Information and Privacy
Commissioner of Ontario
Commissaire à l'information et à la
protection de la vie privée de l'Ontario

VIA ELECTRONIC MAIL

December 20, 2021

Dr. Dubi Kanengisser
Senior Advisor, Strategic Analysis and Governance
Toronto Police Services Board
40 College Street
Toronto, ON M5G 2J3

Dear Dr. Kanengisser:

RE: The Toronto Police Services Board's Public Consultation Regarding its *Use of New Artificial Intelligence Technologies Policy*

On behalf of the Office of the Information and Privacy Commissioner of Ontario (IPC), I am pleased to provide our comments on the Toronto Police Services Board's (Board) draft Use of New Artificial Intelligence Technologies Policy (Policy). We appreciate that the Board has recognized the need for strong governance and oversight of technologies that utilize or integrate artificial intelligence (AI), and commend the Board for taking the necessary step of inviting comment from the public.

The recommendations that follow are intended to clarify and strengthen the Policy and help the Board enhance its ability to govern the use of AI in a manner that protects privacy, freedom of information, and other fundamental rights. As the Policy is designed to apply to a wide range of current and potential future AI applications and technologies, our comments are broad — focusing on key issues, critical principles, and recommended next steps. These comments represent the continued evolution of my office's approach to AI, given the dynamic nature of the field.

The IPC will continue to work with stakeholders to promote the responsible governance of AI in a policing context as part of our strategic priority related to *Next Generation Law Enforcement*. We would welcome further collaborative engagement with the Board and the Toronto Police Service on the Policy and associated procedures.

In the interest of transparency to the people of Ontario, this letter and attachments will be posted on our website.

Sincerely,

Patricia Kosseim
Commissioner

Summary of IPC Recommendations to the Board

Clarifying the focus of the Policy

The Policy would be more effective with clear definitions and a clear scope. This clarity would help the Toronto Police Service (Service) to consistently interpret, apply and comply with the Policy, and help the public to understand the extent of the safeguards created by the Policy.

1. Provide a short, clear definition of AI in plain language.

2. Clarify the definition of bias to focus on outcomes, rather than the cause of a biased output.

3. Include a section related to scope with criteria and examples about which types of AI Technologies and uses are in-scope and which are out of scope, and explicitly extend the application of the Policy to the use of AI Technologies (including free trials) by individual Service members in the course of their work.

4. Explicitly include in the Policy's scope of application AI that relies on de-identified information or information about groups or communities.

5. Clearly require the Service to assess AI Technologies that are already in use by the Service against the Policy's risk framework.

6. Regarding the risk assessment framework:
   a) Clarify that the risk categories to be included in the risk rating scheme should be based on specific categories of harms, their severity and likelihood, rather than applications;
   b) Require that risk assessments focus on the context in which AI technologies will be used, including how multiple AI systems may be used by the Service in ways that cumulatively may lead to new risks; and
   c) Require that the stakeholders involved in the development of the risk rating scheme include members of impacted communities.

Strengthening transparency, accountability and oversight

The Policy includes mechanisms designed to create transparency, accountability and oversight of the Service's use of AI. However, we recommend supplementary measures be adopted.

7. Require a whistleblower mechanism for Service members to report violations of the Policy to the Board anonymously and securely.

8. Require a clear description of roles and responsibilities for conducting and participating in risk assessments under the Policy and direct the Service to ensure that there is interdisciplinary expertise and diverse perspectives involved, as well as reasonable time and resources to contribute to a robust assessment process.

9. Require recordkeeping requirements for AI, having regard to the retention of personal information used by AI Technologies and a duty to document the activities of the Service involving AI.

10. Broaden the information disclosed in the public AI website based on recommendations made in the Law Commission of Ontario's *Regulating AI: Critical Issues and Choices* and the UK government's *Algorithmic Transparency Standard*. This should include proactive disclosure of algorithmic impact assessments, subject to legitimate exemptions under the *Municipal Freedom of Information and Protection of Privacy Act*, or at the very least, summaries of such assessments.

Putting the Policy into practice

The Policy requires the Service to develop numerous processes and procedures to support the Policy's risk management framework. As for human-in-the-loop oversight and AI explainability, we recommend the following:

11. Develop requirements relating to human-in-the-loop processes that include:

    a) A process to determine which aspects of decision-making should be automated and which should be subject to human review. This process should assess:

       i. The role of discretion in the decision;

       ii. The accuracy, performance, and robustness of the AI system being considered; and

       iii. The consequences of an incorrect decision.

    b) Ensuring humans-in-the-loop have the time, resources, information, and capacity needed to effectively review automated decisions or predictions.

    c) Logging and recordkeeping requirements for human review and intervention activities.

    d) Including human oversight metrics among the indicators that must be reported back to the Board.

12. Consider explainability as a mitigating factor and lack of explainability as an aggravating factor when assessing potential harms in the risk assessment process, and set out a certain baseline level of explainability requirements that must be met for all levels of risk.

Moving the Policy forward

While the Policy is an important first step toward effective AI governance, many others steps must follow. Continued engagement with the public and the IPC will be essential.

13. Continue to engage a broad range of public stakeholders, including the IPC, as the Board works towards finalization of its Policy.

14. Ensure that the Service engages an equally broad range of stakeholders, including the IPC, in developing processes and procedures to implement the Policy.

15. Require that criteria be developed to establish the circumstances in which a public consultation should be included as part of the pre-deployment risk assessment process.

IPC COMMENTS ON TORONTO POLICE SERVICES BOARD'S *USE OF NEW ARTIFICIAL INTELLIGENCE TECHNOLOGIES POLICY*

The Office of the Information and Privacy Commissioner of Ontario (IPC) is pleased to provide our comments on the Toronto Police Services Board's (Board) draft Use of New Artificial Intelligence Technologies Policy (Policy).

The IPC provides independent oversight of organizations under the *Municipal Freedom of Information and Protection of Privacy Act* (MFIPPA). That law mandates the IPC to investigate complaints, resolve appeals, review privacy and information management practices, conduct research, and offer guidance and comment on municipal government initiatives as they relate to access, privacy, and by extension transparency and accountability related to information.

This submission builds on our early artificial intelligence (AI) related work, including our:

a) Comments on the Ontario government's consultation on *Ontario's Trustworthy Artificial Intelligence Framework* (Trustworthy AI Comments);

b) Staff-level comments provided to the Board on an earlier draft of the Policy in June 2021;

c) Contributions to the *Draft privacy guidance on facial recognition for police agencies* (Draft FRT Guidance), jointly developed by privacy protection authorities across Canada;

d) *Model Governance Framework for Police Body-worn Camera Programs in Ontario*, which was informed by an in-depth consultation with the Board and the Service; and

e) Comments on the Ontario government's White Paper *on Modernizing Privacy in Ontario* (Private Sector Privacy Comments).

The wider context

Police use of AI has the potential to improve public safety and support data-driven innovation in police services. However, police use of AI also raises important privacy, transparency and accountability-related issues, many of which are set out in our comments below.

The Board's consultation on its Policy comes at a critical juncture for public trust in both AI and policing. A 2021 global survey by Edelman found that only 39% of Canadians believe artificial intelligence technology companies can be trusted to do the right thing, a decline of five percent since 2020.[1] Further, a recent Government of Canada survey found that — more than any sector — Canadians were concerned that AI would lead to negative outcomes for law enforcement.[2] A

[1] Edelman (2021) "2021 Edelman Trust Barometer Tech Sector Report." https://www.edelman.com/sites/g/files/aatuss191/files/2021-03/2021%20Edelman%20Trust%20Barometer%20Tech%20Sector%20Report_0.pdf

[2] Innovation, Science and Economic Development Canada (May 2021). "Views of Canadians on Artificial Intelligence: Final Report." https://ised-isde.canada.ca/site/public-opinion-research/en/views-canadians-artificial-intelligence-final-report

2019 interview-based study of Canadians' views on AI conducted by Deloitte observed that racial and gender bias in AI systems is "dominating public sentiment."[3]

Feelings of distrust and experiences of discrimination associated with policing have also been shared with the Board in surveys and town halls conducted in recent years.[4] It is clear that the stakes are high for the Board, the Toronto Police Service (Service), and the public. AI systems must be operated in a manner that respects privacy, fosters transparency, and advances equity. They should be used to reduce, not exacerbate, existing patterns of discrimination, including in relation to patterns of over policing and under policing.

We appreciate the Board's efforts to prohibit the use of AI Technologies that pose risks which the Board identifies as unacceptable, and to develop a framework and governance model that takes a risk-centric view. This is similar to the approach in the European Union's proposed *Artificial Intelligence Act*, which ties governance requirements to AI risk levels.[5] The Policy's core concept — subjecting higher risk uses of AI to more thorough safeguards — is a reasonable conceptual foundation for governing this rapidly evolving area. However, as our recommendations indicate, further work lies ahead in order for the Policy to fulfill its purpose.

Our comments that follow are divided into four categories:

1. Clarifying the focus of the Policy
2. Strengthening transparency, accountability and oversight
3. Putting the Policy into practice
4. Moving the Policy forward

1. Clarifying the focus of the Policy

As indicated in our Trustworthy AI Comments, an effective AI governance model requires clear definitions and a clear scope. Clarity will help the Service consistently interpret, apply and comply with the Policy. Clarity will also help the public and regulators understand the requirements of the Policy to help inform reasonable standards and expectations.

[3] Deloitte (2019). *Canada's AI Imperative: Overcoming Risks, Building Trust.* https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/deloitte-analytics/ca-overcoming-risks-building-trust-aoda-en.pdf

[4] See, e.g. Toronto Police Services Board (August 202). "'I Don't Want to Live In Fear': Voices from the Toronto Police Services Board Town Hall Meetings – Interim Summary.". https://tpsb.ca/consultations-and-publications/publications-list/send/2-publications/633-town-hall-interim-summary; Toronto Police Services Board (2019). "Perceptions of the Toronto Police and Impact of Rule Changes under Regulation 58/16: A Community Survey." https://tpsb.ca/consultations-and-publications/publications-list/send/2-publications/612-perceptions-of-the-toronto-police-and-impact-of-rule-changes-under-regulation-58-16-a-community-survey

[5] European Commission. "Regulatory framework proposal on artificial intelligence." https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

1.1. Clarify definitions of key terms

Defining “artificial intelligence” (or similar terms, such as automated decision-making) is a difficult task. Even so, the definition adopted by the Board is unnecessarily complicated because it mixes technical characteristics about the technologies with statements about the scope of the Policy. The definition also includes “any goods or services whose procurement, deployment or use require that a privacy impact assessment be conducted”, which potentially brings a wide range of non-AI goods and services within the scope of the Policy.

We recommend that the Board amends the Policy to provide a short, clear definition of AI using plain language.

The policy’s definition of “bias” could be simplified by focusing on the outcome rather than the cause of bias. It may not always be clear when a flawed output is causally related to a flaw in the design of an AI technology or training data. For instance, training data may accurately represent data that was generated in an inequitable situation. Is such training data — or any outputs that emerge from it — ‘flawed’ under this definition? By focusing the definition on the outcome — for example, “output that consistently either misidentifies certain types of subjects or ascribes them with characteristics that disadvantage them based on illegitimate grounds (e.g., [*Ontario Human Rights*] *Code* protected grounds)” — those applying or seeking to understand the Policy will not be limited in calling out bias only when they can demonstrably point to its cause.

We recommend that the Board clarify the Policy’s definition of bias to focus on outcomes, rather than the cause of a biased output.

1.2. Clarify who and what are covered in a dedicated scope section of the Policy

The definition of AI Technology narrows the Policy’s scope by limiting the term to goods and services “which collect or use information about members of the public” and for which decisions are made “pertaining to the information or the members of the public to which it pertains.” As the scope of the Policy relies on the term ‘members of the public’, the Policy should make it clear who is excluded and/or who is included within the meaning of that term. There should be no doubt about whether ‘members of the public’ includes victims of crime, complainants, persons of interest, suspects, people in custody, Service members themselves, or job applicants to the Service.

Ambiguity about the Policy’s scope could lead to uncertainty about whether or not tools such as those to detect altered fingerprints,[6] crack mobile device passcodes,[7] monitor Service member stress levels,[8] or streamline recruiting processes at the Service would be subject to the Policy.

[6] Seffers, G. (Sept 8, 2020). “FBI Upgrades Biometric Technologies.” *AFCEA Signal*. https://www.afcea.org/content/fbi-upgrades-biometric-technologies

[7] O’Kane, J. (Nov 5, 2021). “RCMP wants to use AI to learn passwords in investigations, but experts warn of privacy risks.” *The Globe and Mail*. https://www.theglobeandmail.com/business/article-rcmps-plan-to-use-ai-to-learn-passwords-in-investigations-has-privacy/

[8] Fussell, S. (May 8, 2019). “The Push to ‘Predict’ Police Shootings.” *The Atlantic*. https://www.theatlantic.com/technology/archive/2019/05/how-machine-learning-can-help-prevent-police-shootings/588937/

The lengthy definition of the term “New AI Technology” outlines five different circumstances in which an AI Technology would be deemed to be ‘new’. These cover never-before-used technologies as well as technologies already in use by the Service that are being considered for a novel use, being augmented by AI, enhanced with additional data sets, or new data sets created for use by an AI Technology. We appreciate the Board’s effort to outline the broad range of situations where new AI might be used. However, a dedicated section of the Policy that explains its scope of application would more clearly communicate which uses of AI are covered by the Policy.

A clear scope is critical to enabling the Board, the Service, and members of the public to understand the limits of this governance framework. Without a clear scope, members of the Service could mistakenly consider a technology out of scope — such as free trials of software offered by companies.[9] Similarly, the public could be surprised to learn that some major AI initiatives go unassessed.[10]

We recommend that the Policy include a section defining its scope. The section should include clear criteria and examples describing what types of AI Technologies and uses are in-scope and what technologies and uses are out of scope. The Policy’s scope of application should explicitly extend to include the use of AI Technologies (including free trials) by individual Service members in the course of their work.

1.3. Ensure de-identified information and group privacy issues are in-scope

AI systems that do not process personal information may still impact the values that privacy seeks to protect. For instance, AI enables inferences made about a small sample of individuals to be generalized to entire communities or groups. Decisions based on AI outputs can be made at a group level that can affect an individual’s autonomy, self-development, and other liberties, all without relying on any identifiable information about a particular individual.[11]

For instance, gunshot detection systems or de-identified crime statistics that feed into AI systems need not rely on personal information or make decisions about identifiable individuals. However, these systems will generate alerts that may direct additional police attention to the communities in which they are located. This in turn may inform decisions that affect patrol routes and community risk levels, bringing certain neighborhoods under more intense police scrutiny. As a result of increased surveillance, police may uncover more crimes among marginalized communities relative to other neighborhoods that are not being surveilled as actively, or may focus unduly on certain types of crimes as opposed to others, such as cyberattacks or white collar crimes.

The Policy includes some consideration for group privacy in its statement of purpose, which establishes an objective to preserve the privacy, rights, and dignity of individuals and communities.

[9] See, e.g., Mac, R. et al (2021). “Police In At Least 24 Countries Have Used Clearview AI. Find Out Which Ones Here.” *Buzzfeed News.* https://www.buzzfeednews.com/article/ryanmac/clearview-ai-international-search-table

[10] See, e.g. Cardoso, T. and Curry, B. (Feb 7 2021) “National Defence skirted federal rules in using artificial intelligence, privacy commissioner says.” *The Globe and Mail*. https://www.theglobeandmail.com/canada/article-national-defence-skirted-federal-rules-in-using-artificial/

[11] See, e.g. Kammourieh, L. et al (2017). “Group Privacy in the Age of Big Data.” *Group Privacy: New Challenges of Data Technologies*. Taylor, L., Floridi, L. & van der Sloot, B., eds. Philosophical Studies Series, Springer. p.p. 37-66.

However, the Policy's definition of AI Technology does not make clear if de-identified information is in scope. Nor is it clear whether information about one group of individuals which is analyzed to make decisions about another individual, group of individuals, or a whole community, is in scope.

We recommend that the Policy explicitly includes in its scope of application AI that relies on de-identified information or information about groups or communities.

1.4. Clearly require a retroactive assessment of AI currently in use by the Service

We appreciate the Board's commitment to require the Service to stop using AI identified as extreme risk as soon as it is identified (and by no later than 2024), and to report to the Board about the existing uses of high or medium risk AI.

However, technologies will only be categorized at a certain risk level if they are in fact assessed. To assure the public that the necessary assessments will take place, the Policy should explicitly require the Chief of Police to ensure that existing technologies, goods, and services are screened to determine if they have an AI component, and if they do, that they be assessed for risk. Making this retroactive assessment requirement explicit will help the Service understand its obligations and foster greater public trust that the Policy will address risks associated with AI presently in use.

We recommend the Policy more clearly requires the Service to assess AI Technologies currently in use by the Service against the Policy's risk framework.

1.5. Rework the risk rating scheme

In our Private Sector Privacy Comments, we expressed the view that automated decision-making systems — a term which generally includes AI — should be subject to some form of algorithmic impact assessment. We also suggested linking accountability and oversight requirements to assessed risk levels. We are pleased to see that the Policy has made significant efforts to incorporate this approach.

The Policy includes a risk rating framework that includes risk tiers for minimal, low, moderate, high, and extreme risk AI Technologies. However, the different risk tiers do not have definitions. Instead, they are illustrated by examples of circumstances that the Board would assess at a certain risk level. These examples cover different categories ranging from harms, to flaws in data, to oversight limitations, as well as specific applications of technologies or use cases. Different risk tiers focusing on different categories make it difficult to assess in a practical and methodical way.

For instance, the Policy includes "mass surveillance defined as the indiscriminate covert monitoring of a population or a significant component of the population" as an extreme risk technology. While the IPC agrees that covert mass surveillance should not be conducted, overt mass surveillance also poses major risks. The Policy does not make clear whether such surveillance would be permitted or what level of risk it would pose.

We recognize that the Board's Policy does not establish the risk rating, but instead requires the Chief of Police to establish risk categories for new AI Technologies in consultation with experts and stakeholders. However, in our opinion, the current framing may lead to the creation of a risk assessment framework that is difficult to use in practice when comparing the risks of different AI products intended to be used for the same purpose, or to compare the adequacy of different mitigation methods in reducing a risk.

Rather than relying exclusively on examples of AI applications, the proposed risk category scheme should focus on specific categories of harms, their severity, and their likelihood of occurring. Harms to individuals and groups could include mistreatment, discrimination, loss of autonomy, interruption of personal relationships and private contexts, psychological harms, or loss of trust in law enforcement and the administration of justice. Examples could then illustrate how the severity and impact of a type of harm may vary across risk levels.

The Service should also assess risk in context, as opposed to assessing each technology in isolation. A contextual assessment would focus on how the technology may be used alongside other technologies and practices by the Service, and assess the potential cumulative harms in that light. We also note that the communities that are likely to interact with AI (i.e. as subjects of monitoring or decision-making) may have important perspectives that could inform what constitutes harms, and how they are experienced. This can serve as a crucial input into the development of a meaningful risk rating scheme.[12]

As such, we recommend that the Board:

a) **Clarify that the risk categories to be included in the risk rating scheme should be based on specific categories of harms, their severity and likelihood, rather than applications;**

b) **Require that risk assessments focus on the context in which AI technologies will be used, including how multiple AI systems may be used by the Service in ways that cumulatively may lead to new risks; and**

c) **Require that the stakeholders involved in the development of the risk rating scheme include members of impacted communities.**

2. Strengthening transparency, accountability and oversight

While the Policy includes numerous provisions to inform the Board about the Service's use of AI, we recommend additional measures to ensure the Board and the public have expanded access to meaningful information about the uses of AI and their risks.

[12] Moss, E. et al (2021). "Assembling Accountability: Algorithmic Impact Assessment for the Public Interest." *Data & Society*. https://datasociety.net/library/assembling-accountability-algorithmic-impact-assessment-for-the-public-interest/

2.1. Expand internal oversight and accountability provisions

In 2020, the Global Privacy Assembly (GPA) adopted a resolution co-sponsored by the IPC that outlines twelve measures to promote accountability in the development and use of AI.[13] In our Trustworthy AI Comments, we referred to the GPA resolution and recommended the Ontario government consider several mechanisms to ensure adequate oversight of AI, including:

- Procedures to ensure AI systems are not put to use for new purposes without re-assessment;
- Ensuring strong documentation and recordkeeping practices are in place;
- Whistleblower mechanisms to enable reports of policy violation without fear of reprisal;
- Funding, training, and business roles dedicated to human intervention and oversight; and
- Independent review by existing oversight bodies including the IPC, Ontario Human Rights Commission, and the Ontario Ombudsman.

We are pleased to see the Policy already includes numerous oversight provisions designed to promote the accountable use of AI, including:

- A requirement to train service members to identify AI Technologies.
- Required reports to the Board for high and moderate-risk AI Technologies that must be approved by the Board before the technologies may be used.
- Required notice to the Board that the Service intends to use a low-risk AI Technology.
- Required indicators to monitor the performance of high and moderate-risk AI Technologies for effectiveness and unintended consequences that inform a required post-deployment report to the Board.
- A public website that functions as an inventory of high, moderate, and low-risk AI Technologies used by the Service.
- A five-year review cycle to report to the Board if existing high, moderate, or low-risk AI Technologies have been put to novel uses that may substantially change the data collected or used.
- A method for members of the public to submit concerns to the Board about AI Technologies in use and a requirement that the Executive Director assess and report to the Board about those concerns and necessary action (e.g., recommendations that technologies assessed by the Service as low or minimal risk be reassessed).

The above provisions will help the Board oversee the introduction and ongoing use of AI by the Service. The Policy, however, does not provide for an internal whistleblower mechanism. Such a mechanism can provide an additional source of information to the Board, outside of the formal reporting structure.

[13] Global Privacy Assembly (October 2020). "Adopted Resolution on Accountability in the Development and Use of Artificial Intelligence." *42nd Closed Session of the Global Privacy Assembly*. https://globalprivacyassembly.org/wp-content/uploads/2020/11/GPA-Resolution-on-Accountability-in-the-Development-and-Use-of-AI-EN.pdf

Additionally, a recent study of algorithmic impact assessments noted that these assessments typically require coordination from various specialized actors within institutions.[14] Moreover, as recently indicated by the National Institute of Standards and Technology:

> "Studies have shown that one of the most important determinants of a team's ability to confront issues of harmful bias in AI is that team's diversity. Less diverse teams have a harder time reducing unintended bias in their machine learning models than teams that are made up of members that come from a wide range of genders, ethnicities, and backgrounds. Furthermore, teams that do not represent the perspective of communities impacted by AI systems have a harder time predicting and mitigating potential harms that the system causes to those communities."[15]

This highlights the importance of ensuring that algorithmic impact assessments are adequately resourced and have access to inter-disciplinary experts and range of perspectives and experiences required to effectively conduct assessments.[16] The Policy does not, at present, require that the Service define clear roles and responsibilities or ensure diversity in the expertise, perspectives and experiences that come to bear on the assessments conducted by the Service. Nor does the Policy require that reasonable time and resources be dedicated to the process.

Finally, we note that the Policy does not include specific recordkeeping obligations for AI Technologies beyond the required content of the reports to the Board and the public AI inventory. Personal information used by AI must be retained in accordance with MFIPPA in support of an individual's right of access to their personal information.

Beyond personal information, a general duty to document and retain information used to make decisions, for policy development, or for program delivery is a prerequisite for institutional transparency and accountability. With respect to AI, this should include:

- Design documentation;
- The approach taken to training the system, including information about training data;
- The information used as inputs to the system;
- Explanations of how an AI Technology made a given prediction or decision;
- What information was presented to — and the actions taken by — 'humans-in-the-loop' when they reviewed or overrode an automated decision; [17]
- Incidents where the system was found to exhibit bias or lack of robustness;

[14] Ada Lovelace Institute, AI Now Institute and Open Government Partnership (2021). *Algorithmic Accountability for the Public Sector*. Available at: https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/

[15] National Institute of Standards and Technology (2021). *Summary Analysis of Responses to the NIST Artificial Intelligence Risk Management Framework (AI RMF) - Request for Information (RFI).* https://www.nist.gov/system/files/documents/2021/10/15/AI%20RMF_RFI%20Summary%20Report.pdf

[16] For an overview of the types of competent personnel that may be valuable to involve in an AI risk assessment, see e.g., CIO Strategy Council (2020). "Artificial Intelligence: Ethical design and use of automated decision systems." *CAN/CIOSC 101:2019*. https://ciostrategycouncil.com/standards/101_2019

[17] We address human-in-the-loop and explainability requirements in additional detail in sections 3.1 and 3.2.

- The results of technical or security testing conducted on the system; and
- System and security event logs.

We recommend that the Policy require a whistleblowing mechanism for Service members to report violations of the Policy to the Board anonymously and securely.

We also recommend that the Board require the development of a clear description of roles and responsibilities with respect to conducting and participating in risk assessments under the Policy. The Policy should also direct the Service to ensure that there be interdisciplinary expertise and diverse perspectives involved, as well as reasonable time and resources to contribute to a robust assessment process.

We additionally recommend that the Board require recordkeeping requirements for AI, having regard to the retention of personal information used by AI Technologies and a duty to document the activities of the Service involving AI.

2.2. Publicly disclose additional information about AI Technologies

One of the most significant challenges with respect to the use of AI is transparency. AI models are often so complex that even the developers and operators of AI models can have difficulty in describing the specific reasoning or steps taken by a system to make particular decisions. This challenge becomes very important in contexts where procedural fairness is essential, such as law enforcement.

Other transparency challenges with AI pertain to the data used to train the AI model. Of particular concern are the means by which the data was obtained, whether the data is representative of the real world setting in which the AI model will be used, and whether the data was generated through modern or historical practices that exhibit bias against one or more groups. Transparency can help the users of AI systems and the public determine if efforts have been taken to ensure training data was collected legally and ethically, and if steps were taken to reduce statistical biases in the data.[18]

Another key challenge is ensuring that individuals who interact with, or are otherwise exposed to, AI Technologies are aware that AI is being used to make predictions, to classify or otherwise make decisions about them or affecting them, and what information is being used.[19]

To address some of these challenges, transparency is critical to most AI ethics frameworks and policies.[20] Among the policy proposals intended to bolster transparency is the notion of public AI

[18] See, e.g. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). "Datasheets for datasets." *Communications of the ACM*, *64*(12), 86-92. https://arxiv.org/pdf/1803.09010.pdf

[19] E.g. as outlined in the Ontario Government's 'No AI in Secret' proposed Trustworthy Artificial Intelligence Framework principle. https://www.ontario.ca/page/ontarios-trustworthy-artificial-intelligence-ai-framework-consultations

[20] A study that examined the topics addressed in 112 AI ethics frameworks from public, private, and non-governmental organizations around the world found that transparency was the second-most frequently addressed topic after general social responsibility. See D. Schiff, J. et al, "AI Ethics in the Public, Private, and NGO Sectors: A

registers, which are public websites that disclose information about the AI systems in use by government. The Law Commission of Ontario has published a comprehensive discussion about the types of information that AI registries should contain.[21] Relatedly, the UK government published an Algorithmic Transparency Standard that describes required and optional attributes that should be included in a public disclosure of an algorithmic tool.[22] The Ontario government and Government of Canada have also committed to publishing inventories of AI and algorithmic systems[23] and the results of algorithmic impact assessments,[24] respectively.

We recognize that the Policy includes provisions designed to disclose information about the Service's use of AI Technologies. In particular, the Policy requires that a public website that functions as an AI register be established. This public website would document what high, moderate, and low-risk technologies are in use. With respect to high and moderate-risk technologies more specifically, such a register would include information about the purpose of the technology, how the technology is used, what information is collected, and the expected usage context.

While the public website is a vital transparency measure, we agree with the Law Commission of Ontario that algorithmic impact assessments for AI Technologies should be also publicly and proactively disclosed.[25] We recommend that the Board proactively disclose algorithmic impact assessments subject to existing access to information exemptions under MFIPPA that the Board can rely on, where necessary, to protect confidential information. At the very least, the Policy should require that a summary of any such assessments be prepared and published that includes sufficient information to alert the public to the risk rating, the character of the risks involved, and the mitigation strategies to be employed should the AI Technology be implemented.

We recommend that the Policy broaden the information disclosed in the public AI website based on recommendations made in the Law Commission of Ontario's *Regulating AI* report and the UK government's Algorithmic Transparency Standard. This should include proactive disclosure of algorithmic impact assessments, subject to legitimate exemptions under MFIPPA, or at the very least, summaries of such assessments.

Review of a Global Document Collection," in *IEEE Transactions on Technology and Society*, vol. 2, no. 1, pp. 31-42, March 2021, preprint: https://doi.org/10.36227/techrxiv.14109482.v1.

[21] Thomas, N., Chocla, E., & Lindsay, S. (2021). "Regulating AI: Critical Issues and Choices." *Law Commission of Ontario*. https://www.lco-cdo.org/wp-content/uploads/2021/04/LCO-Regulating-AI-Critical-Issues-and-Choices-Toronto-April-2021-1.pdf

[22] UK Central Digital and Data Office (2021). *Algorithmic Transparency Standard*. https://www.gov.uk/government/collections/algorithmic-transparency-standard

[23] Government of Ontario. *Data Catalogue: Artificial Intelligence and Algorithms*. https://data.ontario.ca/group/artificial-intelligence-and-algorithms

[24] See, e.g. Government of Canada (2021). *Algorithmic Impact Assessment - ArriveCAN Proof of Vaccination Recognition*. https://open.canada.ca/data/en/dataset/afc17416-3781-422d-a4a9-cc55e3a053c8

[25] Thomas, N. (Sept 2021). "Letter to TPSB re: AI Policy." *Law Commission of Ontario*. https://tpsb.ca/images/consultations/AI/LCO_Letter_to_TPSB_re_AI_Policy.pdf

3. Putting the Policy into practice

The Policy requires the Service to develop numerous processes and procedures that are essential to putting the Policy's risk management framework into practice. In this section we highlight requirements that will need to be defined and offer preliminary recommendations on what they should address.

3.1. Establish requirements for effective human intervention

The role of a 'human in the loop' is often discussed as a critically important measure to ensure that decisions about individuals are not made solely on the basis of an algorithmic system. In our Trustworthy AI Comments, we emphasized the importance of taking a risk-based approach to determining human oversight and intervention requirements. We reflected on the nuance of this issue in our Private Sector Privacy Comments, where we noted that human oversight is not a cure-all to address all algorithmic harms.

In the context of law enforcement, if predictive policing systems make predictions based on training data about policing in a community, the system may reproduce biases associated with historical policing practices.[26] Simply putting a human in the loop does not solve this problem. The human-in-the-loop will need the appropriate experience or training to identify and mitigate bias. And they will need support to avoid deferring to the judgement of AI systems which they may see as more objective than their own thinking — in a phenomenon labelled 'automation bias.'[27] At the same time, humans-in-the-loop will need to be aware of how human biases and decision-making have resulted in the inequitable situations that produce the data in the first place and which may cause or perpetuate bias in AI systems.

There should be clear objectives in the Policy for human oversight and intervention. These should be supplemented by criteria and a methodology for assessing the efficacy of a human-in-the-loop process in meeting those objectives. These goals, criteria, and methods can then inform training content for Service members who play the role of a human-in-the-loop, inform procurement requirements, and help determine how human review and intervention might be tested, measured, or monitored. The criteria should pay particular attention to the role of discretion in the decision-making process being considered for automation by the AI system. In circumstances where a significant degree of discretion is required, it may not be appropriate for AI to make even the initial recommendation, while in other cases it might be. Care must also be taken to ensure individuals responsible for exercising human oversight have the time, resources, information, and capacity

[26] Richardson, R., Schultz, J. and Crawford, K. (2021). "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *NYU Law Review Online*. https://www.nyulawreview.org/online-features/dirty-data-bad-predictions-how-civil-rights-violations-impact-police-data-predictive-policing-systems-and-justice/

[27] For a discussion of the limitations of human-in-the-loop, see, e.g. Green, B. and Kak, A. (June 15, 2021) "The False Comfort of Human Oversight as an Antidote to A.I. Harm." *Slate*. https://slate.com/technology/2021/06/human-oversight-artificial-intelligence-laws.html

needed to effectively review decisions. If such supports are not in place, human oversight provides limited benefit.[28]

In the Policy, human oversight is not presented as a control used to mitigate other risks. Instead, it is discussed in the context of defining risk levels, such that the absence of, or difficulty in exercising, human oversight in and of itself is treated as though it is an independent harm. In our view, human-in-the-loop should be included instead as a required or recommended mitigation measure, depending on the risk assessment.

We recommend that the Board develop requirements relating to human-in-the-loop processes that include:

a) **A process to determine which aspects of decision-making should be automated and which should be subject to human review. This process should assess:**

   i. **The role of discretion in the decision;**

   ii. **The accuracy, performance, and robustness of the AI system being considered; and**

   iii. **The consequences of an incorrect decision.**

b) **Ensuring humans-in-the-loop have the time, resources, information, and capacity needed to effectively review automated decisions or predictions.**

c) **Logging and recordkeeping requirements for human review and intervention activities.**

d) **Including human oversight metrics among the indicators that must be reported back to the Board.**

3.2. Establish requirements for meaningful explainability

Another important AI accountability measure is ‘explainability.’ AI explainability is the degree to which an AI system can reveal its inner workings to describe the information, criteria, and reasoning it used in making a decision or coming to a conclusion. Explainability can help support meaningful human-in-the-loop processes by providing the information required by a human to review decisions made by the system.

There is not yet a clear standard for what constitutes a meaningful explanation.[29] However, the purpose of the AI system, the class of people who might seek an explanation and for what purposes (i.e. the intended audience), and the risks associated with a failure to explain an output may constitute a reasonable starting place for determining the explainability requirements for a given system.[30] This is in line with the Draft Facial Recognition Technologies Guidance, which requires

[28] See, e.g. Green, B. (2021). “The Flaws of Policies Requiring Human Oversight of Government Algorithms.” *Working paper*. http://dx.doi.org/10.2139/ssrn.3921216

[29] See, e.g. Arrieta A.B. et al (2020). “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI.” *Information Fusion (58)*. Preprint retrieved from: https://arxiv.org/abs/1910.10045

[30] See, e.g. Amarasinghe, K. et al (2021) “Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions.” Working paper available at: https://arxiv.org/abs/2010.14374

that the performance of a facial recognition technology be explainable insofar as it should be subjected to testing to determine false positive and false negative rates, as opposed to requiring a comprehensive explanation about *how* the technology identified an individual.

The Policy indicates AI Technology that is not fully explainable should be considered an extreme risk. As with our discussion of human-in-the-loop, however, it would be better to see explainability as a mitigation measure for certain risks, and the lack of explainability as an aggravating factor for others. It may be reasonable to expect that an extreme risk technology should be meaningfully explainable, but it does not follow that any AI system becomes an extreme risk if it cannot be meaningfully explained. Conversely some form of explainability is just as important for many low-risk AI technologies. A certain baseline of explainability helps to support meaningful responses to FOI requests, requests for access to personal information, and protects against risks to procedural fairness and accountability. The use of an AI technology that diminishes the overall transparency of the Service, or makes it more difficult for the Service to respond to freedom of information requests would be harmful for those reasons.

We recommend that the Policy consider explainability as a mitigating factor, and lack of explainability as an aggravating factor when assessing potential harms in its risk assessment process, and that the Policy set out a certain baseline level of explainability requirements that must be met for all levels of risk.

4. Moving the Policy forward

In 2020, the Board made a commitment to build public confidence and address systemic racism,[31] including by consulting with experts and communities, enhancing transparency, and providing for independent auditing and greater service accountability. These commitments are echoed in the Policy, whose guiding principles include the need for public trust in policing and to preserve the dignity and privacy and other rights of individuals and communities.

The IPC commends the Board's commitment to develop an AI governance framework in a transparent manner that allows for public input. However, as we have highlighted in our submission, there remain outstanding issues that require revisions to the Policy, as well as detailed work on the operational procedures and processes required to support the Policy.

4.1. Continue to consult the public on the development of the procedures

The Policy requires that the Service develop, in consultation with experts and stakeholders, procedures and processes for the assessment of New AI Technologies. We note that many procedures and processes will need to be developed, including:

- Guidance for Service members to identify an AI Technology;
- Requirements for recordkeeping;
- Requirements for effective human intervention;

[31] Toronto Police Services Board. (2020) "Police Reform in Toronto: Systemic Racism, Alternative Community Safety and Crisis Response Models and Building New Confidence in Public Safety." *Toronto Police Services Board Report*. https://www.toronto.ca/wp-content/uploads/2020/09/8e5a-public_agenda_aug_18.pdf

- Requirements for meaningful explainability;
- One or more risk assessment methodologies;
- Required mitigation measures for various risks;
- Guidance on how to identify and monitor unintended consequences; and
- Criteria for assessing data quality.

We recommend that the Board continue to engage a broad range of public stakeholders, including the IPC, as it works towards finalization of its Policy.

We also recommend that the Board take steps to ensure that the Service engages an equally broad range of stakeholders, including the IPC, in developing processes and procedures to implement the Policy.

4.2. Include the IPC and public consultation in the risk assessment methodology

A global study of public sector algorithmic accountability policies found that few policies meaningfully incorporate participation from the public and external stakeholders.[32] Meaningful public engagement is essential to ensure that the perspectives of those affected by AI Technologies as well as experts with technical and legal backgrounds can be incorporated into AI initiatives and help identify and develop risk mitigation strategies for dealing with risk.

In our Private Sector Privacy Comments, we recommended that consultations with the IPC and with the public should be considered for automated decision making-systems that might pose high risks.

We appreciate that a role for the IPC during the risk assessment process is present in the Policy. The required report to the Board for high and moderate risk technologies includes both reference to consultations with the IPC as well as analyses required by the IPC. [33] We would welcome being consulted by the Service on new AI initiatives (or assessments of existing systems) at the earliest opportunity, including reviewing privacy impact assessments or algorithmic impact assessments.

The Policy requires that public consultations be conducted for high risk AI Technologies after deployment, and that their results be included in a follow-up report to the Board. While such consultations could be valuable in identifying the impacts of the AI Technology on public trust as well as potentially surface unintended consequences, public consultation should also play a key role during the assessment process itself.

While we are not suggesting that every new AI Technology should necessarily be subject to public consultation, **we recommend the Policy require that criteria be developed to establish the circumstances in which a public consultation should be included as part of the pre-deployment risk assessment process.**

[32] Ada Lovelace Institute, AI Now Institute and Open Government Partnership (2021). *Algorithmic Accountability for the Public Sector*. Available at: https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/

[33] Sections 5(g) and 5(j) of the Policy, respectively.