# The Office of the Privacy Commissioner of Canada
## *Consultation: Proposals for ensuring appropriate regulation of artificial intelligence*

March 12, 2020

*Written Comments*[1] *of*
*Amba Kak, Director of Global Strategy and Programs, AI Now Institute, NYU*
*Rashida Richardson, Director of Policy Research, AI Now Institute, NYU*

Thank you for the opportunity to provide comments on the Office of the Privacy Commissioner of Canada (OPC) request for comments on proposals for ensuring appropriate regulation of artificial intelligence (AI), including amendments to the Personal Information Protection and Electronic Documents Act (PIPEDA). Our comments proceed in the same order of the proposals and are designated with relevant headings.

**PROPOSAL ONE: DEFINITION OF AI AND PIPEDA APPLICABILITY**

In this section while we provide general definitional guidance around the concept of AI we would caution against an entirely parallel legal regime for AI within PIPEDA. Firstly, any protections for AI systems should be *in addition* rather than to the exclusion of existing data protection rules. There are already multiple points at which the PIPEDA applies to data-related activities within AI and automated decision making systems and puts in place critical safeguards. More fundamentally, as we explain in our response to proposal two, there may be limits to trying to address all risks and harms associated with AI, especially exclusion and discrimination, within the data protection framework.

**Guidance on Defining AI for Regulation**

AI has many definitions, and can include a wide range of methods and processes, including machine learning. AI has also grown in popularity as a marketing term, designating computational products and services that may or may not fit within the definitions of AI commonly accepted by the AI research field. Whatever technical capabilities and methods an AI system relies on, it is important for AI to be understood as more than the sum of its technical approaches. It is also developed out of the dominant social practices of engineers and computer scientists who design the systems, and the industrial infrastructure and market incentives of companies that operate and sell those systems. Thus, a more complete definition of AI includes technical approaches, social practices and industrial power.

[1] The authors extend gratitude to Meredith Whittaker, Inioluwa Deborah Raji, and Ben Green for their feedback and insights while drafting these comments.

We recognize that adopting a comprehensive and precise definition for regulatory purposes is challenging, so we offer the following recommendations for consideration. First, a definition of AI should emphasize and center the impact and potential effects of the technology rather than centering the underlying technical mechanisms or methodologies. Many existing legal definitions of AI and related systems, such as automated decision-making or automated decision systems, tend to lead with and emphasize the technical mechanisms or methodologies that facilitate the various capabilities or functions of the technology. This focus on the technical details is not ideal because it can reinforce automation bias and legitimize certain technical solutions over alternatives,[2] or ignore autonomous systems already in place and their present risks.[3] Considering the diverse applications and variety of sectors that AI is used, definitions that do not center public concerns and known risks can hamper the development of appropriate and effective regulatory practices.[4]

Second, a definition of AI must demonstrate awareness of the context in which the technology is operationalized. Context is important for definitional precision and enforceability. Since AI includes a constellation of processes and capabilities that can be related yet distinct (i.e. ranking versus classification), the definition must evaluate or speculate current and future applications to ensure these distinctions are not attenuated. Failure to appreciate these subtleties can lead to a reductive definition, which can impede enforcement or make harmful technologies legally permissible. Reductive or context-agnostic definitions can also hinder compliance because different stakeholders may interpret the definition, in part or whole, differently, so there must be particular care given to cross-cultural or cross-discipline semantics and sector-related assumptions.

Finally, operational context is also significant when considering potential exemptions to a law or regulation. For example, after 18 months of deliberation the New York City Automated Decision System Task Force (Task Force) could not reach consensus about which applications warranted exemption. The Task Force's report suggested that spreadsheets may be an application warranting exemption because of its quotidian functions;[5] however, spreadsheets facilitate a variety of computational functions and in some contexts have been used to automate decision-making that produced harmful results.[6] Therefore, appreciating the context that AI is operationalized in is imperative to understanding the full scope of risks and opportunities that must be contemplated when assessing effective regulation.

---

[2] Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (2018); Ruha Benjamin, Race After Technology (2019).
[3] P.M. Krafft, et al., Defining AI in Policy versus Practice, 6 (2020), https://arxiv.org/pdf/1912.11095.pdf.
[4] P.M. Krafft, et al., Defining AI in Policy versus Practice, 6 (2020), https://arxiv.org/pdf/1912.11095.pdf.
[5] New York City Automated Decision Systems Task Force Report, 26 (2019), https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf.
[6] Jay Stanley, Pitfalls of Artificial Intelligence Decision Making Highlighted in Idaho ACLU Case (2017), https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case.

**Applicability of PIPEDA to AI**

PIPEDA covers processing of personal information (i.e. information about an identifiable individual), which is a component of most AI technologies, and to that extent PIPEDA does and should apply to data processing activities. Personal data can be embedded in the model,[7] training data (used to train the model), input data (during use), and output data (results of the models like labels and/or prediction targets). Personal data can also manifest indirectly as proxies.[8] Often training data might be composed of pseudonized or aggregated data, which would appear difficult to link to a particular named individual. However, strict delineations of such aggregate data as non-personal may not be advisable in these cases, given the rapid development of de-anonymization and re-identification technologies. A 2019 study by Lus Rocher et al[9] found that "even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR." Entities should therefore conduct diligence around the likelihood and ease with which data can be manipulated to reveal personal information while evaluating the applicability of PIPEDA. The ramifications of the personal data threshold are discussed further in Proposal Two.

Once data processing activities within an AI system come within the scope of PIPEDA, there are a series of substantive protections around collection, purpose and storage limitations that will apply. As explored below in Proposal 6, these are important regulatory tools to ensure accountability vis-a-vis use of personal data in AI systems. In addition, the PIPEDA principles around data security and privacy by design will mandate institutional and technical best practices to safeguard against data breaches[10] which is vital when managing systems with large amounts of personal data and points of access.

---

[7] Personal data can be "embedded" in the model in situations where it is an input into the model and therefore its representations in the model will also capture aspects of personal data, for example biometric data embedded in image search models. *See* also UK ICO, Guidance on the AI auditing framework: Draft for Consultation 2020 https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

[8] *See* Bernard Harcourt, Risk as a Proxy for Race (2015), https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=3568&=&context=faculty_scholarship&=&sei-redir=1&referer=https%253A%252F%252Fwww.google.com%252Furl%253Fq%253Dhttps%253A%252F%252Fscholarship.law.columbia.edu%252Fcgi%252Fviewcontent.cgi%253Farticle%25253D3568%252526context%25253Dfaculty_scholarship%2526sa%253DD%2526ust%253D1584030454076000%2526usg%253DAFQjCNEZ6-jVG3JN7aK3Cn2awbINRoJftg#search=%22https%3A%2F%2Fscholarship.law.columbia.edu%2Fcgi%2Fviewcontent.cgi%3Farticle%3D3568%26context%3Dfaculty_scholarship%22.

[9] Lus Rocher et al, Estimating the success of re-identifications in incomplete datasets using generative models, https://www.nature.com/articles/s41467-019-10933-3/

[10] *See* Betsy Swan, Facial-Recognition Company That Works With Law Enforcement Says Entire Client List Was Stolen (2020), https://www.thedailybeast.com/clearview-ai-facial-recognition-company-that-works-with-law-enforcement-says-entire-client-list-was-stolen.

## PROPOSAL TWO: PROMISE AND LIMITATIONS OF A DATA PROTECTION LENS

AI includes a variety of processes and technologies that perform or inform decision making, and this in turn can produce discrimination, exlcusion, and hypervisibility, harms that are typically not addressed by traditional data protection or privacy rights framing. Because these interests and harms often relate to power dynamics, some emerging legal frameworks to address the risks of AI are shifting the focus from threshold questions, like whether the data is personal, to the outcomes produced or facilitated by AI, or the impact of an AI-informed decision. For example, the Algorithmic Accountability Act in the United States Congress attempts to regulate AI with an accountability framework.[11] This legislation requires companies to evaluate the privacy and security of consumer data as well as the social impact of their technology, and includes specific requirements to assess discrimination, bias, fairness and safety. Though it may not be practically possible to explicate all potential risks and interests in regulatory and enforcement regimes generally, or within the authority of the OPC, including broader frameworks, such as accountability, can help capture categories of harm and interests that are related yet not specific to data processing.

Moreover, an expansive view of data protection is also necessary to adequately assess the social validity of a technology and address any concerns regarding legal responsibilities for harm. Most existing legal or regulatory frameworks lack specificity on who bears legal responsibility when a harm occurs. In some cases this ambiguity has resulted in a party that merely implemented the technology bearing responsibility, despite having no power or authority in the development or design of the technology or practical ability to avoid foreseeable harm.[12] Thus, remedial interventions must also include accountability and power analysis if they are to adequately assess whether a technology should exist, and which stakeholders are best suited to redress harm.

As the OPC focusses on AI systems, we would encourage drawing inspiration from how other jurisdictions have applied the data protection legal and enforcement framework to address broader categories of harms. The General Data Protection Regulation (GDPR), for example, includes a regulatory mandate that all data processing should be "fair, lawful, and transparent."[13] Regulators like the United Kingdom's Information Commissioner's Office (ICO) have interpreted this "fairness" requirement as a directive for all data processors to evaluate discriminatory impacts like whether "the system is sufficiently statistically accurate and avoids discrimination."

---

[11] Algorithmic Accountability Act, H.R. 2231, 116 Congress (2019-2020), https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf.

[12] Madeleine Clare Elish, Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (2019), https://estsjournal.org/index.php/ests/article/view/260/177.

[13] *See* European General Data Protection Regulation, Article 5(1).

[14] Data protection tools like "Privacy by design"[15] assessments and Data Protection Impact Assessments (DPIA),[16] in particular, have been important vehicles to operationalize a more holistic evaluation of AI systems.[17] Both of these requirements are triggered at the design phase (before the system is implemented) and can be powerful tools to encourage self-reflexivity around the broader social impact of AI systems and to structure these systems in ways that minimize harm. The identification of risks in the DPIA should go beyond data security and data provenance, to include potential discriminatory impacts on historically disadvantaged communities.

DPIA's can also be mobilized to demand more participation of directly impacted communities in the risk identification process by soliciting the perspectives of those that would be directly affected by AI through diverse and inclusive consultation. Critically, the DPIA exercise must enable communities to reject technologies that they don't want and are not comfortable with, and should not be undertaken with a predetermined commitment to eventually implementing a given AI system. Where the risks identified cannot be sufficiently mitigated, or where the concerns of the affected community remain unresolved, there should scope within the assessment to stop or prevent the system from being deployed altogether.

In the long run, however, the data protection framework is likely to come up against hard limits when addressing algorithmic accountability, especially to the extent it proceeds from an individualistic rather than a collective understanding of privacy and harm. The personal data threshold, as discussed, routinely breaks down in the context of AI systems that often make sensitive inferences about people (or the communities they are part of) based on discrete non-personal data categories. More broadly, this legal threshold focus is a somewhat incongruous lens when the objective is to minimize harms that emanate from algorithmic profiling, which is often on the basis of classes, aggregates, and patterns.[18] In *Overseers of the Poor*, John Gilliom has powerfully argued that the prevailing privacy paradigm *"relies on and maintains the idea of the autonomous individual and the idea of surveillance as mere visitation"* whereas in fact modern surveillance categorizes all individuals into frames for bureaucratic analysis translating all *"ongoing actions into tactics of compliance, evasion, and above all,*

---

[14] UK ICO, Guidance on the AI auditing framework: Draft for Consultation (2020), https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

[15] *See* European General Data Protection Regulation, Article 25. The 2018 Canadian House of Commons Report on PIPEDA reform clearly recommends that PIPEDA be amended to include Privacy By Design as a critical component https://www.ourcommons.ca/Content/Committee/421/ETHI/Reports/RP9690701/ethirp12/ethirp12-e.pdf.

[16] *See* European General Data Protection Regulation, Article 35.

[17] Michael Veale et al, Some HCI Priorities for GDPR-Compliant Machine Learning (2018) , https://arxiv.org/abs/1803.06174.

[18] *See* Kate Crawford & Jason Schultz, Big Data and Due Process: Toward a Framework to Redress Predictive Privacy, Boston College Law Review 2014 https://lawdigitalcommons.bc.edu/cgi/viewcontent.cgi?article=3351&context=bclr; Taylor, L., Floridi, L., van der Sloot, B. eds. Group Privacy: new challenges of data technologies, 2017.

*calculation*."[19] Often it is the assessment of AI technologies in the welfare context that melds privacy with broader disparities of social and economic justice and triggers a community (rather than individualistic) approach. In 2017, for example, the Indian Supreme Court while evaluating the lawfulness of a biometric ID system for access to welfare noted that dignity (rather than "the right to be left alone") was in fact the core value protected by the right to privacy. When someone is denied her welfare benefit because a technical system (biometric recognition) failed to identify her as deserving, the judges recognized a violation of privacy in this denial of both dignity and autonomy.[20] These broader notions of privacy go beyond a limited focus on personal identifiers and individual impact that can provide the foundations for a more holistic regulatory approach to AI systems. We therefore recommend the OPC evaluate alternative frameworks for complementary provisions and approaches that can be operationalized with the existing PIPEDA framing yet capture the broader categories of harm, interests, and remedies that are not adequately addressed by PIPEDA.

## PROPOSAL THREE: THE RIGHT TO OBJECT

### (GDPR) Right To Object To Processing In PIPEDA

At the outset, it is worth noting that Article 21 Right to Object to Processing, as it exists in the GDPR, applies only to non-consensual data processing, i.e. where the legal basis of the processing is *not* consent but rather the performance of a contract[21] or carried out in the "public interest"[22] or as part of "a task carried out by official authorities"[23] among others. In these situations the GDPR safeguards individual choice and autonomy even where some actors may argue there are overriding legal interests to process data. This is noteworthy because the PIPEDA does not contain analogous non-consensual grounds. If and when the PIPEDA does expand its grounds of processing similar to the GDPR, we would recommend that the right to object to processing be included. Where consent has been obtained, the right and ability to withdraw consent would fulfil the same function and so the right to object may not be necessary.

That said, we would not overemphasize the practical value of a right to object in these contexts of non-consensual data processing. It is unclear how this right can be meaningfully exercised without detriment to the individual, especially where there is an extreme imbalance of power and resources between the individual and the party undertaking the data processing. Take the example of a government ADS that processed data without taking individual consent, relying

[19] John Gilliom, Overseers of the Poor (2001), https://www.press.uchicago.edu/ucp/books/book/chicago/O/bo3626685.html, See also Simone Browne, Dark Matters: On the Surveillance of Blackness (2015), https://read.dukeupress.edu/books/book/147/Dark-MattersOn-the-Surveillance-of-Blackness.
[20] K.S Puttaswamy v Union of India, Supreme Court of India 2017, https://indiankanoon.org/doc/127517806/.
[21] *See* European General Data Protection Regulation, Article 6(b).
[22] *See* European General Data Protection Regulation, Article 6(e).
[23] *See* European General Data Protection Regulation, Article 6(e).

instead on the GDPR ground of a task carried out by a public authority. While the right to object might provide an individual the avenue to opt out of having her data processed in the AI system, this does nothing to address the impact that such choice will have on the individual. If she chooses to object, will this result in denial of services? Under the GDPR's Recital 42, consent should not be regarded as freely given if the subject is *"unable to refuse...without detriment"*.[24] Similar safeguards should be built-in to the logic of the right to object.

More fundamentally, this calls attention to the limits of the right to object model, especially when it comes to government data processing and use of AI systems. It is critical to supplement these rights with accountability mechanisms that ensure that AI systems are designed to protect and uphold meaningful choice. It is also worth noting that many kinds of government AI (like predictive policing systems) might be out of the scope of this protection entirely because they involve non-individualized, community-wide assessments.

**Prohibition Against Solely Automated Decisions And Human-in-the-loop Requirements**

GDPR's Article 22 operates as a general prohibition (and *not* a right, as the OPC suggests)[25] against decisions based solely on automated processing that have legal or similar impact on individuals. Even in the limited situations where purely automated decisions are permitted under the GDPR,[26] the law requires that there are ways to request for meaningful human intervention.

While these efforts are well meaning, we would caution against regulating based on a rigid distinction between "solely" automated decisions versus decisions that are informed, aided or supported by algorithms. In practice, these distinctions are slippery and the fact that there is human intervention in the final decision does not address major concerns over opacity or control and should not be automatically presumed to be at a lower risk level. In fact, where AI systems are used as "decision making aids", research by Ben Green and Yiling Chen demonstrates that humans are often unable to accurately evaluate the quality or fairness of the predictions made. People fail to rely more heavily on accurate predictions compared to inaccurate predictions, and often respond to predictions in biased and inaccurate ways.[27] This follows from a large body of research showing that people struggle to effectively interpret, use, and oversee algorithms when making decisions.[28]

---

[24] See European General Data Protection Regulation, Recital 42.
[25] Michael Veale & Lillian Edwards, Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling, 2017 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3071679.
[26] For example if the data processing is justified under the ground of performance of contract, or explicit consent of the data subject.
[27] See Ben Green & Yiling Chen, Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments, https://www.benzevgreen.com/wp-content/uploads/2019/02/19-fat.pdf; Ben Green & Yiling Chen, The Principles and Limits of Algorithm-in-the-Loop Decision Making, https://www.benzevgreen.com/wp-content/uploads/2019/09/19-cscw.pdf.
[28] Megan T. Stevenson, Assessing Risk Assessment in Action (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3016088; Berkeley J. Dietvorst, et al., Algorithm

It should not be assumed that having a person overseeing an algorithm means that there is sufficient quality control over the algorithm's decisions or predictions. It is often assumed that when algorithms are decision making aids, the people who make the final decisions will provide an important quality control on a model's determinations. Yet such behavior requires people to evaluate the quality of determination, to calibrate their decisions based on these evaluations, and to respond to the algorithm's output in a timely manner and without bias—abilities that research indicates people do not reliably possess.[29] Governance must therefore consider the full sociotechnical system of the human-algorithm relationship, rather than consider the algorithm or human in isolation.

We would encourage the OPC to adopt a regulatory regime that incentivizes AI systems and human-algorithm interactions that enhance real capacity for human oversight - and restrict the use of AI systems where such oversight cannot be meaningful. In sensitive social domains, where governments are adopting AI systems to determine the allocation of welfare benefits/services or deciding criminal justice outcomes, the consequences of overestimating human oversight could have consequences on basic civil liberties. In other high risk domains too like self-driving cars or automated pilots, research has found over-reliance of drivers[30] or pilots[31] on automated systems led to complacency and a degradation in manual skills eventually putting human life at risk. These examples highlight the complex (and non-binary) relationship between automation and human intervention, and the need to exercise the highest level of caution and reflexivity when implementing these systems in social domains. For these reasons, we recommend that AIAs place equal emphasis on human-algorithm interaction i.e. how people take action (or fail to) on the basis of any specific prediction or result from the AI system.

We also recommend that AIAs include an internal assessment of the knowledge differentials or inefficiencies that limit accountability and contribute to their inability to adequately assess and anticipate problems that may arise from such systems. The UK ICO's recent draft auditing framework has some useful guidance on documenting these limits on human capacity to

---

Aversion: People Erroneously Avoid Algorithms After Seeing them Err (2015), https://psycnet.apa.org/fulltext/2014-48748-001.html; Amirhossein Kiani, et al., Impact of a deep learning assistant on the histopathologic classification of liver cancer (2020), https://www.nature.com/articles/s41746-020-0232-8.

[29] Ben Green & Yiling Chen, The Principles and Limits of Algorithm-in-the-Loop Decision Making, https://www.benzevgreen.com/wp-content/uploads/2019/09/19-cscw.pdf.

[30] John Markoff, Google's Next Phase in Driverless Cars: No Steering Wheel or Brake Pedals, 2014 https://www.nytimes.com/2014/05/28/technology/googles-next-phase-in-driverless-cars-no-brakes-or-steering-wheel.html.

[31] House Committee on Transport & Infrastructure, The Boeing 737 MAX Aircraft: Costs, Consequences, and Lessons From its Design, Development, and Certification , 2020 https://transportation.house.gov/imo/media/doc/TI%20Preliminary%20Investigative%20Findings%20Boeing%20737%20MAX%20March%202020.pdf.

engage with the AI systems.[32] They recommend documenting not just potential risks emanating from these systems, but also the capacity of those interacting with the system to recognize such risks. Where risks and strategies of mitigation (if they exist) are identified, they encourage creating a knowledge base that can be drawn upon by others interacting with the system. *Confronting Black Boxes: The Shadow Report of the New York City Automated Decision Systems Task Force* also discusses how automated decision system vendors can support efforts to increase government capacity to audit and evaluate these systems. It recommends that government agencies should require vendors to provide more training materials for agency staff to understand the system, in addition to requiring the vendor to collaborate with the agency in developing public-education materials and engaging the public.[33]

**PROPOSAL FOUR: RIGHT TO AN EXPLANATION IN THE PIPEDA**

Explanation rights are increasingly becoming a common feature in algorithmic accountability and data protection legislation and regulations;[34] however, there is less consensus or certainty on what such explanations should entail. Explanations are social and contextual, which means that the information transferred in an interaction is both dependent on and relative to the party responsible for providing the explanation and the recipient of that information.[35] Therefore, legally mandated explanations must account for the positionality of both parties, and the power, knowledge, and resource asymmetries that accompany such positions. Additionally, any legally mandated explanation must specify a timeframe when an explanation must be provided,[36] and the OPC must have authority to enforce non-compliance.

Explanations for individuals should include but are not limited to: (1) the types of decisions or situations being subjected to automated processing; (2) factors involved in a decision relying on automated processing operations (e.g.behavioral data; socioeconomic indicators; legally

---

[32] UK ICO, Guidance on the AI auditing framework: Draft for Consultation 2020, https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

[33] Rashida Richardson, ed., Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force (2019), https://ainowinstitute.org/ads-shadowreport-2019.html.

[34] *See, e.g.,* California Consumer Privacy Act (2018), https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375; A Local Law to amend the administrative code of the city of New York, in relation to reporting on automated decision systems used by city agencies (2019), https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4265421&GUID=FBA29B34-9266-4B52-B438-A772D81B1CB5&Options=&Search=.

[35] Tim Miller, Explanation in Artificial Intelligence: Insights from social sciences (2019), https://www.sciencedirect.com/science/article/pii/S0004370218305988.

[36] See,e.g., United Kingdom Information Commissioner's Office, Right to Object, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-object/ (providing one month for a response for an explanation); United States Department of Justice, *Responding to Requests* (2018), https://www.justice.gov/archives/open/responding-requests (requiring a response within 20 days with procedures to extend the response period for "unusual circumstances).

defined categories of data;[37] location data); (3) descriptions of the types of data used in automated processing; (4) a legible description of the methodology and mechanism underlying the automated processing (e.g. “this technology employs a linear regression model to predict who will succeed”); (5) a description of how the automated decision is being used by humans to make a decisions (e.g. specific details about a public official or employee is interpreting or applying the outputs of an automated process); (6) a description of potential legal or other significant effects or consequences of automated processing. When the automated processing operations are run or facilitated through a government agency or authority, additional explanation requirements should exist. Specifically, these explanations should include: (1) how use of automated processing operations is related to the government agency’s mission or interests, and (2) a description of relevant appeal or redress processes.

Operationalizing these explanation requirements will require robust documentation practices in the development and use of automated processing systems. Robust documentation is necessary to understand the various decisions and choices made during design, which can involve normative value judgments or appear more objective due to pretextual explanations.[38] When government agencies are responsible for providing an explanation for an automated processing system acquired from a third-party vendor, we recommend that the agency require the vendor to develop such explanations as part of the procurement contract.[39] Additionally, considering the variety of structural barriers associated with request-based explanation models, governments should evaluate approaches to proactively making such information public and accessible.

## PROPOSAL SIX: WORKABILITY OF DATA MINIMIZATION & PURPOSE LIMITATION IN THE CONTEXT OF AI and IMPACT ON AI PROGRESS

The legal principles of purpose limitation, purpose specification and data minimization not only “work” in an AI context, but we would argue they create a healthy tension with other goals of AI systems. Fundamentally, they require reflexivity, prudence and proportionality with respect to any use of data. This is *more*, not less, valuable in the context of AI systems. We acknowledge that perfect compliance to these principles might sometimes be challenged by technical realities, as the OPC points out, and the enforcement regime should proactively take these into account and issue guidance to clarify their application. Overall however, as we argue below, the limits they put in place and the reflexivity they demand of private companies and government agencies is essential and should not be dispensed with.

---

[37] Where relevant legal or regulatory regimes determine forms of classification or which entities are subject to such classification. For example, labor laws create and define labels such as “employer” and “employee” and these classifications are encoded into datasets or algorithmic models in accordance.
[38] Roel Dobbe, Thomas Krendl Gilbert, Yonatan Mintz, Hard Choices in Artificial Intelligence: Addressing Normative Uncertainty through Sociotechnical Commitments (2019), https://arxiv.org/pdf/1911.09005.pdf.
[39] Rashida Richardson, ed., Confronting Black Boxes: A Shadow Report of the New York City Automated Decision System Task Force, p.21 (2019), https://ainowinstitute.org/ads-shadowreport-2019.pdf.

In a blog from the UK ICO's office, researchers Reuben Binns and Valeria Galle explain a variety of techniques that can be used in the training and inference stage of developing machine learning-based AI that process as little personal data as possible while still producing a functional AI model.[40] These principles can also put brakes on harmful uses of AI to begin with - take the example of facial and other forms of biometric recognition. Data minimization would demand that sensitive facial data is only collected when absolutely necessary and where no less intrusive forms of data collection can fulfil the objective. Similarly, purpose limitation would require that any purposes for which these systems are used should be strictly specified and documented in advance of deployment, which means analyzing facial data for additional purposes as technology evolves (like to predict race, ethnicity, or emotion) would not be permitted. These are some of the ways that data minimization and purpose limitation can be operationalized to limit if and when (not just how) these technologies are developed and deployed, at a time when there is widening concern around the proliferation of these technologies in sensitive social domains like public places, employment, schools and the criminal investigation process. Applying the analogous Article 5 of the GDPR, the Swedish Data Privacy Authority (DPA) recently found that processing facial data of children went beyond the data minimization principle and was disproportionate to the purpose sought to be achieved.[41] If registering attendance was the goal, the Swedish DPA concluded that this was a disproportionate means to achieve this goal and far less intrusive means could have been deployed. It is easy to see how this ruling might translate and be influential to other domains such as employment or the criminal justice system.

The other core concern raised by the OPC is that the implementability of these principles is challenged because machine learning systems are black boxes that do not lend themselves to clear predetermined purposes. This understanding stems from the premise that AI systems and their outcomes are entirely technologically determined. We would argue that how these systems are developed and what goals they are optimized to achieve is heavily determined by the social and institutional priorities of the organizations that develop and implement them, like efficiency, or profit maximization. These institutional choices are extremely relevant to any robust understanding of what and how an AI system functions, and should be scrutinized under the purpose limitation test. Moreover, where the purpose of an AI system does change from the use for which consent was originally sought, then obtaining fresh consent is a necessary procedural and substantive safeguard. The importance of triggering legal rights when the context of data processing changes was demonstrated in the recent controversial Clearview AI reporting, which revealed that people's face data was being collected from social media and then used for an

[40] Reuben Binns & Valeria Gallo, Data minimization and privacy preserving techniques in AI systems (2019), https://ico.org.uk/about-the-ico/news-and-events/ai-blog-data-minimisation-and-privacy-preserving-techniques-in-ai-systems/.
[41] Sofia Edwardson, How to interpret Sweden's first GDPR fine on facial recognition in school (2019), https://iapp.org/news/a/how-to-interpret-swedens-first-gdpr-fine-on-facial-recognition-in-school/.

app that was accessed by many public, private, and law enforcement agencies, without their knowledge.[42] As the UK ICO notes "the fact that some data might later in the process be found to be useful for making predictions is not enough to establish its necessity for the purpose in question, nor does it retroactively justify its collection, use or retention."[43]

The next concern raised by the OPC is the potential negative impact of data protection norms on Canada's technical and economic AI progress. We urge the OPC to unequivocally oppose this false binary, especially at a time when it is being used to lobby against these laws or to dilute data protection provisions where they exist.[44] While it may be true that some of the most successful consumer-facing AI companies today are also those that have been able to collect the largest and most granular datasets, it is the privacy harms caused by the very same companies that have prompted strict data collection norms globally. Rather than see these regulatory frameworks as "slowing down" technological or business progress, we would argue they set the boundaries within which such innovation should happen and set the terms of technological progress within broader social and political goals.

**PROPOSAL SEVEN: ADDITIONAL LEGAL GROUNDS FOR PROCESSING BEYOND CONSENT**

We endorse the recommendation of the 2018 Canadian House of Commons Standing Committee on Access to Information, Privacy and Ethics Report that *"consent remain the core element of the privacy regime, but that it be enhanced and clarified by additional means, when possible or necessary".*[45] The crisis of faith around consent stems from the lack of meaningful consent in a wide range of private and public activities due to imbalance of bargaining power and imperfect information and cognition. The solution to this, as noted by the HoC Committee, is not to abandon the framework of consent, but to create structural solutions that depend less on the choices individuals make and instead put in place accountability mechanisms that collectively enhance privacy.[46]

In the context of AI, we have noted that consent does play a role in ensuring notice and due process around the use of personal data in these systems, including the ability for individuals to

---

[42] Kashmir Hill, Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich(2020), https://www.nytimes.com/2020/03/05/technology/clearview-investors.html.
[43] Reuben Binns & Valeria Gallo, Data minimization and privacy preserving techniques in AI systems (2019), https://ico.org.uk/about-the-ico/news-and-events/ai-blog-data-minimisation-and-privacy-preserving-techniques-in-ai-systems/.
[44] *See generally* Graham Webster & Scarlet Kim, The Data Arms Race is no Excuse for Abandoning Privacy (2018), https://foreignpolicy.com/2018/08/14/the-data-arms-race-is-no-excuse-for-abandoning-privacy/.
[45] House of Commons Canada Standing Committee on Access to Information, Privacy and Ethics, Towards Privacy By Design (2018). https://www.ourcommons.ca/Content/Committee/421/ETHI/Reports/RP9690701/ethirp12/ethirp12-e.pdf.
[46] Ibid.

refuse or to later withdraw consent. In an environment where there is little known about the types of AI being developed, consent can be a critical check on ensuring broader individual and public scrutiny. This is especially true at a time when personal data (including face or other identifying images) are routinely used to train AI systems and then identify people without their knowledge. As a recent example, several lawsuits were filed against the Clearview AI company under the Illinois Biometric Information Privacy Act (BIPA)[47] proceeding from the claim of lack of consent for the photographs used to train and then populate the facial recognition app. If consent had been obtained, it would have been an opportunity to scrutinize the system that created multiple layers of privacy concerns.

As the OPC considers adding additional grounds of processing to the PIPEDA similar to the GDPR, these should not operate as an escape clause for firms to avoid seeking consent. In the GDPR for example, each additional ground is narrowly scoped and the broadest ground i.e. "legitimate interests" of a business has a clear three step test of showing necessity, proportionality, and only proceeding with such a ground where consent cannot be meaningful. This documentation of why consent was not appropriate is an important check to ensure that firms do not abuse this broadly worded ground.

## PROPOSALS NINE AND TEN: DATA TRACEABILITY AND DOCUMENTATION REQUIREMENTS

Data traceability and documentation requirements are important mechanisms and practices to ensure algorithmic transparency and accountability by providing greater clarity on a system's underlying logic and design, as well as creating more opportunities for correction. The OPC should consider regulatory requirements and incentive structures that encourage private vendors to adopt documentation practices such as, data sheets,[48] fact sheets,[49] and model cards,[50] as part of the development and design process. A combination of regulation and incentives are necessary because documentation is both an artifact and a process. Emerging research and practical guidelines on documentation often include templates or questions that can be adapted by features of the system, which serves the artifact, but it is also a part of the design process that is context specific and requires constant evolution to scale.[51] Thus, a combination of regulatory approaches is necessary to capture this duality because relying on companies to comply with abstract or evolving principles can result in an artificial accountability

[47] Al Fowerbaugh et al, BIPA may apply to Clearview AI's creation of biometric database (2020), https://www.law360.com/articles/1244493/bipa-may-apply-to-clearview-ai-s-creation-of-biometric-data.
[48] Timnit Gebru, Datasheets for Datasets (2020), https://arxiv.org/abs/1803.09010.
[49] Matthew Arnold, et al., FactSheets: Increasing Trust in AI Services through Supplier's Declaration of Conformity (2019), https://arxiv.org/abs/1808.07261.
[50] Margaret Mitchell, Model Cards for Model Reporting (2019), https://arxiv.org/abs/1810.03993.
[51] Jingying Yang, *Bridging AI Principles to Practice with ABOUT ML* (2020), https://www.partnershiponai.org/bridging-ai-principles-to-practice-with-about-ml/; Partnership in AI, *About ML*, https://www.partnershiponai.org/about-ml/.

ceiling, where efforts towards robust accountability are stymied by mass adoption of less rigorous or incomplete standards.

Moreover, while data traceability and documentation are effective approaches to ensuring greater transparency, they are not the panacea for corrective and accountability concerns.[52] This is particularly true for sensitive social domains, like policing and immigration,[53] where the efficacy or value of transparency is fettered by the need for broader structural reform. Simply knowing where data comes from, and how it is constructed, would not fix the problematic practices that served to create the data in the first place. For example, in our recent law review essay, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, we examined the use of predictive policing in United States jurisdictions with documented histories of racially biased and unlawful policing practices and policies. We found that these policing practices and policies skew police data so that it does not accurately represent actual crime trends or rates but rather reflects the department's policing practices, policies, and priorities. The data is an artifact of flawed practices, not an accurate reflection of crime activity. The reliance on this flawed data in a predictive model produced a confirmation feedback loop where in some of the jurisdictions the predictive policing system's forecasts predominantly targeted the same demographic that was disproportionately affected by the police department's unlawful and biased practices. In most of the jurisdictions reviewed, the implicated police departments were required to comply with various practice and policy reforms, but few of these reforms contemplated how police data is shared and used by private companies or other government agencies, and how it informs AI technologies applied in law enforcement contexts. This lack of oversight in institutional reform efforts can inadvertently limit the reach and efficacy of data traceability and documentation approaches. Conversely, data traceability and documentation practices that fail to contemplate the social and political contexts of relevant datasets and data-reliant decisionmaking, can serve to distort structural problems or impede necessary structural or institutional reforms. Given these concerns, we recommend the OPC consult with government and civil society organizations, like the Office of the Independent Police Review Director in Ontario, who can offer important insights on systemic problems in sensitive social domains and potentially serve as a partner in developing holistic interventions.

## PROPOSAL ELEVEN: EMPOWERING THE OPC TO ISSUE BINDING ORDERS AND FINANCIAL PENALTIES

The proposal to empower the Privacy Commissioner of Canada with coercive law enforcement authority, primarily through order-making powers and imposition of financial penalties, is an important improvement to PIPEDA that will ensure compliance and additional forms of redress

[52] Mike Ananny & Kate Crawford, *Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability* (2018), https://doi.org/10.1177/1461444816676645.
[53] Citizen Lab, Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System (2018) ,https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf.

to potential rights violations. Though PIPEDA and similar data protection laws from other jurisdictions provide a private right of action as a form of redress for violation of privacy and other rights, relying on individual challenges on compliance or private litigation as a primary means of curtailing or mitigating noncompliance with PIPEDA is not adequate. In addition to recommending a coercive law enforcement authority, we also provide suggestions for how to reform the existing private right of action model.

Litigation in a federalist system is oriented towards mediating power between state actors, not individuals and institutions, which means that litigation as a primary remedial or deterrence mechanism is limited. In the United States, we observed some of these limitations from analyzing legal challenges to the use of automated processing systems in our *Litigating Algorithms 2019 US Report* and our *2018 Litigating Algorithms Report*.[54] We found that the issues being litigated are often entangled with structural or systemic problems in government agencies and society (e.g. discrimination and chronic divestment), so litigation as the sole remedial mechanism depends on the legitimacy and perogatives of the judiciary as well as the disposition of potential plaintiffs. For example, the BIPA provides a private right of action for the failure to provide notice and obtain consent for the collection, sale, purchase, storage or distribution of biometric data.[55] Although BIPA was enacted in 2008, there has been great uncertainty regarding standing to bring BIPA claims. Last year, the Illinois Supreme Court held that an individual does not have to demonstrate actual harm to establish a claim under BIPA;[56] yet questions persist as to whether class action plaintiffs can bring BIPA claims in federal courts.[57] In addition to this looming uncertainty, the ability to pursue a lawsuit under BIPA requires financial resources to hire a private attorney and the ability to find a lawyer with expertise or interest in pursuing BIPA claims, which are high enough burdens that can serve to preclude individuals and communities at greatest risk for potential violations from seeking redress. Though the 2019 Illinois Supreme Court decision stated BIPA is intended to be a preventative measure to incentivize data protection; the inherent barriers to using this remedial measure may stymie the efficacy for compliance enforcement. Therefore, we also encourage the OPC to explore BIPA-style regulation with a private right of action, standing to sue for statutory

[54] AI Now Institute, NYU Law Center on Race Inequality and the Law, Electronic Frontier Foundation, Litigating Algorithms: Challenging Government Use of Automated Decision Systems (2018), https://ainowinstitute.org/litigatingalgorithms.pdf; Rashida Richardson, Jason M. Schultz, & Vincent M. Southerland, Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems (2019), https://ainowinstitute.org/litigatingalgorithms-2019-us.html.
[55] Washington and Texas have similar laws regulating the collection and storage of biometric data, but they do not provide a private right of action for statutory damages.
[56] *Rosenbach v. Six Flags Entertainment Corporation, et al.*, 2019 IL 123186 (Ill. 2019).
[57] There is currently a circuit split regarding whether statutory violations are sufficient to confer constitutional standing ("Article III standing") following the United States Supreme Court decision in *Spokeo v. Robins*, 136 S. Ct. 1540 (2016). In January 2020, the Supreme Court denied *certiorari* to an appeal of a BIPA class action lawsuit against Facebook Inc, which means uncertainty regarding Article III standing remains.

violations, statutory fines for each violation, and public resources to support indigent plaintiffs.[58] A multi-faceted enforcement regime can help incentivize compliance and provide financial redress for aggrieved parties.

[58] Though we acknowledge some of the deficiencies with BIPA and private litigation as remedial and law enforcement measures, BIPA claims have led to significant monetary settlements for aggrieved parties so an amended version addressing the aforementioned concerns can still be an useful intervention.