GETTING AHEAD OF THE CURVE:

Meeting the challenges to privacy and fairness arising from the use of artificial intelligence in the public sector

OMBUDSPERSON
BRITISH COLUMBIA

Yukon
Ombudsman

OFFICE OF THE
INFORMATION &
PRIVACY COMMISSIONER
FOR BRITISH COLUMBIA

Yukon
Information
and Privacy
Commissioner

Joint Special Report No. 2
June 2021

June 2021

The Honourable Raj Chouhan
Speaker
Legislative Assembly of British Columbia
Victoria BC V8V 1X4

The Honourable Jeremy Harper
Speaker
Yukon Legislative Assembly
Whitehorse YT Y1A 2C6

Dear Hon. Chouhan and Hon. Harper:

It is our pleasure to present the report *Getting Ahead of the Curve: Meeting the Challenges to Privacy and Fairness Arising from the Use of Artificial Intelligence in the Public Sector*. This is a joint report of our respective offices and is presented pursuant to:

- s. 31(3) of the British Columbia *Ombudsperson Act*
- s. 42(1) of the British Columbia *Freedom of Information and Protection of Privacy Act*
- s. 36(1) of the British Columbia *Personal Information Protection Act*
- s. 31(2) of the Yukon *Ombudsman Act*, and
- s. 117(3) of the Yukon *Access to Information and Protection of Privacy Act*

Sincerely,

Jay Chalke
Ombudsperson
Province of British Columbia

Michael McEvoy
Information and Privacy Commissioner
Province of British Columbia

Diane McLeod-McKay
Ombudsman and Information and Privacy Commissioner
Yukon Territory

CONTENTS

Contributors

Alexander Agnello: Policy Analyst, B.C. Office of the Ombudsperson

Ethan Plato: Policy Analyst, B.C. Office of the Information and Privacy Commissioner

Sebastian Paauwe: Investigator and Compliance Review Officer, Office of the Yukon Ombudsman and Information and Privacy Commissioner

MESSAGE FROM THE OFFICERS

With the proliferation of instantaneous and personalized services increasingly being delivered to people in many areas in the private sector, the public is increasingly expecting the same approach when receiving government services. Artificial intelligence (AI) is touted as an effective, efficient and cost-saving solution to these growing expectations. However, ethical and legal concerns are being raised as governments in Canada and abroad are experimenting with AI technologies in decision-making under inadequate regulation and, at times, in a less than transparent manner.

As public service oversight officials upholding the privacy and fairness rights of citizens, it is our responsibility to be closely acquainted with emerging issues that threaten those rights. There is no timelier an issue that intersects with our respective mandates as privacy commissioners and ombudsman, than the increasing use of artificial intelligence by the governments and public bodies we oversee.

The digital era has brought swift and significant change to the delivery of public services. The benefits of providing the public with increasingly convenient and timely service has spurred a range of computer-based platforms, from digital assistants to automated systems of approval for a range of services – building permits, inmate releases, social assistance applications, and car insurance premiums to name a few. While this kind of machine-based service delivery was once narrowly applied in the public sector, the use of artificial intelligence by the public sector is gaining a stronger foothold in countries around the world, including here in Canada. As public bodies become larger and more complex, the perceived benefits of efficiency, accessibility and accuracy of algorithms to make decisions once made by humans, can be initially challenging to refute.

Fairness and privacy issues resulting from the use of AI are well documented, with many commercial facial recognition systems and assessment tools demonstrating bias and augmenting the ability to use personal information in ways that infringe privacy interests. Similar privacy and fairness issues are raised by the use of AI in government. People often have no choice but to interact with government and the decisions of government can have serious, long-lasting impacts on our lives. A failure to consider how AI technologies create tension with the fairness and privacy obligations of democratic institutions poses risks for the public and undermines trust in government.

In examining examples of how these algorithms have been used in practice, this report demonstrates that there are serious legal and ethical concerns for public sector administrators. Key privacy concerns relate to the lack of

transparency of closed proprietary systems that prove challenging to review, test and monitor. Current privacy laws do not contemplate the use of AI and as such lack obligations for key imperatives around the collection and use of personal information in machine-based systems. From a fairness perspective, the use of AI in the public sector challenges key pillars of administrative fairness. For example, how algorithmic decisions are made, explained, reviewed or appealed, and how bias is prevented all present challenging questions.

As the application of AI in public administration continues to gain momentum, the intent of this report is to provide both important context regarding the challenges AI presents in public sector decision-making, as well as practical recommendations that aim to set consistent parameters for transparency, accountability, legality and procedural fairness for AI's use by public bodies. The critically important values of privacy protection and administrative fairness cannot be left behind as the field of AI continues to evolve and these principles must be more expressly articulated in legislation, policy and applicable procedural applications moving forward.

This joint report urges governments to respect and fulfill fairness and privacy principles in their adoption of AI technologies. It builds on extensive literature on public sector AI by providing concrete, technology-sensitive, implementable guidance on building fairness and privacy into public sector AI. The report also recommends capacity-building, co-operation and public engagement initiatives government should undertake to promote the public's trust and buy-in of AI.

This report pinpoints the persistent challenges with AI that merit attention from a fairness and privacy perspective; identifies where existing regulatory measures and instruments for administrative fairness and privacy protection in the age of AI fall short and where they need to be enhanced; and sets out detailed, implementable guidance on incorporating administrative fairness and privacy principles across the various stages of the AI lifecycle, from inception and design, to testing, implementation and mainstreaming.

The final chapter contains our recommendations for the development of a framework to facilitate the responsible use of AI systems by governments. Our recommendations include:

- The need for public authorities to make a public commitment to guiding principles for the use of AI that incorporate transparency, accountability, legality, procedural fairness and the protection of privacy. These principles should apply to all existing and new programs or activities, be included in any tendering documents by public authorities for third-party contracts or AI systems delivered by service providers, and be used to assess legacy projects so they are brought into compliance within a reasonable timeframe.
- The need for public authorities to notify an individual when an AI system is used to make a decision about them and describe to the individual in a way that is understandable how that system operates.
- Government promote capacity building, co-operation, and public engagement on AI. This should be carried out through public education initiatives, building subject-matter knowledge and expertise on AI across government ministries, developing capacity to support knowledge sharing and expertise between government and AI developers and vendors, and establishing or growing the capacity to develop open-source, high-quality data sets for training and testing Automated Decision Systems (ADS).
- Requiring all public authorities to complete and submit an Artificial Intelligence Fairness and Privacy Impact Assessment (AIFPIA) for all existing and future AI programs for review by the relevant oversight body.

- Special rules or restrictions for the use of highly sensitive information by AI.

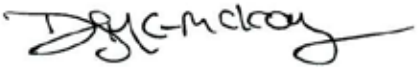
It is the responsibility of the public sector to adopt such emerging technologies only where these fundamental values are protected rather than abdicate responsibility by leaving such matters to the developers and owners of the technology. This report aims to bring attention to the fact that the question of AI is not just a question for computer scientists but a question for the whole of society. Policymakers, regulators, and technology developers all have important roles to play in ensuring that AI in government is consistent with good governance and the public interest. Fair government processes, decisions, services, and respect for privacy are non-negotiable aspects of this good governance package.

We look forward to careful consideration of this guidance by public bodies and will continue to actively monitor developments in this emerging field.

Sincerely,

Jay Chalke
Ombudsperson
Province of British Columbia

Michael McEvoy
Information and Privacy Commissioner
Province of British Columbia



Diane McLeod-McKay
Ombudsman and Information and Privacy Commissioner
Yukon Territory

INTRODUCTION

AI and public governance

Artificial intelligence (AI) has increasingly been used over the last ten years, and the use of AI is projected to be more widespread in the next fifteen. Global spending on AI was 12.4 billion USD in 2019 and is expected to reach 232 billion USD by 2025.[1] As part of Canada's national AI strategy, the federal government has invested 355 million CAD to develop synergies between retail, manufacturing, infrastructure, and information and communications technology sectors to build intelligent supply chains through AI and robotics.[2]

While interest and investment in AI have soared in the last decade, the use of AI in government is not new. There has been a long-standing trend of using automated processes 'behind the scenes' to support faster and more efficient public service delivery.[3] Since the late 1990s, the U.S. Postal Service has been using machine vision methods to recognize the handwriting on envelopes and automatically route letters.[4] Today, AI in government is increasingly public-facing. Virtual assistants are available 24/7 to help people navigate government processes and to access services, while natural language processing technologies can produce answers to questions directed at Parliamentarians.[5]

With the proliferation of digital applications and platforms across the economy, people expect more responsive and tailored services when they interact with their government. Peter Tyndall, former President of the International Ombudsman Institute and the Ombudsman of the Republic of Ireland, argued that one of the biggest challenges facing independent oversight offices and core government alike is the expectation of speedy results and high levels of interactivity with external clients: "They expect to interact with public services as they do with Amazon or Facebook, to communicate as they do on WhatsApp."[6]

---

1 Khube Mag: KPMG's Innovation Publication, "Intelligent automation edition" (April 2019) at 8.

2 "Supercluster invests in AI's economic potential for Canadians" *Government of Canada* (14 January 2020) online: <https://www.canada.ca/en/innovation-science-economic-development/news/2020/01/supercluster-invests-in-ais-economic-potential-for-canadians.html>; "CIFAR Pan-Canadian Artificial Intelligence Strategy" *CIFAR* online: <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>.

3 Hila Mehr Hila, "Artificial Intelligence for Citizen Services and Government" *Harvard Kennedy School* (2017) online (pdf): <https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf>.

4 *Ibid* at 5.

5 Jessica Mulholland, "Artificial Intelligence Will Help Create a More Responsive Government" *Government Technology* (Jan/Feb 2017) online: <https://www.govtech.com/opinion/Artificial-Intelligence-Will-Help-Create-a-More-Responsive-Government.html>; Charlene Chin, "Japan trials AI for parliament use" *GovInsider* (7 December 2016) online: <https://govinsider.asia/innovation/japan-trials-ai-for-parliament-use/>.

6 Síndic de Greuges de Catalunya and IOI-Europe, "AI & Human Rights: Ombudsmanship challenges, roles and tools (March 2-3, 2020) [available at: <https://www.youtube.com/watch?v=_dnPWUwR1eM> at 52:00-54:00 mins].

AI is being touted as the solution to these growing expectations. However, ethical and legal concerns are now being raised as governments in Canada and abroad are experimenting with AI technologies in high-stakes decision-making without regulation and, at times, in a covert manner.[7] The fairness and privacy issues resulting from the use of AI are well documented, with many commercial facial recognition systems and assessment tools demonstrating bias and augmenting the ability to use personal information in ways that infringe privacy interests.[8]

Similar privacy and fairness issues are brought on by the use of AI in government. People have no choice but to interact with government and the decisions of government can have serious, long-lasting impacts on our lives. A failure to consider how AI technologies create tension with the functions and obligations of democratic institutions poses risks for the public and undermines trust in government. This should bring our attention to the fact that the question of AI is not just a question for computer scientists but a question for the whole of society. Policymakers, regulators, and technology developers all have important roles to play in ensuring that AI in government is consistent with good governance and the public interest. Fair government processes, decisions, services, and respect for privacy are non-negotiable parts of the good governance package.

This report provides government officials and technology developers with detailed guidance on incorporating fairness and privacy obligations into the design, implementation, testing, use and mainstreaming of AI.

Defining AI

There is no single definition of AI. One common feature among authoritative definitions is that they describe AI in terms of its capacity, functions, and goals. Professors David Poole, Alan Mackworth and Randy Goebel define AI as "any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals."[9] This definition is often invoked in academic literature because it is broad enough to accommodate an ongoing discussion on the growing functions and uses of AI.[10] Professor John McCarthy, who coined the term AI, defined it as "the science and engineering of making intelligent machines."[11] Popular media articles on AI tend to subscribe to McCarthy's definition, framing AI in terms of the development of machines that can perform tasks normally requiring (human) intelligence, such as visual perception, speech recognition, language translation and decision-making. In the field of AI, "intelligence" is generally defined as the capacity to respond to opportunities and challenges in context.[12] "Artificial" implies that the device, intelligent machine, or rational agent has a human originator.[13]

---

8 In 2019, Reuters reported that Amazon's AI-powered recruiting tools penalized resumes with the word "women's" (e.g., "women's chess club captain") and it downgraded graduates of two all-women's colleges. See Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women" *Reuters* (10 October 2018) online: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>). For an excellent overview of race-based bias in technology, see Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Medford, MA: Polity, 2019).

9 David Poole, Alan Mackworth & Randy Goebel, *Computational Intelligence: A Logical Approach* (New York: OUP 1998).

10 One question being vigorously debated is whether the field of AI aims at building systems that *think or act like humans*, or systems that *think or act rationally.* See Stuart Russell & Peter Norvig, *Artificial Intelligence: A Modern Approach 3rd edition* (Saddle River, NJ: Prentice Hall, 2009); see also John Haugeland, *Artificial Intelligence: The Very Idea* (Cambridge, MA: MIT Press, 1985); Patrick H. Winston, *Artificial Intelligence*, (Reading, MA: Addison-Wesley, 1992).

11 John McCarthy, "What is Artificial Intelligence?" *Stanford University* (12 November 2007) online: <http://www-formal.stanford.edu/jmc/whatisai.pdf> at 2.

12 Stuart Russell & Eric Wefald, *Do the Right Thing: Studies in Limited Rationality* (Cambridge, MA: MIT Press, 1991).

13 This phrasing accounts for the fact that AI could theoretically create itself or other AI. The use of the term *human originator* makes it clear that at some point in the AI lifecycle, there was human creation or intervention.

The advantage of these 'goal-oriented' definitions of AI is that they align well with existing definitions that are tailored to the goals and functions of government. The Government of Canada's "Directive on Automated Decision Making" frames AI as "automated decision systems" (ADS): "any technology that either assists or replaces the judgment of human decision-makers." This broad definition is meant to cover the administrative decisions of government.

For this report, we will borrow the concept of ADS to zero in on AI technologies deployed in administrative decision-making, where there is the potential to have a significant and adverse impact on individuals.[14]

AI and social responsibility

AI is human-made. Human interaction occurs at every level of the AI lifecycle. Professor Joanna Bryson observes that even AI that is trained using highly automated techniques has still gone through a number of important human decision points, including the choice of algorithm and training and test data sets; the determination of the point at which AI is considered adequately trained; who will be subject to AI implementation and mainstreaming, and under what conditions; and whether testing will continue after implementation.[15] Therefore, we don't abdicate human responsibility when we are working with or being assisted by technology that can automate processes or operate autonomously.[16] AI systems "must be understood as composites of nonhuman actors woven together with human actors such as designers, data-creators, maintainers, and operators into complex sociotechnical assemblages."[17] The mere fact that AI 'learns' and makes decisions autonomously does not displace human responsibility in designing AI, putting it into motion, and allowing it to continue to operate in public settings.

The regulatory challenge

Regulatory intervention is necessary. The regulatory challenge is deciding how to adapt or modernize existing regulatory instruments to account for the new and emerging challenges brought on by government's use of AI. The increasing automation of government decision-making undermines the applicability or utility of existing regulations or common law rules that would otherwise apply to and sufficiently address those decisions.

---

[14] We use the term "ADS" to refer specifically to automated decision systems as defined by the Treasury Board of Canada's Directive on Automated Decision Making. The terms "AI systems" and "AI technologies and/or techniques" are used to refer to the application of AI more broadly.

[15] Joanna Bryson, "The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation" in Markus D. Dubber, Frank Pasquale & Sunit Das, eds, *The Oxford Handbook of Ethics of AI* (OUP: 2020) at 6.

[16] *Ibid*.

[17] Mike Ananny, "Towards an Ethics of Algorithms" Science, Technology & Human Values 41:1 (2015): 93-117; Nicholas Diakopoulos, "Transparency" in Markus D. Dubber, Frank Pasquale & Sunit Das, eds, *The Oxford Handbook of Ethics of AI* (OUP: 2020) at 198.

CHAPTER 1: HISTORY AND CORE CONCEPTS

Development of AI in two waves

AI has developed in an ebb and flow pattern marked by two waves. This is due in part to the fact that the development and success of AI depend on big advances in computing power. It is also driven by the fact that we are generating data at an unprecedented speed: "[o]ur world is undergoing an information Big Bang, in which the universe of data doubles every two years and quintillions of bytes of data are generated every day."[18] AI systems and techniques are becoming more and more adept at processing and maintaining extremely large and growing data sets, understanding data at a minute scale, and 'learning' from repeated exposure to example data. AI's ability to efficiently and effectively capture, store, organize and analyze troves of data has helped spur or at least accelerate the field of big data. At the same time, big data is critical for training and testing AI systems to perform classification and prediction tasks or make decisions.

Tracing the development of AI is important for pinpointing persisting issues with the technology that merit our attention from fairness and privacy perspectives. This approach also helps us understand the nature and scope of such issues:

- is the issue inherent to a specific type or iteration of an AI technology or technique, or a more 'global' issue with AI?
- Is there a solution on the horizon?

This line of questioning will help us identify issues worth focusing on in our analysis and recommendations.

The first wave (1960s-90s)

The first wave of AI (1960s-90s) is characterized by technologies and techniques that enable reasoning over narrowly-defined problems. First wave systems are based on clear and logical rules. The developer creates a rules-based algorithm – a series of mathematical instructions for transforming informational input into an output – that is applied to a defined knowledge base, and a logical conclusion is derived based on the algorithm's instructions.[19]

[18] In 2020 alone, an estimated 59 zettabytes of data will be "created, captured, copied, and consumed," enough to fill about a trillion 64-gigabyte hard drives. See "IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data" *IDC* (08 May 2020) online: <https://www.idc.com/getdoc.jsp?containerId=prUS46286020>; see also Cameron F. Kerry, "Protecting privacy in an AI-driven world" *Brookings* (10 February 2020) online: <https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/>.

[19] These systems are also known as knowledge-based systems or expert systems.

The BC Civil Resolution Tribunal's Solutions Explorer is a good example of a government application of first wave AI. The Civil Resolution Tribunal – the first online dispute resolution system in the world to be fully integrated into the justice system – uses simple logical rules to provide guided pathways to give people tailored legal information and tools via its Solution Explorer platform.[20]

However, given that first wave systems can only apply their knowledge base as the rules-based algorithm instructs, they are limited in their ability to handle 'novel' situations.

The second wave (2000s to present)

The second wave of AI (2000s to present) is characterized by machine learning (ML) algorithms that are developed and optimized "through statistical analysis of large datasets of historical examples."[21] Instead of programming AI systems to follow precise rules, second wave systems powered by ML 'learn' through the continual adjustment of its programming parameters to optimize their performance at various prediction, classification and decision-making tasks.[22]

These abilities allow second wave systems to conduct real-time, dynamic tasks such as speech and text recognition and transcription, facial recognition, and even piloting autonomous vehicles and drones. This is a significant improvement in capabilities over first wave systems, which are unable to execute tasks beyond the set rules and knowledge base that govern their operation. That said, there are serious unresolved issues that exist in second wave AI.

Second wave systems can perform statistical evaluations that are blind to the context-sensitive nature of natural language, signifiers, symbols, or memes. One well-known example of this is the controversy surrounding Microsoft's AI chatbot called Tay.[23] Tay was created to learn from teenagers through plain language conversations. Users started pranking Tay, telling the chatbot that 9/11 was an 'inside' job, that immigrants are the bane of America, among other disinformation. Tay then began disseminating this disinformation, spreading conspiracy theories and offensive views on Twitter. Microsoft shut down Tay 16 hours after it was launched. This example illustrates that the quality of the example data that a second wave system is trained on (or exposed to) determines what the system deems statistically significant. An ML model running on an accurate or unbiased algorithm that is being fed skewed data can produce a biased result.

Another significant issue with second wave systems is that they have difficulty with mapping the steps an ML algorithm took to transform informational input into a final outcome. Deep learning (DL), a subset of machine learning, is underpinned by a deep artificial neural network (ANN) modelled after the human brain. An ANN is made up of:

- artificial neurons that receive input;
- hidden layers consisting of mathematical equations to transform input; and

[20] See Shannon Salter, "Online Dispute Resolution and Justice System Integration: British Columbia's Civil Resolution Tribunal" (2017) 34 Windsor Y B Access Just 112; see also Darin Thompson, "Creating New Pathways to Justice Using Simple Artificial Intelligence and Online Dispute Resolution" (2015) 2:1 Intl J Online Dispute Resolution.

[21] David Spiegelhalter, *The Art of Statistics: Learning from Data* (Pelican, 2019) at 144; see also Simon Deakin & Christopher Markou, "From Rule of Law to Legal Singularity" in Simon Deakin & Christopher Markou, eds, *Is Law Computable? Critical Perspectives on Law + Artificial Intelligence* (Hart 2020) at 2 and 35-36.

[22] *Ibid*.

[23] Jane Wakefield, "Microsoft chatbot is taught to swear on Twitter" *BBC* (24 March 2016) online:<https://www.bbc.com/news/technology-35890188>.

- artificial synapses that connect artificial neurons together by transferring the output of one neuron to act as the input of other neurons.[24]

An input is supplied to the first layer of neurons, which transforms it and feeds the output as an input to the next layer of neurons through synapses. Each layer of neurons repeats this process until the final output is generated.[25] Learning through a 'deeply-layered' ANN model refers to 'learning' through a model characterized by several layers of neurons that divide the network into several cascading stages of calculation.[26]

Fully mapping the process of DL can require mapping out a cascading chain reaction among thousands of artificial neurons involved in the generated outcome. For this reason, DL engenders the issue of interpretability,[27] which makes it difficult for people to observe and measure causal relationships within an AI system to assess how the system generates an outcome. For example, Google's AlphaGo, which conducts DL through deep ANN, could be said to lack interpretability because its programmers could not determine how it produced the strategies for the ancient game of *Go* that defeated the human grandmaster in 2016.[28]

Moreover, training and testing ML systems such as DL is a very data-intensive undertaking. Depending on the nature and complexity of the task at hand, the complexity of the model, the level of performance sought, and the quality of the data available, it can take enormous amounts of data to produce valid models or yield even small optimizations to their performance.[29] This raises potential issues regarding purposeful and limited data collection, which we will examine in Chapter 3.

This report focuses on the current second wave statistical systems that are being deployed to assist or supplement government decision-making. Known issues with second wave systems (e.g., opacity and bias) can have serious and negative impacts on the ability to interpret and review government decisions. This is of particular concern for independent government oversight offices tasked with reviewing the fairness or legality of decisions made by government bodies. But it also poses significant risks to the public. For example, bias in automated decision systems (ADS) can produce uneven outcomes for people who most often interact with government (e.g., welfare recipients, detainees, etc.) and algorithmic opacity makes it more difficult to appeal an ADS decision in an informed manner.

This report is equally focused on the privacy risks associated with the need for vast amounts of data to develop and even achieve incremental improvements to ML and DL models.

In this next chapter, we will examine these issues in greater depth through well-documented use cases.

---

[24] Simon Deakin & Christopher Markou, "Ex Machina Lex: Exploring the Limits of Legal Computability" (2019) at 7-12 [available online: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3407856>].

[25] *Ibid*.

[26] *Ibid*.

[27] In the context of AI, interpretability is generally understood as the degree to which a causal relationship within an AI system can be observed and measured to inform decisions or predictions about that system. See Diogo V. Carvalho, Eduardo M. Periera & Jaime S. Cardozo, "Machine Learning Interpretability: A Survey on Methods and Metrics" (2019) 8:8 Electronics at 5-7.

[28] David Silver et al, "AlphaZero: Shedding new light on the grand games of chess, shogi and Go" *DeepMind* online: <https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/>.

[29] Theophano Mitsa, "How Do You Know You Have Enough Training Data?" *Towards Data Science* (22 April 2019) online: <https://towardsdatascience.com/how-do-you-know-you-have-enough-training-data-ad9b1fd679ee>; see also Malay Haldar, "How much training data do you need? *Medium* (28 November 2015) online: <https://medium.com/@malay.haldar/how-much-training-data-do-you-need-da8ec091e956>.

CHAPTER 2: THE DANGERS OF AI IN PUBLIC GOVERNANCE

This chapter examines three case studies to demonstrate how AI can undermine procedures and outcomes as they relate to fairness and privacy rights. These case studies cover the use of risk assessment instruments in criminal sentencing, biometrics and AI, and social credit systems.

Opacity and algorithmic bias in risk assessment instruments

The use of risk assessment instruments (RAI) has been employed for decades in many aspects of criminal justice decision-making. RAI are used in at least 44 countries by police, probation officers, and psychologists to assess the risk of criminal reoffending.[30] These decades-old tools have been rebranded as AI and are increasingly being used to inform judges concerning probation, sentencing and parole decisions.

A 2016 Wisconsin Supreme Court decision, *Loomis v Wisconsin*, disputed the trial court's use of a closed-source RAI, called COMPAS, in the sentencing of Eric Loomis – who pleaded guilty to eluding police and driving a stolen vehicle.[31] Developed by Equivant (formerly Northpointe), COMPAS – or the Correctional Offender Management Profiling for Alternative Sanctions – purports to predict the likelihood of an offender reoffending. It works through a proprietary, closed-source algorithm that analyzes answers to a 137-item questionnaire.[32] Loomis alleged that the trial court's use of COMPAS in his sentencing infringed on his right to an individualized sentence.[33] Moreover, he argued that the closed-source, protected nature of COMPAS's algorithm prevented him from challenging the system's decision criteria with respect to their scientific validity and accuracy.[34]

---

30 Jay P. Singh et al, "International perspectives on the practical application of violence risk assessment: A global survey of 44 countries" (2014) 13:3 International Journal of Forensic Mental Health.

31 Mitch Smith, "In Wisconsin, a Backlash Against Using Data to Foretell Defendants' Futures" *The New York Times* (22 June 2016) online: <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html>; Brief of Defendant-Appellant at 1–3, *State v Loomis*, No. 2015AP157-CR (Wis. Ct. App. Sept. 17, 2015), 2015 WL 1724741, at *iii–2.

32 Ed Yong "A Popular Algorithm Is No Better at Predicting Crimes Than Random People" *The Atlantic* (17 January 2018) online: <https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/>.

33 *Loomis*, 881 N.W.2d at 756-7.

34 *Ibid*.

The decision went on appeal before the Wisconsin Supreme Court. The Supreme Court found that Loomis could have verified the accuracy of the information used in sentencing because COMPAS uses only publicly available data and data provided by the defendant.[35] Regarding the question of individualized sentencing, the Court found that while COMPAS provides only aggregate data on the likelihood of recidivism for groups similar to the offender, the Court's decision was still sufficiently individualized as the Court had the discretion and information needed to properly weight the assessments and deviate from them where appropriate.[36]

The methodology used by COMPAS to produce the assessment was disclosed to neither the Court nor Loomis. Loomis could have made inferences about the factors COMPAS considered based on the questionnaire and public information COMPAS captures. But given that COMPAS uses a proprietary, closed-source algorithm that is protected by trade secret and was kept secret in this trial, Loomis had no way to know exactly which factors were considered and how they were weighed by the RAI in arriving at his risk assessment. Similarly, the Court lacked the information it needed to tailor their considerations of COMPAS's assessment if it could not scrutinize the series of steps that COMPAS took to arrive at Loomis's assessment.

In an interview with The New York Times, Northpoint general manager Jeffrey Harmon explained that the company's algorithms will remain proprietary because they are a core part of its business. Harmon also downplayed the importance of algorithmic transparency: "It's not about looking at the algorithms. It's about looking at the outcomes."[37] This is highly problematic from a fairness standpoint. Disregard for the quality of the decision-making process is antithetical to fair administrative decision-making as it is a serious challenge to one's ability to understand and evaluate government action. It is not possible to reconstruct how an RAI came to the assessment that it did without the ability to evaluate the information, steps and strategy that underpin the decision. Interpretability – the degree to which a causal relationship within an AI system can be measured and inform predictions made about that system – and explainability – the degree to which the internal processes of an AI system or the methods or techniques used in the application of that system can be described in human terms – is critical for evaluating the overall fairness of an automated decision system (ADS).[38]

The Supreme Court's decision in *Loomis v Wisconsin* also did not consider an ongoing debate regarding the accuracy of COMPAS and other RAI. A 2018 study of COMPAS by Julia Dressel and Hany Farid in *Science Advances* raises doubt about the efficacy of the RAI, showing that COMPAS is no better at predicting an offender's risk of reoffending than a random layperson recruited from the internet (67 percent accuracy rate).[39] Moreover, Dressel and Farid were able to create an algorithm that could predict

[35] *Ibid* at 761-2.
[36] *Ibid* at 764-5.
[37] Mitch Smith, "In Wisconsin, a Blacklash Against Using Data to Foretell Defendants' Futures" *The New York Times* (22 June 2016) online: <https://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html>.
[38] Diogo V. Carvalho, Eduardo M. Periera & Jaime S. Cardozo, "Machine Learning Interpretability: A Survey on Methods and Metrics" (2019) 8:8 Electronics at 5-7; Leilani H. Gilpin et al, "Explaining Explanations: An Overview of Interpretability of Machine Learning" Computer Science and AI Laboratory (MIT, 2019).
[39] Julia Dressel and Hany Farid's study judged COMPAS to have a 67 percent accuracy rate for predicting the risk of recidivism. See Ed Yong "A Popular Algorithm Is No Better At Predicting Crimes Than Random People" *The Atlantic* (17 January 2018) online: <https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/>; see also Julia Dressel & Hany Farid, "The accuracy, fairness, and limits of predicting recidivism" (2018) 4:1 Science Advances.

recidivism just as accurately as COMPAS by using just two data points – the subject's age, and their number of previous convictions.[40] Other researchers have arrived at similar results.[41] Dressel and Farid's algorithmic reconstructions show the apparent lack of computational sophistication behind COMPAS's assessment: if a person is young and has a number of prior convictions, they would be judged by COMPAS as a high risk for reoffending.

Dressel and Farid argue that the issue isn't that COMPAS is unsophisticated *per se*, but that it has likely reached a peak in sophistication for RAI with this outcome of interest.[42] When Dressel and Farid designed more complex algorithms that employed more data points, they still weren't able to improve on their initial model that used just age and prior convictions.[43] This suggests that, at this time, algorithmic modelling lacks predictive power when it comes to predicting an individual's risk of reoffending. Their research is a warning that we should take care not to deploy AI systems that lack sufficient predictive or explanatory power, especially in high-stakes decision-making.

Biometrics and privacy

As will become clear below, AI-driven use of biometric data by government has recently garnered significant public attention and criticism. The sensitive nature of this kind of personal information raises significant privacy concerns.

The Office of the Information and Privacy Commissioner for British Columbia (OIPC BC) defines biometrics as "the technology of measuring, analyzing and processing the digital representations of unique biological data and behavioral traits such as fingerprints, eye retinas, irises, voice and facial patterns, gaits, body odours and hand geometry."[44] More colloquially, biometrics refers to the measurement of life.

Facial recognition technology

Facial recognition technology (FRT) is a form of biometrics that can identify or authenticate individuals by comparing their facial features against a database of known faces to find a match. The process can be broken down into three steps. First, the computer finds facial features in a digital image, video frame or other representation. It then creates a numeric representation of the face based on the relative position, size, and shape of identified facial features. Finally, this numeric "map" of the face in the image is compared to a database of identified faces, for example, a driver's licence database.[45] Below we examine two use-cases of FRT to illustrate the privacy challenges with the technology.

The first case is Clearview AI, an American technology company that developed and provided app-based FRT software to law enforcement agencies. Canadian police forces, including the Royal Canadian Mounted Police (RCMP),

[40] *Ibid*.

[41] A paper by Elaine Angelino et al in the *Journal of Machine Learning Research* found that simple, transparent, and more interpretable algorithms such as a linear regression algorithm based on a person's age, sex, and prior convictions could predict recidivism just as accurately as COMPAS. See Elaine Angelino et al, "Learning Certifiably Optimal Rule Lists for Categorical Data" (2018) 18: 234 Journal of Machine Learning Research.

[42] Ed Yong "A Popular Algorithm Is No Better at Predicting Crimes Than Random People" *The Atlantic* (17 January 2018) online: <https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/>.

[43] *Ibid*.

[44] Office of the Information and Privacy Commissioner for British Columbia, "Investigation into the Use of Facial Recognition Technology by the Insurance Corporation of British Columbia" (2011 BCIPC No. 5) Report F12-01 [available at: <https://www.oipc.bc.ca/investigation-reports/1245>], citing Btihaj Ajana, "Recombinant Identities: Biometrics and Narrative Bioethics" (2010) 7 *Bioethical Inquiry* at 238.

[45] Office of the Privacy Commissioner of Canada, *Automated Facial Recognition: In the Public and Private Sectors* (Gatineau: 2013).

have used Clearview's product. Clearview AI's application uses a database of over 3 billion images scraped from the internet. The application automatically collects images of people's faces from employment sites, news sites, and social networks including Facebook, YouTube, Twitter, Instagram and Venmo, without authorization from these platforms.[46] An artificial neural network (ANN) uses biometrics to analyze facial features from digital images or videos that are scraped from these sites. Clearview AI's software identifies key facial features (e.g., the distance between your eyes) to develop a mathematical formula that is a person's facial signature. This signature is then compared to a database of identified faces to find a match. A joint investigation by the privacy commissioners of Canada, Alberta, British Columbia and Quebec found that Clearview AI's collection was done without the consent of individuals and, even if consent had been obtained, was not reasonable in the circumstances.[47]

An earlier use-case of FRT is the Insurance Corporation of British Columbia's (ICBC) use of its FRT database of driver's licence photos to assist law enforcement agencies in identifying individuals suspected of crimes. Most notably, ICBC offered to use its FRT database to assist the Vancouver Police Department in identifying suspects in the 2011 Stanley Cup riots. An OIPC BC investigation on this issue found that ICBC's stated use of FRT – to combat driver's licence fraud – did not allow ICBC to use that database for a collateral purpose of law enforcement without a warrant or court order.[48]

As these use-cases illustrate, the improper collection and use of biometric data raises significant privacy concerns for citizens. It is also worth noting that the very nature of the way biometrics operates presents a threat to individual privacy. The unique identifier being used in biometrics is a person's body. Avoiding identification and surveillance will become more difficult as visual surveillance technologies performing remote biometric identification become more pervasive and invasive.[49] The ability to choose what one shares about oneself helps us to define the boundaries of what we share with others, even in public settings. The Supreme Court of Canada in *R v Jarvis* (2019) recognized that people have a reasonable expectation of privacy, even in public spaces; people do not lose this expectation simply by walking out their front door.[50] For example, one reasonably expects to be recorded in a bank but does not expect to have their biometrics read upon touching the handle of the bank's entrance door. Similarly, FRT operates in public settings in ways that may undermine our reasonable expectation of privacy.[51] The standard terms of service that mediate digital consent are absent. We are often not made aware that we are being observed or recorded, how and why we are being observed or recorded, what biometric data or other information is being collected in the process and how it is being used.

---

46 Kashmir Hill, "The Secretive Company That Might End Privacy as We Know It" *The New York Times* (18 January 2020) online: <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>.

47 Office of the Information and Privacy Commissioner for British Columbia, "Joint investigation of Clearview AI, Inc." [available online: <https://www.oipc.bc.ca/investigation-reports/3505>]. <https://www.priv.gc.ca/en/opc-actions-and-decisions/ar_index/202021/sr_rcmp/>

48 Office of the Information and Privacy Commissioner for British Columbia, "Investigation into the Use of Facial Recognition Technology by the Insurance Corporation of British Columbia" (2011 BCIPC No. 5) Report F12-01 [available at: <https://www.oipc.bc.ca/investigation-reports/1245>].

49 Privacy International, "Visual Surveillance Technology" online: <https://privacyinternational.org/learn/visual-surveillance-technology#:~:text=Surveillance%20cameras%20and%20facial%20recognition,spaces%20and%20to%20identify%20people>.

50 *R v Jarvis* [2019] 1 SCR 488 at para 40. In *R v Jarvis*, at para 27, the Supreme Court of Canada eschews an "unduly narrow, location-based understanding of privacy."

51 Office of the Privacy Commissioner of Canada, "Cadillac Fairview collected 5 million shoppers' images: Customers not aware that their sensitive biometrics information was gathered" (29 October 2020) [available online: <https://www.priv.gc.ca/en/opc-news/news-and-announcements/2020/nr-c_201029/>].

Another challenge with biometrics is function creep. This occurs when a process or system intended for one purpose is subsequently used for a new or originally unintended purpose. In the context of personal information (PI), function creep refers to a change in use that is done without the knowledge or consent of the individual the information is about.[52] Both the ICBC and Clearview AI examples above illustrate how PI collected or disclosed for a particular purpose (driver licence fraud prevention and social media posts, respectively) is then repurposed for something entirely different without legal authority. Other biometric information that is already being collected and used by both public and private actors (e.g., heart rate data, MRI scans, blood type, DNA sequencing, etc.) is at risk of function creep as public and private databases can become linked and more biometric data is consolidated and more readily available for various uses.[53]

The aggregation of AI technologies and China's social credit systems

We have covered the dangers of using AI technologies and techniques to predict certain outcomes of interest (the likelihood of recidivism), inform high-stakes decision-making (criminal sentencing) and facilitate law enforcement activities (policing) in the justice system. The use of AI in these areas has generated a significant amount of public attention and criticism.[54] Another area of concern is the move away from AI usage that is limited in scope, towards a pan-society aggregation of AI. This move is characterized by the increasingly widespread use and aggregation of AI in the public and private domains to improve the collection and consolidation of data, insights, and other advantages across different platforms. Technology giants can acquire early-stage competitors at will. Through commercial acquisition, these giants can obtain new AI technologies and techniques, big data sets and business insights that further bolster their monopoly-like positions.[55] This allows them to interact with a broader market through a wider range of digital platforms and services, and by doing so, gain insight into consumer behaviour. Notable examples include Google and Amazon's ability to pool data from a range of internet-connected devices, such as smart glasses, wireless cameras, and voice-controlled smart speakers, by acquiring the companies that make them.[56] Similarly, AI has made data collection, analysis and sharing across government entities more efficient and ubiquitous, chipping away

---

[52] Office of the Information and Privacy Commissioner for British Columbia, "Investigation into the Use of Facial Recognition Technology by the Insurance Corporation of British Columbia" (2011 BCIPC No. 5) Report F12-01 at para 44 [available at: <https://www.oipc.bc.ca/investigation-reports/1245>].

[53] While clear legal authority for such linking of private sector and public sector databases is not apparent in Canada, this has been proposed in other jurisdictions such as China, as described below. We worry that the temptation to link private and public sector databases will likely only increase in the future given the potential advantages of doing so.

[54] See Julia Angwin et al, "Machine Bias" *Propublica* (May 23, 2016) online: <https://perma.cc/ZWX5-6BZP>; see also "Police can't use ICBC facial recognition to track rioters" *CBC* (16 February 2012) online: <https://www.cbc.ca/news/canada/british-columbia/police-can-t-use-icbc-facial-recognition-to-track-rioters-1.1207398>; finally see Karen Hao, "AI is sending people to jail – and getting it wrong" *MIT Technology Review* (21 January 2019) online: <https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/>.

[55] United States Subcommittee on Antitrust, Commercial and Administrative Law of the Committee on the Judiciary, "Investigation of Competition in Digital Markets" (2020) [available at <https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf>].

[56] Darrell Etherington, "Google acquires smart glasses company North, whose Focals 2.0 won't ship" *TechCrunch* (30 June 2020) online: <https://techcrunch.com/2020/06/30/google-acquires-smart-glasses-company-north-whose-focals-2-0-wont-ship/>.

[57] This includes the existence of AI-ready data governance infrastructure, appropriate safeguards, transparency, monitoring and accountability mechanisms.

at 'data islands' between public bodies. Under the right conditions,[57] this can promote better social outcomes by, for example, improving the detection of breast cancer,[58] more accurately predicting traffic congestion and car accidents,[59] and addressing tax compliance inefficiencies.[60] However, without a socially responsible purpose and adequate safeguards, the aggregation of AI and data can translate to the enhancement of government's ability to conduct surreptitious activities, such as blanket surveillance, and tracking and scoring of individuals.

This is an emerging reality in China. The Communist Party of China's 2014 Plan for the Construction of the Social Credit System (the Plan) is a formal announcement of the country's pan-society digital reform ambitions. The centrepiece of the Plan is the development of social credit systems (SCS). These systems can function as pan-society incentive mechanisms by gathering "public credit" information from commercial enterprises, different levels of government and even directly from individuals to penalize activities and behaviours that are deemed trust-breaking and reward those that are deemed trust-keeping.[61] SCS are supported by AI-powered surveillance infrastructure including facial recognition technology, actuarial assessment instruments, public credit score rating models, among other social management tools, to create "mechanisms of positive and negative reinforcement [...] intended to create a citizenry that continually engages in automatic self-monitoring and adjustment of its behavior."[62]

There are various SCS that target and operate on individuals by demographic by singling out individuals as consumers, business owners or local government officials.[63] Think of SCS as a network of various subsystems that target and compel specific actors to engage in or avoid certain activities and behaviours.

Local and provincial governments in China have built "Public Credit Information Platforms" to consolidate data that is generated from the public management functions of various departments.[64] These government SCS use that data along with social management tools to induce moral and law-abiding behaviour. Moreover, they can publicly identify individuals and enterprises who engage in illegal or otherwise "untrustworthy" activities. They can also impose penalties and restrictions that have serious social and economic ramifications

---

58 Neil Savage, "How AI is improving cancer diagnostics" *Nature* (25 March 2020) online: <https://www.nature.com/articles/d41586-020-00847-2>.

59 L.D. Tavares, "Detecting Car Accidents Based on Traffic Flow Measurements Using Machine Learning Techniques" in Ioannis Hatzilygeroudis & Jim Prentzas, eds, *Combinations of Intelligent Methods and Applications. Smart Innovation, Systems and Technologies* (Springer, 2011).

60 "Artificial Intelligence in Taxation" *Centre for Public Impact* (2018) online: <https://www.centreforpublicimpact.org/assets/documents/ai-case-study-taxation.pdf>.

61 Eunsun Cho, The Social Credit System: Not Just Another Chinese Idiosyncrasy" (n.d.) *Journal of Public & International Affairs* [available online: <https://jpia.princeton.edu/news/social-credit-system-not-just-another-chinese-idiosyncrasy>]; Bruce Sterling, "Chinese Planning Outline for a Social Credit System" *Wired* (3 June 2015) online: <https://www.wired.com/beyond-the-beyond/2015/06/chinese-planning-outline-social-credit-system/>.

62 Genia Kostka, "China's social credit systems and public opinion: Explaining high levels of approval" (2019) 21:7 New Media & Society at 1568.

63 Martin Chorzempa et al, "China's Social Credit System: A Mark of Progress or a Threat to Privacy?" (2018) online: *Peterson Institute for International Economics* <https://www.piie.com/system/files/documents/pb18-14.pdf> at 2.

64 Nesta, "The AI Powered State: China's approach to public sector innovation" (2020) [available at: <https://media.nesta.org.uk/documents/Nesta_TheAIPoweredState_2020.pdf>].

for individuals, ranging from being disqualified from obtaining bank credit or government subsidies to bans on purchasing flights or train tickets.[65]

The most widely used SCS are those run by commercial companies.[66] Increasingly, commercial enterprises in China are entering into commitments with government to restrict access to services and their platforms of, or by individuals and entities who have low public credit scores. For example, the ride-hailing service Didi may prevent drivers and riders from using its platforms. Ant's Sesame Credit score, used for several financial services, is negatively affected by social credit status. At the time of writing, commercial SCS offer users a wide range of benefits including qualification for personal credit loans, easier access to sharing-economy services (e.g., renting of a car or bike), fast-tracked visa applications, preferential treatment at hospitals, and free health check-ups.[67] Being banned from these platforms blocks access to a range of services and products that are important for social and economic mobility.

The full implementation of pan-society SCS in China crucially depends on the aggregated use of AI technologies and techniques, enhanced data collection, sharing and analysis techniques, as well as the breaking down of silos between government and private sector platforms and databases. At present, SCS are still largely localized and operating in silos, they are not implemented in every part of China, and opting out of SCS is possible but may effectively lead to retaliatory measures such as having to pay higher rates for products and services.[68] However, this is all slated to change in 2021 when the Communist Party of China plans to introduce sweeping SCS legislation to make these systems more comprehensive and practically mandatory.[69] Partially in response to these developments, the European Union's High-Level Expert Group on AI has recommended that the EU ban systems that automatically 'rate' individuals.[70]

The experience of SCS in China identifies AI's ability to make more powerful and intrusive the surveillance and social management apparatus of the State. The administrative actions and decisions of government already have incalculably significant impacts on people's lives. Aggregating AI technologies and big data across government enhances surveillance, tracking, and data collection resulting in the State expanding its reach. Certainly, positive outcomes in the form of more coordinated service delivery and evidence-based policymaking

---

65 Genia Kostka, "China's social credit systems and public opinion: Explaining high levels of approval" (2019) 21:7 New Media & Society; Marianne Von Blomberg, "The Social Credit System and China's Rule of Law" (2018) 2 Mapping China Journal.

66 In 2015, the People's Bank of China gave permission to eight enterprises to develop social credit pilots. The most common commercial SCSs are Sesame Credit, developed by Ant Financial Services Group, an affiliate of Alibaba, and Tencent Credit, developed by Tencent Holdings.

67 Rogier Creemers, "China's social credit system: an evolving practice of control" (2018) at 27 [available at: https://ssrn.com/abstract=3175792]; Genia Kostka, "China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval" 21:7 New Media & Society.

68 Social Credit Watch, "Trivium Primer: Understanding China's Social Credit System" (2019) [available at: <http://socialcredit.triviumchina.com/wp-content/uploads/2019/09/Understanding-Chinas-Social-Credit-System-Trivium-China-20190923.pdf>].

69 Christina Zhou & Bang Xiao. "'We Are Basically Living Naked': The Complicated Truth about China's Social Credit System" *ABC News* (2 January 2020) online: <www.abc.net.au/news/2020-01-02/china-social-credit-system-operational-by-2020/11764740>; Dutch Ministry of Foreign Affairs, "Country of origin information report: China" (2020) [available at: <https://www.government.nl/binaries/government/documents/reports/2020/07/01/country-of-origin-information-report-china-july-2020/COI+Report+China.pdf>].

70 European Commission: AI High-Level Expert Group, "Policy and Investment Recommendations for Trustworthy AI" (2019) European Commission at 20.

are more possible with this aggregation. However, as we saw with the example of ICBC's driver licence database being offered for collateral law enforcement purposes, there is a temptation for government entities in our jurisdictions to share data and capabilities in a non-justified manner for non-sanctioned purposes. Thinking about future dangers, more pressure may be brought to bear on governments to use AI technology to monitor adherence to public health directives to manage the spread of COVID-19. We may also see AI used by governments in the future to deliver services and provide health care as a measure to realize cost savings to relieve the pressure of debt accumulated during the pandemic.

Key takeaways and insights

This chapter was a detailed examination of public governance challenges with the current second wave statistical systems that are being deployed to assist or supplement government decision-making. It canvassed the challenges arising from the experimentation with automated decision systems (ADS) in high-stakes decision-making, the shift towards a whole-of-government and a society-wide approach to AI, all combined with increased opportunities for enhanced information sharing and collaboration between private AI vendors and public sector entities. While exploring these issues, we derived insights and points of analysis that informed this report's recommendations.

Risks and dangers of aggregating AI-driven technologies

- Aggregated AI technologies have the potential to be more intrusive than the sum of their parts. The society-wide aggregation of big data and AI may, if not checked by adequate regulation and enforcement, result in non-justified and non-sanctioned use of technological capabilities and data across government functions.
- Big data repositories do not always remain separate. The risk of data that is collected, used, and retained for a particular purpose (to prevent driver licence fraud) being repurposed for something entirely different (to assist police in identifying suspects in a riot) without the consent of the data subject grows as AI-enabled tools become more capable and accessible.

Importance of proper review, testing and monitoring of ADS

- Closed-source, trade protected algorithms are a significant barrier to proper review, monitoring and testing of risk assessment instruments (RAI) and other ADS. Judges, administrative decision-makers, and other public decision-makers lack the information they need to tailor their considerations of RAI if these instruments offer no reliable means to reconstruct ADS decisions.
- Full algorithmic transparency is not always warranted. However, in cases where transparency is warranted (e.g., in high-stakes decisions) and comes into tension with trade secret protections, systems might be made available for closed review to specific actors that are both legally bound and in a position of authority for assessing the system.[71] Transparency does not have to be an all-or-nothing affair; practically speaking, transparency is about producing information that promotes the effective governance and accountability of a system.

Defining what could be made transparent about AI systems

- Transparency is critical to the effective governance of automated systems. Policymakers need to articulate the range of data that could be made available about such systems, even proprietary, closed-source systems that are protected by trade secrets.
- At minimum, those responsible for an automated system should be required to disclose that an algorithmic process is taking place, the level and nature of human involvement in this process, the data that is used in training or operating the system, and the algorithmic model and the inferences that it draws.

Deploying automated systems in a measured, targeted manner

- There is a real risk that some AI systems reach a peak or limit in predictive or explanatory power for certain outcomes of interest. It is important to be aware of such risks and refrain from using AI systems where this is the case.

[71] Danielle K. Citron & Frank Pasquale, “The Scored Society: Due Process for Automated Predictions” (2014) 84:8 Washington Law Review; see also Paul B. de Laat, “Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?” (2018) 31 Philosophy & Technology.

CHAPTER 3: REGULATING AI: CHALLENGES AND BEST PRACTICES

At the outset of this report, we framed the regulatory challenge ahead as the challenge of deciding how to adapt or modernize existing regulatory tools to account for the challenges brought on by government use of AI. An intuitive place to start is by evaluating existing regulations that are meant to address, or can be tailored to meet, these issues. This chapter evaluates existing regulatory measures for administrative fairness and privacy protection, and identifies areas where regulations are challenged in applying to AI systems and where they can be enhanced. In doing so, we extract insights that will inform this report's recommendations.

Regulation and innovation: Earning public and private sector trust and buy-in

Regulation is often pitted against innovation. The idea behind this narrative is that regulation is simply a barrier to innovation. The "move fast and break things" norm of disruptive innovation encourages developers to postpone thinking about safeguards.[72] What is not often discussed is that safeguards by way of regulation are often key to carving out a space for AI to operate with the trust of the public and the business community; the latter playing a significant role in driving the innovation sector.

AI developers and vendors should realize that while there is a great deal of excitement surrounding AI, there is also a great deal of apprehension, including from businesses. According to a 2018 study by IBM's Institute for Business Value, 82 percent of all businesses surveyed, and 93 percent of high-performing businesses surveyed, are now considering or moving ahead with AI adoption as they are attracted by its ability to drive revenues, improve customer service, lower costs, and manage risk.[73] However, the same study showed that 60 percent of those companies fear liability issues and 63 percent say they lack the skills to harness AI's potential.[74]

This study demonstrates that trust in accountability mechanisms surrounding AI, including mechanisms for predictably determining who is accountable and when, is crucial to

[72] Nick Stratt, "Zuckerberg: 'Move fast and break things' isn't how Facebook operates anymore" CNet (30 April 2014) online: <https://www.cnet.com/news/zuckerberg-move-fast-and-break-things-isnt-how-we-operate-anymore/>.

[73] IBM Institute for Business Value, "Shifting toward Enterprise-grade AI" (2018) at 2 [available online: <https://www.ibm.com/downloads/cas/QQ5KZLEL>].

[74] *Ibid.*

business's adoption of AI. This study also shows that a considerable number of big businesses lack the capacity to fully harness AI and require guidance on responsible AI use. Building trust and guiding AI activity requires a sophisticated and robust regulatory framework that can cover, among many other issues, accountability mechanisms as well as permissible and impermissible uses of the technology.

As we highlight throughout, AI is shifting from being piloted technology deployed in narrow circumstances to technology that plays a front-and-center role in administrative decision-making and broader governance functions. Given that the scope of AI's application is potentially all of society, it becomes a society-wide issue, not just an issue for the innovation or technology sector. There is a responsibility on regulators and policymakers to ensure that good governance and individual rights and interests are not undermined by AI.

AI and privacy

Challenges with current laws

Public sector privacy law and AI: Yukon's ATIPP Act and BC's FIPPA

Public sector privacy is regulated by the *Access to Information and Protection of Privacy Act* (ATIPP) in the Yukon and the *Freedom of Information and Protection of Privacy Act* (FIPPA) in BC. Originally passed in 1992 and 1995 respectively, this legislation is based on a transactional and paper-based model of personal information (PI). Both have been amended in the intervening years, and a new ATIPP came into force at the beginning of 2021. As will become clear from the discussion below, these Acts, and other similar public sector privacy legislation across Canada, are not calibrated to address the unique risks and challenges that AI use poses.

AI usage is limited to a very narrow application[75]

The legal authority for collection[76] most widely used by public bodies requires that PI only be collected if it relates to and is necessary for carrying out a program or activity of the public body (PB). Similar provisions for use[77] and disclosure[78] must also be for a purpose consistent[79] with the PB's program or activity. Any adoption of AI by a PB will need to meet these requirements, which may prove challenging for the following reasons:

- Meeting the threshold of "necessary" is a high bar[80] and it may be difficult for a PB to justify or predict the amount of PI an AI system would need to run a program that previously ran without AI. (If AI enhances outcomes for either the PB or the individual, does this justify the additional collection of PI?)
- The limitations and safeguards in both FIPPA and ATIPP were not designed with AI in mind. As such, these laws contain no obligation for transparency in automated processes, nor the right of an individual to object to determinations made against them by automated decision systems (ADS) and the like.

---

75 *Access to Information and Protection of Privacy Act* (Yukon), (SY2018, c.9; amended by SY2019, c.15) at ss. 15, 21, 25 [ATIPP]; see also *Freedom of Information and Protection of Privacy Act* (British Columbia) (RSBC 1996 c. 165) at ss. 26, 32, 33, and 34 [FIPPA].

76 ATIPP, *supra* note 75 at s. 15(c)(i); FIPPA, *supra* note 75 at s. 26(c).

77 ATIPP, *supra* note 75 at s. 21(a); FIPPA, *supra* note 75 at s. 32(a).

78 ATIPP, *supra* note 75 at s. 25(a); FIPPA, *supra* note 75 at s.33.2(a).

79 Consistent purpose is defined as one that has a reasonable and direct connection to that purpose and is necessary for performing the statutory duties of or operating a program or activity of the public body; see ATIPP, *supra* note 75 at s. 21(b); FIPPA, *supra* note 75 at s. 34.

80 The standard is not so strict as to only permit usage where it would be impossible to operate a program or carry on an activity without the personal information. But the standard is not met where the use of personal information would be a mere convenience or advantage. See *Board of Education of School District No. 75 (Re)*, 2007 CanLII 30395 (BC IPC) at paras 48-49.

- Processing by AI may create new PI about an individual.[81] This constitutes an indirect collection. There is currently no authority for this specific type of indirect collection.

The above challenges can be mitigated by clearly documenting and assessing the potential risks and benefits of adopting any AI.

PI cannot be readily disclosed between public bodies for the purpose of improving AI

Even though FIPPA and ATIPP have been amended to include a specific provision for data linking, there still is no explicit authority in either Act to harness or interlink PI for the *purpose* of training an AI. Nor are there provisions regulating this kind of use in either Act. Any program using AI will still need to be able to clearly identify data flows and comply with existing collection, use, and disclosure requirements as outlined above.

The effective use of AI depends on having sufficient data for training and testing purposes. The protection of PI, instead, depends on strong principles and limitations whenever a new use of PI emerges. FIPPA and ATIPP keep PI 'siloed', meaning the use of PI is limited to the PB which collected it. As such, the PI cannot be readily shared with other PBs. Both Acts contain provisions[82] for data linking for the delivery of integrated services which may allow for AI to draw on multiple data sources containing PI to deliver a service. However, both legal and technical safeguards are needed to ensure a potential increase in the flow of PI does not erode rights such as the ability to correct and access information, or erode PI protections by creating more exposure as data linking could result in more copies of the data.

Opacity of AI decisions

Deep learning or other black-box techniques[83] pose challenges for compliance. Depending on the implementation of the learning mechanism, it might be difficult or impossible in certain cases (e.g., when leveraging neural network learning techniques) to comply with the following requirements in ATIPP and FIPPA:

- Requirements to document decisions: When PI is used by or on behalf of a PB to make a decision that directly affects an individual, the PB must retain that PI for one year to allow the affected individual a reasonable opportunity to obtain access to that PI.[84] The new ATIPP Act expands this retention requirement[85] beyond the PI of the individual affected by a decision, to also include the information used to make the decision as well.
- Requirements for accuracy of PI: PBs must make every reasonable effort to make sure that PI used is accurate and complete. Implicit in this requirement is the ability for the PB and the individual affected to be able to review data points used, including PI, and correct any inaccurate PI.[86]
- Time limits for response to access to information requests: Both ATIPP and FIPPA contain strict time limits for responding to access to information requests.[87] These include the obligation to produce and share

[81] For example, processing by AI may lead to the creation of an underlying hash value in biometric processing or the creation of a credit score.
[82] ATIPP, *supra* note 75 at ss. 27, 29; FIPPA, *supra* note 75 at 33.2(d); Freedom of Information and Protection of Privacy Regulation (BC Reg 155/2012) at s. 12.
[83] See Chapter 1 of this report for more information on deep learning (DL), artificial neural networks (ANN) and black-box AI systems.
[84] FIPPA, *supra* note 75 at s. 31.
[85] ATIPP, *supra* note 75 at s. 22(b).
[86] ATIPP, *supra* note 75 at s. 22(a); FIPPA, *supra* note 75 at ss. 28-29.
[87] ATIPP, *supra* note 75 at s. 50; FIPPA, *supra* note 75 at s. 7.

with the requestor machine-readable records within a certain timeframe.[88] There is a risk that the timeliness requirement might not be met if the PB has not, in advance, devised a way to produce such records for ADS decisions.

All the above requirements are difficult to observe where a PB would have difficulty reverse engineering and explaining an AI decision in human terms. The purpose of these requirements is to ensure meaningful access to, and right to request a correction of PI about oneself.[89] If an individual cannot understand how a decision about them was made, or whether or how certain pieces of their PI were used to make the decision, how could they realistically exercise their access to information rights?

To provide guidance to PBs considering ADS, clear rules are essential. These rules should provide for proactive transparency – a requirement that individuals be given access not only to their own PI, but also an ordinary-language explanation of how the decision was made. A 2020 report by Professor Ignacio Cofone, commissioned by the federal privacy commissioner, recommended the following definition for a "meaningful explanation" of an ADS decision: "an explanation that allows individuals to understand the nature and elements of the decision to which they are being subject or the rules that define the processing and the decision's principal characteristics."[90] This definition would provide guidance to PBs for meeting their documentation, accuracy and access obligations in the context of ADS.

Insufficient mechanisms for ensuring a right of access to, and a right to request correction of, personal information[91]

During the lifecycle of AI decisions, there are several opportunities for failure during the processing of the PI. Data input can be manipulated, data repositories can be stolen or changed, an ADS can become compromised, and outcomes could be forged. Even though BC's FIPPA and Yukon's ATIPP address PBs' responsibilities regarding the protection of PI, like with the obligation to use correct information, the balance shifts when using an ADS. We might require more defined standards regarding the digital infrastructure upon which the ADS operates than what is currently provided for in BC's FIPPA and Yukon's ATIPP Act. Modernized requirements should include the standardization and auditing of security requirements and an obligation to be able to prove, with non-repudiation, the integrity of ADS decisions made that are based on PI or affect an individual.

Third-party technologies used by public bodies

PBs may use third-party products to create their ADS. Recent privacy impact assessments of AI-powered products such as Office 365 highlight the risk of disclosure of (personal) information that comes with the vendor collecting data about the use of their products.[92] Regulations should place due diligence requirements on PBs acquiring third-party AI products that conduct these types of hidden information flows so there is transparency about any collections, use or disclosures of PI by the third-party.

---

88 ATIPP, *supra* note 75 at ss. 65(3) and 65(4); FIPPA, *supra* note 75 at s. 6.

89 Indeed, this is stated in both acts under the relevant purpose sections: *ATIPP* s. 1(1)(b) and *FIPPA* s. 2(1)(b).

90 Ignacio Cofone, Policy Proposals for PIPEDA Reform to Address Artificial Intelligence Report" (November 2020) Office of the Privacy Commissioner of Canada: [available online: <https://priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/pol-ai_202011/>].

91 ATIPP, *supra* note 75 at s. 33; FIPPA, *supra* note 75 at s. 30.

92 "Data protection impact assessments DPIA's Office 365 ProPlus, Windows 10 Enterprise, Office 365 online and mobile apps" *Rijksoverheid* (2019) online: <https://www.rijksoverheid.nl/documenten/rapporten/2019/06/11/data-protection-impact-assessment-windows-10-enterprise>.

HIPMA and Canadian health sector privacy laws

Similar to other health sector legislation in Canada, Yukon's *Health Information Privacy and Management Act* (HIPMA) is consent-based legislation that operates in both the private and public sectors. As such, the same consent principles discussed in the below section on private sector legislation apply but are tailored to the provisions of care in the Act.[93] HIPMA facilitates the disclosure of personal health information (PHI) between custodians for health-related purposes. Challenges specific to AI in healthcare are addressed below.

Third-party services that leverage AI

Third parties deliver much of the technology used by the health-care sector. Aside from privacy impact assessment requirements, the current legislation does not specify due diligence requirements for custodians acquiring products that leverage AI for the provision of health care. Moreover, a thorough understanding of the technical aspects of AI as they impact personal health information (PHI), or decisions made about health care based on PHI, is not explicitly required of the custodian.

Insufficient information security mechanisms to protect personal health information

As compared to some privacy laws in Canada, HIPMA and Yukon's new ATIPP contain relatively detailed information security requirements that are similar. The primary threshold in HIPMA is the reasonableness standard and ATIPP requires PBs to implement measures that are appropriate to protect PI. Neither provides for an express standard but requires one to follow industry standards. Because of the security risks associated with using AI to process highly sensitive PHI, specific safeguards for information security should apply. Adopting or adapting such a standard,[94] would create prerequisites for a secure infrastructure relative to the sensitivity of the PHI used, and the purpose for which it is used (providing healthcare services).

The reliance of medical professionals on AI to analyze medical imagery and form diagnoses is a good example of why these specific standards are needed. The unsettling findings of a 2019 study on the malicious tampering of 3D medical imagery using deep learning raises the information security stakes. Use of a specified standard for information security opens the way for meaningful audits of AI systems, and the infrastructure used to access these systems and their results, in a healthcare context.

The limitation principle

In HIPMA and other health privacy legislation collection, use and disclosure of PHI is restricted by limitation principles.[95] Even if consent can be obtained and information may be used, the information must be limited, meaning: 1) if other, non-identifying information suffices, no PHI may be used; and 2) PHI may be used only as far as it is necessary for the purpose for which it is being used. Given this limitation, it follows that in cases where non-identifying data can be created that serves the purpose for which originally PI would have been necessary, it is no longer necessary to use any identifying information. Where AI needs to be trained, it will be hard to argue why the information used cannot be anonymized, or, better yet, replaced by synthetic data, preferably at the source, before collection takes place.

---

93 *Health Information Privacy and Management Act* (Yukon) at ss. 32-46 [HIPMA].

94 ISO 27799:2016 (Health informatics – Information security management in health using ISO/IEC 27002) [available online: <https://www.iso.org/standard/62777.html>].

95 HIPMA, *supra* note 93 at ss. 13-18.

Private sector laws: PIPEDA and PIPA

While this report is focused on AI in public sector governance, engagement with private sector privacy laws is necessary because public bodies are likely to contract with commercial AI vendors. A full analysis of the challenges of regulating private sector AI is outside of the scope of this report, but this section seeks to outline the high-level challenges of doing so.

The applicable private sector legislation in British Columbia is the *Personal Information Protection Act* (PIPA). In the Yukon, the federal *Personal Information Protection and Electronic Documents Act* (PIPEDA) applies. PIPEDA applies everywhere in Canada where a province or territory has not enacted private sector privacy legislation that is substantially similar to PIPEDA, which is why businesses operating in the Yukon are subject to PIPEDA, and those in BC are subject to PIPA.

Both PIPEDA and PIPA regulate the collection, use and disclosure of personal information under two important requirements:

- for the purposes that a reasonable person would consider appropriate in the circumstances; and
- with the consent of the individual.

An organization cannot proceed without meeting both of the above requirements. This means that even if meaningful consent is obtained, the collection, use, or disclosure may only be for the purposes that a reasonable person would consider appropriate in the circumstances.

Consent can be express or implied, depending on the circumstances.[96] Consent functions as a way for individuals to protect their privacy by giving them control over how their personal information (PI) is handled in the private sector, specifically with respect to what PI businesses can collect, how they can use it, and to whom they can disclose it. An often-cited concern is that consent in the digital age has become a "meaningless, procedural act because users encounter so many different, long, and complicated terms of service that do not help them effectively assess potential harms or threats."[97] AI can exacerbate this issue. A 2018 report by the Canadian House of Commons Standing Committee on Access to Information, Privacy and Ethics expressed general concern about the lack of transparency in AI systems and how that could undermine meaningful consent. The Standing Committee found that "users have little information about how they work, the data they collect and how they are used."[98]

Even in circumstances where consent is obtained, organizations' data practices must still meet the standard of what a reasonable person would consider appropriate in the circumstances.[99] In other words, an organization can be in violation of privacy laws even where consent is obtained. Regulating AI using these foundational concepts of consent and reasonableness, as they currently operate, presents challenges.

---

[96] Clause 4.3 of Schedule 1 and section 7 of the *Personal Information Protection and Electronic Documents Act* lists the exceptions to the consent requirement; sections 6-9 of BC's *Personal Information Protection Act*.

[97] Meg Leta Jones & Elizabeth Edenberg, "Troubleshooting AI and Consent" in Markus D. Dubber, Frank Pasquale & Sunit Das, eds, *The Oxford Handbook of Ethics of AI* (OUP: 2020) at 373.

[98] Standing Committee on Access to Information, Privacy and Ethics, "Towards Privacy by Design: Review of The Personal Information Protection and Electronic Documents Act" (February 2018) at 25 [available online: <https://www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-12/page-ToC>].

[99] Personal Information Protection and Electronic Documents Act (Canada), (S.C. 2000, c. 5) at s. 5(3); Personal Information Protection Act (British Columbia), (SBC 2003, c. 63 at ss. 11, 14, 17.

First, the nature of consent has changed from the time these laws were implemented. Meaningfully consenting to all potential uses of our PI is more complicated today. In a 'smart' environment filled with Internet of Things (IoT) devices, the usual terms of service that mediate digital consent to screen-based data exchanges are largely absent. The "viability of opting out in these situations puts significant pressure on the legitimacy of consent to many of the aspects of our digital infrastructure."[100]

Second, even if broad consent is obtained, the collection, use and disclosure that follows must still be reasonable in the circumstances. Individuals may have consented to have their photos on social media, and even to certain collateral uses by various social media platforms, but some uses such as being included in Clearview AI's FRT application will be considered unreasonable regardless of whether consent is obtained.[101]

Administrative fairness and AI

Democratic, rule-of-law societies expect government to deliver public services in an open, accountable, equitable, and overall fair manner. As an independent voice for fairness, Ombuds offices investigate complaints and inquiries regarding the administrative decisions, practices, and services of public bodies. Under the *Ombudsperson Act* (BC) and the *Ombudsman Act* (Yukon), our Offices are tasked with investigating whether decisions, recommendations, acts, or omissions of public bodies of government are counter to the principles of administrative fairness. In the Canadian context, administrative fairness encompasses principles derived from individual legal rights, constitutional principles, best practices in governance and Canadian societal values and expectations.[102] The principles of administrative fairness, covered in fuller detail in Chapter 4, demand a fair decision-making procedure, fair decisions, and fair service by administrative decision-makers, such as boards and tribunals, commissions, and government executives where they exercise ministerial discretion.

The issues with ADS noted above can frustrate Ombuds' investigatory capacity and negatively impact a would-be complainant's ability to identify an issue of administrative fairness.

Opacity is a key problem with predictive, second wave systems and can make it difficult to reconstruct or retrace the path that an ADS took to render a decision. As we discussed in Chapter 2, ADS are opaque if they are a closed, proprietary system protected by trade secret. They can also be opaque as a result of their design. In either case, without the ability to evaluate the information, steps, and strategy that underpin the decision, Ombuds are unable to evaluate the process taken by the ADS to arrive at a decision outcome.

Moreover, even in cases where an ADS decision is not opaque *per se*, an Ombuds' investigatory power can be frustrated if piecing together every automated decision becomes a resource-intensive enterprise. This point underscores the need for government bodies to maintain a sufficiently detailed audit trail of the steps its ADS took to arrive at its decisions. Moreover, a detailed audit trail that is explainable in plain language terms is also important to give a person subject to a government decision an understanding of the basis upon which the automated decision was made.

[100] Meg Leta Jones & Elizabeth Edenberg, "Troubleshooting AI and Consent" in, Markus D. Dubber, Frank Pasquale & Sunit Das, eds, *The Oxford Handbook of Ethics of AI* (OUP: 2020) at 369.
[101] Office of the Information and Privacy Commissioner for British Columbia, "Joint investigation of Clearview AI, Inc." at paras 73-79 [available online: <https://www.oipc.bc.ca/investigation-reports/3505>].
[102] BC Ombudsperson, "Strategic Plan 2016-2021" (2016) Special Report No. 37 at 4 [available online: <https://bcombudsperson.ca/assets/media/2016-2021-Strategic-Plan-Special-Report-No-37.pdf>].

A related issue is access to the data on which the automated decision is based. A third-party data set upon which the decision is based might be 'closed', and this can frustrate the Ombuds' power to obtain information that they believe is relevant to conduct a full investigation.[103] Even in cases where third parties are forthcoming about the *type* of data contained in the data set, this still can have the effect of removing the assessment of relevance away from the Ombuds.

Overall, ADS employed by government bodies must be able to fully account for the investigatory powers of Ombuds.

Current and proposed approaches to AI

Both in Canada and abroad, the question of how to regulate AI has already attracted significant attention. Below is an overview of efforts to regulate AI.

Treasury Board of Canada Secretariat – Directive on Automated Decision-Making

In April 2018, the Treasury Board of Canada Secretariat (TBS) released a white paper on Responsible Artificial Intelligence in the Government of Canada. The paper committed the TBS to develop "a tool by which institutions can assess the degree of automation that is appropriate for their program" based on the degree of potential impact on individuals and society.[104] This set the stage for an evolving regulatory framework for AI at the federal level, which is centred on the Directive on Automated Decision-Making (the Directive) and its associated Algorithmic Impact Assessment (AIA).

The Directive is intended to ensure that an automated decision system (ADS), defined as "any technology that either assists or replaces the judgement of human decision-makers," is deployed in a manner that leads to more efficient, procedurally fair, consistent, and interpretable decision-making. The Directive mandatorily applies to any ADS used by federal government institutions, including systems implemented before the Directive. It requires the incorporation of fairness requirements into autonomous administrative decision-making by way of impact assessments, public reporting, verifiability, and auditing requirements. However, the Directive does not actually articulate and develop specific principles or requirements that track fairness in autonomous administrative decision-making by public bodies. It treats administrative fairness as though it were a self-defined or self-contained concept.

The AIA is a questionnaire meant to quantify the impact of an ADS – ranked on a four-level impact scale – and determine the appropriate degree of intervention.[105] Different levels of assessed impact will require varying degrees of intervention. For example, a decision with an impact level assessment of 1 does not require notice, peer review or direct human involvement in the decision. By contrast, decisions considered level 4 require two independent expert peer reviews, a

[103] Ombudsperson Act (British Columbia), (RSBC 1996, c. 340) at s. 15; Ombudsman Act (Yukon), (RSY 2002, c.163) at s. 15.

[104] An ADS that receives a Level 1 impact assessment requires minimal monitoring and testing, and no human failsafe. By contrast, an ADS that receives a Level 4 assessment will require two independent expert peer reviews, a public plain language notice, a human intervention failsafe, and periodic training. It was reported that an ADS deciding on a person's authorization to enter and leave the country will be immediately flagged for a Level 3 or 4 assessment. See "Algorithmic Impact Assessment" Government of Canada (03 June 2020) online: <https://open.canada.ca/aia-eia-js/?lang=en>; see also "The Government of Canada's Algorithmic Impact Assessment" *Medium* (7 August 2018) online: <https://medium.com/@supergovernance/the-government-of-canadas-algorithmic-impact-assessment-take-two-8a22a87acf6f>.

[105] "Canada's New Federal Directive Makes Ethical AI a National Issue" *Digital* (8 March 2020) online: <https://www.borndigital.com/2019/03/08/canadas-new-federal-directive-makes-ethical-ai-a-national-issue>.

public plain language notice, a human intervention failsafe, and periodic training courses.[106] For example, an ADS that makes decisions regarding a person's authorization to enter and leave the country will immediately be flagged for a level 3 or 4 assessment. The federal government has, through the Directive and the AIA, taken a leading role in Canada on creating a regulatory framework that aims to match the risk(s) of an ADS with a proportionate level of oversight and safeguards.

Advantages

- The Directive and AIA are fit-for-purpose. The criteria they use are meant to directly promote core administrative law principles (e.g., transparency, accountability, legality, and procedural fairness). Moreover, they apply to "any system, tool, or statistical model used to recommend or make an administrative decision about a client," except for any national security system.[107]
- The source code of the AIA is open source and distributed under a MIT licence,[108] a simple permissive licence with conditions only requiring preservation of copyright and licence notices.
- The AIA is continually undergoing public consultation with the understanding that technology is rapidly developing. Public consultation and workshops appear to happen on a weekly basis and new iterations of the AIA are posted to Github.

Shortcomings

- While the Directive mandatorily applies to any ADS used by federal government institutions, the Directive and the AIA do not appear to create enforceable rights for individuals who are adversely impacted by a decision of an ADS approved under the Directive, such as the right to appeal the automated decision made against them.
- The AIA appears to rely heavily on self-assessment and self-reporting by public bodies seeking to employ ADS, including commercial AI systems. This raises potential concerns regarding adequate reviews and safeguards. In completing the AIA, a government body might be relying on a risk assessment done by the private developer or vendor. A government body that fails to fully ascertain the risks with an ADS might engage in a lower level of review than is called for.
- The AIA questionnaire features a combination of more objective questions (e.g., "Does this system confer a legal status that is otherwise required to receive a benefit or service? Can this system result in granting or restricting access to a premises or network?") and more subjective questions (e.g., "Have *appropriate* strategies been developed to manage the risk that outdated or unreliable data is used to make an automated decision? Is the scope of the system *clearly* reflected in project documentation?" [emphasis added]).[109] The more subjective criteria introduce vagueness into the equation, and this can affect the usefulness of the assigned values in the AIA, as they are subjective.
- While the Directive and AIA purport to incorporate fairness in autonomous administrative decision-making, neither tool develops specific principles or requirements that track the demands of administrative fairness.

[106] *Ibid*.
[107] Government of Canada, "Directive on Automated Decision-Making" s. 5.2.
[108] Massachusetts Institute of Technology (MIT) licence [available online: <https://opensource.org/licences/MIT>].
[109] Algorithmic Impact Assessment questionnaire [available online: <https://canada-ca.github.io/digital-playbook-guide-numerique/views-vues/automated-decision-automatise/en/algorithmic-impact-assessment.html>].

EU proposed harmonized rules on Artificial Intelligence

On April 21, 2021, the European Commission published its proposal for a regulation on harmonized rules for artificial intelligence (EU AI Regulation).[110] The European approach is a cumulation of years of consultation and research and complements the automated decision-making provisions already found in the EU data protection law, the General Data Protection Regulation (GDPR).[111]

Much like the GDPR, the EU AI Regulation will undoubtedly become a benchmark for liberal democracies worldwide. The proposal regulates AI systems[112] through a classification model that rates them as "prohibited," "high," and "lower" risk. Prohibited systems are those considered to be a clear threat to the safety, livelihoods and rights of people. These include systems which manipulate behaviour to circumvent users' free will; social scoring systems by governments; and real-time biometric identification systems (except in extremely narrow and authorized circumstances).[113] High-risk systems include those that use critical infrastructure; provide educational or vocational training, employment services and essential services; or conduct law enforcement, migration processes, and the administration of justice and democracy.[114] These high-risk AI systems are subject to strict obligations before they can be put on the market, including:

1. Adequate risk management systems to continually evaluate the compliance;[115]
2. Requirements for high quality data and data governance for training, validation and testing data;[116]
3. Technical documentation and record-keeping requirements to ensure all necessary information is present to assess compliance, including the algorithm(s) used;[117]
4. Record-keeping and logs to allow for traceability of results;[118]
5. Transparent information as to allow users to interpret the system's output;[119]
6. Human oversight sufficient to allow natural persons to effectively oversee the system;[120] and
7. Robustness, security and accuracy, including appropriate measures to protect against cybersecurity threats.[121]

The preamble to the EU AI Regulation clearly draws upon the potential harms that can arise from the unregulated use of AI in a free and democratic society, many of which are identified in this report, such as the risks of discriminatory outcomes due to bias,[122] challenges of opacity

[110] "Proposal for a Regulation laying down harmonised rules on artificial intelligence" <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence>
[111] Articles 4 and 22 of the European Union General Data Protection Regulation contains robust protections related to automated processing of personal information to profile individuals and the right to object to such processing.
[112] Defined in the EU AI Regulation as "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with."
[113] EU AI Regulation, art. 5.
[114] *Ibid*, at art. 6(2) and annex III.
[115] *Ibid*, at art. 9.
[116] *Ibid*, at art. 10.
[117] *Ibid*, at art. 11 and annex IV.
[118] *Ibid*, at art. 12.
[119] *Ibid*, at art 13.
[120] *Ibid*, at art. 14.
[121] *Ibid*, at art. 15.
[122] *Ibid*, at preamble and clause 33.

in administrative justice,[123] and government social credit scores.[124] Beyond the world of data protection, the EU has recognized through this regulation the need for standalone legislation to address the society-wide impacts of ADS.

Quebec Bill 64

In June 2020, Quebec introduced Bill 64, An Act to modernize legislative provisions as regards the protection of personal information, to modernize the province's privacy law regime for the private and public sector. In their current state, the proposed changes create positive obligations on public and private sector organizations to inform an individual of the use of ADS, maintain an audit trail of ADS decisions and provide an individual with the right to appeal a decision made exclusively by ADS.

Entities that use ADS in decision-making would be required to inform individuals of its use at the time the decision is made. Bill 64 will require organizations to have an adequate process in place to track how their ADS works. It would also require them to maintain an audit trail of ADS decisions.[125]

This obligation is perhaps a measure intended to overcome the lack of algorithmic transparency in closed-source, proprietary systems. It might give an individual impacted by ADS a better chance of being able to appeal a decision made by such a system in a meaningful manner with sufficient precision. This aspect of the law is an improvement over the TBS Directive. The Directive requires disclosure and audits only where AI-enabled services are provided to the Government of Canada. Bill 64 requires entities to have disclosure and auditing processes regardless of to whom they provide services.

At the affected individual's request, an entity that makes decisions based *exclusively* on the automated processing of the individual's information must share what personal information (PI) was used to render the decision, the reasons, and the key factors that led to the decision, and inform the individual of the right to have corrected PI used to render the decision.[126] The organization would also be required to allow the individual to submit observations for a review of the decision.[127]

Public and private sector organizations that use technology that allow them to identify, locate, or profile[128] individuals would be required to inform the individuals in question of the use of that technology and, if applicable, allow them to deactivate the functions that identify, locate, or profile them.[129]

Consider this requirement in light of facial recognition technology and other AI technologies that collect and use personally identifiable information. Bill 64 would reinforce the consent requirement that exists under PIPEDA and possibly go further by allowing individuals to deactivate the functions that identify, locate or profile individuals. However, it's unclear

[123] *Ibid*, at preamble and clause 40.
[124] *Ibid*, at preamble and clause 17.
[125] Bill 64 (Quebec): An Act to modernize legislative provisions as regards the protection of personal information, s. 102 [available online: <http://www.assnat.qc.ca/en/travaux-parlementaires/projets-loi/projet-loi-64-42-1.html>].
[126] *Ibid*.
[127] *Ibid*.
[128] Under Bill 64, "profiling" means the collection and use of personal information to assess certain characteristics of a natural person, in particular for the purpose of analyzing that person's work performance, economic situation, health, personal preferences, interests, or behaviour.
[129] "Strengthening Privacy for the Digital Age: Proposals to modernize the Personal Information Protection and Electronic Documents Act" *Government of Canada* (21 May 2019) online: <https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html>. The Personal Information Protection and Electronic Documents Act (PIPEDA) currently does not provide data subjects such a right but the federal government is considering introducing such a right as part of its efforts to modernize the law.

how this right of deactivation would work with technologies that do not readily give users the ability to deactivate such functions, such as facial recognition systems that operate in public or commercial spaces.

Federal initiatives: Bill C-11 and consultations on AI

On November 12, 2020, the federal Office of the Privacy Commissioner released their Regulatory Framework for AI: Recommendations for PIPEDA Reform which was the result of extensive stakeholder consultation. It stated in part that any AI law should:

- Allow personal information to be used for new purposes towards responsible AI innovation and for societal benefits;
- Authorize these uses within a rights-based framework that would entrench privacy as a human right and a necessary element for the exercise of other fundamental rights;
- Create provisions specific to automated decision-making to ensure transparency, accuracy, and fairness; and
- Require businesses to demonstrate accountability to the regulator upon request, through proactive inspections and other enforcement measures through which the regulator would ensure compliance with the law.[130]

On November 17, 2020, the federal government introduced Bill C-11: Digital Charter Implementation Act, 2020.[131] This Bill repeals Part 1 of PIPEDA in its entirety and replaces it with a new *Consumer Privacy Protection Act* (CPPA).

The CPPA helpfully includes a definition of ADS[132] but the accompanying substantive provisions governing ADS are much weaker than both the GDPR and Quebec's Bill 64. Critically, these provisions only require an organization to provide a "general account" of the organization's use of ADS "to make predictions, recommendations or decisions about individuals that could have significant impacts on them."[133] Only after a request by an individual does an organization have the obligation to provide an explanation and a description of the decision and how it was made.[134] This lack of meaningful transparency and redress is of significant concern from a fairness and privacy perspective. With no notification, an individual would have no way of knowing that ADS was used in a decision about them.

The OIPC BC recommended in its submissions[135] to the Special Committee to Review the Personal Information Protection Act of the BC Legislature for PIPA to require notification, disclosure of the reasons and criteria used, and receive objections from individuals. This recommendation is repeated in this report.

[130] "A Regulatory Framework for AI: Recommendations for PIPEDA Reform" *Office of the Privacy Commissioner of Canada* online: <https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/>.
[131] Second Session, Forty-third Parliament, 69 Elizabeth II, 2020.
[132] Section 2 of the *Consumer Privacy Protection Act*, (Bill C-11) defines "automated decision systems" as "technology that assists or replaces the judgment of human decision makers using techniques such as rules-based systems, regression analysis, predictive analytics, machine learning, deep learning and neural nets."
[133] *Ibid*, 62(2)(c).
[134] *Ibid*, s.63(3).
[135] OIPC BC, "Submission to the Special Committee and Supplemental Submission to the Special Committee to Review the Personal Information Protection Act"

Insight

The goals of studying the regulatory landscape are to acknowledge why the regulation of AI is necessary, to validate our understanding of how administrative fairness and privacy principles recognized in existing regulations should apply to AI and identify changes that need to be made to regulations to ensure that AI respects these principles. We identified important tensions to resolve between fairness and privacy principles and automated processes. Moreover, paying particular focus to AI-specific regulatory instruments such as Canada's Directive on AI, though non-binding regulation, presages much-needed enhancements to the current regulations.

AI regulation crucially promotes innovation through public and private sector trust and buy-in

According to a 2018 study by IBM's Institute for Business Value, 60 percent of all companies surveyed fear liability issues arising from the use of AI and 63 percent say they lack the skills to harness AI's potential. Trust in accountability mechanisms surrounding AI, including mechanisms for predictably determining who is accountable and when, is crucial to business's adoption of AI. Building trust and guiding AI activity requires a sophisticated and robust regulatory framework that can cover, among many other issues, accountability mechanisms as well as permissible and impermissible uses of AI.

AI complicates consent

The lack of information and transparency combined with the different, long, and complicated terms of service make it difficult for users to effectively assess potential harms or threats before consenting to the collection or use of their personal information (PI). Even with transparent and meaningful information, the ability to meaningfully consent can still be undermined by the use of AI-powered Internet of Things (IoT) devices in 'smart' environments. In such environments, the usual terms of service that mediate digital consent (e.g., screen-based data exchanges) are absent. An example of this are AI assistants, which generally fail to inform their users of changes to their terms of service and privacy policies.[136]

[136] The authors scoured the web for examples of AI systems updating their users when changes are made to the system's terms of service or privacy policy but have found no such examples. The authors also asked Siri and Google assistant to provide a notification the next time either of these documents are updated, but neither appeared to understand the question.

Gap between ADS regulation and fairness principles

The Treasury Board of Canada Secretariat's (TBS) Directive on Automated Decision-Making is intended to ensure that ADS are deployed in a manner that leads to more efficient, procedurally fair, consistent, and interpretable decision-making. It purports to incorporate fairness requirements into autonomous administrative decision-making by way of impact assessments, public reporting, verifiability, and auditing requirements. However, the Directive does not actually articulate and develop specific principles or requirements that track fairness in autonomous administrative decision-making by public bodies. It treats administrative fairness as though it were a self-defined or self-contained concept.

Neither the TBS Directive nor Bill C-11 creates enforceable rights for individuals who are adversely impacted by a decision of an ADS, such as an explicit right to appeal the automated decision made against them. Meaningful rights are required to adequately ensure fairness in ADS.

Tensions between privacy legislation and how AI operates

Both public and private sector privacy legislation assumes individual transactions. This fails to consider aggregate impacts made possible through AI-enabled processing. For example, the authority for collection of information by law enforcement may be permitted at an individual level, but the impact of tools like predictive policing on society are far more concerning.[137]

There is a chicken and egg issue when it comes to AI meeting the "necessity threshold" stipulated in privacy legislation for the legal use of AI. Where programs and services are delivered without the use of AI, it is not, strictly speaking, necessary to use AI. If AI use is not "necessary," the AI is not allowed to use more PI than is required by the government program or activity, even if this would improve service or program delivery (e.g., reduce fraud, draw efficiencies that could be reinvested in the program or service, etc.).[138] Furthermore, under this understanding, AI may not use PI as training data where it is not necessary to the achievement of the program outcome. Overall, this is a highly-restrictive regulatory framework that stands in the way of program and service improvement and may prove to be increasingly impractical to comply with.

[137] For an excellent review of this topic, see Kate Robertson, Cynthia Khoo, and Yolanda Song, "To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada" *Citizen Lab* (1 September 2020) online: <https://citizenlab.ca/2020/09/to-surveil-and-predict-a-human-rights-analysis-of-algorithmic-policing-in-canada/>.

[138] An exception to this restriction is made where such outcomes are an explicitly stated program or service outcome.

Privacy legislation inadequately addresses our digital reality

Privacy legislation in Canada was introduced before the explosion of the digital age. Current legislation is based on transactions between an individual and a commercial enterprise or public body, which create straightforward relations between clearly-identified parties. In the digital age, users are often engaging in information-sharing with many parties and third-parties and for an array of purposes that are not contemplated by the user when granting consent.

Legislation must be flexible enough to keep up with fundamental changes in socio-technical developments impacting privacy. Due to the rapid increase in investment in and application of new uses for AI, provisions must be amended or created to maintain protection and the right to access.

The legislation discussed does not provide enabling provisions and restrictions needed to harness AI in a controlled manner. Current laws provide insufficient or unclear authorization for its use in, for example, research, development, and improvement of public sector ADS. They also fail to establish appropriate safeguards for problem areas in AI, such as black-box systems and third-party service providers of ADS.

CHAPTER 4: PROPOSED SOLUTIONS

Chapter 2 was a detailed examination of public governance challenges with current second wave predictive AI systems being deployed in government decision-making.[139] It canvassed the challenges arising from the experimentation with these systems in high-stakes decision-making, the shift towards a whole-of-government and a society-wide approach to AI, all combined with increased opportunities for enhanced information sharing and collaboration between private AI vendors and public sector entities. Chapter 3 was an evaluation of existing regulatory measures and instruments for administrative fairness[140] and privacy protection in the age of AI. The chapter identified areas where regulations fall short of AI systems and where they can be enhanced.

This chapter brings together these insights and points of analysis to provide detailed, implementable guidelines on incorporating administrative fairness and privacy obligations across the different stages of the AI lifecycle, from inception, design, testing, implementation, and mainstreaming.

The recommendations that follow are informed by scrutiny of current and anticipated AI use-cases and an analysis that pinpoints areas where automated processes come into tension with regulatory frameworks. However, this guidance is not meant to be applied "pro forma" or "as is," as it cannot realistically attend to all the features of the circumstances it attempts to anticipate. These guidelines are necessary, but they are not sufficient requirements for fair and privacy-centric use of AI. They will require revision as the technology evolves.

Fairness by design

Ensuring that AI-enabled government decision-making is held to the same administrative fairness standards as human-based processes requires thoughtful consideration of how fairness-by-design principles should factor into the AI lifecycle. The goal of elucidating a fairness-by-design framework with respect to AI is to translate the established requirements of administrative fairness to the context of AI decision-making.

Ombuds in Canada recently collaborated on the development of a fairness-by-design tool.[141] The tool sets out what is required to achieve fairness in decision-making by public bodies and guides

[139] See Chapter 1 of this report for a detailed discussion on the development of AI in three waves.
[140] Administrative fairness refers to fairness in the various dimensions of administrative decision-making, namely decisions that are not legislative or broadly based on policy direction. This includes decisions from a wide range of administrative decision makers, including boards and tribunals, commissions, and government executives where they exercise ministerial discretion.
[141] Various Ombuds Offices in Canada, "Fairness by Design: An Administrative Fairness Self-Assessment Guide" [available online: <https://www.yukonombudsman.ca/yukon-ombudsman/for-authorities/resources>].

them through the process of embedding fairness into program and service delivery. This model can be used as a reference to build fairness into the design of AI.

Fair procedure

Administrative decision-makers in government (e.g., boards and tribunals, commissions, regulatory agencies, government officials when they exercise ministerial discretion) must follow a fair procedure in making decisions. Administrative decisions are decisions of government that are not legislative or broadly based on policy direction. Flowing from administrative decisions is a duty to act fairly and make procedurally fair decisions. This duty exists as a safeguard for people in their interactions with government, as decisions made by administrative bodies can have a serious and long-lasting impact on individuals' lives.[142]

Below are four requirements of a fair procedure that must be met in every case:

1. **Adequate notice:** the person affected by the decision must be given adequate information to be able to participate meaningfully in the decision-making process (e.g., informed of the key issues in the decision process).
2. **Fair hearing:** the person affected is given a reasonable opportunity to present their case or to respond to the facts presented by others. Moreover, the decision-maker has genuinely considered what the person has presented to them when making their decision.
3. **Absence of bias:** the decision-maker decides with impartiality and independence. The term "impartiality" refers to the state of mind or attitude of the decision-maker and demands that there be no bias on this level, either real or perceived. Independence demands that the decision-maker not have ties with anyone that could lead to a reasonable doubt about their impartiality.[143]
4. **Justifiability:** the exercise of public power must be justified, intelligible and transparent, not in the abstract, but to the individuals subject to it.[144] This does not always require formal reasons and may also be justified in relation to the constellation of law and facts that are relevant to the decision.[145]

The scope of these four requirements is determined by contextual factors such as the gravity of the power being exercised, the party affected, the consequences and the impact of the decision or action at play and any relevant efficiency factors. For example, revoking a doctor's medical licence attracts a higher degree of procedural fairness[146] than a municipal building permit decision. The doctor's hearing would be closer to a court-style proceeding with witnesses and rules of evidence, where the building permit is assessed by a single inspector by way of application. These differences reflect what is justifiable, proportionate, and practical given the nature of the issue at play and its context, the importance of the decision to the individual affected, the legitimate expectations of the parties and the public purpose that the decision-making process is fulfilling in each context.[147]

A fair decision-making procedure requires:

- Advance notice of a decision and adequate information about the decision-making process and criteria;

[142] For the rules of procedural fairness to apply, the nature of the decision must be administrative; decisions that are legislative or broadly based on policy decisions are not required to be procedurally fair.
[143] *Committee for Justice and Liberty v Canada (National Energy Board)*, [1978] 1 S.C.R. 369 at p. 394.
[144] *Canada (Minister of Citizenship and Immigration) v Vavilov,* 2019 SCC 65 at para 95.
[145] *Ibid,* at para 105.
[146] *Dr. Q v College of Physicians and Surgeons of British Columbia*, [2003] 1 S.C.R. 226, 2003 SCC 19.
[147] *Baker v Canada (Minister of Citizenship and Immigration)*, [1999] 2 S.C.R. 817 at paras 21-27.

- An explanation of the organizational goal, purpose, or intent of the decision-making process;
- Review or appeal mechanism(s) that reflect the nature of the decision and its potential impact on a person's rights or interests;
- Open and clear communication throughout the decision-making process;
- Giving reasons for the decision where a duty to give reasons exists;
- Confidence that the decision-maker has conducted a rigorous review and is impartial; and
- Rendering a decision within a clear and reasonable timeframe.[148]

Integrating these fair procedure requirements in automated administrative decision-making will require:

- Publicly available, plain language descriptions and information about any ADS system that:
  1. Explains the organizational goal, purpose, or intent of the ADS, including the intended uses and out-of-scope uses as envisioned by the designers;[149]
  2. Details what the system is doing as it interacts with persons (e.g., "looking" at our faces to gauge our expressions, pooling personal information from various sources, etc.);
  3. Explains how the components of the ADS work to enable or support lawful decision-making, including how criteria for automated processing and the processing itself are consistent with the decision-making criteria found in law and regulation;
  4. Provides a description of the data used to train and test the system (i.e., detailing the type of personal information being used and from what sources) and a link to the de-identified training and test data if the data is public domain so that users can understand the basis upon which decisions are reached; and
  5. Gives advance notice to individuals that an ADS will be used to render a decision, along with clear steps on how the decision will be made.
- Giving users the means to appeal an ADS decision by:
  1. Providing users with a meaningful, plain language explanation of the steps and processes undertaken to arrive at a decision in their case; and
  2. Making publicly available, in plain language, reports, recommendations or other results arising from testing, monitoring, training, or auditing processes, so that people can contest an ADS decision with information regarding known or potential system issues.[150]
- Building confidence and trust in the quality of ADS decisions by:
  1. Ensuring that systems undergo periodic review, testing, and monitoring, and administrators undergo training as required:
     - **Review:** All ADS should be subject to risk assessments and systems deemed a substantial risk should, before implementation and mainstreaming, undergo peer review by several independent, well-positioned experts

[148] BC Ombudsperson, "Fairness by Design: An Administrative Fairness Self-Assessment Guide" (July 2019).
[149] Batya Friedman & Helen Nissenbaum, "Bias in Computer Systems" (1996) 14:3 ACM Transactions on Information Systems.
[150] Cynthia Rudin & Joanna Radin, "Why Are We Using Black Box Models in AI When We Don't Need To?" (2019) 1:2 *Harvard Data Science Review*.

from independent oversight bodies, government ministries or agencies, academia, or NGOs with the relevant capacity and expertise;

- **Testing:** The ADS, and its training and test data, should be fit-for-purpose and tested for relevance, accuracy and unintended data biases that may unfairly impact outcomes;
- **Monitoring:** The processes and outcomes of an ADS should be periodically monitored to ensure compliance with applicable legislation, regulation and to safeguard against unintended outcomes;
- **Education:** The administrator of an ADS should be educated in the design and functionality of the system on a reoccurring basis;[151]
- **Evaluation and public reporting:** Existing safeguards for the ADS, including the measures above, should undergo an independent and continuous evaluation and any findings of concern should be reported and made publicly available as soon as is possible.

Human control is critical to fairness in AI. Human intervention during the AI design cycle and human monitoring of AI in its operation ensures the system is performing as anticipated (human-on-the-loop). Similarly essential is establishing what tasks or responsibilities humans transfer to AI and ensuring the ability to override a decision made by AI (human-in-command).

Government use of closed-source, proprietary AI systems that cannot undergo the review, testing and monitoring outlined due to trade secrets is particularly problematic. Closed-source proprietary technologies are not only a barrier to adequate independent review, testing and monitoring but they also imply closed-source updates, which could (inadvertently) introduce new bias, errors, or mechanisms that entrench the software developers' worldviews. Full technical transparency may not always be warranted. But in cases where it is warranted (e.g., in high-risk decision-making) and comes into tension with trade secrets, systems should be made available for closed review to specific recipients that are both legally bound and in a position of authority for assessing the system, such as Ombuds offices and privacy commissions. Transparency does not have to be an all-or-nothing affair; practically speaking, transparency includes producing information that promotes the effective governance and accountability of a system.

Fair decision

Case law also imposes an obligation on administrative decision-makers to give *adequate* reasons for their decisions, which is different from the procedural fairness requirement to give reasons.[152] The reasons that underpin the decision must be based on "an internally coherent and rational chain of analysis that is justified in relation to the facts and law that constrain the decision maker."[153] In other words, there must be a rational connection linking the relevant evidence and the decision maker's arguments and conclusions, including a clear explanation of how relevant legislation, regulation or policy was followed and applied. Decision-makers should also be able to explain that evidence was rejected and why it was rejected. Not all decisions require written reasons. Many administrative decisions are made absent

[151] Treasury Board Secretariat of Canada, "Directive on Automated Decision Making: Appendix C – Impact Level Requirements" (2019).
[152] *Newfoundland and Labrador Nurses' Union v Newfoundland and Labrador (Treasury Board)*, 2011 SCC 62 separates procedural review for failure to provide reasons from substantive review for reasonableness.
[153] *Canada (Minister of Citizenship and Immigration) v Vavilov* 2019 SCC 65 at para 85.

written reasons but still based on criteria that must be demonstrable and fair. Where a written decision is required, the decision-maker must address in writing the major evidence they relied on (or rejected) to make the decision.

A fair decision:

- Must be made by a person who has the legal authority to make the decision;
- Must be in accordance with the applicable law and policy, and must take into consideration the appropriate degree of discretion that is afforded to the decision-maker by law;
- Must be made based on evidence and free from bias;
- Must not be oppressive, unreasonably burdensome, or improperly discriminatory.[154]

Integrating these fair decision requirements into autonomous administrative decision-making will require AI developers to design AI systems with an auditing function that is capable of:

- Identifying authorized decision-makers under the applicable legislation and the version of the system used to render the decision;
- Pinpointing all decision points or recommendations generated by the system;
- Linking decision points or recommendations within the system's logic to relevant law or policy;
- Generating a notification of the decision, including a statement of reasons, where required;
- Integrating change control processes to track modifications to the system's operations;
- Detailing the level and nature of human involvement in the decision-making process, logging instances where a human override of the system has occurred and identifying the natural person involved; and
- Incorporating the full discretion afforded to administrative decision-makers by law to leave an appropriate level of space for human judgment.

Imposing an obligation on organizations to track how their ADS works and maintain an audit trail of ADS decisions is a recommended measure for overcoming the lack of algorithmic transparency in closed-source, proprietary systems. This gives the individual impacted by ADS a better chance of appealing a decision made by ADS in a meaningful manner with sufficient precision. This also ensures that the bodies that review a decision can evaluate the fairness of the decision by examining the process used to arrive at the decision and outcome.

ADS should not apply law and policy to the exclusion of individual cases by, for example, adopting a one-size-fits-all approach to highly discretionary, context-driven cases. Human intervention and decision-making should be required where the ADS would otherwise fail to exercise an appropriate level of discretion.

The transparency of an ADS could be enhanced by way of a requirement to identify the natural persons responsible for engineering, maintaining, and overseeing the design, operation, testing and updating of the system and its dataset(s), with the idea that these persons might feel a greater sense of responsibility if their name and reputation are at stake.[155] It is also crucially important to ensure that there is a public body who can be held legally accountable for a decision made by an ADS.[156]

[154] BC Ombudsperson, "Fairness by Design: An Administrative Fairness Self-Assessment Guide" (July 2019).
[155] Nicholas Diakopoulos, "Accountability in Algorithmic Decision Making" (2016) 59:2 Communications of the ACM.
[156] Nicholas Diakopoulos & Sorelle Friedler, "How to Hold Algorithms Accountable" *MIT Technology Review* (2016) online: <https://www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/>.

Accountability would be further enhanced if the relevant professionals involved in the development of an ADS, such as software engineers, were subject to an enforceable code of ethics backed up by a statutory professional body.[157]

Fair service

Fair service regards how public bodies treat people who access their services. The principle of fair service poses the notion of 'user' interests as an obligation for the person responsible for providing the service. In the AI context, automated decision systems (ADS) should operate in a 'human-centric' manner. Developers and administrators of ADS should carefully consider and, where appropriate, integrate public feedback to manage continuous improvements as part of making sure that the service is fit-for-purpose, sufficiently individualized and does not produce uneven impacts or discriminatory outcomes.

Fair service includes:

- Equitable treatment and fair consideration of people's needs and circumstances in the delivery of the service;
- Keeping the lines of communication open and taking the time to understand stakeholders' perspectives; and
- Accepting responsibility for mistakes, providing apologies, and fostering a culture of ongoing quality review and continuous service improvement.[158]

Integrating these fair service requirements in autonomous administrative decision-making will require the following steps:

1. In designing and developing an AI system:
   - Ensure that algorithmic decision-making is appropriate for the proposed domain of application and does not run a foreseeable risk of producing bias, discriminatory outcomes, infringing on any other individual rights, negatively impacting public health or safety, or amplifying digital inequalities;[159]
   - Envision and design an ADS that can be more easily updated and maintained to facilitate continuous system improvement;
   - Train and test AI systems using only quality data that is fit-for-purpose, and be transparent about the data's accuracy, completeness, timeliness, update frequency, and uncertainty;
   - Consider the use of synthetic data where possible to reduce privacy risks;[160] and
   - Engage with the public at the initial stages of the design and development of AI that is going to be used on the public as it is helpful for anticipating unintended consequences early on and builds trust with the public if they know how they will be impacted.
2. Before deploying an AI system that will be used by the public:
   - Make sure that algorithms have gone through adequate training and tests to develop their predictive capacities;[161]

[157] Forum on Information and Democracy, "Working Group on Infodemics: Policy Framework" (November 2020) [available online <https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf>].
[158] BC Ombudsperson, "Fairness by Design: An Administrative Fairness Self-Assessment Guide" (July 2019).
[159] Uneven access to AI technology runs the risk of intensifying digital inequalities. The tendency to favor efficiency and personalization in service design based on the 'user' may lead to the development of public services that provide unfair advantages to people who fit the 'user' profile. See BC Ombudsperson, "Looking Ahead: Symposium on the Future of the Parliamentary Ombudsman Functions and Services" (2019) at 23-26.
[160] It is important to bear in mind that synthetic data does not improve or guarantee the accuracy or overall quality of the initial data.
[161] The training and testing data should be a representative sample that captures various nuances of the population set.

- Make sure that an ADS can perform its intended function with a high degree of predictive or explanatory power;
- Have ready processes for genuine consultation with internal and external stakeholders to ensure that the AI system will meet the needs of target groups and garner a "social licence to operate";[162] and
- Implement accessible mechanisms for people to raise concerns and appeal decisions made by AI.

3. Once the AI system is deployed:
   - Evaluate behaviours and outcomes as each new algorithm is introduced and continue to monitor them once a program is established to understand longer-term effects;
   - Introduce adequate mechanisms to collect, respond to and integrate critical feedback from users of AI systems for the purposes of ongoing quality review and continuous service improvement; and
   - Earmark resources for maintenance and improvement of the ADS.

Fair AI starts with the people who design it. Algorithms do not simply mirror the world; they reconstruct and alter it. In the words of sociologist Donald MacKenzie, an algorithm is "an engine, not a camera."[163] Fashioning AI systems that are beneficial to all and prevent uneven access to or impact of AI, starts with the direct involvement of people from many walks of life. The future of technology and who benefits from it will depend on who builds and implements it and who utilizes it or is subject to it.[164]

Privacy rights and AI

As detailed in Chapter 3, AI raises a host of privacy challenges. Existing rules are not nimble enough to account for AI uses that would improve program and service delivery. At the same time, current regulation is inadequate. We propose the following recommendations to improve the responsiveness of legislation and compliance tools to meet this challenge.

Rights-based approach to privacy

A robust rights-based approach to privacy is missing from Canada's privacy laws at the federal, provincial and territorial levels. Unlike other jurisdictions where they have recently modernized privacy law (e.g., the GDPR and EU AI Regulation), there is currently no Canadian law in force that addresses rights or obligations relating directly to AI. As discussed, Quebec's Bill 64 goes further than Canada's Bill C-11 in this regard. Neither law is in force yet.

A modern interpretation of the right to privacy as a human right is necessary for the exercise of other fundamental rights. At a minimum, privacy legislation should be amended to include the right to notification that ADS is used, an explanation of the reasons and criteria used, and the ability to object to the use of ADS.

Adjusting compliance provisions and tools

For compliance purposes, government and the private sector should be required to assess the privacy impacts before implementing AI technology. This obligation should be ongoing and verifiable through proactive audits by regulators once the technology is deployed. Some controls

[162] BC Ombudsperson, "Stem to Stern: Crown Land Allocation and the Victoria International Marina" (2018).
[163] Donald MacKenzie, *An Engine, Not a Camera: How Financial Models Shape Markets* (Cambridge, MA: MIT Press, 2007).
[164] For more details on various issues related to ethical data collection and transformation, see Nicholas Diakopoulos, "Ethics in Data-Driven Visual Storytelling," in Nathalie Henry Riche et al, eds, *Data-Driven Storytelling* (Boca Raton: CRC Press, 2018).

and obligations are already present in legislation, such as the need to complete privacy impact assessments (PIA).

- PIA regulations, templates and tools may need to be crafted to address AI-specific concerns, including the creation of an Artificial Intelligence Fairness and Privacy Impact Assessment (AIFPIA). This should include conditions that trigger the obligation to complete a PIA for systems that leverage AI to process PI and rules about when an AIFPIA must be conducted. The process should include a requirement to conduct security threat risk assessments and incorporate algorithmic impact assessment components[165] specific to ADS and the processing of PI. It should also require transparency and mandate the review of AIFPIAs by the appropriate oversight bodies.
- Industry-specific ethical standards should be implemented and include specific provisions regarding the processing of personal (health) information by AI. Such standards could be implemented as a form of co-regulation in which the industry has a significant measure of discretion in choosing standards, but once they are agreed upon are mandatory and enforceable.

Standards for security safeguards, including third-party processing

AI can play a role in collecting, transmitting, processing, and destroying PI and needs to be designed with adequate safeguards for processing PI. The current lack of explicit standards alongside the risk imposed by the use of third-party systems makes current requirements inadequate.

- The use of third-party solutions for ADS and other AI processing of PI must be balanced by requirements for transparency regarding this processing, including reporting on, and explicit standards for, security safeguards.
- This could include an obligation on public bodies to have the third parties they contract with prove compliance with their product or service with the security standard. For example, this could be done by means of demanding that third parties are (security standards) certified, and periodically validating the certification when these third parties process sensitive PI.
- Compliance with standards is no silver bullet, but it does provide a certain baseline and proof of due diligence. Compliance can be supported with proactive measures such as bug bounty programs and penetration testing of AI products or services.

Oversight of de-identified and synthetic data

With the compilation of massive amounts of data in recent years – some of which is publicly available – the de-identification of PI alone is an increasingly weak safeguard for the protection of privacy. Even when a name is stripped from a dataset, a combination of unique data points can be used to identify an individual with a high degree of certainty.[166] If a dataset used for cross-reference contains a name, re-identification can be performed. Even if the dataset contains no name, the dataset still constitutes PI and is still a compliance risk to the controller because new datasets may become available that then enable re-identification.

[165] See Chapter 3 of this report for an explanation regarding Algorithmic Impact Assessments (AIA).
[166] See, as examples, browser fingerprinting (https://panopticlick.eff.org/) or the reidentification of the Netflix user preference database (https://www.securityfocus.com/news/11497).

An alternative to de-identification is the use of synthetic data.[167] To derive accurate synthetic data, initial use of de-identified PI is required. If properly legislated, a provision that enables organizations to use PI to create synthetic data may enable organizations to meet legitimate business purposes without compromising privacy. This provision could reinforce *privacy by default* and *embedded privacy.*

- De-identification of PI is an increasingly weak mechanism for the protection of privacy and should be phased out of legislation.
- Using PI for synthetic data creation can strengthen privacy protection when using AI for legitimate business purposes. A special provision that, under specific circumstances, authorizes the use of PI for this purpose may enable AI usage while maintaining privacy protection.
- If authorized, synthetic data can enhance limitation principles. Currently, several pieces of legislation[168] spell out the requirement to use non-PI if this suffices for the intended purpose. If a statistically significant dataset exists, synthetic data can be created and the processing of PI (besides the creation of the synthetic data) is no longer needed, limiting the exposure of the actual PI.
- Arguably, de-identified or synthetic data is no longer personal information which could result in an oversight gap. Defining this process and explicitly extending oversight to privacy commissioner offices will address this issue.

Prohibitions and restrictions

As shown in Chapter 2 of the report on dangerous AI use-cases in the public sector, AI enhances the ability of governments and organizations to collect and analyze personal information (PI) and to act on this information. These use-cases illustrate the potential for abuse and misuse when leveraging the power of AI. An excellent precaution against the most malicious uses of AI is prohibiting or tightly restricting the creation of repositories of certain sensitive PI or sensitive combinations of PI that AI could draw on. Repositories containing religious affiliation or ethnicity have been used in the past to facilitate crimes against humanity long before the introduction of AI.[169] However, AI can facilitate the use of such repositories for nefarious purposes, a recent example being the use of biometric data and facial recognition technology in repressing the Uighur population in China.[170] In this context, AI has been used to identify and create lists of people deemed suspicious, and it is reported that more than fifteen thousand Xinjiang residents were placed in detention centers during a seven-day period in June 2017 after being flagged by AI.[171]

- Governments of liberal democracies have an obligation to protect current and future generations against the long-term impact of the creation of government or corporate-owned big-

[167] Alexander Watson, "Deep dive on generating synthetic data for Healthcare" *Medium* (2020) online: <https://medium.com/gretel-ai/deep-dive-on-generating-synthetic-data-for-healthcare-41acb4078707>.
[168] HIPMA, *supra* note 96 at s. 15, *Personal Health Information Protection Act* (Ontario), (S.O. 2004, c. 3, Sched. A) at s. 30(1).
[169] Examples include "a comprehensive population registration system for administrative and statistical purposes" that was created by Dutch authorities in the years before Nazi occupation of the Netherlands during World War II. This system was then used by the Nazis to effectively round up and deport Jewish and Roma populations to destruction camps in Germany, Austria and Poland. Seventy-three percent of Dutch Jews were killed by the end of the war, compared to 40% in Belgium and 25% in France, where such comprehensive registration systems did not exist. For other examples. See Zara Rahman, "Dangerous Data: The Role of Data Collection in Genocides" *The Engine Room* (21 November 2016) online: <https://www.theengineroom.org/dangerous-data-the-role-of-data-collection-in-genocides/>.
[170] Lindsay Maizland, "China's Repression of Uighurs in Xinjiang" *Council on Foreign Relations* (30 June 2020) online: <https://www.cfr.org/backgrounder/chinas-repression-uighurs-xinjiang>.
[171] *Ibid.*

data repositories detailing the lives of citizens in terms of their biometric data, (political) opinions, trade union membership, ethnicity, religion, and sexual orientation. It's important and necessary to codify this essential liberal principle in privacy legislation by means of rights protection against such surveillance.

- To prevent the (incidental) creation of such lists, it should be prohibited to create and retain data repositories containing these 'off-limits' categories of PI (it is likely that repositories detailing religious affiliations already exist in Canada).[172] For example, it's inappropriate for a political party to hold lists detailing, for instance, the religious affiliation, ethnicity, or skin colour of their constituents. Article 9 of EU GDPR prohibits the processing of this category of sensitive PI by default. EU member countries may develop mechanisms to create exceptions in specific circumstances.[173] Privacy laws in Canada should include similar provisions to protect against the risks associated with the processing of this highly sensitive PI.

Recently, the BC Government announced plans to introduce legislation that "will help reduce systemic discrimination and pave the way for race-based data collection essential to modernizing sectors like policing, health care and education."[174] An example of similar legislation was recently passed in Ontario,[175] which includes specific provisions for data governance. With the appropriate protections in place, significant social benefits can accrue from the collection of this type of information and our offices look forward to assisting with any projects that advance this goal.

Review of legislation

AI has become a mainstream phenomenon over the past two decades. According to Moore's Law, we should expect the rate at which computational processing power advances to increase with each passing year. As discussed in Chapter 1, processing power is a core ingredient for the advance of AI capabilities. To keep up with technological developments, the legislated timelines at which legislation is to be reviewed must be short enough to address significant changes in technology and their impact on society. Depending on the speed such developments reach, governments may have to consider models of continuous development of legislation as a solution to keep up with such rapid change.[176]

[172] "Privacy commissioner looking into cards sent to Jewish homes by PM" *CBC* (11 October 2007) online: <https://www.cbc.ca/news/canada/privacy-commissioner-looking-into-cards-sent-to-jewish-homes-by-pm-1.663605>.
[173] For example, a church should be able to have a member list.
[174] Mandate letter of Parliamentary Secretary Rachna Singh, online: <https://www2.gov.bc.ca/assets/gov/government/ministries-organizations/premier-cabinet-mlas/minister-letter/singh_mandate_2020_jan.pdf>.
[175] *Anti-Racism Act* (Ontario), S.O. 2017, c. 15.
[176] One example is the use of the incremental development model of AGILE in software development [available at: <https://www.tutorialspoint.com/agile/agile_primer.htm>].

CHAPTER 5: THE NEED FOR A WIDER APPROACH

The emergence of AI is no longer relegated to science fiction but growing to become an inescapable part of human life in the information age. Harnessing the challenges associated with this will require action across government and sectors.

Beyond silos

Narrow legislative changes or increased enforcement powers on their own will not address the larger challenges that come with the emergence of Tech giants. A lack of healthy competition in the Tech sphere may result in a lack of market influence on privacy as a property of services or products.

1. As AI is usually designed, developed, deployed, and sold as a service or software by Tech giants, relative fines give credibility to the importance of privacy by discouraging the design of AI that violates privacy rights for profit.
2. The relatively high monetary value of relative fines makes seeing privacy as just "the cost of doing business" an unattractive proposition. In order for these fines to be a substantial deterrent, they need to be commensurate to the financial power of the organization in question.
3. To be able to monitor compliance with purpose limitation, proper consent, transparency, and the right to object, privacy laws should include a provision requiring these organizations to implement an internal complaint mechanism with whistleblowing protection. To effectively monitor compliance, these organizations must be equipped with adequate technical and legal expertise to interpret, and investigate issues based on said provision.
4. Legislation gives regulators power to proactively review compliance with privacy laws and investigate privacy concerns. The effectiveness and timeliness of compliance monitoring and enforcement will be enhanced if such a provision is introduced. As well, compliance, in general, will be enhanced by providing regulators with the power to issue orders and fine companies who engage in non-compliance. These controls are needed to counterbalance monopoly power.
5. There will be less incentive for companies to make privacy a priority of an AI product or service where they hold *de facto* monopolies. Early-stage competitors have fallen prey to 'killer acquisitions' and become part of the growing 'Facebook conglomerate' (e.g., WhatsApp, Instagram, Oculus, etc.). In such a scenario, individuals and organizations alike cannot choose a more privacy sensitive version of a social media platform like Facebook.[177] Effective anti-trust regulations are important for opening the market to privacy-sensitive technologies.

[177] Facebook serves as an example, but a similar case could be made for other digital platforms and products that have few or no realistic alternatives.

The above bullets are requirements for creating an ecosystem in which public bodies can safely acquire AI technologies for government use. On the other hand, failing to build in the incentives for the design of AI that uphold privacy will either result in the acquisition of flawed technologies that violate privacy rights, or significantly slow and reduce the adoption of AI for beneficial purposes.

Strengthening expertise, enhancing diversity in AI across government

As the pace of technological advancement increases public servants will increasingly require rigorous technical knowledge of AI technologies and techniques. This will be important to not only negotiate complex agreements with AI vendors and other contractors, but also to help the public service capitalize on the benefits, and anticipate the risks, of deploying AI. At the same time, vendors and engineers will require a nuanced understanding of good governance norms, privacy rights, and the underlying legal and constitutional framework in which they are embedded, to identify how to undertake system design, testing, and implementation in ways that are consistent with our laws and system of governance. In addition, the development of an AI workforce in government that is diverse in educational background and reflects the make-up of the population is an important safeguard against unethical practices, bias, and groupthink. Hiring policies and practices must recognize that people's lived experiences contribute to the trajectory that technology follows.

These measures and initiatives must happen on an ongoing basis. There may be certain technological shifts that signal the need for additional expertise, upskilling, and other capacity-building measures (e.g., the (proposed) adoption of a new AI system, a known cybersecurity threat, etc.). That said, these measures should not be implemented merely as a reaction to shifts and risks that arise from technology. These changes need to be planned and carried out prospectively and proactively because they improve institutional capacity to anticipate and recognize these shifts and risks in the first place.

Preparing oversight bodies to co-operate on cross-mandate issues

AI in government may use personal information for making a decision about an individual. As indicated above, any use of PI in government decision-making must comply with privacy laws and the decision made must be fair. The use of AI to process PI will, therefore, merge the work of Ombuds and privacy commissioners. Given this, it would be beneficial for these oversight bodies to work jointly on AI reviews and investigations. Due to the specific nature of the use of AI to process PI, there may also be a need to work with human rights and anti-trust oversight bodies, depending on the matter under review or investigation. The legislation governing these oversight bodies should be revised to facilitate this joint work.

- Privacy and Ombuds laws should be amended to include provisions that facilitate joint compliance work, such as review of Artificial Intelligence Fairness and Privacy Impact Assessments (AIFPIAs) and fairness by design, audits, and joint investigations. This would allow privacy commissioners and Ombuds offices to leverage expertise, including technical, avoid the duplication of efforts, and draw efficiencies.
- Privacy and Ombuds laws need to be amended to facilitate the compliance work necessary to evaluate the use and impacts of processing PI through an ADS including the ability to audit for compliance and implementation of the fairness by design principles.

Other laws governing oversight of human rights and market competition should also be updated to include the ability to collaborate with

privacy commissioners and Ombuds offices on compliance reviews and investigations.

Promoting open, high-quality data

Fair AI critically requires open, high-quality data sets for training and testing AI systems. It also requires the capacity to develop new data sets. Open means that initiatives are taken to make data sets available to the public in easy-to-access formats. The purpose of collection and intended use should be relayed to the public so that they can assess the data set and the planned use of the data set in an informed way. The following should be considered in any regulations or best practices regarding the use of data in the public sector:

- High-quality means a data set is fit-for-purpose and industry-standard anonymization techniques are used to sever a data set from the identity of the data contributor to prevent any future re-identification.
- The ability to continuously develop new data sets is important so that AI systems can be trained and tested on up-to-date data, as this is key to improving the system's predictive or explanatory power.
- Machine learning is data-hungry and deep learning through artificial neural networks is data-ravenous. Those developing deep learning applications will need hundreds of thousands of cases to develop and test new tools. The level of transparency and the quality of training and test data is connected to the quality of outcomes generated by AI.

Public education on AI

Broad and inclusive public education is required to give people a true understanding of what can be done through AI to promote more informed discussions about what rights and interests are affected by the technology, and how society ought to treat people's digital information, rights and interests.

A big part of this is navigating ideas and expectations about what AI *is* doing and realistically *can* do. This kind of discussion provides the background information and conditions needed to begin an informed, society-wide discussion on what AI *should* be permitted to do. It is particularly important to include in this discussion diverse perspectives about how current practices impact people and what specific interests are at stake for particular groups.

As AI expands its social reach, the public will demand better explanations on the specific social and economic impacts and outcomes of AI. Digital literacy and timely and detailed updates on AI initiatives are a start to give people the tools to make informed choices that reflect their interests with respect to technology. Government transparency on AI opens the possibility for an informed critical analysis from well-positioned and legitimate sources (e.g., media, civil society, academia), and this sets the right tone for broader public engagement on the subject.

RECOMMENDATIONS

In this chapter, we provide recommendations for the establishment of a framework to facilitate the responsible development and use of an ADS by government public bodies that use AI to deliver public services.

Recommendations

Public guiding principles of AI

1. Before commencing its first AI project following the release of these recommendations, each public authority should make a public commitment to guiding principles for the use of AI that incorporates transparency, accountability, legality, procedural fairness and protection of privacy. These principles should:
    a. apply to all existing and new programs or activities;
    b. be included in any tendering documents by public authorities for third-party contracts or ADS delivered by service providers; and
    c. be used to assess legacy projects to be brought into compliance within a reasonable timeframe.

Transparency

2. If an ADS is used to make a decision about an individual, public authorities must notify and describe how that system operates to the individual in a way that is understandable.
3. All public authorities designate and identify individuals within the public authority who are responsible for engineering, maintaining, and overseeing the design, operation, testing and updating of any ADS.
4. All ADS should include robust and open auditing functionality with enhanced transparency measures for closed-source, proprietary datasets used to develop and update any ADS.
5. Wherever possible, public authorities should use synthetic or de-identified data in any ADS.

Capacity-building and public engagement

6. To promote capacity building, co-operation, and public engagement, government must:
    a. Undertake public education initiatives to improve general knowledge of the impact of AI and other emerging technologies on the public, on organizations that serve the public, their stakeholders, and their routine service delivery;
    b. Build subject-matter knowledge and expertise on AI across government ministries;
    c. Build capacity to support knowledge sharing and expertise between government and AI developers and vendors;

d. Build the capacity to develop open-source, high-quality data sets for training and testing ADS; and

e. Build capacity for ongoing training of ADS administrators.

Scrutiny and oversight

7. Privacy legislation be amended to include:

a. A requirement that all public authorities complete and submit an Artificial Intelligence Fairness and Privacy Impact Assessment (AIFPIA) for all existing and future AI programs for review to the relevant oversight body;

b. The right to notification that ADS is used, an explanation of the reasons and criteria used, and the ability to object to the use of ADS;

c. The explicit inclusion of service providers to the same obligations as public authorities;

d. Stronger enforcement powers in both the public and private sector including adequate authority for oversight bodies to review AIFPIAs, investigate non-compliance, make binding orders, issue appropriate fines, and publicly report findings;

e. Special rules or restrictions for the processing of highly sensitive information by ADS; and

f. Shorter legislative review periods of 4 years.

8. Legislation be reviewed to ensure oversight bodies are able to review AIFPIAs and conduct investigations regarding the use of ADS alone or in collaboration with other oversight bodies.

GLOSSARY

Algorithm: A procedure or set of instructions for transforming informational input into output.

Artificial Intelligence (AI) (as an adjective): Any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.[178]

Artificial Intelligence (AI) (as a noun; academic discipline): The science and engineering of making intelligent machines.[179]

Artificial Intelligence (AI) model: A mathematical algorithm that is trained to reason over data and replicate a decision-making process and/or decision that an expert human decision-maker in the same conditions would make.

Automated decision system (ADS): Any technology that assists or replaces the judgement of human decision-makers using techniques such as rules-based systems, regression analysis, predictive analytics, machine learning, deep learning and neural nets.[180]

Big data: A term describing the development or existence of many large and comprehensive data repositories. Big data is characterized by both the volume and speed of the creation of new data.

Biometrics: The technology of measuring, analyzing, and processing the digital representations of unique biological data and behavioral traits such as fingerprints, eye retinas, irises, voice and facial patterns, gaits, body odours and hand geometry.[181] More colloquially, biometrics refers to the measurement of life.

Bug bounty program: A program set up by an organization that awards a reward (usually money) to individuals who report flaws in product designs to the organization. Reports are vetted and if a bug is determined to meet preset criteria, a reward is given proportionate to the severity of the bug found. Bug bounty programs are mostly applied in software development but may also be used to harden other types of engineering or process designs.

---

[178] David Poole, Alan Mackworth & Randy Goebel, *Computational Intelligence: A Logical Approach* (New York: OUP 1998).

[179] John McCarthy, "What is Artificial Intelligence?" *Stanford University* (12 November 2007) online: <http://www-formal.stanford.edu/jmc/whatisai.pdf>.

[180] Bill C-11, *An Act to enact the Consumer Privacy Protection Act and the Personal Information and Data Protection Tribunal Act and to make consequential and related amendments to other Acts*, 2nd Sess, 43rd Parl, 2020, cl 2 of the *Consumer Privacy Protection Act* (as introduced at First Reading by the House of Commons on 17 November 2020).

[181] Investigation Report F12-01, para 35 citing Btihaj Ajana, "Recombinant Identities: Biometrics and Narrative Bioethics" Bioethical Inquiry (2010) at 238.

Deep learning (DL) and Artificial Neural Networks (ANN): Deep Learning (DL) is a subset of machine learning (see definition below) which uses techniques that simulate human neural networks known as Artificial Neural Networks (ANN). DL refers to learning through an ANN determined by several layers of neurons that divide the network into distinct stages of calculation.[182]

Explainability: The degree to which the internal processes of an AI system or the methods or techniques used in the application of that system can be described in human terms.[183]

Facial recognition technology (FRT): A technology that can identify or authenticate individuals by comparing their facial features with a database of known faces and looking for a match.

Human-on-the-loop approach to AI: Refers to the capability for human intervention during the AI design cycle and human monitoring of AI while operational to ensure it is performing as anticipated.

Human-in-command approach to AI: Refers to the capability to retain control over AI at all times to be in command of how AI is used in our everyday lives by establishing what tasks or responsibilities we transfer to AI and ensuring the ability to override a decision made by AI.

Interpretability: The degree to which a causal relationship within an AI system can be observed and measured to inform decisions or predictions about that system.[184]

Learning: In the field of AI, learning generally refers to the characteristics of a system that enables the continual and autonomous adjustment and optimization of its programming parameters, due to repeated exposure to example data when directed towards specific tasks.[185]

Machine Learning (ML) algorithms: Algorithms that are developed and optimized through the statistical analysis of large datasets of historical examples.[186] AI systems powered by ML algorithms typically 'learn' through the continual adjustment of mathematical parameters and data retention and error correction techniques to optimize their performance at various prediction, classification, and decision-making, among other tasks.[187]

Personal information: Information about an identifiable individual.

Risk Assessment Instrument (RAI): An actuarial instrument used to predict the likelihood of an outcome of interest. For example, RAIs that purport to predict an offender's likelihood of re-offending have been used by judges in probation, sentencing and parole decisions.[188]

[182] Juergen Schmidhüber, "Deep learning in neural networkers: An overview" (2015) 61 Neural Networks [available online: <https://arxiv.org/abs/1404.7828>].
[183] Leilani H. Gilpin et al, "Explaining Explanations: An Overview of Interpretability of Machine Learning" (2019) MIT Computer Science and AI Laboratory.
[184] Diogo V. Carvalho, Eduardo M. Periera & Jaime S. Cardozo, "Machine Learning Interpretability: A Survey on Methods and Metrics" (2019) 8:8 Electronics at 5-7.
[185] Robert Dale, "Law and Order: NLP in Legal Tech" (2019) 25:1 Natural Language Engineering at 211-212; Simon Deakin & Christopher Markou, "From Rule of Law to Legal Singularity" in Simon Deakin & Christopher Markou, eds, Is Law Computable? Critical Perspectives on Law + Artificial Intelligence (Hart 2020) at 2 and 35; Pedro Domingos, "A Few Useful Things to Know About Machine Learning" (2012) 55:10 Communications of the ACM at 82-83.
[186] David Spiegelhalter, *The Art of Statistics: Learning from Data* (Pelican, 2019) at 144.
[187] Charu C. Aggarwal, *Data Mining: The Textbook* (Springer 2015) at 1.
[188] Jay P. Singh et al, "International perspectives on the practical application of violence risk assessment: A global survey of 44 countries" (2014) 13:3 International Journal of Forensic Mental Health.

Social Credit System (SCS): System of positive and negative reinforcement intended to foster a citizenry that continually engages in self-monitoring and adjustment of its behaviour.[189] This system can gather "social credit" information from commercial enterprises, different levels of government and even directly from individuals to penalize activities and behaviours that are deemed trust-breaking and reward those that are deemed trust-keeping. SCS is supported by AI powered surveillance infrastructure, actuarial assessment instruments, among other social management tools.

Synthetic data: A type of anonymized data used as a filter for information that would otherwise compromise the confidentiality of certain aspects of data. Personal information is removed by a process of synthesis, ensuring the data retains its statistical significance. To create synthetic data, techniques from both the fields of cryptography and statistics are used to render data safe against current re-identification attacks.

Training and test data: In the context of AI systems, training data is used to build up and improve the performance of the system. A test set is used to evaluate the AI system built. Typically, a single dataset is divided into a training set and test set.

[189] Samantha Hoffman, "Programming China: The Communist Party's autonomic approach to managing state security" (2017) 44 Merics China Monitor at 1-12.

OMBUDSPERSON
BRITISH COLUMBIA

PO Box 9039 Stn Prov Govt | Victoria, BC V8W 9A5
Call: 250-387-5855 (Victoria) or 1-800-567-3247 (Rest of BC) | Fax: 250-387-0198
Email: info@bcombudsperson.ca
bcombudsperson.ca

PO Box 9038 Stn Prov Govt | Victoria, BC V8W 9A4
Call: 250-387-5629 | Toll-free in BC: 1-800-663-7867
Email: info@oipc.bc.ca
oipc.bc.ca

3162 Third Avenue, Main Floor | Whitehorse, Yukon Y1A 1G3
Call: 867-667-8468 | Toll-free: 1-800-661-0408 ext. 8468 | Fax: 867-667-8469
Email: info@yukonombudsman.ca
yukonombudsman.ca