Hacking AI Governance:
Exploring the Democratic Potential of Canada's Algorithmic Impact Assessment

Nicholas Gertler

A Thesis
in
The Department
of
Communication Studies

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts (Media Studies) at
Concordia University
Tiohtià:ke (Montreal), Quebec, Canada

August 2023

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Nicholas Gertler

Entitled: Hacking AI Governance:

Exploring the Democratic Potential of Canada's Algorithmic Impact Assessment

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Media Studies)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

                                                              Chair

_____
              Dr. Alessandra Renzi

                                                             Examiner

_____
              Dr. Alessandra Renzi

                                                             Examiner

_____
              Dr. Martin French

                                                              Supervisor

_____
              Dr. Fenwick McKelvey

Approved by:

_____
Dr. Monika Kin Gagnon, Chair, Department of Communication Studies

_____
Dr. Pascale Sicotte, Dean Faculty of Arts and Sciences

August 28th, 2023

Abstract

Hacking AI Governance:

Exploring the Democratic Potential of Canada's Algorithmic Impact Assessment

Nicholas Gertler

Amid growing concern over the adoption of artificial intelligence systems, algorithmic impact assessments (AIAs) have increasingly been proposed as a means of measuring and mitigating the impacts of AI. Proposed AIA methods vary significantly in their approaches, but even within this heterogeneous group, the AIA tool released by the Government of Canada in 2019 stands out. This AIA tool—an open-source, online questionnaire platform—represents one of the first-ever attempts at putting principles of "responsible AI" into practice. In this research-creation thesis, I explore Canada's AIA tool as a media object, looking at the online questionnaire as a strategic opportunity to intervene in the growing debates about AI governance. Building on methods in critical making and civic hacking, this project includes the creation of both a critical guide to the AIA tool (aia.guide) and a series of "AIA hackathon" workshops designed to explore the tool's use by the Government of Canada and its potential in the broader AI governance context. Informed by a deep ambivalence over the technology (Bucher 2019), I argue that the AIA tool is largely performative but also represents an important site for tactical intervention. In particular, I argue that collaborative processes of questionnaire design may prove to be effective methods for participatory and community-based AI governance.

Acknowledgments

Table of Contents

Table of Figures

Introduction

How do you *do* "responsible AI"? Or "ethical AI"? Or "trustworthy AI"? Or whatever it is that we're calling it these days. Consider the *Montreal Declaration for a Responsible Development of Artificial Intelligence*. Announced in 2017, the Montreal Declaration outlines ten principles for responsible AI: well-being, privacy and intimacy, solidarity, democratic participation, equity, diversity inclusion, respect for autonomy, prudence, responsibility, and sustainable development (Université de Montréal 2018). To my ear, those principles seem sound; I too would like the development of advanced artificial intelligence systems to uphold the principles of equity, democratic participation, and so on. But how should those principles be upheld? The full text of the Declaration offers some further guidance on what is meant by each principle, but it does not indicate how those principles should be put into practice.

In July 2017—four months before the Montreal Declaration was announced—Michael Karlin published a blog post titled "Responsible AI in the Government of Canada: a Sneak Peek." Karlin worked for the Treasury Board of Canada Secretariat (TBS), the branch of the Government of Canada primarily responsible for overseeing the federal public service itself. The post brought together three related concepts: algorithms, AI, and automated decision-making.[1] Karlin outlined what he saw as both the potential and the risks for government use of artificial intelligence; government use of AI could mean more timely, efficient, and informed delivery of services, but poorly designed systems could ultimately undermine trust in government (Karlin 2017).

In that same blog post, Karlin announced the development of a "Digital Disruption White Paper" on the use of AI in Government. The white paper would inform the Government of Canada's position on its own use of AI and, exceptionally, it was going to be drafted in the open;

---

[1] For this project I use artificial intelligence and automated decision-making interchangeably. The TBS defines automated decision-making as, "Any technology that either assists or replaces the judgment of human decision-makers. These systems draw from fields like statistics, linguistics and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets." (Treasury Board of Canada Secretariat 2023). For reasons I discuss, the kinds of technology assessed by an algorithmic impact assessment mattered less to this project as it developed. For more on the concept of algorithms, see Beer 2017; Zarsky 2016; and Goffey 2008.

Algorithmic Impact Assessment

ⓘ Information in the AIA is only stored locally on your computer, and the Government of
Canada does not have access to the information you place into the tool. If you wish to keep your work, please save the data
locally for future use by using the 'Save' button. You can reload a previously saved AIA form using the 'Upload JSON File'
button.

[Save] [Upload JSON File] [Start Again]

Navigate to a Specific Page (Out of 13)

Section 3: Risk Profile ▲▼

Page 3 of 13

Risk Profile

Is the project within an area of intense public scrutiny (e.g. because of privacy concerns) and/or frequent litigation?

○ Yes
○ No

Are clients in this line of business particularly vulnerable?

○ Yes
○ No

Are stakes of the decisions very high?

○ Yes
○ No

Will this project have major impacts on staff, either in terms of their numbers or their roles?

○ Yes
○ No

Will the use of the system create or exacerbate barriers for persons with disabilities?

○ Yes
○ No

[Previous] [Next] [Complete]

✓ Impact Level: 1   Current Score: 0   Raw Impact Score: 0   Mitigation Score: 0

Figure 1: A screenshot of the "risk profile" section of the Treasury Board's AIA.

anybody (but mostly experts in relevant fields) would be able to see and comment on the actual draft, of the actual white paper, as it was being written. Thus began the multi-year, quasi-open policymaking process which led to the development of the Government of Canada's algorithmic impact assessment—the AI governance framework and software tool which has been my object of study, and the focus of this project, for the past two and a half years.

My research-creation thesis explores this algorithmic impact assessment (AIA) as a media object, looking at the online questionnaire as a strategic opportunity to intervene in the growing debates about AI governance. I am informed by a deep ambivalence over the technology (Bucher

2019). This AIA tool is largely performative, yet its significance marks an important tactical intervention. My project has tried to hack the AIA through a critical guide and workshop designed to use and reflect on the tool's potential, as well as new possibilities for AI governance.

This particular algorithmic impact assessment is significant because it is one of the first fully realized attempts at operationalizing the concept of "responsible AI." Indeed, to the best of my knowledge, it is *the* first such attempt in a public sector context. Coming at a time when high-level AI ethics guidelines and statements of principle were being published left, right, and center (Jobin, Ienca, and Vayena 2019), Canada's Treasury Board released its AIA as part of a concrete, mandatory policy for regulating government use of automated decision-making systems.

In practical terms, the TBS AIA is a glorified online questionnaire that government departments should complete when developing an automated decision-making system (ADM). Respondents answer questions about the proposed system—such as "Are clients in this line of business particularly vulnerable" or "Will the system be making decisions or assessments that require judgement or discretion?" The TBS AIA is designed to be used in conjunction with the Treasury Board's *Directive on Automated Decision-Making*. The Directive mandates when departments must complete an algorithmic impact assessment and what impact-mitigating steps need to be taken depending on the AIA's outcome. Most of the AIA's questions contribute to a score corresponding to a series of risk levels in the Directive. Each risk level comes with a specific set of requirements, such as whether the ADM system can render decisions without a human-in-the-loop and whether (or to what extent) users of an ADM system must be informed of that fact.

This system has not proven particularly effective at regulating government use of AI. While the Directive is a mandatory policy instrument, it also features no meaningful enforcement mechanism. As a result, there have already been several high-profile instances where systems that should have required AIAs were developed and released without one (Cardoso and Curry 2021; Keung 2023). Moreover, even when an AIA is completed, there is ample reason to doubt whether a self-administered questionnaire can be a meaningful check on the development of complex

algorithmic systems. Tellingly, the TBS AIA only faintly resembles other proposals for how to conduct algorithmic impact assessments.

As you may notice, however, these problems are largely a product of how the Government of Canada uses the TBS AIA. The tool being self-administered and there being no enforcement mechanism are both facts stemming from the *Directive on Automated Decision-Making*. These problems are not inherent to the tool itself.

The Directive is a reasonably standard policy instrument. It is only one of sixty-odd directives in the Treasury Board's policy suite. But the TBS AIA is much more unique. Not only was the TBS AIA one of the first attempts at putting responsible AI into practice, but the tool also brings together open-source software and public policy in a way that makes it all the more remarkable.

Research-Creation Overview
With that in mind, this project has served as a way to explore the potential of the TBS AIA as a standalone governance tool. In particular, I have sought to explore two theories as to how this tool might prove useful in the broader context of AI governance:

1. That by hacking the TBS AIA—working *with* the tool—it could be re-used, re-conceptualized, or re-purposed to empower communities and help broaden the definition of what constitutes an "impact of AI."

2. That in the process of working with the TBS AIA—working *through* it—the materiality of the tool might ground otherwise abstract discussions about "artificial intelligence" and facilitate participation in AI governance.

The primary venue for my exploration of the tool was a series of "AIA hackathon" workshops, which I discuss in detail in section 4. Building on models of civic hacking and critical making (discussed in section 2), these workshops served as spaces where participants could work directly with the AIA tool and modify it to support its use in contexts other than government. The collaborative process of envisioning and building alternative configurations of the tool also offered opportunities to discuss practical issues in AI governance. As such, the workshops functioned as a means of assessing both of my theories about the TBS AIA's potential.

The second output of this project is aia.guide—a website that acts as a resource for using and understanding the TBS AIA. The guide—which I discuss in greater detail in section 3—is a collation of my background research for this project, presenting it in accessible language with an explicit focus on empowering users to think about how they could make the AIA their own.

I initially developed the guide as a companion to the hackathon workshops; it was a means of overcoming the Treasury Board's scattered and arcane documentation for the Directive and the AIA tool. But the guide also became more than the sum of its parts, and the process of writing aia.guide became an essential method for thinking through the archive of materials I had collected on the history, inner workings, and use of the TBS AIA. As Kim Sawchuk and Owen Chapman so aptly describe in their seminal work on research-creation, "Knowledge is not separate from the practice of inscription; indeed, it is through the ways we iterate our projects back to ourselves that we come to know" (Chapman and Sawchuk 2012, 18). Synthesizing my archive of research materials, and especially writing in the descriptive mode of a guide, put the ambiguities of the TBS AIA and the *Directive on Automated Decision-Making* into stark relief. In this way, aia.guide evolved from being a straightforward support document for my "AIA hackathon" workshops into a distinct piece of my research-creation practice exploring the TBS AIA.

These two outputs build on three key features of the TBS AIA which—I believe—make it worthy of further consideration and experimentation. They are:

1. its "thing-ness";
2. its open-source foundation;
3. and its origins in government.

The "thing-ness" of the TBS's AIA comes in two primary forms: the AIA tool— hosted by the Government of Canada and accessible by anybody with an internet connection—and the PDFs the tool outputs to document responses. These "things" are media objects that can help make explicit otherwise abstract discussions of algorithmic media. This may be especially important when it comes to improving participation in and forming publics around the governance of algorithmic

and AI-based systems.[2] Fenwick McKelvey (2014), drawing on Walter Lippmann's model of democratic participation, highlights the importance of mediators that can make the processes of algorithmic media visible. I believe that the TBS AIA may be able to function as one such mediator, offering the possibility of forming publics around AI-based systems.

The Treasury Board's AIA platform is also open-source—released under the highly permissive MIT License—and its source code is hosted in a public GitHub repository. As such, while all instantiations of the TBS AIA share this "thing-ness," the TBS AIA is not a singular "thing." The fact that the Treasury Board made both the AIA tool and its questionnaire open-source not only allows for its modification and repurposing in different contexts, but also encourages it. This feature of the AIA creates an opening, and my work in this project has been guided by an interest in exploring the potential created by the tool being open-source.

Finally, the potential of the tool being open-source is reinforced by the fact that the tool is both developed and used by the Government of Canada. Being open-source allows for the tool to be repurposed, but this does not guarantee that repurposed versions would be taken seriously. The Government's use of the tool, however, confers a certain legitimacy to this model of algorithmic impact assessment as an approach to AI governance.

Ultimately, this project is engaged with ongoing debates over the governance and regulation of AI. By deconstructing and disassembling TBS AIA, both of these outputs serve as methods for understanding and critically engaging a tool that lies at the heart of Canada's approach to regulating algorithmic systems. But both the AIA hackathons and aia.guide also offer a means of reassembling those pieces. In doing so, this project ultimately offers an approach to developing new participatory methods for AI governance and regulation.

---

[2] I have found John Durham Peters' concept of "logistical media" particularly useful for considering the TBS AIA as a media object. "Logistical media," Peters writes, "arrange people and property into time and space" (2013, 41). This definition is broadly consistent with approaches to media from scholars such as Marshall McLuhan, Harold Innis, and Sarah Sharma, who emphasize media's organizational capacity. As Sharma writes, a media's content "is secondary to the spatial and temporal effects that are produced at the level of culture" (2008, 458)

Section 1: Literature Review and Theoretical Perspective

Situating the TBS AIA

The Government of Canada may have been the first to implement an algorithmic impact assessment, but they were not the first to propose one. In the early stages of the tool's development, the TBS AIA was not ever referred to as an AIA; that language came later in the development process, putting the TBS's work more directly in conversation with proposals from academia and civil society. The two most notable proposals to this effect were from the AI Now Institute in the US and Nesta in the UK, both of whom released their proposals in February 2018—part-way through the TBS AIA's development (Karlin 2018).

The Treasury Board's questionnaire-based system for algorithmic impact assessment is quite different from other approaches. The model proposed by the AI Now Institute (AI Now Institute 2018; Reisman et al. 2018) places a much greater emphasis on informing the public and communities about ADM systems that may impact them. The proposal's authors underscore that algorithmic impact assessment's core purpose is to engage those affected by ADMs. To that end, they do not set out specific assessment questions. Instead, they argue that questions should be domain-specific and developed by particular departments. In the AIA the TBS developed, every department and agency responds to a standard set of questions.

Nesta's proposal—outlined in a blog post by Eddie Copeland titled "10 Principles for Public Sector Use of Algorithmic Decision Making" (Copeland 2018)—more closely resembles the model that the Treasury Board eventually developed. The Nesta proposal never actually uses the term "algorithmic impact assessment." However, principles like using a risk score to classify ADM systems and mandatory disclosure statements when users encounter ADM systems were directly taken up by the Treasury Board in the *Directive on Automated Decision-Making*.

The parts of Nesta's proposal that would require more government resources—either by way of additional development time or by requiring more funding—were not taken up by the TBS. Some aspects of what Nesta proposed, like the suggestion that governments develop an insurance scheme to compensate those unjustly affected by the ADM system, were simply outside the scope

of the TBS's policy mandate; other parts of their proposal were adopted by the Treasury Board but in extremely limited ways. For example, Nesta proposed that a sandboxed version of the algorithm should be released for testing by external auditors. The TBS included a peer-review requirement in the Directive, which is arguably a partial implementation of this principle. But it is implemented in such a way as to remove the open and public aspects, which are core to the principle in Copeland's proposal.

If we return to how responsible AI principles should be implemented, neither the Nesta nor AI Now's proposals directly answer that question. They do begin approaching an answer; unlike the Montreal Declaration, Nesta and AI Now considered their proposals' real-world implications and political feasibility. But, as I have noted, what makes the TBS AIA so interesting is that it is one of the first fully realized attempts at answering *how* responsible AI could be operationalized in a public sector context. This is an impressive feat, but it came at a substantial cost.

Developing an AIA that (1) fit within the Treasury Board's preexisting mandate and (2) that the public service could easily adopt became *the* central focus of the development process. As Michael Karlin put it when we spoke in 2022, "We wanted a lightweight algorithmic impact assessment that would take people no more than half an hour to do. So we wouldn't interrupt their development flow too much, but still provide transparency" (Karlin 2022). The prioritization of creating something feasible and minimally disruptive, however, is also what led to many of the tool's greatest weaknesses.

When comparing the Treasury Board's AI governance framework with the AI Now Institute and Nesta proposals, the major omissions boil down to consultation and substantive public input. The TBS did not adopt the core elements of these proposals—direct engagement with affected people and communities—which would have introduced substantially more friction into the impact assessment process. These elements are undoubtedly messy, time-consuming, and more challenging to implement. But they are also fundamentally important to most theorizations of how algorithmic impact assessment processes function.

Indeed, I would argue that many of the TBS AIA's problems stem from this deep-rooted bias to action. The compromises that make the AIA easily digestible by its users within the public service have also been the primary sites of critique. As Moss et al. argue in their report on algorithmic impact assessment for the public interest, one of the core issues with the TBS AIA is that it fails to establish a meaningful "forum for accountability" (Moss et al. 2021, 32). In the context of impact assessments more generally, a forum is a site that can "allocate responsibility for potential consequences of such systems and demand changes in their design, deployment, and operation" (Moss et al. 2021, 15). For a forum to provide a meaningful accountability mechanism, it needs to be able to require substantive changes to the system being assessed. But there is an inherent tension between creating a meaningful forum and the desire to create a lightweight tool that can be completed in half an hour.

Theoretical Perspectives: Policy and Media
As we have seen, the TBS AIA does not fit neatly into many categories. It is called an AIA, but it does not resemble other AIAs. It is a Government of Canada policy instrument, but it is also a piece of open-source software. This fact, in particular, makes it challenging to theorize the TBS AIA, in large part because the TBS AIA is an oddly recursive object to study. On the one hand, it is a tool designed to govern algorithmic systems. On the other hand, it is itself a (simple) algorithmic system; the AIA tool takes responses to a questionnaire and uses a set of procedures to output an impact score. The TBS AIA is an algorithm that governs algorithms that govern. This fact places the TBS AIA at the intersection of two literatures: one on algorithmic regulation and the other on algorithmic governance.

In this case, "algorithmic regulation" broadly refers to the regulation or governance *of* algorithms. In response to the increased prevalence of algorithms in everyday life and concerns over algorithmic power, the question of how best to regulate and reign in that power has increasingly become a matter of concern. As such, algorithmic regulation largely focuses on how algorithmic systems should be regulated and who should regulate them (Saurwein, Just, and Latzer 2015).

"Algorithm" is, of course, an extremely broad term. There are enough resemblances between different types of algorithmic systems for the catchall term "algorithmic regulation" to be meaningful. However, concrete recommendations and frameworks are usually specific to particular types or uses of algorithms. The concept of an "algorithmic impact assessment" emerged as one approach to regulating algorithms that either supplement or supplant decisions that would otherwise have been human-made (Moss et al. 2021). As such, the Treasury Board's *Directive on Automated Decision-Making*—which uses, but is ultimately distinct from, the TBS AIA—is a clear example of algorithmic regulation.

Whereas algorithmic regulation focuses on the governance *of* algorithms, algorithmic governance focuses on governance *by* algorithms. Here, I am largely drawing on the understanding of algorithmic governance put forward by Christian Katzenbach and Lena Ulbricht. They define algorithmic governance as "a form of social ordering that relies on coordination between actors, is based on rules and incorporates particularly complex computer-based epistemic procedures" (Katzenbach and Ulbricht 2019, 2).

Versions of this idea, of course, far predate Katzenbach and Ulbricht. In his book *Code and Other Laws of Cyberspace*, Lawrence Lessig (2002) famously argued that "code is law." Lessig focused on how code acted in "cyberspace" as a governing structure—much as a constitution might elsewhere. Lessig's argument is relatively narrow in the current context; I do not want to suggest that algorithms and code are fully synonymous, or that algorithmic governance necessarily involves code. But Lessig's argument has undoubtedly been influential in structuring the body of literature on algorithmic governance. It helped establish the concept that code and code-like objects like algorithms could be regulatory instruments rather than simply being regulated (Mackenzie 2006).

Both these concepts—algorithmic regulation and algorithmic governance—are fundamental to understanding the TBS AIA. Again, the TBS AIA is an algorithm that governs algorithms that govern. Governance *of* and governance *by* algorithms are both germane to the discussion of this tool. But most of the existing literature on the TBS AIA considers it only in relation to the

Directive on Automated Decision-Making and, by extension, it is only seriously considered through the lens of algorithmic regulation. Teresa Scassa, for example, has written extensively on the Directive and the TBS AIA. Scassa's work on the Directive's history and relationship to administrative law is particularly relevant, as she draws attention to ways in which the Directive's grounding in long-standing administrative law principles creates blind spots in assessing harms caused by algorithmic systems (Scassa 2021). But here governance *by* algorithms is only considered as the object of the TBS AIA, in terms of the algorithmic systems it regulates. Little attention has been paid, by Scassa or anybody else, to the TBS AIA tool as a form of algorithmic governance itself.

Moreover, little attention has been given to governance by algorithm in policy studies. This gap in the literature reflects a general emphasis in policy studies and related disciplines on the textual and discursive elements of policy (Fairclough 2013). What is significant about algorithmic governance, however, is the introduction of non-discursive elements—algorithmic systems—into policy itself. Some more recent work in critical policy studies has begun to address this issue. Regine Paul, for example, has suggested that critical policy scholars ought to draw on science and technology studies and theories of sociotechnical systems to understand, critique, and challenge the role of AI in public policy. Paul writes, "The methodological implication for critical analyses of AI-policy-interactions is that both technical and human affordances, restraints, and roles in policy practice—and dominant interpretations thereof—must enter the equation" (Paul 2022, 502). Paul's proposed approach is helpful for developing critiques of particular instances of AI or algorithmic systems in policy. However, it fundamentally aims to integrate AI or algorithmic systems into a discursive approach that cements it in a singular context.

As I have noted, however, one of the key features of the TBS AIA is that it is open-source. By definition, the AIA tool being open-source means that it can exist in different contexts and as a part of different sociotechnical systems. And unlike text-based policy instruments, it is not wholly dependent on human interpretation to have effects when translated to different contexts. The TBS AIA is both a policy and a media object.

As such, I have found approaches that understand policy as assemblage to be better suited to understanding the TBS AIA. As the policy scholars John Clarke, Dave Bainton, Noémi Ledvai, and Paul Stubbs argue in their book *Making Policy Move*, an assemblage approach to policy emphasizes that *any* policy comprises heterogeneous elements. These elements can be brought together in more fixed configurations, but any particular assemblage is always contingent (Clarke et al. 2015). Tess Lea makes a similar argument in *Wild Policy*, where she writes that "Policy, whether understood as narrative, artifact, embodiment, surround sound, or state effect, is a mobile assemblage, composed of different forces, materials, and actors, wending through time and space" (Lea 2020, 26).

Approaching policy as an assemblage provides a framework for considering a policy's shifting meanings and non-discursive elements without artificially separating those elements from the sociotechnical systems they are embroiled in at any given time. In this case, it provides an opportunity to isolate the media object—the TBS AIA tool—and consider how it might function in different contexts as part of different policy assemblages. But this approach would also suggest that this process of isolation and translation might provide insights into the particular assemblage that is the Government of Canada's regulation of automated decision-making systems.

Section 2: Research-Creation Methods

My goal with this project was to explore the potential of the TBS AIA tool as an object that might help facilitate broader, more substantive participation in AI governance. As I have noted, there are two ways in which I considered this to be possible: one by working *with* the tool, and the other by working *through* it. These two theories of how the TBS AIA could be mobilized were, in turn, informed by two different approaches to what I would broadly refer to as participatory exploration: civic hacking, and critical making.

Civic Hacking

Policy and hacking can be understood together in two different ways, depending on how you understand hacking. The first way is what we can think of as the conceptual approach. This approach is perhaps best articulated by Sophia Maalsen, who understands hacking as simply "a process of

problem-solving which celebrates experimentation, creativity and openness" (Maalsen 2021, 1). Maalsen argues that we can understand the hack as a state of exception which acts as a means of addressing the rigidities of a system. As such, while Maalsen acknowledges the concept's roots in computing, she argues that this definition of hacking can and is used conceptually—including in analogue contexts.

Within the context of urban studies, Maalsen identifies this ethos in attempts at "hacking policy" through small-scale, iterative policy projects, as well as through "hacking work" and "hacking provision" in the development of co-working spaces and gig economy platforms, respectively. In these cases, hacking is applied as a framework or an ethos that can be applied to different kinds of problems.

The second relevant way to understand hacking relates more directly to a specific set of computer-based practices. In a policy context, this second approach to hacking—what we might call the materialist approach—is explored by Alison Powell in her work on public interest hacking. Drawing on examples such as the open hardware license (OHL) developed at CERN, and the Public Lab's model of participatory collection of scientific data, Powell draws attention to the history of hacking as a model for participatory governance (Powell 2016). This conceptualization of policy or civic hacking is directly rooted in the free and open-source movement. As such, it provides one model for how policymaking, hacking, and open-source software—like the TBS AIA—might interact. This is why the AIA workshops I led for this project were framed as "hackathons." In doing so, I sought to engage the history of open-source software in participatory governance and to draw attention to the possibility of approaching the AIA in this way.

But a materialist approach to hacking also largely focuses on practices within computing, science, and engineering, which limits the scope of who can participate and what can be achieved. Powell identifies this as a crucial weakness in her case studies of civic hacking projects. While the projects that Powell studied generally espoused the idea that their DIY and hacking practices would undermine existing structures of knowledge and authority, they also relied on and replicated those

very same structures. Participation in this model of participatory governance still required the ability to write software or to work with complex scientific data (Powell 2016).

Some of the tensions that Powell identifies can be resolved by adopting a conceptual approach to hacking. If, following Maalsen (2021), we understand hacking as an orientation rather than a specific set of practices, there is substantially more room for people with different expertise to meaningfully participate in processes of iterative experimentation and problem-solving. As such, the actual design of the "hackathon" workshops explicitly emphasized the possibility of engagements and "hacks" of the TBS AIA that were not rooted in code. In this way, the workshops engage a broader conceptualization of "hacking," one which is explicitly oriented towards engaging with expertise beyond software development.

Hacking is still a contested concept at a conceptual level, however. In their critique of hackathons in *Design Justice*, Sasha Costanza-Chock writes that "Like many tech spaces, they [hackathons] tend to be dominated by white, straight, able-bodied, cisgender males, masculinist assumptions about technical competence, universalizing discourse, and solutionism. They are too frequently exclusive, alienating to those who don't already feel comfortable in normative tech culture, and dismissive of the difference, experiential knowledge, and domain expertise of marginalized people" (Costanza-Chock 2020, 24). Parts of Costanza-Chock's critique echo Powell's arguments about the limits of public interest hacking. But Costanza-Chock also goes further, questioning the solutionist ethos ingrained in many hacking practices. They argue that hackathons generally prioritize creating new things rather than maintaining what already exists. Moreover, the need to produce something tangible quickly often creates a dynamic that flattens complex social issues into simple problems that can be addressed with technological or technocratic fixes (Costanza-Chock 2020). The issues that Costanza-Chock raises do not only apply to materialist or computer-based hacking practices; the limits of a solutionist approach apply equally to hacking as a concept.

Hacking as Critical Making

How, then, might we imagine a non-solutionist hackathon in the context of the TBS AIA? While it never refers to "hacking" as such, I have found that Matt Ratto's practice of "critical making" offers a particularly compelling approach.

As Ratto describes it, critical making signals "a desire to theoretically and pragmatically connect two modes of engagement with the world that are often held separate—critical thinking, typically understood as conceptually and linguistically based, and physical 'making,' goal-based material work" (Ratto 2011, 253). Instead of eschewing the element of production that is central to hackathons, critical making offers an alternative framework for understanding the purpose of that production. The focus shifts to the process rather than the outcome. As Ratto puts it, "Critical making emphasizes the shared acts of making rather than the evocative object" (Ratto 2011, 253). In the context of this project, Ratto's work draws attention to the function that working with the TBS AIA might have beyond the outputs that participants would produce.

In practical terms, Ratto breaks down critical making into three stages. The first is a review of the relevant literature and compilation of useful concepts and theories; these ideas are then mapped metaphorically onto material prototypes. In the second stage, groups are asked to build their own prototypes—both as a means of skills development but also as a process of conceptual exploration. Finally, the third stage begins a process of reflection, which loops back into theory. As Ratto notes, however, these stages are more separate in analysis than in practice (Ratto 2011). In practice, the line between prototyping and retheorization is particularly blurry. This is how we can consider working *through* the TBS AIA; as we will see in the outcomes of the AIA hackathon workshops, the processes of prototyping and reflection are intertwined.

One key difference between a traditional hackathon and a critical making workshop is that the objects Ratto centers in his prototyping processes are specific to those workshops; they are metaphorical representations designed to materially explore a set of concepts. In a hackathon— and this project—the object at the center of the prototyping process is readymade. This adds a layer of complexity to critical making, as the relationship between theory and practice is not as direct.

One way to resolve this tension would be to change objects altogether and instead consider the question of participatory AI governance through a purpose-built prototyping process. But doing so would preclude the exploration of how this *specific* AIA tool might be able to mediate or facilitate discussions about issues in AI governance.

To understand how the TBS AIA might function in the critical making process as a ready-made (and expressly political) object, I have found Noortje Marres' work on issue publics and the role of objects in politics to be a particularly useful complement to Ratto's method. Marres draws on the work of John Dewey, who holds that publics in a democratic society are not extant or somehow natural, but instead that they come into being around issues (Marres 2005). Marres has since expanded on this to assert an explicitly materialist understanding of publics and publicity—the latter being "the process in which publics come to 'recognize themselves' and somehow acquire the capacity to act" (Marres 2010, 198). She particularly emphasizes the role of objects and material practices in the formation of issue publics, arguing for an object-oriented understanding of publicity. In this way, her approach stands in contrast to other post-Foucauldian work, which emphasizes the objects' hidden or clandestine effects (Marres 2010).

Marres' approach highlights how the TBS AIA's "thing-ness" might be especially relevant. The AIA mediates between concepts in responsible AI and the practice of AI governance. Through Marres, we can consider this not only as a manifestation or realization of those concepts but also as an object around which publics could form on issues in AI. While the workshops were unlikely to somehow "spark a public into being," they might still provide an opportunity to test the concept's applicability on a smaller scale.

This is particularly important because of the medium specificity that Marres attributes to political objects. Through Dewey, Marres does not only assert that issues instantiate publics but also that the specific contours of the issue—its materiality—will impact the 'shape' of the public (Marres 2010). It is, therefore, crucial in this case to assess not only *if* the TBS AIA can be used to facilitate discussions about AI governance but also *how* it does so.
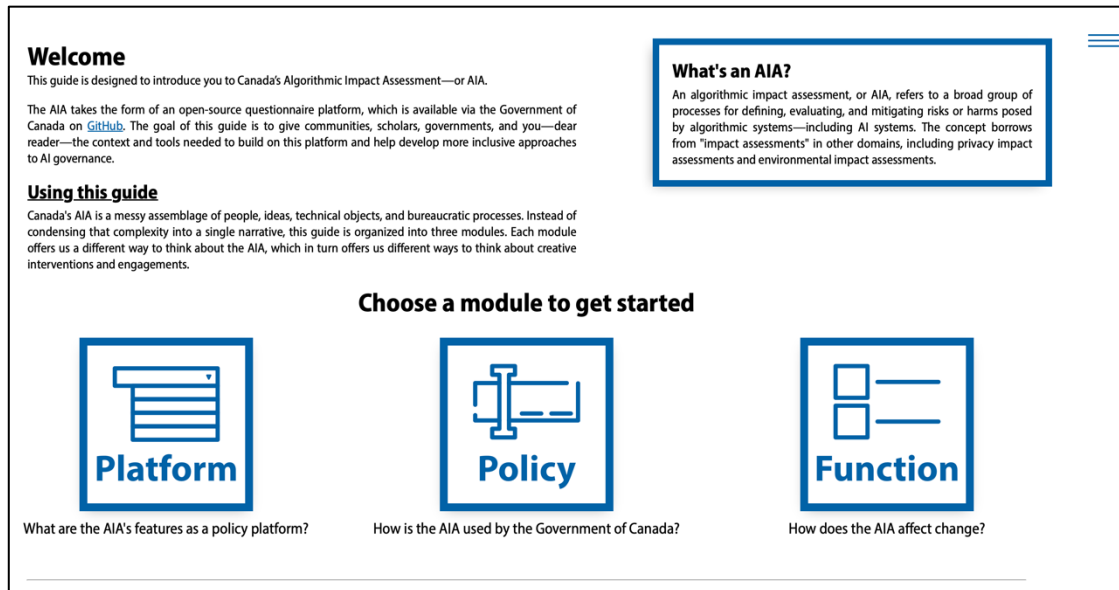
Figure 2: Screenshot of the aia.guide welcome page.

Section 3: Outputs – aia.guide

aia.guide is the first of this project's two "creation" pieces. If we consider Ratto's three stages of critical making, the workshops (which I will—finally—discuss in section 4) act as the second and third stages—prototyping and reflection. These stages are preceded, however, by a process of research, literature review, and metaphorical mapping of concepts onto the eventual prototype (Ratto 2011). aia.guide serves these functions.

As I noted in the introduction, the guide is a website that acts as a resource for using and understanding the TBS AIA. It collates my research for this project and presents it in accessible language with an explicit focus on empowering users to think about how they could make the AIA their own. The process of theory mapping is less metaphorical in aia.guide than what Ratto envisions. As a guide to an existing object, it is necessarily somewhat descriptive. But it is by no means an uncritical reiteration of information, either. My primary intention was for the guide to establish a framework through which workshop participants could think about reinterpreting and prototyping the new versions of the TBS AIA. While the underlying object was indeed readymade, I recognized that if my eventual goal was to have workshop participants build something new from the pieces of this policy media, it could not feel natural in its existing state. It needed to feel strange.

The guide is divided into three modules: Platform, Policy, and Functions. The name of each module is intentionally ambiguous; they are meant to invite questions and encourage exploration rather than provide clear demarcations. Instead of reflecting existing divisions from the policy itself, each module offers a different approach to thinking about the TBS AIA system as a whole. The site has a table of contents in its menu for accessing specific topics, but the primary interface for navigating the site is a set of three large buttons—one for each module—placed after a brief introduction. To give users a jumping-off point, each module's navigation button is captioned with a guiding question that provides insight into the questions the module aims to answer.

Structuring the guide this way allowed the modules to stand alone. A more linear approach, like one structured around the individual pieces of this policy system, would also need to be read linearly. Organizing the guide thematically made the sections into modular pieces that could act as a reference during the workshop. As a result, the guide does not need to be read in its entirety or any particular order.

Platform Module
What are the AIA's features as a policy platform?
The Platform module of aia.guide is primarily an attempt to defamiliarize the TBS AIA. Drawing on the formalist literary theory of Viktor Shklovsky, the concept of defamiliarization originally referred to the ways in which artistic and literary works can make everyday objects or experiences strange. Shklovsky argued that this process of estrangement could provoke a renewed appreciation of the everyday within the audience (Pangborn 2010). Since Shklovsky initially identified defamiliarization as a literary technique, the concept has been rearticulated as a critical methodology. As Genevieve Bell, Mark Blythe, and Phoebe Sengers argue in their work on defamiliarizing the design of domestic technologies, defamiliarization as a method "calls into question our usual interpretations of everyday objects" (Bell, Blythe, and Sengers 2005, 154)

In that spirit, the Platform module lays out the pieces underlying the software that the Treasury Board put together to host the AIA. It includes descriptions of what languages and frameworks the AIA tool is built on, guide to the aia-eia-js GitHub repository and how to (attempt to)

load a local copy of the AIA, and a walkthrough (Light, Burgess, and Duguay 2018) of the features and affordances of the questionnaire system. Many of these "features" would be considered mundane to anybody who has spent a reasonable amount of time on the internet. (Or just anybody who has spent a reasonable amount of time filling out forms, for that matter.) I think it is fair to presume that most users of aia.guide are familiar with the difference between a checkbox and a radio button, for example. Nevertheless, the guide includes detailed descriptions of these elements and their contribution to the Treasury Board's AIA questionnaire system.

The ubiquity of these elements as standard pieces of the web is precisely the reason for this module. In the TBS AIA, each input type represents a design choice (and a policy choice) about what aspects of an algorithmic system can be evaluated. The most notable distinction to this effect is between scorable and non-scoreable inputs. Checkboxes and radio buttons might have different affordances, but ultimately they both produce quantifiable answers. Responses in free-text fields, on the other hand, cannot be straightforwardly quantified. As such, they *cannot* contribute to the score produced by the AIA. This score determines the system's risk level and corresponding requirements under the Directive. It is fundamental to how the AIA is meant to enforce certain practices. Free-text fields—which have proliferated in recent updates to the Treasury Board's AIA questionnaire—do not contribute to this system.

Although they represent significant design choices in the AIA questionnaire's development, the form elements' mundanity makes them fade into the background. One's focus tends to be on the questions being asked, not the inputs under them. Removing the question and presenting the inputs on their merits breaks a habitual and familiar usage pattern. In doing so, it draws attention to the inputs' underlying significance and the general significance of software design to the AIA's role in a policy assemblage.

Policy Module
How is the AIA used by the Government of Canada?
The guide's Platform module sought to distinguish the TBS AIA's approach to algorithmic impact assessment—the questionnaire and scoring model—from the policy context the Treasury Board

deployed it within. The module focuses on the software object and its effects as a mediator. The specifics of *how* the Treasury Board and the Government of Canada use that platform are left to the Policy module.

Of aia.guide's three modules, the policy module is the most straightforwardly factual. It would be challenging to call this a "guide" to the TBS AIA without providing the context it was developed within, or how it is officially used. In this module, I largely attended to the role of the Treasury Board, the content of the TBS AIA questionnaire, and the *Directive on Automated Decision-Making*.

The challenge that the Policy module posed was not what content to include but rather how to present it. Certain subsections—particularly those on the AIA questionnaire and the AIA's scoring system—follow the structure used by the Treasury Board in the Directive and the descriptive landing page for the AIA tool. The main purpose of their inclusion in aia.guide was to bring all this information into one place. But describing the *processes* established by the Directive was less straightforward. All the necessary information is in the Directive, but it is both verbose and in extremely formal language. This may make sense in the context of an official policy document—especially when the expected audience is largely made up of public servants—but it does not make sense for a guide of this sort. I needed something shorter and more accessible.

Summarizing the Directive's processes for aia.guide was also a helpful research exercise. I consider this one of the instances where the process of creating the guide became a part of the research process in and of itself. What became clear is that the Directive only actually outlines one procedure: policymakers are expected to complete an AIA questionnaire, then determine the system's impact level based on the AIA's score, and finally implement any necessary changes to the system based on the requirements corresponding to that impact level. This basic procedure is repeated twice during an automated decision-making system's development: once when the system is being designed and again just before it is implemented. This is, in effect, a small-scale example of the policy translation process (Clarke et al. 2015). And, in line with Clarke et al.'s assemblage approach to policy, the meanings and performances of the same procedure change in each context.

Once again, this information is all available in the *Directive on Automated Decision-Making*. But it is the process of synthesizing that information that makes the underlying structure apparent.

Function Module
How does the AIA affect change?

To explain the Function module of aia.guide, we need to return to the three aspects of the TBS AIA which, to me, made it worthy of consideration and experimentation as an AI governance tool: its "thing-ness," its open-source nature, and the legitimacy conferred to it as a tool originating in government. In the context of aia.guide, the TBS AIA's "thing-ness" is primarily addressed in the Platform module, where I focused on the materiality of TBS AIA as a policy media. The Policy module focuses on the tool's source of legitimacy—its uses within the Government of Canada. aia.guide's function module was my attempt at grappling with the TBS AIA's open-source nature.

When I say "open-source nature," I mean that quite intentionally. The fact that the tool is open-source—the public availability of its source code—is not the focus of the Function module. Instead, the Function module attends to the *conceptual* implications of a policy being open-source.

In this context, "function" refers to the *ways* that the TBS AIA can be used to affect change. In the guide, I liken it to a hypothesis—a theory of cause and effect. "Function" allows for an articulation of how we expect the policy to work, and it provides a falsifiable measure of whether it is living up to its intended purpose.

The utility of this approach to the TBS AIA only became clear to me after spending considerable time on this project trying to understand *how* this policy worked. This was hard precisely because the TBS AIA is riddled with contradictions: it is an enforcement tool without enforcement power; it purports to offer transparency through binary answers to vague questions; and it tries to provide public accountability, but only after the fact. These contradictions were not resolving themselves, and I realized that I needed a framework for understanding how they could all coexist.

Conceptualizing the TBS AIA as having different "functions" provided just such a framework. As a media object, it was important to interrogate how the TBS AIA arranges people and things in time and space (Peters 2013, 41). But as part of a policy assemblage, it was equally

important to attend to the fact that these arrangements, and the meanings they produce, are not fixed (Clarke et al. 2015). Articulating how the TBS AIA works as a hypothesis—expressed in the guide as "The AIA will X by Y"—reveals the different (and sometimes contradictory) theories of change underlying the tool's use in AI governance. It also requires an explicit articulation of those mechanisms and therefore gives a specific measure for testing how effective a given implementation of the tool is.

In addition to providing a framework for critique, my goal in framing the TBS AIA in terms of its functions was to open space for different uses of the tool. Recognizing that the tool can work in several ways—and that it already does so within the Government of Canada—opens conceptual space for thinking about alternative ways that it could be used. This is why, to me, the concept of "function" is an attempt to grapple with the TBS AIA's open-source *nature*. At a technical level, the AIA tool being open-source means that it can be reused in different contexts. But the concept of function draws attention to the different meanings and ways of interpolating the world that it can take on in these different contexts.

With this in mind, aia.guide ends not with a statement but with a question: "What functions can you imagine for Canada's AIA?" That question is accompanied by a fill-in-the-blanks version of the function hypothesis, prompting the user to consider how the AIA might be used to effect change in their context. This is a conclusion for the guide, but it is also an opening for future interventions—including the AIA hackathon workshops—which I will now discuss.

Section 4: Outputs – AIA Hackathons

At a surface level, the goal of the hackathon workshops was to focus on what I have referred to as the TBS AIA's "thing-ness." What distinguishes the Treasury Board's approach to algorithmic impact assessment from other proposals is the materiality of its method. The language of a "hackathon" (recalling Powell's (2016) model of civic hacking) elicits a process of engagement and experimentation with that material object.

By allowing participants to work directly with the AIA tool, my goal was to test the two theories I had developed regarding the tool's potential in AI governance beyond its application in the Government of Canada:

1. That by hacking the TBS AIA—working *with* the tool—it could be re-used, re-conceptualized, or re-purposed to empower communities and help broaden the definition of what constitutes an "impact of AI."

2. That in the process of working with the TBS AIA—working *through* it—the materiality of the tool might ground otherwise abstract discussions about "artificial intelligence" and facilitate participation in AI governance.

I began developing the workshop while researching the quasi-open policymaking process that led to the creation of the TBS AIA. These consultations, and the fact that early versions of the AIA were developed in the open, are the defining features of the TBS AIA's history. But the narrative of public participation stands in stark contrast to the actual politics of expertise which played out during the period of open policy development.

During the Treasury Board's consultation process, draft documents were posted publicly on Google Docs, but members of the Treasury Board team only actively solicited contributions from people they perceived to be experts. The process was nominally open, but the Treasury Board's intention in doing so was only to attract "concentric circles of experts" (Karlin 2022).

Even if we set intention aside, the structure of the consultations also precluded more substantive public participation. The consultations lacked focus or guidance; they asked very generally about issues in AI. This open-endedness ultimately required that people have preexisting knowledge of the field to participate meaningfully.

It is important to note that broad participation is not in and of itself a democratic good. As Charles Thorpe and Jane Gregory argue in their work on public participation in science and technology policy in Britain, consultations and other forms of public engagement often use these venues as sites of cooptation and control of public opinion. In particular, consultations can shape the market for "emerging" technologies by priming participants as future consumers of those

technologies (Thorpe and Gregory 2010). Thorpe and Gregory's examples primarily focus on discourses around nano-technology, but many of the same patterns apply to public engagement on artificial intelligence.

I wondered if the materiality of the TBS AIA might create an opening for participation without substantial preexisting knowledge of AI. But more importantly—and in response to these issues raised by Thorpe and Gregory—I wondered if the tool might allow participants to bring expertise from *different* domains to AI governance; if the AIA tool brings the AI, participants might be able to map expertise and situated knowledges from other fields onto these problems. In this way, the materiality of the TBS AIA—its "thing-ness"—would be what gives it the potential to broaden the definition of what constitutes an "impact of AI." It allows people with a greater range of knowledges to contribute meaningfully to that discourse.
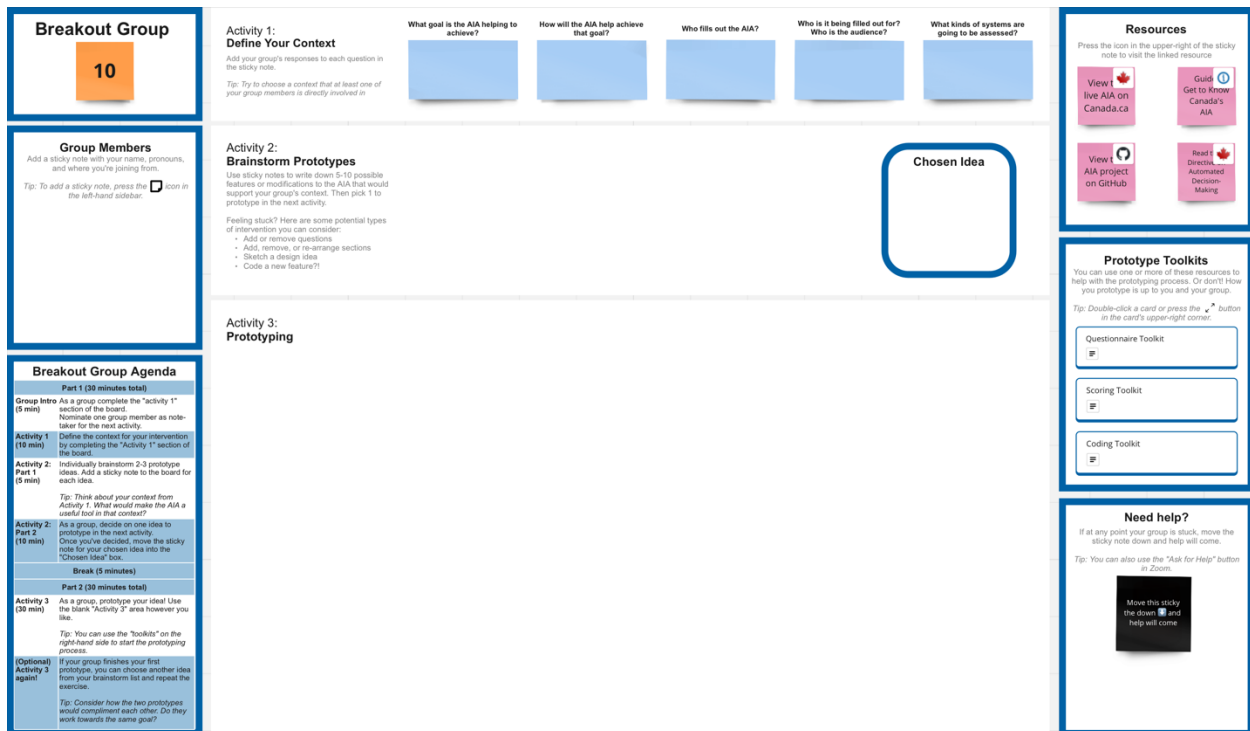

Figure 3: Screenshot of an unfilled board for the AIA hackathon workshops in Miro.

Workshop Design and Procedures

Ultimately, the goal I gave participants in the workshops was straightforward: prototype *some kind* of intervention or modification to the AIA. The workshops were framed as "hackathons," but I was clear in the description and my introduction that no technical knowledge or skills would be

required to participate. Recalling my two theories, my intention in framing the workshops this broadly was to create a space where participants with various skills and backgrounds could actively engage *with* and think *through* the TBS AIA.

I ran the workshop twice: once online and once in-person. But it bears noting that the constraints of running the workshop online proved to be the lowest common denominator. Therefore the overall design of the workshop was influenced by these constraints. In particular, the online workshop needed to be at most 90 minutes long, it needed to be able to accommodate an unknown number of participants, and those participants would be from unknown backgrounds. This meant that the design had to scale easily and that it could not assume that participants had prior expertise in AI. Without a cap on virtual session sizes, it was equally likely that thirty participants would attend, just as it was that three would attend.

Accordingly, the workshop procedure was as follows:

Step 1: AIA Walkthrough
In considering how to introduce the workshop and the TBS AIA to participants, I had to resolve two conflicting priorities. I wanted to give participants a baseline to work off in the hands-on portion of the workshop, which meant I needed to give them some context on the Treasury Board's implementation of the AIA platform. But I also

Figure 4: A screenshot of the IRCC AIA used as a demonstration in the workshops.

wanted to convey that the TBS AIA is a media object with which participants could (and should) interact. It was clear that beginning the workshop with a thirty-minute policy lecture would not have been wise.

With this in mind, the workshops began with a brief general introduction to the TBS AIA. But rather than describing the platform and the questionnaire, I led participants in a walkthrough of the TBS AIA platform using an actual algorithmic impact assessment completed by Immigration, Refugees and Citizenship Canada (IRCC) (2022). I gave participants the link to the Government of Canada's live AIA tool and the JSON file for the completed assessment. They were encouraged to explore the tool on their computers as I led a walkthrough of the platform and the questionnaire.

As I walked through the tool, I made a point of exhibiting how different aspects of the questionnaire and the platform's design affected the outcomes of the AIA. In cases where questions were arbitrary, I highlighted that fact. And in cases where IRCC's responses seem to underplay the impact of the system they were describing, I demonstrated how a different response could impact the risk score assigned to the system. This introduction was meant to ground the TBS AIA in the real world and demonstrate how the AIA platform mediates that world and constructs "AI impacts."

Step 2: Defining Context
Following the walkthrough, participants were divided into groups of 2-4 and introduced to a Miro board, which I designed to guide them through the prototyping stage of the workshop. Miro is a virtual whiteboard platform that allows users to collaborate in real-time via a variety of annotation tools. Users are represented by their mouse cursor position, allowing collaborators to see what another participant is working on. All participants were on the same Miro board, but each group had a dedicated space with a standardized template for completing the rest of the activity. The rest of the workshop was broken into three "activities," each with its own area in the centre column of the Miro board.

The first of these three activities was titled "Define Your Context." As a group, I prompted participants to define an alternative context and use case for the TBS AIA. Following the same logic as aia.guide's "Function," my intention was for participants to define a sort of hypothesis for how the TBS AIA might be used. Groups were prompted to choose a context that at least one group

member was personally familiar with and, for that given context, to answer the following questions:

1. What goal is the AIA helping to achieve?

2. How will the AIA help achieve that goal?

3. Who fills out the AIA?

4. Who is it being filled out for? Who is the audience?

5. What kinds of systems are going to be assessed?

Step 3: Brainstorming
The second activity was titled "Brainstorm Prototypes." Participants were prompted to spend 5 minutes individually brainstorming possible features or modifications to the TBS AIA that might support their group's proposed use or otherwise adapt it to their group's chosen context.

To facilitate the activity, I included a list of possible types of interventions:

- Add or remove questions

- Add, remove, or re-arrange sections

- Sketch a design idea

- Code a new feature?!

After five minutes of individual brainstorming, members regrouped to select a single intervention that would be the focus of the third activity: prototyping.

Step 4: Prototyping
Whereas the first two activities were more prescriptive, the "Prototype" section of the template was a blank slate. This reflects the conceptual approach to "hacking" I adopted for this project; no particular type of intervention was prioritized.

I did, however, provide participants with resources to facilitate different types of interventions. In addition to Miro's built-in illustration and design tools, I included three "prototype toolkits" in the template: a questionnaire toolkit, a scoring toolkit, and a coding toolkit.

In the questionnaire toolkit, I provided instructions for loading the Treasury Board's AIA questionnaire into the visual form builder for Survey.JS—the open-source survey building

software on which the TBS AIA platform is built. Groups could paste the included JSON file into the form builder and then freely add or remove questions, change input types, add descriptions, or otherwise modify the survey questionnaire.

The scoring toolkit included an editable spreadsheet in Google Sheets listing each question and the score the Treasury Board assigned. The scoring system is the TBS AIA's marquee feature, but it was custom-built for the AIA tool. As such, it was the one element of the core survey that could not be viewed
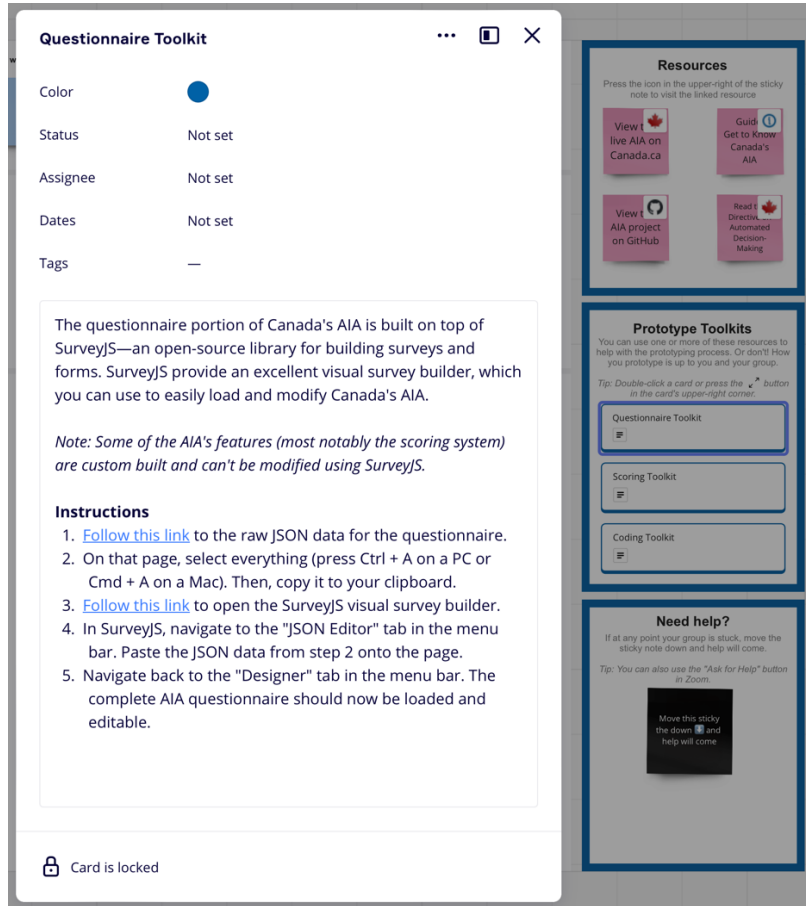

Figure 5: Screenshot of the "Questionnaire Toolkit" from the workshop Miro board.

or modified in Survey.JS. To give participants a way of interacting with the scores and the relative weights assigned to each question, I used a Python script written by the Treasury Board to export the maximum scores for each question to a spreadsheet.

Finally, the coding toolkit included instructions for forking the TBS AIA's GitHub repository and running a local version of the tool, which could be modified directly. In some ways, this was more of a nod to the TBS AIA being open-source than a serious expectation that groups would work with the tool at the code level. It was unlikely that groups could get a local copy of the tool up and running, become familiar with its structure, and make meaningful changes, all in the short time allotted to the workshops. But even if participants had the necessary expertise to run the tool locally, it simply would not have worked. While validating the toolkits before the workshops, I tested whether I could run a local copy myself. While I had been able to get it to work in the past,

I realized that one of the TBS AIA's dependencies had broken in a recent update. This meant the AIA tool could not be run locally until a developer outside the Government of Canada updated their software package. In practice, this was not a serious impediment to the workshops, but it did serve as a stark reminder of what is at stake when policy becomes software.

Section 5: Workshop Results and Discussion

Towards the end of each workshop, I reconvened all the groups to present and discuss each group's prototype. In the first workshop—which took place online at the Mozilla Festival (MozFest)—the discussion quickly turned to issues of domain specificity in algorithmic impact assessment. Two of the teams had considered how the AIA tool might be adapted to assess particular kinds of algorithmic systems. One group addressed uses in the banking sector, and the other discussed assessing the uses of AI in medical devices. They both described nuanced discussions about how the questionnaire might be adapted to these contexts, but neither had produced prototypes as such. It became clear that holding the workshop online, which meant having to form groups quickly and among strangers, made it challenging for some groups to complete the activity as intended.

Another team had more success completing the prototyping. Both group members worked in the non-profit sector, and they considered how this model of algorithmic impact assessment could be used in their field. Their prototyping focused on adding nuance and granularity to the questionnaire in ways that would make the output of the AIA tool clearer and more transparent. They systematically went through the TBS AIA questionnaire and considered how input types for different questions affected the utility of the outputted result. Ultimately, they argued that using a scaled system (1-5) rather than a binary one would provide more insight into a system for a human assessor. Recalling the TBS AIA's "Functions" from aia.guide, this approach focused on how to emphasize what I refer to in aia.guide as the AIA's accountability function; it calls attention to the ways that the AIA's outputs can be used by third parties to understand an algorithmic system and hold implementing organizations to account.

The second time I ran the workshop was at a Concordia summer school on "The Social Life of Artificial Intelligence." While the workshop procedure was substantively the same, the

venue change made for some significant differences. First, the summer school workshop was held in person. This meant that groups could work more collaboratively by sharing one computer. The workshop was also under less time pressure and was prefaced by a class discussion on the history and uses of algorithmic impact assessments beyond just the TBS AIA. Finally, I gave the second workshop after the summer school had been running for about a week. This meant that participants knew each other to some extent beforehand, which translated into easier collaboration and more meaningful prototyping toward the end of the workshop.

As with the MozFest workshop, groups in the summer school workshop largely focused on the AIA questionnaire and scoring system. Again, some groups focused on how this model of AIA might be used in different domains. One group focused on assessing AI systems used in mental health care contexts. Their prototype questionnaire emphasized evaluating how the system being assessed complied with existing best practices in the field in which the system was meant to be implemented. They also drew attention to whether users could opt-out of using a given system. Another group prototyped adding questionnaire modules on impacts in specific areas, including human rights, civil rights, and privacy. Interestingly, their proposed approach mirrored an early version of the TBS AIA. The Treasury Board abandoned this approach because it was considered too challenging to quantify (Karlin 2022).

The second workshop generally resulted in much more direct engagement with the AIA questionnaire. Most groups used the provided "Questionnaire Toolkit" to work with the questionnaire in Survey.JS, which, I think, was facilitated by the fact that the workshop was in-person. Survey.JS's platform has no built-in collaboration tools, so groups in the online MozFest workshop had to share a screen via Zoom to collaborate; in-person participants could simply huddle around one laptop to use the tool. This was also true for other forms of hands-on engagement with the TBS AIA. If a group working online had chosen to work with the code of the AIA platform, they would have faced the same issue. This fact was reflected in the workshop outcomes: groups in the in-person workshop were much more likely to produce meaningful prototypes by the end of the workshop.

Discussion of Workshop Results

While I would still argue that the TBS AIA tool *could* be re-used, re-conceptualized, or re-purposed in more interesting ways, I do not think that the workshops were ultimately able to test this theory. Upon reflection, this is primarily because of the limitations that hosting the first workshop online placed on my workshop design. To account for the fact that groups would be in separate breakout rooms and the potential for high attendance, I designed a workshop template that allowed participants to largely guide themselves through the hands-on portion of the workshop. Each group's Miro board included an overview of the three activities, additional resources, and a schedule for the whole hands-on portion of the workshop. But even so, it was evident in the results that some groups got lost in the early stages of the workshop—particularly when it came to defining a context for their intervention. Testing my first theory relied on the clarity of purpose that the context activity was supposed to provide. Without a defined idea of where the AIA tool might be repurposed or how it might function in that new context, it became challenging for groups to conceptualize and prototype meaningful interventions.

The other element that I think impeded the workshop's ability to test my first theory was the fact that I grounded the introduction in how the Government of Canada uses the TBS AIA. As discussed, I started each workshop with an interactive walkthrough of the AIA tool using a completed AIA from Immigration, Refugees and Citizenship Canada as an example. While the demonstration did serve the intended purpose of grounding the AIA tool, it also seemed to constrain the use cases and functions that participants envisioned for their prototypes.

This problem mirrors an issue that Matt Ratto considers in Critical Making. Ratto describes an early instance of a critical making workshop where he asked participants to build "bristle-bots"—small robots made from a toothbrush head and motor. The overarching goal of the workshop was to explore questions on the efficacy of distance learning. To this end, participants were provided with online resources to help them build their robots, and they could also collaborate with other participants in the room to get help. Ratto intended to have participants reflect on the different affordances of in-person and online learning as experienced through the workshop. But

he ultimately describes how participants found it challenging to map the activity from the workshop—building a bristlebot—onto the concepts they were ultimately asked to reflect on.

Ratto's experience with the bristlebot workshop demonstrates the need for a more direct connection between subject and object in critical making. But I would argue that my experience with the AIA hackathon workshops demonstrates how too direct a connection can also be detrimental. In this case, introducing the AIA tool by placing it within its existing context as an impact assessment tool used by the Government of Canada reinforced the connection between the tool and this specific use case. This made it challenging for participants to test my first theory, which relied on severing that connection.

The workshops were ultimately inconclusive as to what can be done *with* the TBS AIA. But I think the workshops provided much more evidence for what could be achieved by working *through* the TBS AIA. Going into the workshops, my theory was that, by grounding otherwise abstract discussions about AI, working with the TBS AIA tool might facilitate participation in an AI governance discourse. In both workshops, participants' direct engagement with the tool as an object facilitated nuanced discussions about how AI systems should be governed in different contexts. Admittedly, some of these discussions veered more into critique of the Treasury Board's AIA process than I would have hoped. (Again, I think this was rooted in how I demonstrated the tool using an example from the Government of Canada.) But what I want to draw attention to, and what—I think—is ultimately the most interesting outcome from the workshops, is the extent to which the specific format of the TBS AIA tool *as a survey platform* facilitated participants' engagement.

The TBS's decision to format their AIA as a survey platform has been critiqued, and for good reason. For instance, Moss et al. argue that, as a rule, questionnaire-based assessment methods fail to engage with the real-world effects of specific algorithmic systems (Moss et al. 2021). And even for the impacts that the Treasury Board's questionnaire does capture, the questions yield responses so broad as to provide no meaningful insight into—let alone oversight of—a given algorithmic system. These critiques speak to the inadequacy of the questionnaire format in the way

that the Treasury Board uses it. In its original context, it is a fundamentally ineffective governance tool.

And yet, precisely the same format can lead to engagement with AI governance if it is oriented differently. The features that make the TBS AIA an ineffective governance tool in its original context also enabled broader participation in discussions of AI governance during the workshops. As I have already noted, most groups focused their prototypes on modifying the contents of the questionnaire. Their specific modifications and reasoning varied, but a common thread emerged: participants were reflecting on what they themselves would want to know about an AI system being used in a given circumstance. They were contemplating what can broadly be understood as AI governance.

The workshops took the TBS AIA out of its context, changing the meanings and performances that the tool engenders (Clarke et al. 2015). Whereas the questionnaire model forecloses the meaning of AI impact in the Government of Canada's use case, reorienting that model opens the door to new understandings. Changing the mode of engagement to one of critique and questionnaire *design* inverts the AI-expert-oriented power dynamic of the consultations led by the Treasury Board in the AIA's development. Asking for answers requires preexisting expertise about AI, but asking for questions allows for meaningful contributions from a much broader array of knowledges.

Recalling Marres' assertion that the particular contours of issues (political objects) shape the publics that form around them (Marres 2010), it becomes clear that the general "thing-ness" of the TBS AIA is not where its potential lies. Instead, the questionnaire model, *in particular*, makes it useful as an object for grounding discussions of AI governance. The questionnaire format defines the TBS AIA's contours as a political object, and the orientation of different actors to the questionnaire ultimately defines its effects as both a policy and a media.

Section 6: Conclusion

This project has substantially changed since I began researching the TBS AIA in early 2021. For one thing, this was not going to be a research-creation project at all. My initial intention was to

write a decidedly un-creative thesis; I was going to discuss the history of the AIA tool's development, develop a theoretical framework for understanding it as both a policy and a media, and then consider how both of these aspects shape it as a governance tool. That version of this thesis would, I think, still have been valuable. It would have gone further toward developing the kinds of critiques that Moss et al. (2021) made of the TBS AIA, for example.

But the TBS AIA is a slippery object to study. Every time I felt I had a clear understanding of this strange policy object, a new fact or facet would emerge and upend everything. As I came to realize, developing a complete and accurate history of this tool is a near-impossible task. The quasi-open development policy process that produced the TBS AIA suggests that a tempting archive of the policy's development might exist. But in practice, the archive is scattered and incomplete; once public Google Docs have been made private, other files have been taken offline, and the Government of Canada's secretive internal culture puts key non-public documents out of reach.

As Tess Lea reminds us in *Wild Policy*, though, this teleological approach to policy misses the messy bigger picture (Lea 2020). Not only is writing a biography of this particular policy and tool extremely challenging, but it may not be desirable in the first place. Through interviews, access to information requests, and internet sleuthing, I have reconstructed a "good enough" sense of the TBS AIA's history. But these are pieces of a fairly incoherent policy assemblage. It would be dishonest to present them as a complete picture.

I developed the "creation" portion of this thesis to grapple with that incompleteness and incoherence. As I have noted, the TBS AIA is best understood using an assemblage approach to policy (Clarke et al. 2015; Lea 2020) because, among other things, an assemblage approach allows us to isolate one piece—in this case the AIA tool—while remaining attentive to the particular sociotechnical system of its original policy context. It is, in effect, a means of maintaining an ambivalent (Bucher 2019) position toward this technology.

There are undoubtedly critiques to be made of the TBS AIA. Indeed, this project has further developed several of them. Attempting to assess algorithmic impact using a one-size-fits-all, self-administered questionnaire is a flawed approach. And, as I have demonstrated, certain decisions

like including additional unscored questions in more recent versions of the questionnaire have made it all the less effective. The tool's use by the Government of Canada remains deeply questionable and largely performative.

For all these problems, however, I have still chosen not to write a polemic against this policy. Perhaps the key contribution of the TBS AIA—and indeed the thing that drew me to it in the first place—is that it was an early attempt at putting principles into practice. It certainly would have been easy enough to write a purely critical thesis on this tool, but doing so would not have produced many answers about where to go next. By exploring the TBS AIA through research-creation, it has been possible to develop a deeper critical understanding of the tool while also trying to find a path forward.

As for the AIA itself, the place where I see it as having the most potential is as a platform *for* algorithmic impact assessments. The questionnaire approach can be stifling when it is top-down, but it shows tremendous promise as a model for bottom-up consultation processes. In this way, reorienting the TBS AIA shows how we might—in practice—begin to develop inclusive models of AI governance that are not de facto expert-driven. Whether or not those models ultimately make use of this particular tool, hacking the TBS AIA gives us an avenue toward participatory methods for governing AI systems.

References

AI Now Institute. 2018. "Algorithmic Impact Assessments: Toward Accountable Automation in Public Agencies." https://ainowinstitute.org/publication/algorithmic-impact-assessments-toward-accountable-automation-in-public-agencies.

Beer, David. 2017. "The Social Power of Algorithms." Information, Communication & Society 20 (1): 1–13. https://doi.org/10.1080/1369118X.2016.1216147.

Bell, Genevieve, Mark Blythe, and Phoebe Sengers. 2005. "Making by Making Strange: Defamiliarization and the Design of Domestic Technologies." *ACM Transactions on Computer-Human Interaction* 12 (2): 149–73. https://doi.org/10.1145/1067860.1067862.

Bucher, Taina. 2019. "Bad Guys and Bag Ladies: On the Politics of Polemics and the Promise of Ambivalence." *Social Media + Society* 5 (3): 2056305119856705. https://doi.org/10.1177/2056305119856705.

Cardoso, Tom, and Bill Curry. 2021. "National Defence Skirted Federal Rules in Using Artificial Intelligence, Privacy Commissioner Says." *The Globe and Mail*, February 7, 2021. https://www.theglobeandmail.com/canada/article-national-defence-skirted-federal-rules-in-using-artificial/.

Chapman, Owen B., and Kim Sawchuk. 2012. "Research-Creation: Intervention, Analysis and 'Family Resemblances.'" *Canadian Journal of Communication* 37 (1). https://doi.org/10.22230/cjc.2012v37n1a2489.

Clarke, John, Dave Bainton, Noémi Lendvai, and Paul Stubbs, eds. 2015. *Making Policy Move: Towards a Politics of Translation and Assemblage*. Bristol: Policy Press.

Copeland, Eddie. 2018. "10 Principles for Public Sector Use of Algorithmic Decision Making." *Nesta* (blog). February 20, 2018. https://www.nesta.org.uk/blog/10-principles-for-public-sector-use-of-algorithmic-decision-making/.

Costanza-Chock, Sasha. 2020. "Design Sites: Hackerspaces, Fablabs, Hackathons, and DiscoTechs." In *Design Justice*. PubPub. https://design-justice.pubpub.org/pub/yvwnnz7b/release/1.

Fairclough, Norman. 2013. "Critical Discourse Analysis and Critical Policy Studies." *Critical Policy Studies* 7 (2): 177–97. https://doi.org/10.1080/19460171.2013.798239.

Goffey, Andrew. 2008. "Algorithm." In Software Studies: A Lexicon, edited by Matthew Fuller, 15–20. Cambridge.

Immigration, Refugees and Citizenship Canada. 2022. "Algorithmic Impact Assessment – Advanced Analytics Triage of Visitor Record Applications - Open Government Portal." https://open.canada.ca/data/en/dataset/01396e33-2c69-47e5-9381-32e717943b96.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. *Artificial Intelligence: The Global Landscape of Ethics Guidelines*.

Karlin, Michael. 2017. "Responsible AI in the Government of Canada: A Sneak Peek." *Code for Canada* (blog). July 5, 2017. https://medium.com/code-for-canada/responsible-ai-in-the-government-of-canada-a-sneak-peek-973727477bdf.

———. 2018. "A Canadian Algorithmic Impact Assessment." *Supergovernance* (blog). March 18, 2018. https://medium.com/@supergovernance/a-canadian-algorithmic-impact-assess-ment-128a2b2e7f85.

———. 2022 Interview by Nicholas Gertler.

Katzenbach, Christian, and Lena Ulbricht. 2019. "Algorithmic Governance." *Internet Policy Review* 8 (4). https://doi.org/10.14763/2019.4.1424.

Keung, Nicholas. 2023. "How AI Is Helping Canada Keep Some People out of the Country. And Why There Are Those Who Say It's a Problem." Toronto Star. June 2, 2023. https://www.thestar.com/news/canada/how-ai-is-helping-canada-keep-some-people-out-of-the-country-and-why-there/article_ea4dc715-debf-50e1-94c2-fda48a91dd09.html.

Lea, Tess. 2020. *Wild Policy: Indigeneity and the Unruly Logics of Intervention*. Stanford University Press. https://doi.org/10.1515/9781503612679.

Lessig, Lawrence. 2002. *Code: And Other Laws of Cyberspace*. Nachdr. New York: The Perseus Books Group.

Light, Ben, Jean Burgess, and Stefanie Duguay. 2018. "The Walkthrough Method: An Approach to the Study of Apps." *New Media & Society* 20 (3): 881–900. https://doi.org/10.1177/1461444816675438.

Maalsen, Sophia. 2021. "The Hack: What It Is And Why It Matters To Urban Studies." *Urban Studies*, February, 004209802098630. https://doi.org/10.1177/0042098020986300.

Mackenzie, Adrian. 2006. *Cutting Code: Software and Sociality*. Digital Formations, v. 30. New York: Peter Lang.

Marres, Noortje. 2005. "Issues Spark a Public Into Being." In *Making Things Public: Atmospheres of Democracy*, edited by Bruno Latour and Peter Weibel, 208–17. Cambridge, Mass. : [Karlsruhe, Germany]: MIT Press ; ZKM/Center for Art and Media in Karlsruhe.

———. 2010. "Front-Staging Nonhumans: Publicity as a Constraint on the Political Activity of Things." In *Political Matter: Technoscience, Democracy, and Public Life*, edited by Bruce Braun and Sarah Whatmore, 177–209. Minneapolis: University of Minnesota Press.

McKelvey, Fenwick Robert. 2014. "Algorithmic Media Need Democratic Methods: Why Publics Matter." *Canadian Journal of Communication* 39 (4). https://doi.org/10.22230/cjc.2014v39n4a2746.

Moss, Emanuel, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf. 2021. "Assembling Accountability: Algorithmic Impact Assessment for the Public Interest." Data & Society Research Institute.

Pangborn, Matthew H. 2010. "Defamiliarization," December. https://doi.org/10.1002/9781444337839.wbelctv1d001.

Paul, Regine. 2022. "Can Critical Policy Studies Outsmart AI? Research Agenda on Artificial Intelligence Technologies and Public Policy." *Critical Policy Studies* 16 (4): 497–509. https://doi.org/10.1080/19460171.2022.2123018.

Peters, John Durham. 2013. "Calendar, Clock, Tower." In *Deus in Machina: Religion, Technology, and the Things in Between*, edited by Jeremy Stolow, 1st ed, 25–42. New York: Fordham University Press.

Powell, Alison. 2016. "Hacking In The Public Interest: Authority, Legitimacy, Means, And Ends." *New Media & Society* 18 (4): 600–616. https://doi.org/10.1177/1461444816629470.

Ratto, Matt. 2011. "Critical Making: Conceptual and Material Studies in Technology and Social Life." *The Information Society* 27 (4): 252–60. https://doi.org/10.1080/01972243.2011.583819.

Reisman, Dillon, Jason Schultz, Kate Crawford, and Merideth Whitaker. 2018. "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability." AI Now Institute. https://ainowinstitute.org/wp-content/uploads/2023/04/aiareport2018.pdf.

Saurwein, Florian, Natascha Just, and Michael Latzer. 2015. "Governance of Algorithms: Options and Limitations." *Info* 17 (6): 35–49. https://doi.org/10.1108/info-05-2015-0025.

Scassa, Teresa. 2021. "Administrative Law and the Governance of Automated Decision Making: A Critical Look at Canada's Directive on Automated Decision Making." *U.B.C. Law Review* 54 (1): [i]-298.

Sharma, Sarah. 2008. "Taxis as Media: A Temporal Materialist Reading of the Taxi-Cab." *Social Identities* 14 (4): 457–64. https://doi.org/10.1080/13504630802211910.

Thorpe, Charles, and Jane Gregory. 2010. "Producing the Post-Fordist Public: The Political Economy of Public Engagement with Science." *Science as Culture* 19 (3): 273–301. https://doi.org/10.1080/09505430903194504.

Treasury Board of Canada Secretariat. 2023. "Directive on Automated Decision-Making."

https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

Université de Montréal. 2018. "The Declaration - Déclaration de Montréal IA Responsable." https://montrealdeclaration-responsibleai.com/the-declaration/.

Zarsky, Tal. 2016. "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine
   Efficiency and Fairness in Automated and Opaque Decision Making." Science, Technol-
   ogy, & Human Values 41 (1): 118–32. https://doi.org/10.1177/0162243915605575.