very important for the purposes of risk mitigation in specific contexts (e.g. having various protections in place to mitigate lapses and failures in the operation of an airplane or a motor vehicle), this is but one specific and narrow understanding of what "safety" entails. I suggest this approach can be better understood as "risk-mitigation" which contributes to but is not exhaustive of what is necessary for safety. See for example James Reason, "The contribution of latent human failures to the breakdown of complex systems" (1990) 327:1241 Philosophical Transactions of the Royal Society B 475 (who also does not contend that risk mitigation is comprehensive of safety but rather that it contributes to greater overall safety in relation to complex socio-technical systems).

[65] Meaghan G McDowell, "Insurgent Safety: Theorizing Alternatives to State Protection" (2019) 23:1 Theoretical Criminology 43. See also Jessi Lee Jackson & Erica R Meiners, "Fear and Loathing: Public Feelings in Antiprison Work" (2011) 39 Women's Studies Quarterly 270 at 278. E.g. at 278: "the definition of what makes a place or

who have been deemed unsafe.[66] Carceral safety is often enacted through the banishment (to prison or through deportation) of people deemed a risk.[67]

Some AI and robotics governance is explicitly carceral—using technologies to identify and remove people from social contexts in the name of protecting safety; or deeming people who operate technologies in violation of risk-mitigation requirements to be dangerous and in need of state removal for the sake of preserving safety.[68] Other AI and robotics governance might not directly implicate state policing, removal, and imprisonment. However, this governance is still shaped by the insights of this articulation of a narrow vision of safety, contrasted with a more inclusive and comprehensive way of giving meaning to "safety," outlined below. It is helpful to explicitly recognize that eliminating a perceived threat is not the same as building safety. These approaches do not have to be mutually exclusive in AI and robotics governance, so long as risk-mitigation efforts (e.g., braking design requirements for automated vehicles) are but one contribution to other more systematic efforts to build safety, and do not themselves compromise the safety of members of the public.

This narrow vision of safety often operates through the creation of binaries in government policies, such as "safety" vs "fear," or "safety" vs "threat." Through this process, safety is the absence of its opposite (fear, or threat), which implies that removing the opposite, on its own, creates safety. Binaries establish the grounds for policies aimed at "targeting the

---

community safe is most often shaped by absence: absence of violence and intimidation, or in some cases absence of discomfort."

66 Jackson & Meiners, *supra* note 65 at 276: "public safety is approached through attempts to control or remove those bodies, in part through the expansion of prisons." See also Kur & Duffy, *supra* note 60.

67 McDowell, *supra* note 65 at 45 explains, "The idea that safety stems from banishment (via jail, prison, or deportation), mass criminalization, and policing is taken as common sense, despite its deadly consequences." Technology can also play a key role in asserting this carceral view of safety, for instance by keeping those who are socially othered and deemed unsafe out of shared spaces: see e.g. Meghan G McDowell & Nancy A Wonders, "Keeping Migrants in Their Place: Technologies of Control and Racialized Public Space in Arizona" (2009) 36:2 Soc Justice 54.

68 While much drone/automated vehicle regulation is technical in nature, there are many instances where violations of risk-mitigation requirements (thereby becoming "risky") lead to carceral consequences including arrest and the possibility of imprisonment. See e.g. *R v Shah*, 2017 ABPC 259 (accused found guilty of flying in a manner hazardous to aviation safety). See also *R v Minot*, 2011 NLCA 7. And see *Criminal Code* section 77, supra note 50. See also the examples in Section IV below.

sources of fear, or risks, that are *assumed* to be threatening this safety."[69] Scholars Ida Sjöberg and Katarina Giritli Nygren emphasize that in government planning:

> the ambition to produce safety or reduce unsafety often encourages the production of safekeeping and risk management strategies in which safety is often synonymous with the absence of activities and groups that are considered 'disorders' within the desired cityscape, upon which control measures aimed at blocking or removing these groups and activities are introduced as safety strategies.[70]

This quotation echoes an important point emphasized by abolitionist scholars - that a narrow conception of safety can also unduly narrow the scope of *whose* safety matters. The "public" in "public safety" can be varyingly inclusive or exclusive depending on the political circumstances. Inconsistency and exclusion are not supportive of a broad sense of public or community safety.[71] Those deemed to be a risk or a threat become the focus of elimination, and are no longer beneficiaries of "safety" practices and policies. Where the creation of safety means removing people from an environment in ways that render those very people unsafe, their safety does not matter in the scope of that governance decision.[72] Rhetoric that stokes fear of crime, and connects this fear with particular communities or conduct, reinforces a sense that safety equals the removal of members of those communities or people who engage in the proscribed conduct,

---

69 Ida Sjöberg & Katarina Giritli Nygren, "Contesting City Safety: Exploring (un) safety and Objects of Risk from Multiple Viewpoints" (2021) 24 Journal of Risk Research 1251 at 1251, emphasis added. See also J Simon Hutta, "Geographies of *Geborgenheit*: Beyond Feelings of Safety and the Fear of Crime" (2009) 27:2 Environment and Planning D: Society and Space 251. As summarized by Sjöberg and Giritli Nygren, Hutta emphasizes that "safety" amounts to "the formation of a regime of power that operates by staging binaries of "safety" and "fear" in governmental practices, thus creating the grounds for [policing], safekeeping practices and policies, as well as targeting the sources of fear, or risks, that are assumed to be threatening this safety." See also, Hutta, at 254: "The fact that new forms of exclusion and control are articulated through the concepts of safety and fear is not only an effect of the generalised 'concerns about personal safety' but also announces a change regarding urban planning as well as the relations between governmental actors and subjects."

70 Sjöberg & Giritli Nygren, *supra* note 69 at 1253. The authors build on this work through community interviews to better understand "safety" in urban planning and design.

71 See Bell, *supra* note 21.

72 In their study on urban safety efforts in Sweden, Sjöberg and Giritli Nygren's interviews with users of a soup kitchen reveal that: "being targeted as a threat to public safety, and thereby being on the receiving end of different safekeeping practices aimed at enhancing public safety, is strongly associated with feelings of unsafety among those who are constructed as 'risk objects' and do not conform to the desired cityscape." Sjöberg & Giritli Nygren, *supra* note 69 at 1259.

effectively removing those communities from the political scope of the public.[73] This narrative process can occur specifically to galvanize political support for a government seeking to control the state's structures.[74]

Those deemed outside the scope of the "public" who ought to be protected by safety governance can also be rendered unsafe through the neglect or deprioritization of their interests in the development of regulatory frameworks. Thus, this narrow vision of safety can also become incomplete through often-intrinsic prioritization of which risks ought to be mitigated.[75] Anyone whose experience of risk falls outside the scope of the decision-makers' focus will not even experience safety as a freedom from danger or threat, let alone a more comprehensive sense of safety.[76] This is not to suggest that decision-makers ought to address every single risk. Rather, there are other ways to approach the governance goal of creating or protecting "safety," which could help decision-makers think more comprehensively about how to govern automated technologies. A comprehensive understanding of safety can help decision-makers more explicitly articulate governance goals that are in fact narrow in focus, rather than draw upon broad but inaccurate appeals to protecting "public safety." Safety in this usage becomes a floating signifier, and unless decision-makers articulate what exactly they mean in their governance objectives, the ambiguous use of "safety" as a policy goal will always threaten inequity and exclusion for those not deemed within the "public" at any given

---

[73] For further reference on this broad notion see e.g. Ruth Wilson Gilmore, *Golden Gulag: Prisons, Surplus, Crisis, and Opposition in Globalizing California* (Berkeley, CA: University of California Press, 2007); Gilmore, "Geography", *supra* note 5; and this podcast conversation: "Ruth Wilson Gilmore w/ Alberto Toscano and Brenna Bhandar" (28 May 2022), online (podcast): *The Dig* <thedigradio.com/podcast/ruth-wilson-gilmore-w-alberto-toscano-and-brenna-bhandar/> [perma.cc/8VBG-484B]; Beth Richie, *Arrested Justice: Black Women, Violence, and America's Prison Nation* (New York, NY: New York University Press, 2012) at 21; Zach Norris, *Defund fear: Safety without policing, prisons, and punishment*, ed (Boston, MA: Beacon Press, 2021).

[74] See Gilmore, "Geography", *supra* note 5 at 271-72.

[75] See e.g. Brenna Bhandar, "Organised State Abandonment: The Meaning of Grenfell", *The Sociological Review* (4 October 2022), online: <thesociologicalreview.org/magazine/october-2022/verticality/organised-state-abandonment/> [perma.cc/AWT7-RTHR].

[76] I have commented on this issue in the context of the exclusion of privacy from the scope of "safety" in drone regulations. Failing to consider the ways in which the particular affordances of drones will exacerbate privacy invasions, especially in public and shared spaces, leaves those who are at greater risk of privacy invasion in public space vulnerable to unsafety. Greater vulnerability to privacy invasion often arises along intersecting axes of gender, race, class, disability, and cis/hetero-normativity. Ultimately, a failure to view privacy invasion as a "risk" that should be mitigated to attain "safety" renders the law inequitably unhelpful to already privacy-marginalized communities. See Thomasen, "Airspace", *supra* note 52.

moment. Risk-mitigation governance can co-exist with a broader view of safety, but decision-makers must understand their narrow goals as contributing toward, but not exhaustive of, a more comprehensive and inclusive vision, if they are going to actually create safety in relation to automated systems.

B) An Affirmative/Comprehensive Vision of Safety

In contrast to a narrow view of safety focused on elimination are articulations of a range of generative and affirmative conditions that help create personal and public safety. Conditions of safety emerge from the development and strengthening of community bonds and reliable networks of support (including strengthening inter-relationality,[77] interdependence,[78] mutual aid,[79] care,[80] and trust[81]), restorative and transformative practices of justice (as contrasted to punitive and

---

[77] As a helpful explainer, see: Norris, *supra* note 73 at 144: "If safety is anything, it is relationship and connection. […] Real safety happens when we bridge the divides and build relationships with each other, overcoming suspicion and distrust. […] Real safety results from reinstating full humanity and agency for everyone who has been dehumanized and traumatized, so they can participate fully in society"; and at 15: "replace deprivation, suspicion, punishment, and isolation with resources, relationships, accountability, and participation, what taken together I call a 'culture of care'".

[78] See e.g. McDowell, *supra* note 65 at 52-53: interdependence as friendship, kinship, community networks, caring about others, agreeing that no one is to be thrown away, having trust between people; "interdependence [as] an epistemological orientation grounded in trust, mutuality, and shared vulnerability". This can include reciprocity, cooperative exchange, collective housing, neighbourhood safety patrols, free daycare, shared land projects, community emergency resource funds, *etc*.

[79] See e.g., Dean Spade, *Mutual Aid: Building Solidarity During This Crisis (and the Next)*, (London, UK: Verso, 2020); Leah Lakshmi Piepzna-Samarasinha, *Care Work: Dreaming Disability Justice* (Vancouver: Arsenal Pulp Press, 2018); Arthur L Ross, *Communal Solidarity: Immigration, Settlement, and Social Welfare in Winnipeg's Jewish Community, 1882-1930* (Winnipeg: University of Manitoba Press, 2019).

[80] In many ways, see the resources listed in the footnote above. And see e.g., Kaba, *supra* note 8; Maynard & Simpson, *supra* note 56.

[81] See e.g. Nora Samaran, *Turn This World Inside Out: The Emergence of Nurturance Culture* (Chico, CA: AK Press, 2019).

[82] See Defund the Police Coalition (Montreal), "Defund to Abolish: A 400-Year Struggle against Policing in Montreal" in Pasternak et al, *supra* note 56 at 34; Free Lands Free Peoples, "A Brief Introduction to Anti-colonial Abolition" in Pasternak et al, *supra* note 56 at 74. See also Norris, *supra* note 73 at 66: "[A] key component of restorative justice is to view violence prevention as a public health issue. […] Rather than asking: 'What law was broken, who broke it, and how should they be punished?' restorative justice asks, 'Who was harmed? What do they need? Whose responsibility is it to meet those needs?'"; *ibid* at 31:

> Shifting the focus away from crimes to harms means we address actions, policies, and behaviors that are most harmful […] it means that when it comes

eliminatory practices),[82] environmental protection and stability,[83] and opportunities for individual and collective well-being (including opportunities for play, joy,[84] rest, psychological security).[85] Safety is physical, but it is also emotional, intellectual, economic, spatial,[86] and environmental.[87] Creating comprehensive conditions of safety includes building broad public and community capacity to address violence, harm, danger, and/or vulnerability when they invariably arise. This can mean ensuring economic support, health care, child care, and environmental resources, among other considerations, are available as needed and are equitably created and distributed.[88] Similarly, it can entail resourcing community-level governance of different aspects of technology use.[89]

---

time to address harms and keep further harms from happening, we involve far more bodies than merely the law; the players include academics, policymakers, community leaders, historians and community members who are involved in arenas such as public health, epidemiology, urban planning, and social policy.

See also, Amy Goodman, "Angela Davis on Abolition, Calls to Defund Police, Toppled Racist Statues & Voting in 2020 Election," *Democracy Now* (12 June 2020) online: <democracynow.org/2020/6/12/angela_davis_on_abolition_calls_to> [perma.cc/K2PP-D9DV] ("Safety, safeguarded by violence, is not really safety"); Maynard & Simpson, *supra* note 56 at 181 where Simpson articulates Nishnaabeg visions of safety.

[83] See Maynard & Simpson, *supra* note 56.

[84] See McDowell, *supra* note 65 at 54: interviewees discussed play, joy, and communion as mechanisms of safety. Safety sounds like "people having a good time." Interviewees engaged in "re-reading safety as a sensory experience embodied through play, laughter, and kinship" (at 55). Play, for instance, provides opportunities to set one's own rules, to relieve stress that can lead to conflict (at 55). Participants also noted that "people should just do art together" (at 55). See also Dionne Brand, "June", *The Globe and Mail* (June 4, 2007).

[85] See Samaran, *supra* note 81.

[86] Safety is also understood spatially in the sense that one's physical space can contribute to an assessment of safety. See e.g. Lisa Weems, "From 'Home' to 'Camp': Theorizing the Space of Safety" (2010) 29:6 Studies in Philosophy & Education 557; Gilmore, "Geography", *supra* note 5.

[87] See Jackson & Meiners, *supra* note 65 at 279-280: proposing a concept termed "affective safety" which recognizes that feelings of safety or fear are based on more than control of external threats, but rather on "the current state of body and mind within a web of histories, relationships, bodily experiences, and physical environments." See also Froomkin & Colangelo, *supra* note 52.

[88] Jackson & Meiners, *supra* note 65 at 280: Safety is both "fluid and multidetermined". Building capacity to address different forms of violence or vulnerability requires: "economic resources, affirming relationships, care, healthy environments, free time, and opportunities for expression" are relational and based on recognizing full context of actors/broader social relations.

[89] See e.g. Sasha Costanza-Chock, *Design Justice: Community-Led Practices to Build the Worlds We Need* (Cambridge, MA: MIT Press, 2020), in particular the Design Justice Principles at 6-7.

Safety, through an affirmative lens, means people have what they need to live and thrive in community (and in conflict) with one another.[90] To be safe goes beyond one's personal circumstances; recognizing human and environmental interdependence, it requires the well-being of others as well as the environments in which one lives.[91] This means, among other things, recognizing the environmental impact of the extractive practices used to obtain the material resources needed to build many of the automated technologies considered here.[92]

Such a conceptualization of safety may initially appear too broad to directly inform discrete rules within a regulatory framework, such as braking requirements for automated vehicles, or airworthiness or operational requirements for drones. But understanding how safety is *created* helps situate these rules within the wider context of the purported governance goal of "safety," and can help inform the application, adoption, and enforcement of these rules (explored further in Section IV). Furthermore, a comprehensive understanding of what people need to be safe in relation to automated technology could contribute to more

---

90 See e.g. Niklas Möller, Sven Ove Hansson & Martin Peterson, "Safety is more than the antonym of risk" (2006) 23:4 J Applied Philosophy 419.

91 See Ross, *supra* note 79; see also Mia Mingus, "Interdependence (excerpts from several talks)" (22 January 2010), online (blog): *Leaving Evidence* <leavingevidence.wordpress.com/2010/01/22/interdependency-exerpts-from-several-talks/> [perma.cc/9YDE-N5GL].

92 The limited view of the "public" deserving of safety also allows law and policy to overlook the global and planetary impact of technologies adopted in the name of safety. The planetary extraction required to develop robotic systems, the spatial and environmental impact of cloud-based robotics, the impact of extraction encampments on local populations and environments, and of labour standards and working conditions, are all implicated in the devices regulated or used with a view to "safety." A comprehensive vision of public safety would not end at the border when technology is multinational; and might simply require *not* adopting a new technology, especially in the name of safety, where its creation produces global insecurity. See e.g., Anja Nygren, Markus Kröger & Barry Gills, "Global extractivisms and transformative alternatives" (2022) 49:4 J Peasant Studies 1; Joseph Muchiri Githiria & Moshood Onifade, "The impact of mining on sustainable practices and the traditional culture of developing countries" (2020) 10:4 J Environmental Studies & Sciences 394; Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven, CT: Yale University Press, 2021); Kate Crawford & Vladan Joler, "Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources" (7 September 2018), online: *AI Now Institute and Share Lab* <anatomyof.ai/> [perma.cc/3AWQ-9DPU]; see also, Abeba Birhane and Jelle van Dijk, "Robot Rights, Let's Talk about Human Welfare Instead" (Paper delivered at AIES '20, 7-8 February 2020) online: <https://arxiv.org/pdf/2001.05046.pdf> [perma.cc/D6AK-7FCJ].

innovation—both in governance (incorporating a broader range of policies, agencies, and authorities) and in technology.[93] Notably, these more comprehensive understandings of safety also correspond with some emerging recommendations from computer science and design scholars who argue that AI and robotics "safety" must be understood as addressing more than physical harm, including a comprehensive understanding of people's needs.[94]

Abolitionist activist, organizer, and educator Mariame Kaba has written about visions of safety that can exist outside of narrow eliminatory and carceral approaches. Safety, she explains, comes from building relationships with one another, acting in ways to keep one another safe, and importantly, developing means to address harm that do not cause further harm.[95] Kaba envisions safety as conditions where "we have everything we need: food, shelter, education, health, art, beauty, clean water, and more things that are foundational to our personal and community safety."[96] Technology could help create some of these conditions. For example, drones are already prolifically used for art and entertainment, as well as distributing medical supplies, and in education, all of which have been facilitated by amendments to Canada's drone regulations.[97] Decision-makers can also increasingly work to build norms and conditions of

---

93 The idea that comprehensive safety can provide a safety-net for experimentation and risk-taking adds an important insight to thinking about AI and robotics governance. The nature of many algorithmic/automated systems means that unpredictable (and potentially harmful) outcomes might arise that can be hard to mitigate in advance given the complex nature of the system and unpredictability in exactly what this harmful outcome might be. Where everyone can trust that they will be cared for should risk materialize into harm, a "public" might be more willing and able to take on the risks associated with otherwise beneficial AI systems. Safety, through this lens, allows for the more equitable taking of risks in the development of new technologies. See the concept of "emergence" in Ryan Calo, "Robotics and the Lessons of Cyberlaw" (2015) 103:3 Cal L Rev 513, and discussion of risk governance in Margot Kaminski, "Regulating the Risks of AI" (Paper delivered at We Robot Seattle, September 2021), online: werobot2022.com/program/.

94 See e.g. Alberto Martinetti et al, "Redefining Safety in Light of Human-Robot Interaction: A Critical Review of Current Standards and Regulations" (2021) 3 Frontiers in Chemical Engineering 1.

95 See Kaba, *supra* note 8 at 98-99. She explains later at 155, "as an abolitionist, what I care about are two things: relationships and how we address harm. The reason I'm an abolitionist is because I know that prisons, police, and surveillance cause inordinate harm."

96 Kaba, *supra* note 8 at xviii.

97 In particular, by doing away with the Special Flight Operations Certificate application process for a range of drone operations. E.g., Government of Canada, News Release, "Minister Garneau unveils Canada's new drone safety regulations" (9 January 2019), online: <https://www.canada.ca/en/transport-canada/news/2019/01/minister-garneau-unveils-canadas-new-drone-safety-regulations.html> [perma.cc/RMT3-WGFW].

technology use such that people no longer jeopardize one another's safety (looking beyond strictly mitigating specific risks). Transport Canada's gradual introduction of privacy guidance and norms for drone operators could be an example of the agency contributing to this broader vision of safety.[98] Automated and robotic systems that facilitate communication, trust-building, accountability, expressive freedom, joy, environmental protection, health care, and economic support may variously contribute to this vision of safety.[99]

Zach Norris has written of the distinction between a fear-based (narrow) understanding of safety, where safety solely means being free from crime, and what he explains as a care-based model, or a culture of care.[100] Within this latter vision, "safety can replace deprivation, suspicion, punishment, and isolation with resources, relationships, accountability, and participation."[101] As Jessica Evans, Alannah Fricker, and Rajean Hoilett write, "safety takes place through the establishment of effective systems and communities of care, not as an after-the-fact response to lack of care."[102] Nora Samaran writes of a nurturance culture in which safety is forged through trust between people. This trust serves as the safety-net that allows for taking risks, knowing that one will be supported should harm materialize.[103]

---

98 *Supra* note 43.

99 See e.g. Costanza-Chock, *supra* note 89 (on the value of community-led design); Juri Viehoff, "Beyond Justice: Artificial Intelligence and the Value of Community" in Bullock et al, *supra* note 3 (on the use of AI systems to create communities); Ian R Kerr, Jason Millar, Noel Corriveau, "Robots and Artificial Intelligence in Health Care" in Joanna Erdman, Vanessa Gruben, Erin Nelson, eds, *Canadian Health Law and Policy*, 5th ed (Toronto: LexisNexis Canada, 2017) 257 (on the opportunities and risks of AI in healthcare); see also the example of resisting violence through drone use below, in Section IV.

100 In *Defund Fear* (*supra* note 73), Zach Norris refers to this as a culture of care: "There are two ways to think about safety. There is a fear-based way and a care-based way […] The fear-based model defines safety only in terms of being free from *crime* and *criminals*. […] [O]ur current fear-based system paradoxically generates more harm than it prevents, in never-ending cycles of trauma" (at 14-15).

101 *Norris, supra* note 73 at 15.

102 Jessica Evans, Alannah Fricker & Rajean Hoilett, "We Keep Each Other Safe" in Pasternak et al, *supra* note 56 at 64. The authors emphasize that, "research has shown that a lack of access to social, material, physical, and emotional supports and well-being play a significant precipitating, if not determining, factor in" stories where individuals have actually committed violent crimes (they simultaneously explain how most people held in prison in Canada have not committed a violent crime, and many have not yet actually been convicted but are held prior to trial): at 61.

103 See Samaran, *supra* note 81.

Greater state communication with, and more authority for, local communities in various aspects of technology governance are also important in attaining broad experiences of safety. As a contrast to her phrasing of 'carceral safety,' McDowell, in conversation with community research participants, proposes the term 'insurgent safety'[104] to capture a more comprehensive way of envisioning safety. She describes insurgent safety as "those *locally determined* practices and ethics that … reconceptualize safety as a mode of sociality built through interdependence, mutual aid, counter-carceral communication, and play."[105] The state may impede such local practices, and where it does, safety would actually be best achieved through the withdrawal of public control/delegation to local practice.[106] Professor Barry Friedman has also extensively argued for decoupling the notion and objective of "public safety" from policing, and centering community voices in the reimagining of public safety.[107]

Notably, many of these broad and generative conceptualizations are not designed or intended to become legal or statutory definitions of "safety," and may be unworkable in such a context. But I am not introducing these considerations for the purpose of inclusion in statutes. Rather, I propose that legal and policy objectives should be better articulated, and be better designed to *contribute* toward a more comprehensive vision of safety in conjunction with other non-statutory approaches (e.g., education, funding and support for departments and organizations doing comprehensive safety work, exemptions or relaxing of statutory restraints on beneficial uses of technology, *etc*). The vast majority of laws pertaining to robots and AI systems in Canada are focused on mitigation of specific risks, and

---

104 McDowell, *supra* note 65, develops this vision of safety through interviews with members of her local community, who emphasize different priorities over elimination of threats.

105 McDowell, *supra* note 65 at 45, emphasis added.

106 Precisely when and how this interaction occurs in the scope of technology governance in Canada merits further direct consideration (but is beyond the scope of this essay). See e.g., *Costanza-Chock*, supra note 89.

107 This recommendation arises through a vision of police reform, not abolition. Barry Friedman, "Disaggregating the Policing Function" (2021) 169:4 U Pa L Rev 925 at 985-991. See also "Reimagining Public Safety: First Convening Report" (January 2021), online (pdf): *Policing Project & The Justice Collaboratory* <static1.squarespace.com/static/58a33e881b631bc60d4f8b31/t/602d826b6b3233405feabd52/1613595247852/RPS+Session+I+Report.pdf> [perma.cc/H8UR-RP94]. See also Capers, *supra* note 58. (imagining policing in a future informed by both critical race theory and Afrofuturism). While Capers imagines that surveillance and policing continue to exist in the interests of public safety, he contextualizes this within a very different social environment in which subordination is eliminated, and in which criminalization and punishment are approached through a critical race theory-based lens: "the primary concern of Critical Race Theory is eradicating subordination along all racial lines. … The goal of CRT is equality, including along lines of gender, sexuality, class, and disability.": at 136.

should be understood as such. Governance should no longer inaccurately wield "public safety" as a broad and comprehensive concept to override individual or collective conflicting interests in the regulation of automated technologies. While state governance aimed at risk-mitigation is meant to be carried out in the public's interest, it alone does not create public *safety* and should not be understood as doing so. When mischaracterized as creating safety, risk-mitigation rules can be interpreted in ways that actually undermine safety for members of the public, and typically in inequitable and oppressive ways (connecting to an exclusion of various communities and people from the scope of the "public"). This has been exhibited in recent AI and robotics governance examples, as explored in the next section.

4. Rethinking Appeals to "Safety" in AI & Robotics Governance

When the nebulous concept of "safety" is adopted to justify a risk-mitigation framework, it can divert attention away from a range of conditions actually needed to create safety.[108] Or, it can ground state decision-making and policies that compromise the actual conditions of safety for members of the public. This latter scenario can arise out of the ways in which "safety" is given meaning. If safety is understood as being "free from risk," then the concept can be galvanized to remove people, conduct, or conditions deemed a possible risk, even in ways that render those people unsafe, and even in technical regulatory contexts. This section canvasses two examples from outside of Canada where a broad appeal to safety has negatively impacted the safety of members of the very public purported to be protected. Drawing on these examples, I assert that at a minimum, a clearer articulation of law and policy goals, and a more nuanced and inclusive understanding of how safety can be achieved or undermined, are necessary in the development of "safety" governance of automated systems.[109]

A) Drone Regulation and Airspace Safety

This section explains how "safety" has been employed as a justification to enact restrictive drone regulation for the purpose of limiting the use of the technology in ways that promote safety. The examples noted here arose in the context of resistance movements in the U.S. that themselves were aimed at securing aspects of a comprehensive vision of safety. These

---

108 See e.g. Thomasen, "Airspace", *supra* note 52.

109 This, while working toward a state, and a state of being, in which marginalization and oppression no longer serve to galvanize power.

examples arose specifically through the federal laws permitting the Federal Aviation Administration (FAA, comparable to Transport Canada) to enact Temporary Flight Restrictions (TFRs).[110] While this paper focuses on Canadian governance, the U.S. examples are nevertheless instructive in showing how appeals to the notion of airspace or aviation safety can be made in pursuit of political or policing goals under a narrow safety framework similar to the one enacted in Canada.[111]

In the fall of 2016, thousands of Water Protectors and other demonstrators gathered at a camp on Standing Rock Sioux Tribe lands to oppose the development of the Dakota Access Pipeline.[112] The resistance sought to protect Indigenous sovereignty and the water, land, and environments to be impacted by the pipeline development.[113] Throughout this effort, Indigenous journalists and others used drones for a variety of information gathering purposes, including documenting the resistance movement, land, and pipeline construction; and to record instances of police use of force and brutality.[114] On November 20, 2016, law enforcement agents used water cannons, rubber bullets, tear gas, and concussion grenades against demonstrators. Cold water was especially threatening given the freezing temperatures at the time.[115] As this was occurring, journalist Myron Dewey filmed the use of water cannons using a drone and shared the soon-to-be viral footage online, drawing substantial

---

110 US, Federal Aviation Administration, *Temporary Flight Restrictions (TFR) and Flight Limitations*, (Advisory Circular), (AC-91-63D) (Washington, DC: US Department of Transportation, 2015), online (pdf): <faa.gov/documentLibrary/media/Advisory_Circular/AC_91-63D.pdf> [perma.cc/MA2J-D9ME] [FAA Circular].

111 Canadian and U.S. drone regulations are also similar—the two countries have had a cooperative history of regulation, and both are centered around a comparable notion of airspace and aviation safety, with federal government power over regulation. See e.g., on harmonization of statutory language: "Aviation initiatives planned for April 2022—April 2024" (31 March 2022), online: *Transport Canada* <tc.canada.ca/en/corporate-services/acts-regulations/forward-regulatory-plan/aviation-initiatives-planned#toc2> [perma.cc/2ACU-N6FH].

112 See e.g. Nick Estes *Our History is the Future: Standing Rock Versus the Dakota Access Pipeline, and the Long Tradition of Indigenous Resistance* (Brooklyn, NY: Verso, 2019).

113 *Ibid*.

114 See e.g. Gaia Casagrande, Mohamed Amine Khaddar & Stefania Parisi, "Technology and the Local Community: Uses of Drones in #NoDAPL Movement and Dandora Dumpsite Storytelling" (2020) 64:13 American Behavioral Scientist 1906.

115 The Indigenous Environmental Network reported that over 160 demonstrators were injured that night. *Ibid*.

116 Dewey was a member of Walker River Paiute Tribe. See Digital Smoke Signals (https://www.facebook.com/DigitalSmokeSignals/); Casagrande et al, *supra* note 115 at 1912.

public attention to the resistance movement and the police violence.[116] Drone technology specifically enabled the capture and sharing of footage that would otherwise not have been possible.[117] Helicopters or other aircraft are not easily available to individuals, and the aerial vantage point proved significant to the ability to convey what happened to the broader public while remaining safe from police use of force.[118]

This was just one significant instance of drone use at #noDAPL. Earlier that fall, journalist Aaron Turgeon, a member of Rosebud Sioux Tribe, was charged with criminal offences including reckless endangerment and risked up to seven years in prison for allegedly flying his drone near a police surveillance plane. He had not actually flown his drone near the plane or in an "unsafe" manner, as confirmed later by a judge who reviewed the video footage from the incident and found him not guilty.[119] However, the charges against Turgeon highlight the intersection of airspace laws and carceral visions of safety, wherein an allegation of unsafe use necessitates removal not just of the drone, but also the user from both the space and potentially from society for up to seven years. Meanwhile, in at least one other reported instance, law enforcement shot down a drone operated by a protestor, engaging safety concerns in relation to state use of deadly and destructive weapons.[120]

These incidents reflect an overarching vision of safety that is both narrow and carceral. Journalists using drones (in ways that do not risk aircraft collision) for information gathering purposes are threatened with imprisonment. Meanwhile, law enforcement agents reportedly mitigate the perceived risks associated with drone use by using dangerous weapons. In the same context, protestors experiencing physical harm use drones to obtain video footage for transparency and consequently, as I discuss below, lose access to these drone practices under the state rubric of protecting airspace safety. Safety loses any common sense meaning throughout these incidents other than amorphously serving as the justification for state control of the demonstration.

---

117 Letter from Patricia A McNall (Acting Chief Counsel at the Federal Aviation Administration) to Lee Rowland (counsel at ACLU) (17 June 2017) online (pdf): <faa.gov/about/office_org/headquarters_offices/agc/practice_areas/regulations/interpretations/Data/interps/2017/Rowland-ACLU_2017_Legal_Interpretation.pdf> [perma.cc/KZA7-3A36] [FAA-ACLU Letter].

118 Digital Smoke Signals, *supra* note 116.

119 Water Protector Legal Collective, News Release, "Prolific the Rapper Found Not Guilty" (25 May 2017), online: <waterprotectorlegal.org/post/prolific-the-rapper-found-not-guilty> [perma.cc/JZT9-ECUZ].

120 Chris Matyszczyk, "Police shoot down drone at Dakota Access pipeline protest", *CNET* (24 October 2016), online: <cnet.com/culture/police-dakota-access-pipeline-drone-protest/> [perma.cc/8729-SM8C]; FAA-ACLU Letter, *supra* note 107.

Following these occurrences, the FAA instituted temporary flight restrictions above the land where the demonstrations were taking place, near Cannon Ball, North Dakota.[121] The FAA imposes TFRs under the broad objective of protecting safety in relation to the airspace. The FAA explains more specifically that a TFR:

> is a regulatory action issued via the U.S. Notice to Airmen (NOTAM) system to restrict certain aircraft from operating within a defined area, on a temporary basis, *to protect persons or property in the air or on the ground.*[122]

The TFRs permitted police to operate drones and other aircraft in the airspace, but prohibited drone use by anyone else other than one non-local, non-Indigenous reporter.[123] Freedom of information requests following these incidents show that local and state law enforcement portrayed the use of drones by protestors as dangerous to police safety, in support of the request for a TFR from the FAA.[124] The initial flight restriction was enacted without any consultation with local Sioux Tribe leaders. Following later consultation, the flight restrictions were narrowed but continued over the site until December 16, 2016, despite local objections.[125]

This was not the first occasion where the FAA adopted a TFR in response to police concerns about aerial information gathering. During demonstrations in Ferguson, Missouri following the police killing of

---

121 See FAA-ACLU letter, supra note 118 at 1-2.

122 FAA Circular, *supra* note 110, emphasis added.

123 Police use of drones over protest, and to surveil Indigenous communities, has occurred already in Canada too. See Muna Mire, "Canada Funds High-Tech Drone Campaign Against Mohawk Tobacco Trade", *Vice News* (27 March 2014) online: <vice.com/en/article/8x7dmz/canada-funds-high-tech-drone-campaign-against-mohawk-tobacco-trade> [perma.cc/ZM6V-G3TT]. See also Jeffrey Monaghan, "Intense Police Surveillance for Indigenous Land Defenders Contrasts with Laissez-Faire Stance for Anti-Vax Protestors", *The Conversation* (27 October 2021), online: <theconversation.com/intense-police-surveillance-for-indigenous-land-defenders-contrasts-with-a-laissez-faire-stance-for-anti-vax-protesters-169589> [perma.cc/E8HP-8A4W]. That said, I recognize the limits of analogy here as Transport Canada has the ability to interpret its home statute differently and arguably should, given the different jurisdictional contexts, including among other things, Canada's different *Charter* values that can inform statutory interpretation. See e.g., Kristen Thomasen, "Private Law & Public Space: The Canadian Privacy Torts in an Era of Personal Remote-Surveillance Technology" (2022) Doctoral Thesis (unpublished), online: <ruor.uottawa.ca/handle/10393/43746> [perma.cc/M7ZQ-JCZ5].

124 Jason Koebler & Sarah Emerson, "FOIA: How Police Convinced the FAA to Put a No Fly Zone Over Standing Rock", *Vice* (27 September 2017), online: <vice.com/en/article/d3yx3a/foia-how-police-convinced-the-faa-to-put-a-no-fly-zone-over-standing-rock> [perma.cc/24QR-HNLQ].

125 FAA-ACLU Letter, supra note 118.

Michael Brown, the FAA granted a TFR in response to a law enforcement request for the specific purpose of "keep[ing] the media out."[126] That instance was focused on news helicopters. Nevertheless, it echoes this vision of airspace safety that seemingly prioritizes the protection of law enforcement agents from journalist video footage. This is a particularly troubling vision of safety given that, like #noDAPL, the Ferguson demonstrations were specifically calling for *more* public safety, in that instance specifically in relation to police.

Following the #noDAPL TFRs, journalist Myron Dewey and others explained the importance of drones to the resistance at Standing Rock, including providing a sense of safety and security to demonstrators on the ground below:

> Drones level the playing field … They get us out of jail, they have saved us from having to get close to police to document what they're doing and to document that, while these atrocities and abuses of power are happening, work as usual on the pipeline has been happening.[127]

Based on reports from drone operators, drones had been used at #noDAPL as part of an insurgent safety practice (to apply McDowell's terminology) wherein demonstrators sought to keep each other safe by creating a

---

[126] On August 9, 2014, police officer Darren Wilson killed 18-year-old Michael Brown. In the course of the public demonstrations that followed, local law enforcement asked the FAA to adopt a TFR over the area. A consequent Freedom of Information request by the Associated Press revealed transcripts of phone conversations in which it was apparent that the TFR was requested by police in order to "keep the media out" of the scene. Conversation transcripts revealed an explicit derogation from the goal of safety, in a way that compromised the role of the press in reporting to the public about politically significant events, in this case specifically responding to police violence. See reproductions of the transcripts here: Associated Press, "Ferguson police asked for no-fly restrictions to keep news helicopters away", *The Guardian* (2 November 2014), online: <theguardian.com/us-news/2014/nov/02/ferguson-police-no-fly-restrictions-missouri-news-helicopters-michael-brown-shooting> [perma.cc/P7FP-V2EV]; Associated Press, "Transcripts of FAA phone conversations about Ferguson no-fly order", *The Guardian* (2 November 2014), online: <theguardian.com/us-news/2014/nov/02/transcripts-of-faa-phone-conversations-about-ferguson-no-fly-order> [perma.cc/7KMV-WX42].

[127] Jason Koebler, "The Government is Using a No-Fly Zone to Suppress Journalism at Standing Rock", *Vice News* (30 November 2016), online: <vice.com/en/article/yp3kak/the-government-is-using-a-no-fly-zone-to-suppress-journalism-at-standing-rock> [perma.cc/3VM7-ZA62]: Rhianna Lakin also comments in this piece: "They [drones] provide a sense of security to the water protectors to know that they're in the air documenting the truth … It provides truthful and accurate documentation of what's happening, so we can take the statements the Morton County Sheriff is putting out and verify it." These observations are significant in other contexts as well, where technologies have facilitated wider collective awareness of state violence.

record of encounters with police, and generating public visibility and the possibility for accountability in the event of violence. The FAA terminated this safety practice, paradoxically guided by the governance objective of "safety."

This paradox can of course be reconciled through reference to the different ways in which the "public" and "safety" are envisioned. The FAA is responsible for protecting a *narrow* eliminatory vision of airspace safety, one that may be more accurately understood as a focus on collision-mitigation, through state control over airspace. The FAA has authority and responsibility to prevent and remove airspace threats that engage air traffic, or bodily or property harm, including through TFRs. As emphasized above, this is not the same as creating conditions of safety. While the TFR over #noDAPL was enacted in response to reported concerns about the safety of police operations (which were, ironically, a primary source of bodily harm for demonstrators), people on the ground in Standing Rock became *less* safe through the adoption of the TFR. Indigenous peoples and, in many cases, demonstrators have long been removable by the state from the scope of the "public" that benefits from federal state safety measures—this is not a one-off fluke in FAA reasoning, but a feature of the colonial history of North America.[128]

This is not meant to suggest that the alternative is unrestricted drone use—requirements that mitigate crashes, and consider privacy and other concerns, factor into a broader understanding of what it means for drones to be operated safely. But a governance framework that recognized its own narrow focus on collision-mitigation as *but one* component of airspace safety might operate differently here. It might demand more evidence of actual unsafe flights before shutting down drone operations (or, would be open to receiving evidence showing a temporary TFR could promptly be cancelled). It might encourage local consultation prior to adopting a TFR, seeking to understand *why* drones are being used, and by whom, consideration if pilots have flight training and experience, enacting local or federal rules that could guide how drones should be operated in this situation rather than a complete prohibition, *etc*. I am not suggesting that the FAA or Transport Canada must single-handedly enact a comprehensive vision of safety. However, understanding TFRs, and drone rules more generally, as risk-mitigation options that do not themselves

[128] One may view this as a case of *terra/airspace nullius*—drawing on the colonial conception of *terra nullius*, the legal principle that presumes the land in North America as unoccupied despite many Indigenous nations living on the land long prior to and throughout the period of colonization. *Terra nullius* is used to justify the acquisition of a territory through the state's occupation of it. In this case, the FAA occupied total control of the airspace, on the basis of federal claim of control over the airspace, regardless of any pre-existing or superseding claims to that air. See e.g. Maynard & Simpson, *supra* note 56.

create safety, in conjunction with genuine governance commitment to comprehensive and inclusive safety, could generate a different experience of safety and of a safe airspace.

B) Algorithmic Policing and Public Safety

As noted above, safety is sometimes cited as a reason for state agencies to adopt automated technologies. This has been the case with state adoption of various automated technologies for policing and border control, including facial recognition and predictive policing systems. Above, I have noted some of the concerns raised about the inaccuracies present in facial recognition systems.[129] Inaccuracies in these systems can unjustifiably compromise public safety by bringing members of the public into contact with the harms of false arrest and imprisonment. But even imagining these systems as accurate, their use can compromise the safety of many members of the public.[130]

For example, the use of surveillance systems at borders has been framed as a safety measure for keeping dangerous people and goods out of a country, while paradoxically bringing targeted individuals into unsafe contact with police and threatening deportation to destinations where they may be unsafe.[131] Medina Bazargali briefly explains this paradox in relation to FindFace, the Russian facial recognition company working with the Russian state that claims to make Moscow "safe."[132] This system is used to identify, detain, and deport Central Asian migrants.[133] Bazargali

---

[129] See note 27.

[130] See also Danielle K Citron & David Gray, "Addressing the Harm of Total Surveillance: A Reply to Professor Neil Richards" (2013) 126 Harv L Rev Forum 262.

[131] See Molnar & Gill, *supra* note 28 ("The Canadian government is also investing in new border security and identification efforts, many of which have been criticized by immigration lawyers and civil liberties advocates alike. They include an expansion of efforts to collect fingerprints of foreign nationals and a move toward developing a 'Known Traveller Digital Identity concept' to facilitate 'pre-vetting risk assessment and security procedures,' for example, by allowing for risk-based immigration lanes. In July 2018, it was also revealed that CBSA has employed the use of private third-party DNA ancestry services such as Familytreedna.com and Ancestry.com to establish the nationality of individuals subject to potential deportation" at 22); see also Petra Molnar on the impact of algorithmic decision-making at the border: "Governments' Use of AI in Immigration and Refugee System Needs Oversight", *Policy Options* (16 October 2018), online: <policyoptions.irpp.org/fr/magazines/october-2018/governments-use-of-ai-in-immigration-and-refugee-system-needs-oversight/> [perma.cc/8463-PWNW];"Technological Testing Grounds and Surveillance Sandboxes: Migration and Border Technology at the Frontiers" (2021) 45:2 Fletcher Forum World Affairs 109. See also Thomasen, "Examining the Constitutionality of Robot Enhanced Interrogation" in Calo, Froomkin & Kerr, *supra* note 15.

[132] See Medina Bazargali, "Safety" in Ilan Manouach & Anna Engelhardt, eds, *Chimeras: Inventory of Synthetic Cognition* (Onassis Foundation, 2022) at 98-101.

emphasizes that "it is evident that the term "safety" doesn't mean universal safety for all, it means safety for a particular group of Slavic-looking residents of Russia, opposing them to the "dangerous" Central Asian subjects." However, the adoption of a surveillance infrastructure, made powerful through automation and underlying AI-systems, threatens the safety of dominant groups too. As Bazargali cautions, the same technology used by the Russian state to deport migrants can then be used to identify and police peaceful Russian protestors: "safety becomes defined as an infrastructure of tools for control, a double-edged sword created artificially in order to balance senses of fear and benefit in society, while harming everyone involved."[134] Support from some members of the public for the use of a "safety" tool in one context can lead, through mission creep, to a loss of safety for that very public later on.

Algorithmic surveillance tools threaten to entrench a narrow eliminatory vision of safety premised on the identification and removal of people categorized as dangerous or threatening. Such systems are not designed for building comprehensive safety. The adoption of these tools for surveillance does little if anything to remedy the root causes of any public safety issues purportedly being addressed; despite sometimes requiring a substantial procurement budget that could be used in other, safety generating, ways. Many algorithmic surveillance and policing tools adopted in the name of public safety belie both the limits of the scope of the "public" who are deserving of safety, and the scope of what it means to be "safe." Tools of this nature should either not be adopted, or should be subject to specific and explicit use restrictions developed under a rubric that works toward building a comprehensive form of public safety.

5. A Framework for Considering "Safety" as an Objective in AI & Robotics Governance

Despite the caution emphasized throughout this essay, safety is an important guiding principle in AI and robotics governance. Building on the preceding discussion, I suggest that at least three governance responses are merited: more precise articulation of governance goals, avoiding inaccurately broad appeals to "(public) safety" unless a genuinely inclusive notion of safety is being sought; that the state cease to use automated technologies in ways that undermine community safety, particularly

---

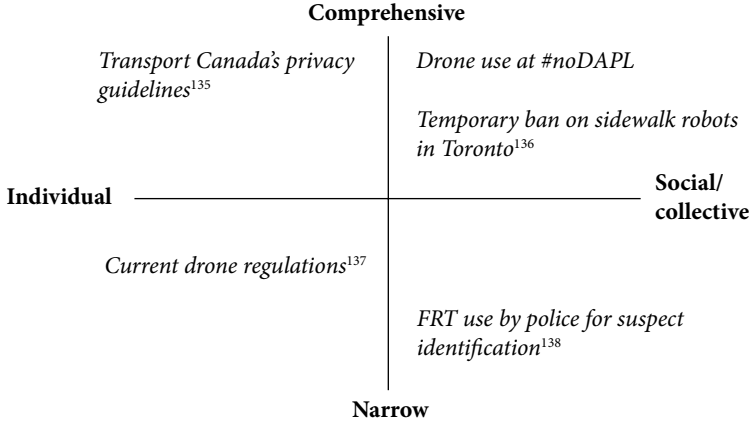
133 See Bazargali, *supra* note 131 at 99; "Safety" is as fluid as the category of Other: at 100.

134 See Bazargali, *supra* note 131 at 99-100: When technology like FRT can be used to engage in policing *after* a protest, police violence can also then be taken out of the scenes of the protest itself, shifting the public attention and scrutiny away from the violence, thus rendering it even more unsafe to participate.

under the guise of promoting public safety; and that governance of AI and robotics in Canada continue to actively build toward a more comprehensive notion of "safety" in relation to automated technologies, including through non-state governance bodies.

In the interests of better articulating governance goals, I propose the below framework for assessing the goals of robotics and AI governance. I caveat this framework by recognizing that in a federal system with various regulatory agencies, operationalizing comprehensive visions of safety can be complex. Drawing on the above discussion though, I suggest that it is necessary for decision-makers to, at least, be explicit about what a law or policy can and will actually achieve. This would aid in identifying potentially paradoxical scenarios where narrowly-focused regulation undermines a more inclusive vision of safety. It would also aid with identifying the gaps arising in governance meant to builds toward a comprehensive safety in relation to automated technologies, optimistically allowing for other modes of governance to fill those gaps. Finally, this framework is germane given the history of inequitable appeals to "public safety," the strong potential for automated technologies to affect individual and collective rights and interests, and contemporary examples that illustrate the already amorphous and political nature of "safety" in AI and robotics governance.

Figure 1



[135] These guidelines look beyond physical harm to consider the other ways that a socially significant technology can cause harm. Furthermore, they are not just focused on mitigating specific risks but appear to encourage a culture-shift/norms around privacy awareness, even in public space. The guidelines call on operators to understand and consider their impacts on others. While they are currently relatively brief, following the guidelines could contribute to trust-building, with and within the drone community. Offering norm-setting guidelines, as opposed to criminalizing operators, also enables a constructive rather than eliminatory approach to governance. These guidelines focus on *individual* concerns

The quadrants each engage different considerations:

- The 'narrow' quadrants on the bottom reflect a vision of "safety" as being free from specific risks, including physical injury, and property damage/loss. The examples cited in these quadrants seek to mitigate these risks through restrictive regulation and/or removal of the threat.

- The 'comprehensive' quadrants along the top include policies that aim to *create* or protect stronger community bonds and trust, conditions that promote physical and psychological integrity, expression, access, well-being, provide support in response to harms that do materialize, *etc*. These policies can make risk-mitigation less urgent or necessary, as they seek to address the reasons why people harm one another, and as they strengthen people's ability to address and cope with harm when it does occur.

---

and protections of privacy, which is why they are in the left quadrant, though implicitly through trust building and norm setting there are also collective considerations at play that engage the upper right quadrant. See Transport Canada, "Privacy Guidelines", *supra* note 43.

136 This policy is close to the horizontal axis. Even though it is removing the robot from the streets (elimination), the ban is adopted in the interests of creating comprehensive sidewalk safety, through which the robots may later be introduced. City Council noted the inequitable negative impact that the introduction of sidewalk robots could have on the public's use of the public sidewalk, rendering the accessibility and safety concerns here broad and comprehensive (seeing safety as more than just ensuring the robot does not collide with someone, and instead acknowledging that its mere presence or possibility of presence might make public access less safe and enjoyable for a range of communities). The policy is collective in the sense it recognizes how everyone benefits from broad and equitable use of public space, including the sidewalk. Additionally, this limit was adopted upon hearing from affected members of the community. It is designed to be temporary, leaving flexibility to later introduce technologies that enhance public space and public accessibility. For more detail, see Bylaw, *supra* note 25.

137 The existing drone regulations (this does not include the privacy guidelines which are not part of the regulations) focus on "safety" as predominantly synonymous with collision mitigation. This is a focus on a specific and narrow aspect of safety—mitigation of a specific kind of risk. It is concerned with the harm (physical or property) affecting people as individuals (as opposed to our collective social interrelationships).

138 This use of technology focuses on identifying, for the sake of removing, dangerous/threatening individuals (other uses of FRT have been proposed, but this example focuses on identification in the context of policing). Such removal of risk or danger is done for the purpose of a collective notion of "public" safety (not inclusive of the whole public, but more than individual). It is not just about keeping specific individuals safe but keeping all of the people who are considered part of the "public" safe from those deemed threatening. This policy is the lowest on the vertical axis.

- The 'individual' quadrants on the left include governance that focuses on individual behaviours (e.g., how individuals act in relation to automated technology, or in relation to each other via automated technology), and/or individual rights and interests (e.g., claims to physical integrity, individual privacy, *etc*).

- The 'collective' quadrants on the right include governance that works toward public or communal benefit, group goals, social norms, *etc*—goals that extend beyond the personal interests of individuals.

Governance appeals to "public safety" falling within the bottom quadrants of the grid could and should likely be more precisely articulated, and require correction when this re-articulation reveals a paradoxical approach that undermines community safety. Policies on the top of the grid may also be differently articulated but will fall within the broader range of considerations that arise from a more comprehensive approach to safety. While governance might appropriately fall in any of these quadrants, drawing from the above discussion, I make a normative appeal for governance to aim or contribute toward the upper quadrants.

Regulating with a view to supporting inclusive, comprehensive conditions of safety requires more work on the part of decision-makers. It necessitates finding out what people need to be and feel safe in particular environments where systems operate, or in relation to how and by whom those systems are used, and then finding ways to meet those needs and balance competing claims.[139] Toronto city council's temporary ban on sidewalk robots reflects this process, wherein City Council heard from affected parties and determined that the presence of robots on sidewalks might undermine the safety of the sidewalk space for many, which is a concern for the public as a whole. Engaging this process can also reveal (if not already obvious) the ways in which "safety" measures may fail to render spaces or communities safe, and in may in fact enact *more* harm.[140]

---

[139] Danielle Sered, *Until we Reckon: Violence, Mass Incarceration, and a Road to Repair* (New York, NY: The New Press, 2019). At page 13, Sered puts forward the following questions to consider: "Who is being hurt? Whose lives are at stake? What do they need to heal and be safe? Have we asked them what they want? What do they say when we do? When we ask those questions, we find not only fundamental challenges to our current thinking about who survivors are and what they want; we also find the seeds of a way forward to sustainable and lasting safety."

[140] For example, Sjöberg & Giritli Nygren, *supra* note 69 conducted group and individual interviews with community members from a range of backgrounds about safety in their community. While they noted that state practice focused on removing people thought to be "unsafe", various respondents emphasized other ways in which they wished for their community to be and feel safe. For example, at 1257, they indicate

This process was perhaps, to a limited extent, reflected through the FAA's decision to narrow its TFR over the #noDAPL demonstrations following consultation with local Sioux Tribe authorities.

Comprehensive approaches to safety should be of interest to agencies that seek to promote more social uses of automated technologies. When a technology improves conditions of safety and/or is used within inclusive conditions of safety, it may be more readily accepted by the broader communities who can now realize its benefits. Furthermore, a comprehensive vision of safety does not require entirely eliminating all risk. Because governance would focus on making people and environments more capable of managing harm, without creating more harm, such an approach may actually allow for more freedom to take risks with respect to innovation.[141] In other words, pursuing a more comprehensive and inclusive vision of safety could offer a supportive social framework allowing for bolder leaps of faith in relation to innovation and the use of AI and robotic systems. Building toward comprehensive and inclusive conditions of safety in relation to automated technologies should be the backdrop to any policies that aim to encourage technological innovation.

6. Conclusion

How governments regulate automated technologies is not inevitable, neutral, nor inherently equitable. "Safety," as a governance goal, should be considered more critically and be situated within its historic and contemporary context. Recognizing that "safety" can be understood as multi-dimensional and relational can offer practical steps forward when approaching robotics and AI governance. When evaluating existing or proposed laws and policies, one might consider for example:

- What vision of safety is being enacted? Whose safety is being protected? Who, if anyone, is excluded from the benefits of safety? Who, if anyone, is harmed by this law/policy framing or proposed use of technology?

- What do people need? (Has the decision-maker asked and listened? Are there local practices already in place to meet these needs?) Can the automated technology help meet those needs? Who should oversee its governance to successfully attain that goal?

---

that "risk and unsafety" could pertain to the structure and inaccessibility of the physical environment (cobblestones, icy streets, high curbs) or hegemonic norms of whiteness and heterosexuality.

141 See e.g. Kaminski, *supra* note 93 on how regulators address risk from automated systems.

That automated technologies can perpetuate social inequities is now well-established. Nonetheless, inequitable state governance of automated technology is but one possibility. Explicitly identifying governance goals does not on its own create safety in relation to automated technologies, but it does dissipate the mirage that a particular policy or law creates "safety" for the whole public, when in fact it might do the opposite. More importantly, AI and robotic governance should work toward a vision of "safety" in relation to automated technologies that strengthens community bonds, interpersonal and interagency trust, accountability, support, and care while mitigating (and certainly not creating more) harm and violence. Technology could contribute to these comprehensive notions of safety, but the goal must be explicit, inclusive, and comprehensive.