CANADIAN CIVIL LIBERTIES ASSOCIATION

ASSOCIATION CANADIENNE DES LIBERTES CIVILES

Consultation Submission regarding Ontario's Trustworthy Artificial Intelligence Framework

June 4, 2021

Canadian Civil Liberties Association
90 Eglinton Ave. E., Suite 900
Toronto, ON M4P 2Y3
Phone: 416-363-0321 x 253

www.ccla.org

The Canadian Civil Liberties Association (CCLA)

The Canadian Civil Liberties Association ("CCLA") is an independent, national, nongovernmental organization that was founded in 1964 with a mandate to defend and foster the civil liberties, human rights, and democratic freedoms of all people across Canada. Our work encompasses advocacy, research, and litigation related to the rights to privacy, freedom of expression, freedom of association, and equality rights, all rights which are potentially engaged within the scope of artificial intelligence applications that have been or may be implemented in both the public and private sector. The range of rights potentially engaged necessitates a strong, effective, human-focused, and enforced legally-binding framework for AI tools spanning the full lifecycle, including conception, design, implementation, use and assessment.

Overview and comments regarding the consultation framework

This submission is the CCLA's response to the Ontario consultation to develop an appropriate framework to guide accountable, safe and rights-based use artificial intelligence in the province.

CCLA notes that Artificial Intelligence, or "AI" is a term that serves as colloquial shorthand for a diverse range of technological systems, and while we have used the generic term throughout, we recognize that there are different affordances in different models and systems that ultimately would necessitate nuance in definition and analysis of the accountability and transparency provisions, the trustworthiness and/or safety of any given system, and the privacy impacts of AI. While it makes sense for a framework to be inclusive and expansive, it would be beneficial to explore a wide range of scenarios as part of your development process to allow discussions at a suitable level of granularity to ensure that the full range of current and anticipated AI applications can be reasonably addressed in this principled approach and to explore whether additional elements are required for enhanced and fulsome protections for Ontarians.

Ultimately, a principles-based framework is a significant first step, but for it to be effective it will also require a whole-of-government approach to mandatory compliance and effective enforcement provisions, integrated into the requisite approval processes, procurement, development and design procedures, and control systems of provincial government institutions.

CCLA would also highlight that the proposed made-in-Ontario private sector data protection statute, if it moves forward, should include appropriate statutory protections relevant to AI. There will also very likely be a need for corresponding amendments to the relevant public sector laws to provide the legal foundation required to support the framework under discussion. In particular it will be imperative that appropriate privacy protection be provided for de-identified data for genuinely trustworthy AI development and use.

Our submissions align with the three commitments addressed in the consultation. Under each commitment, we have included discussion of some of the necessary steps to bring these commitments to life. We have not ranked the potential actions as part of this submission, as requested in the survey; it is our position that the actions proposed are all necessary, along with more. Taking commitment one as an example ("No AI in secret"), we would contend that transparency at the stage of data collection is as important as transparency in use, which are both essential in order to create accountability for potential bias. Allowing any of those actions to drop because the survey data shows it ranked "2" or "3" would undermine the completeness and potential efficacy of your framework.

Commitment 1: No AI in Secret

Be fully transparent when using algorithms to interact with the public (e.g. rules to require the public be informed if they are interacting with a machine or have decisions made about them by an algorithm)

CCLA agrees that being fully transparent when using algorithms to interact with the public is essential to ensure that uses of automated-decision making are fair and appropriate. Without knowledge that an automated decision is being made, there can be no meaningful consent to its use, and no meaningful recourse should an individual wish to challenge that decision based on either bias (discussed further below) or correctness. To give effect to this principle, there will need to be clear rules, supported ultimately in legislation, that specifies not just notice that an automated decision-making system is in place, but ensures that such notice is prominent, clear, and includes information regarding avenues for complaints or appeals regarding the decisions.[1]

Create accountability for the use of AI in the government by giving people rights to address potential biases created by the AI (e.g. right to explainability, right to contest, and right to opt out)

CCLA believes that Ontarians must have a meaningful right to address potential biases created by AI systems. The negative effects of automated processing have the potential to cause real harm: algorithms are not neutral and often import the biases of their designers or encourage

[1] This is similar to the Treasury Board's guideline for AI use, that the Assistant Deputy Minister is responsible for "[p]roviding notice through all service delivery channels in use that the decision rendered will be undertaken in whole or in part by an Automated Decision System as prescribed in Appendix C," and "[p]roviding notices prominently and in plain language, pursuant to the Canada.ca Content Style Guide." See: "Directive on Automated Decision-Making," *Government of Canada Website* at 6.2 (date modified: 1 April 2021), online at: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

existing discrimination.[2] CCLA supports a right to explanation and increased transparency when individuals interact with or are subject to automated decision-making. In practice, this would require any company or government body using AI to process personal data to clearly explain where automated processing has been used, the logic behind the decisions of the automated processing, and verify or ideally publish some version of any required or voluntary assessments. These may, depending on the application, include a Privacy Impact Assessment, AI Impact Assessment, and, for particularly sensitive systems, a Human Rights Impact Assessment.[3]

Providing Ontarians this right would increase public awareness of AI decision-making, in addition to enhancing transparency and intelligibility. It will also serve as a form of potential check on the purposes of automated decision making, to the extent that the descriptions are a meaningful representation of the working of the systems. It appropriately places the burden of ensuring that those processes do not have a discriminatory impact, and explaining how bias is avoided, on the public or private body using the algorithm. Public trust in algorithmic decision-making would also potentially be enhanced, provided that the explanations are indeed complete and meaningful, since individuals could be confident that they knew what factors went into each decision-making process.

Given the opacity and potential discriminatory effects of automated decision-making, CCLA also supports a legal right to object to automated decision-making and to be free from such decision-making, subject to limited exceptions. Included in that right should be the right to request human intervention, to contest any automated decision that has been taken, and the opportunity to express the objector's point of view on the automated decision. Similarly, there should be a legal right to opt out of automated processing, including profiling, without having to actively object. Exceptions to that right might include situations where explicit meaningful consent has been obtained, where an automated decision is necessary for a contract that was freely entered into, or when automated decision-making is prescribed by law.

Provide clarity and transparency to the public on how Ontario collects data for use in algorithms (e.g. explore options to update provincial notices of collection to inform the public if data collected is used to develop algorithms for decision-making)

[2] See, for e.g., Gideon Mann, & Gideon; Cathay O'Neill "Hiring Algorithms Are Not Neutral." *Harvard Business Review* (9 December 2016), online at: https://hbr.org/2016/12/hiring-algorithms-are-not-neutral.

[3] The Treasury Board of Canada's directive on Automated Decision Making has a publishing requirement that makes government departments responsible for "[p]ublishing information on the effectiveness and efficiency of the Automated Decision Systems in meeting program objectives on a website or service designated by the Treasury Board of Canada." See: "Directive on Automated Decision-Making," *supra* note 1 at 6.5.1.
The publication of the final results of the AI impact assessment must be "in an accessible format and in both official languages on the Open Government portal." See: "Algorithmic Impact Assessment Tool," Government of Canada Website (last modified 1 April 2021) online at: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html.

CCLA supports providing clarity and transparency to the public on how Ontario collects data for use in its algorithms. As technology advances, government's increasingly have more access to personal information and data, which has resulted in a tension between government transparency and the protection of the public's right to privacy. Minimizing the amount of personal information that the government of Ontario collects, uses, and retains is essential to ensuring that Ontarians have trust in government practices.

Further, transparency on how and when government departments have access to personal data is necessary if consent is going to be meaningful. This is particularly important in the context of "breaking down information siloes" across government ministries or departments, a process often identified as necessary to facilitate appropriate training of AI-enabled applications, but which carries inherent privacy risks and runs contrary to the privacy principle of purpose specification. As a general guideline, no information that a resident of Ontario is obliged to share with the province in order to receive a benefit or entitlement, or in other mandatory contexts, should be repurposed absent meaningful consent to "opt-in" to such uses.

Commitment 2: AI Use Ontarians Can Trust

Deliver recommendations on ways to update Ontario's rules, laws and guidance to strengthen the governance of AI, including whether to adopt a risk-based approach to determine when which rules apply.

It is obviously important to update rules, laws and guidance to strengthen the governance of AI and such recommendations must be carefully considered and widely consulted on with a range of stakeholders.

CCLA takes the position that an appropriate legal privacy/data protection framework is necessary; while rules and guidance can be helpful, they will be inadequate without the force of law behind them and should be considered a support rather than a substitute for legislation. A supportive policy framework should include consideration of appropriate rules for the procurement, implementation, and use of AI in a provincial context. This must include enforceable whole-of-government compliance mechanisms.

Bias and discrimination in automated decision-making is well documented, as AI systems routinely replicate existing discrimination and tend to have differential impacts on racialized

individuals.[4] When systemic biases permeate data sets, biases become embedded in and perpetuated by the algorithm, which further impacts individuals and communities who have been the subject of historic discrimination. Given the potential adverse and discriminatory effects of automated-decision making,[5] CCLA supports a risk-based approach to the governance of AI that is both proportionate and effective. A risk-based regulatory approach to AI was recently adopted by the European Union ("EU") in the *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence*.[6] This approach imposes specific regulatory restrictions and safeguards that differentiate between uses of AI that create unacceptable risks, high risks, and low or minimal risks, with a particular focus on the impact that AI systems have on rights, safety, and health. In a similar vein, CCLA advocates for a risk-based approach to AI that imposes regulatory burdens to AI systems that are proportionate to the potential risks that any given AI system generates. This includes taking into account the risk and threats that AI systems pose to health and safety and to the fundamental human rights of Ontarians.

Assess whether to use an algorithmic assessment tool as a way to measure risk, security, and quality.

CCLA agrees that AI systems must be assessed for risk, security, and quality. To implement transparency around AI decision-making and ensure that AI systems are not discriminatory or bias, AI systems require regulatory mechanisms that eliminate or reduce the risks associated with reliance on automated decision-making. CCLA supports the implementation of mandatory AI impact assessments, which will mitigate risk in a structured manner while enhancing trustworthiness to the extent that such assessments are demonstrably thorough and publicly accessible.

Recently, the Treasury Board of Canada implemented a mandatory algorithmic risk assessment tool. This tool assesses the risk level of an automated decision-system, including consideration of the capabilities of the systems design; transparency of the algorithm; classification of the automated decision; the impact the automated decision has on freedom, health, economy or environment; and the data source and type.[7] Similarly, the EU Proposed Regulations provide a

[4] See: Jacquelyn Burkell, "The Challenges of Algorithmic Bias," (working paper) *Law Society of Ontario Special Lectures*, Ontario: University of Western (2019), online (pdf): https://ajcact.openum.ca/files/sites/160/2020/08/The-Challenges-of-Algorithmic-Bias-.pdf.
[5] See, for e.g., Hansa Srinivasan, "ML-fairness-gym: A Tool for Exploring Long-Term Impacts of Machine Learning Systems," *Google AI Blog* (5 February 2020), online at: https://ai.googleblog.com/2020/02/ml-fairness-gym-tool-for-exploring-long.html; Andrew Burt, "How to Fight Discrimination in AI" *Harvard Business Review* (28 August 2020), online at: https://hbr.org/2020/08/how-to-fight-discrimination-in-ai.
[6] European Commission, *Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* (21 April 2021), 2021/0106 (COD) online at: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence [EU Proposed Regulations].
[7] "Algorithmic Impact Assessment Tool," *supra* note 3.

robust framework for assessing the risks associated with algorithmic decision making and ensuring that AI systems are in compliance with the legal requirements, including the implementation of risk management systems; data and data governance; documentation and record keeping; transparency; human oversight; robustness, accuracy, and security.[8]

AI Impact Assessments should be a legal precondition for the use of AI systems, should encompass human rights protections, and should be published and publicly accessible.

Ensuring processes are in place so that algorithms are continuously tested and evaluated for bias/risk and whether audits or human oversight controls are needed.

CCLA takes the position that ensuring that processes are in place so that algorithms are continuously tested and evaluated for bias and risk is essential. To give effect to this principle, audits and human oversight controls are necessary and must be implemented in the design, development, implementation, and operational phases of the AI system. In addition, CCLA advocates for a regulatory and enforcement regime that will provide accountability for compliance with these tests and evaluations, including consequences for non-compliance.

CCLA takes the position that routine audits should be mandatory and human oversight controls are often necessary to mitigate potential threats and biases that AI systems pose to human rights. The EU Proposed Regulations provide a framework that requires human oversight for high-risk AI systems.[9] This includes ensuring that human oversight is either identified and built into the high-risk AI system; or identified by the provider of the high-risk AI system and that are to be implemented by the user.[10] CCLA adopts this approach and advocates for the continued human

[8] EU Proposed Regulations, *supra* note 6 at Chapter 2.
[9] EU Proposed Regulations, *supra* note 6 at Article 14.
[10] See: EU Proposed Regulations, *supra* note 6 at Article 14(3). The EU Proposed Regulations also require that the AI system enables individuals to whom human oversight is assigned "to do the following, as appropriate to the circumstances: (a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible; (b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons; (c) be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available; (d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system; (e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a "stop" button or a similar procedure. 5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons."

oversight of AI systems for such systems for the period that the algorithm is in use, in order to mitigate the risks that these systems pose to safety, health, and human rights.

Given that automated decision-making has the potential to generate arbitrary and/or discriminatory results, human oversight is required to ensure that the decisions being made are reasonable and free from discrimination, particularly in novel situations.[11] Human intervention also serves a security function in helping to detect attempts at manipulation: "at present, "data poisoning" and adversarial examples represent ways for malicious actors to exploit AI's inability to think contextually".[12]

Commitment 3: AI That Serves All Ontarians

Embed equity and inclusion in the use of data and digital tools by requiring organizations to take steps to mitigate potential harms (e.g. data set requirements, documentation requirements for traceability, accountability provisions)

Ontarians cannot trust AI unless equity and inclusion are embedded in the use of data and digital tools and organizations must be required to take steps to mitigate discriminatory impacts due to bias throughout the AI system lifecycle. There is widespread recognition of the necessity to include data set requirements, documentation requirements for traceability, and accountability requirements in regulations regarding AI systems for this reason.[13]

As noted under "Commitment 2: above, AI systems are routinely criticized for having biases within their code or emerging from their training data sets which ultimately leads to discriminatory results. These systemic biases are often unintentionally embedded in AI systems, nevertheless these biases are harmful and have a disproportionate impact on vulnerable and marginalized populations.[14] Equally important, although less often discussed, is the risk that discriminatory impacts can emerge from decisions before the system is rendered into code and trained, at the point of specifying the nature and expectations for the system. In her book

[11] Will Douglas Heaven, "Our weird behaviour during the pandemic is messing with AI models," *MIT Technology Review* (11 May 2020) online at: https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/.
[12] Robert Mazzolin, "Artificial Intelligence and Keeping Humans 'in the Loop,'" *Centre for International Governance Innovation* (23 November 2020) online at: https://www.cigionline.org/articles/artificial-intelligence-and-keeping-humans-loop/; Heaven, *ibid*.
[13] See: Ignacio Cofone, "Policy Proposals for PIPEDA Reform to Address Artificial Intelligence Report," *Office of the Privacy Commissioner of Canada* (November 2020), online at: https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/pol-ai_202011/.
[14] See: Susie Lindsay, Jesse Beatson, & Nye Thomas, "Legal Issues and Government AI Development," *Law Commission of Ontario* (15 March 2021) online at: https://www.lco-cdo.org/wp-content/uploads/2021/03/LCO-Govt-AI-Workshop-Report-%E2%80%94-March-2021.pdf.

"Automating Inequality," Virgina Eubanks documents a range of systems which were inherently flawed because they worked exactly as they were conceived—but the initial conceptions prioritized goals such as "efficiency" in systems ultimately meant to support vulnerable people, leading in some cases to devastating impacts including loss of essential medical supports.[15] The discriminatory effects of AI systems are often compounded by the lack of transparency in automated processing – AI decision-making has often been described as a "black box" that even the AI's designers may not be able to explain.[16]

All bodies, public or private sector, who create, implement, and use AI systems must be subject to compliance obligations that require documentation of the processes and decisions made during the design, development, deployment, and operational phases of AI system lifespan**.** For transparency to meaningful, it is essential that when a decision is made by an AI system that system creator is able to "explain" or answer how it made its determinations, the processes that were involved, and ultimately why the decision was made. There may be resistance to transparency and the right to an explanation based on the argument that algorithms are highly valuable forms of intellectual property ("IP") that should remain proprietary. However, there is widening agreement that the kind of descriptions necessary for transparency should be able to be produced with due consideration for IP.[17]

Engage with sector leaders and civil society to develop a standard for "trustworthy AI" and a process to certify that vendors are meeting the government's standard

CCLA supports the principle that sector leaders and civil society should be included in the development of a standard for "trustworthy AI" and the development of a process to certify that

[15] Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018).
[16] See, for e.g., Bathaee, Yavar. "The Artificial Intelligence Black Box and the Failure of Intent and Creation" *Harvard Journal of Law & Technology,* Volume 31, Number 2 Spring 2018, available at: https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathaee.pdf.
[17] See: Carvalho, Diogo V, Eduardo M Pereira & Jaime S Cardoso. "Machine Learning Interpretability: A Survey on Methods and Metrics" (2019) 8:8 Electronics 832 at 8, online at: <http://dx.doi.org/10.3390/electronics8080832>; Kartik Hosanagar and Vivian Jair, "We Need Transparency in Algorithms, But Too Much Can Backfire," *Harvard Business Review* (25 July 2018), online at: https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire; Joel Nantais, " Transparency in Government AI," *Towards Data Science* (3 August 2019) online at: https://towardsdatascience.com/transparency-in-government-ai-7c871a9cc219. See also: EU Proposed Regulations, *supra* note 6 at 11. The EU Proposed Regulations note that "increased transparency obligations will also not disproportionately affect the right to protection of intellectual property (Article 17(2)), since they will be limited only to the minimum necessary information for individuals to exercise their right to an effective remedy and to the necessary transparency towards supervision and enforcement authorities, in line with their mandates [. . .] [w]hen public authorities and notified bodies need to be given access to confidential information or source code to examine compliance with substantial obligations, they are placed under binding confidentiality obligations."

vendors are meeting those government standards. At the same time, it is necessary to note that civil society actors have considerably fewer resources and greater capacity restrictions than industry sector leaders; care must be taken to ensure civil society participants are supported to participate in a standards process, which typically is lengthy, time-consuming, and granular, thus requiring a degree of immersion in the topic that is resource intensive.

It is also important to caution that standards and compliance regimes that are created by those who will ultimately be governed by them are often weaker than they should be and standards of self-certification are insufficient. Therefore, while CCLA appreciates the value of standards and certification processes, to the extent that they are openly developed and publicly accountable, we ultimately advocate for protections also embedded in statue, not standards alone.

Assess whether the government should prohibit the use of AI in certain use cases where vulnerable populations are at an extremely high risk

CCLA believes that the government should prohibit the use of AI in certain use cases where vulnerable populations are at an extremely high risk, however the threshold for what constitutes "extremely high risk" will require consultation with Indigenous, Black, and other communities who regularly experience systemic discrimination in order to set that boundary appropriately, given the serious threats that automated-decision making may pose to vulnerable populations.

Particularly relevant to this discussion is the risk-based approach adopted by the EU Proposed Regulations. Central to the EU's approach to AI governance is a prohibition on harmful AI practices that contravene fundamental EU values.[18] Similarly, the CCLA advocates for the adoption of a risk-based approach that prohibits AI practices that contravene Canada's *Charter*[19] values, including privacy, freedom of expression, democratic freedoms, freedom of assembly, and freedom from discrimination. CCLA proposes a probation on AI practices that have the potential to manipulate persons or exploit specific vulnerable populations in a manner that is likely to cause harm.[20] CCLA also advocates for the prohibition on other manipulative or

[18] EU Proposed Regulations, *supra* note 6 at 11.
[19] *Canadian Charter of Rights and Freedoms*, Part I of the *Constitution Act, 1982*, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11 [*Charter*].
[20] EU Proposed Regulations, *supra* note_ at 13-14. The EU Proposed Regulations prohibit any "AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm;" "AI system that exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm;" "AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading to either or both of the following: (i) detrimental or unfavourable treatment of certain natural persons or

exploitative practices based on AI, including law enforcements use of facial recognition software for mass surveillance.[21]

CCLA is thankful for the opportunity to make submissions on this important topic. Should additional explanation be deemed helpful in your process, we will be happy to discuss these matters further.

Brenda McPhail, PhD
Director, Privacy, Technology & Surveillance

Leslie Schumacher
JD Candidate 2022, *Windsor Law*

whole groups thereof in social contexts which are unrelated to the contexts in which the data was originally generated or collected; (ii) detrimental or unfavourable treatment of certain natural persons or whole groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;" "the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives:
(i) the targeted search for specific potential victims of crime, including missing children; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; (iii) the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence […]" See: Article 5.

[21] Of particular concern is Canadian law enforcement's use of Clearview AI, a company that uses AI technology to match faces to a database of over three billion facial images scraped from the internet. Clearview AI is a clear example of how AI can be used to threaten human rights with deadly effectiveness. See: Kashmir Hill "The Secretive Company That Might End Privacy as We Know It," *The New York Times* (18 January 2020), online at: https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html; David Burke, "Use of facial recognition technology by police growing in Canada, as privacy laws lag," *CBC News* (10 February 2020), online at: https://www.cbc.ca/news/canada/nova-scotia/facial-recognition-police-privacy-laws-1.5452749.