Dubi Kanengisser
c/o Toronto Police Services Board
40 College St.
Toronto ON M5G 2J3

December 15, 2021

To the members of the Toronto Police Services Board:

Re: Public Consultation on the Toronto Police Service Board's Use of Artificial Intelligence Technology Policy

We write to you as members of Aggregate Intellect, a global community of AI practitioners, researchers, and enthusiasts founded in Toronto in 2018. We write to you as data scientists, data analysts, machine learning engineers, AI consultants, and tech entrepreneurs, who are intimately aware of the potential of new AI technologies, especially their potential for harm. But most importantly, we write to you as concerned citizens, who wish to ensure that the use of AI technologies in policing is to the benefit of all.

We applaud the TPSB's call for public feedback on the draft Policy, and its explicit commitments to transparency, fairness, equity, privacy, and public consultation. The kinds of assessments and reporting that this Policy sets out provide a powerful framework for ensuring the appropriate and trustworthy implementation of AI technologies by the Toronto Police. Yet there are also several aspects of the current Policy that we believe must be improved. In particular:

- **§1–2: Review and Assessment of New AI Technologies.** While we understand the need for the Policy not to be overly prescriptive in its definition of its risk categories, we find the current descriptions unsatisfactory. Specifically:
    - No AI technology "where training or transactional data is known to be of poor quality, carry bias, or where the quality of such data is unknown" (§1c(ii)1) should *ever* be considered for use, and thus should be deemed Extreme Risk, not High Risk. Any AI technology based on poor quality or biased data is inherently compromised. (Furthermore, the quality of the data must be assessed by an independent and disinterested party of specialists; see our comments on §3–9 below.)
    - It should be made explicit that any data collected through social media is data known to be poor quality and carry bias, and thus should not be considered a viable source.

- It should be made explicit that any data acquired through data brokers may be difficult to verify that it is *not* "known or thought to be illegally sourced" (§1c(i)5) or that it is *not* "known to be of poor quality [or] carry bias" (§1c(ii)1). Therefore, any technology utilizing data from brokerages or whose manufacture required data from brokerages cannot be classified as below High Risk.
- No AI technology that assists in "identifying, categorizing, prioritizing or otherwise making decisions pertaining to members of the public" (§1c(iv)2) should be deemed Low Risk. Automating such actions through technology, even with the inclusion of a human-in-the-loop, is an intrinsically risky activity, and should be categorized as such by the Policy.
- Any technology that has created a significant cause of concern in the community, or that has been explicitly rejected by Toronto citizens, should be automatically considered as satisfying the definition of Extreme Risk and be subject to the relevant portions of the Policy.

- **§3–9: Board Approval and Reporting Prior to Procurement, Utilization and Deployment.** While we appreciate the Policy's inclusion of a formal review and approval process, there are several ways in which it could be improved. Specifically:
  - The Policy requires that the Chief of Police will "conduct a risk assessment of the AI technology" (§4), but there are no checks in place to ensure that this assessment is conducted by qualified experts and relevant stakeholders. In particular, we recommend that the Policy define what it means by "experts and stakeholders" (§1), and require that the specific list of experts and stakeholders to be consulted for any particular assessment be submitted to and approved by the Board before the assessment is made. We further recommend that the Policy require that the experts to be consulted always include an independent and disinterested party of subject matter specialists who are knowledgeable about the social risks of AI technology, and that the stakeholders to be consulted always include those in the community with experience of the kind of policing under consideration.
  - In particular, since the Policy's reporting requirements depend on the risk level that an AI technology is assigned, this assignment needs early and independent verification. Including this as a mere part of the approval process, as in §5c and §7, is insufficient. The situation described in §13b(i), where public concern "demonstrates that an AI technology was erroneously assessed as of a lower risk level than appropriate" *only after* the technology has already been reviewed and deployed, should not be allowed to happen. It stands to reason, then, that there be a minimum set of report requirements as defined in this policy, including a

detailing of how the erroneous classification was missed, and how the assessment will change in the future.

- **§10–14: Monitoring and Reporting.** While we concur that it is essential for specific monitoring and reporting requirements to be included in this Policy, the current requirements do not go far enough. Specifically:
    - Monitoring should not be limited to "until 12 months after full deployment" (§10), and reporting should not be allowed to wait until "within 15 months of full deployment" (§11). Given the ever-present risk of model drift – the degradation of a model's predictive performance over time, due to changes in the environment that violates the model's assumptions – AI technologies should be *continuously* monitored throughout the tenure of their deployment; and given the real hazards of AI technologies, reporting should be presented within 3 months of the initiation of deployment and repeated every 3 months after that.
    - As much as possible, the results of such reporting should be made public and transparent, so that citizens can remain informed about the deployment of new AI technologies and be able to provide meaningful public feedback.
    - Reporting should include details about the ownership structure of any vendors and any relevant changes in their partnerships and governance, as well as any changes in their policies around data management.
- **§13: Public Feedback.** Ensuring that the public is able to provide feedback on new AI technologies is an essential element of this Policy; as such, the accommodation for public feedback must be expanded for this provision to be meaningful. Specifically:
    - Currently, the only opportunity for public feedback comes *after* the deployment of a new AI technology. A public consultation should also be included in the initial report (§5), and specifically with regards to the technology's risk level assignment. In particular, the public should always be considered to be a stakeholder in any AI technology "identifying, categorizing, prioritizing or otherwise making decisions pertaining to members of the public" (§1c(iv)2).
    - Relatedly, the "public engagement strategy" described in §8 should be expanded. It is not sufficient to merely "inform the public of the use of the new AI technology". The public should always be considered as a stakeholder in the deployment of any new AI technology, and as such, the public's opinions should be collected and taken into consideration throughout the review process, including direct consultation with representative groups of citizenry as is appropriate. The public should also be explicitly informed of their ability to consent (or not to consent) to the use of their data by this technology.

- In addition, a process should be established whereby members of the public can petition to have certain AI technologies and/or vendors removed from consideration.

- **Definitions.** In several instances we recommend that the definitions provided in the current Policy be revised and clarified. Specifically:
  - "Bias": Bias should not be defined as "consistently flawed output", as this presents an overly narrow, mathematical definition of bias. Rather, its definition should be revised to make explicit reference to other kinds of data bias, including but not limited to historical bias, representation bias, measurement bias, and population bias. More broadly, the definition of bias should be connected to the ways in which systems may make decisions that are unjust, as determined by the public interest.
  - "Explainability": High Risk Technologies are defined to include those technologies "where a system cannot be fully explainable in its behaviour" (§1c(ii)4); however, "explainability" is never defined in the Policy. If left undefined, this potentially sets a bar that most AI technologies would not pass – or worse, sets the *wrong* bar, which would not actually ensure that the technology's behaviour is meaningfully justified – and so we recommend that this term be explicitly defined, along with the related concepts of "interpretability" and "justifiability", and that its limitations in practice be noted. An active involvement of many stakeholders from social sciences, computer science, civil society, and industry is recommended.
  - "Human-in-the-loop": It should be specified that the human confirming any decisions or classifications made by the technology must be *qualified* to make this confirmation. We find it concerning and inadvisable that technologies "where the 'human-in-the-loop' may have difficulty identifying bias or other decision failures of the AI" are classified as merely of moderate risk (§1c(iii)1). An unqualified human-in-the-loop is no better than no human-in-the-loop at all, and should not be taken to give the technology any additional credibility.

- **Guiding principles.** Lastly, while we applaud the Policy's commitment to avoid "the potential unintended consequences to the privacy, rights, freedoms and dignity of members of the public, and to the equitable delivery of police services to the public", we believe that the policy should go further in its guiding principles and explicitly commit itself to supporting technologies that *improve* equity, transparency, accountability, and the public interest, and that proactively mitigate systemic issues in how police services have been and continue to be deployed.

In conclusion, we would like to reiterate our appreciation of this draft Policy, and our gratitude to the TPSB for proactively seeking out feedback from the public at this stage. We hope our feedback will prove useful as the TPSB works to revise this Policy, and we are happy to provide further consultation upon request.

Sincerely,

**Disclaimer: The views presented in this response are the individuals' opinions and do not reflect the views of their employers listed below*

Dr. Willie Costello, Data Scientist, Shopify; AI Ethics Stream Owner, Aggregate Intellect

Dr. Somaieh Nikpoor, AI Ethics Stream Owner, Aggregate Intellect

Dr. Amir Feizpour, CEO, Aggregate Intellect

Daria Aza, Data Analyst, Manulife

Anh Dao, Co-founder, Healthtek

Dr. Aaron Maxwell, Policy Analyst

Sai Kumar Nerella, Machine Learning Engineer, Talem Health Analytics

Jay Sheldon, Lead Consultant @ Significance.ai, President @ Dialography

Karthik Bhaskar, Data Scientist, CIBC

Indrani Bhattacharya, Data Scientist, StatusNeo

Dr. Abhishek Bihani, Data Scientist, Tignis

Dr. Yony Bresler, AI Researcher, Crater Labs

Sina Dibaji, Data Scientist, ApplyBoard

Logan Emerson, Technical Consultant, UiPath

Kiran Kadekoppa, CTO & Co-founder, HUEX Labs

Ammar Khan, Community Growth Manager, Aggregate Intellect

MB Jallow (Luka), Product Manager, MarsCrowd

Soomin Aga Lee, Manager of Data and AI Governance, PointClickCare

Frankline Ononiwu, Data Scientist, TD Bank

Suhas Pai, CTO, Bedrock AI

W. Richard Yu, Lead, Data & Digital Product, Jack.org