Dubi Kanengisser, PhD
Senior Advisor, Strategic Analysis and Governance
Toronto Police Services Board
40 College Street
Toronto, ON M56 2J3
dubi.kanengisser@tpsb.ca

December 15, 2021

Dear Mr. Kanengisser,

The Equity, Inclusion & Human Rights Unit of the Toronto Police Service (the "Service") appreciates the opportunity to provide feedback on the Toronto Police Services Board draft *Use of Artificial Intelligence Technology* Policy (the "Policy"). The Policy represents a crucial first step in ensuring transparency, fairness and accountability with respect to the use of AI systems in policing by the Toronto Police Service.

The Equity, Inclusion & Human Rights Unit

The Equity, Inclusion & Human Rights Unit is the first of its kind in Canadian policing, and was created by the Toronto Police Service in 2019 to champion a progressive equity agenda for the Service. The unit is a Center of Excellence led by a team of subject matter experts, utilizing best practices in the promotion of inclusion and human rights for our members and our communities.

Comments on the Policy

1. Guiding principles:

    a. The Policy should rest on principled grounds that are clearly defined so that the public and the Service understands what they mean. There are sets of principles being identified internationally to guide AI/ML ethics and this could be good starting place to look at: https://montrealethics.ai/the-proliferation-of-ai-ethics-principles-whats-next/.

    b. Suggest making it unambiguous that technologies implemented for recruitment, hiring, compliance, etc. also fall within the purview of this Policy, by adding "Service members" in the second to last sentence where it says "… the potential unintended consequences to the privacy, rights, freedoms and dignity of members of the public." Also, the definition of "AI technology" (see below) references only members of the public.

2. Meaningful engagement:

    a. The Policy lacks clear requirements for ensuring the public, especially impacted communities, are meaningfully consulted. There are currently two places identified for community consultations: (Sec. 5 g.) in reports by the Chief on *any* consultations conducted and (Sec. 8) develop and implement an engagement strategy to *inform* the public prior to deployment. There are no explicit requirements for engagements with affected communities to inform the assessment, monitoring, evaluation or development of policies and procedures for the use of AI by the Service.

    b. (Sec. 1) does not specify a role for communities in the development of procedures and processes to review and assess new AI tech risk levels. Suggest that having communities impacted by the criminal justice system at the table would be immensely valuable to inform how risks are identified and determined.

    c. Consistent and regular public engagement is necessary to ensure technology use and impacts remain within acceptable parameters, especially for high or medium/moderate risk AI for the legitimacy and transparency of such procedures and processes.

    d. To allow for meaningful engagement, it is recommended that the Policy have a plain language version to aid in engagement with those who may not possess a technical understanding of the technologies or have language barriers.

3. Defining terms:

    a. “Bias” is defined too narrowly. It should mention how flawed outputs are also affected by transactional data. The data may also lead to improper or discriminatory conclusions about places, things, or other concepts that go beyond misidentification of a subject. It would be useful to emphasize the systemic aspect of bias by adding “directly or indirectly”— although it appears to be implied, it is crucial to ensure that this is understood.

    b. The definition of “AI technology” includes “any goods or services… require that a privacy impact assessment be conducted in advance.” Consider if this is too broad and would bring in any data-related activities under the Policy, if the Policy is not intended to apply to internal activities (*see* #1(b) of this outline). May be helpful to clarify what is *not* considered AI technology, with examples, for the purposes of the Policy.

    c. Under “Policy of the Board” in (Sec 1 ii.) under “1” where it says “carry bias,” suggest adding “or potential for bias” which emphasizes that that bias is not always explicit or obvious on its face, but results in disparate impacts.

d. Suggest certain terms or phrases be more clearly defined or clarified to avoid inconsistent or overly subjective application of criteria. For example, under “Policy of the Board” in (Sec 1 ii.): what does “quality” mean exactly when referring to “where the quality of such data is unknown”? What is meant exactly by “malicious actors”? For example, harms can also be done by well-intentioned actors.

e. The Policy mentions that “extreme risk” AI technologies will not be used by the Service. The example provided is facial recognition using “illegally sourced” data. The Policy should define, with examples where possible, “illegally sourced” data and similarly define “legally sourced” data. It is also important for the Policy to require the Service to not only assess data that is legally or illegally sourced, but whether such data use is ethical.

4. <u>Risk levels</u>:

a. This is a good mechanism for identifying and defining AI risk; however, AI/ML tools and their applications make simple risk categories difficult to implement. Consider that there may be different criteria that are meaningful to different types of AI technologies, depending on how they work, the data they rely on, and their application. For example, criteria for AI risk that rely on police administrative data (i.e. general occurrence data) could be very different from those relying on biometrics (i.e., face, fingerprints, voice, gait, etc.) and those based on service-oriented interactions (i.e., seeking information, requesting records, background checks, etc.). Further to this point, if using police administrative data for enforcement purposes, bias that may exist within this data should also be taken into consideration to reduce over-policing, as well as under-policing, communities. These nuances may be worked out in more specific Service policies, but if there is general Board guidance on how to scope this appropriately, this would be a good opportunity to do so.

b. Suggest that the criteria are grouped into separate areas of risk, such as a risk matrix that may include: i) *conditions* (organizational, technical and environmental) that shape risk, ii) potential adverse *impacts* and iii) risk of *misuse*, including tampering and unauthorized access, particularly as these relate to privacy, human rights, and other public interests. Having a risk matrix helps the public and the Board to better assess and track important aspects of risk.

5. Review and monitoring periods:

    a. AI/ML technology is a fast-moving field with constant changes and the impacts of its deployment cannot be disconnected from the social context, which is dynamic. Setting review periods of five years is too long, and a one-year limit to monitoring deployment is too short. Suggest yearly monitoring of AI/ML use and impacts and attendant review of risks based on any technical, data, and environmental changes that are likely to affect its function and impacts.

    b. Suggest explicitly stating a regular review period of the Policy itself, for the same reasons as per above. For example, the Board's Body Worn Camera Policy has an annual reporting requirement, and its other policies require a three-year review schedule. The review period could be set with a caveat of "or more frequently as needed," defining the conditions that might trigger an earlier review.

    c. To balance against resources/capacity constraints by the Service, consider requiring the Service to report on a regular basis following deployment of technologies that fall within the "High Risk" and "Medium Risk" categories, in tandem with other suggestions for improvement listed in this document, including #6 ("Transparency") below. The frequency of reporting could vary for specific applications, with specific conditions that might trigger a review. These details should also be made public.

6. Transparency:

    a. There are areas for improvement to make Service use of AI/ML more transparent while balancing public safety concerns. Given most AI technology used in policing are proprietary third-party developments, what level of detail would be required under (Sec 5 e.) to support the Board and the public to understand the models (and assumptions) underpinning the AI technology?

    b. There may be instances where facial recognition technologies may be appropriate. The policy identifies an example of linking biometrics to personal identifiers under high-risk technologies. To aid public understanding, it is recommended that be expanded upon to include when facial recognition technologies may be used (for example, in relation to certain criminal offences, missing persons, etc.), which is currently not clear in the policy. This will help distinguish the use of facial recognition by the Service from large-scale information gathering and the extreme-risk example of indiscriminate covert monitoring resulting in mass surveillance. Clarifying what AI facial recognition technologies are, and making the distinction between what they will and will not be used for, will assist with public understanding, transparency, and trust.

c. The Policy is not clear in its expectations for “human-in-the-loop” to ensure the human is not simply rubber stamping AI outputs or applying discretion to over-ride results in a way that reinforces biases. What are the mechanisms to challenge the results generated by AI or how it was interpreted and applied by decision-makers?

7. Governance and accountability:

   a. It’s not clear what the requirements are for governance of AI technologies within the Service (the Policy mentions only *reporting* “steps taken”) and the Board’s oversight responsibilities). For example, is the Service required to seek technologies from third-party providers that have a record of ethical AI or have undergone bias audits? Who is responsible for vetting the risk levels assigned (and which triggers different degrees of reporting and engagement requirements)? How would the Board determine what high/medium risk technologies to approve?

   b. Suggest a clause speaking to minimization in the use of AI where human intervention would accomplish the same objectives/outcomes without the risks and unintended impacts of AI technologies. This can be accomplished, for example, by weighing the balance of efficacy over efficiency, particularly for high risk AI tech used for decision-making in individual cases. Consider that for situations where human and AI efficacy is comparable, the question of accountability may be harder to address for AI than for humans.

   c. Suggest to include requirements for a clear point of contact within the Service and procedures for members of the public to inquire about AI use, raise concerns, make complaints, and inquire about or challenge its application in their specific case. This would help create mechanisms for the public to enact their privacy and due process rights.

   d. Suggest requiring that the Service publicly release on its website the governance/procedures as it relates to AI technology and how it will be used in easily accessible formats.

Sincerely,

Dr. Mai Phan, Laura Flyer & Nicole Rebelo

Equity, Inclusion & Human Rights Unit
Toronto Police Service