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

PAR COURRIER ÉLECTRONIQUE

Le 20 décembre 2021

Monsieur Dubi Kanengisser, Ph. D.
Conseiller principal, analyse stratégique et gouvernance
Commission de services policiers de Toronto
40, rue College
Toronto (Ontario) M5G 2J3

Objet : Consultation publique de la Commission de services policiers de Toronto sur sa *Politique concernant l'utilisation des nouvelles technologies d'intelligence artificielle*

Monsieur,

Au nom du Bureau du commissaire à l'information et à la protection de la vie privée de l'Ontario (CIPVP), j'ai le plaisir de vous faire part de nos observations sur le projet de Politique concernant l'utilisation des nouvelles technologies d'intelligence artificielle (la « politique ») de la Commission de services policiers de Toronto (la « Commission »). Nous nous réjouissons que la Commission reconnaisse la nécessité d'une gouvernance et d'une surveillance solides des technologies qui utilisent ou intègrent l'intelligence artificielle (IA), et nous la félicitons d'avoir pris l'initiative de solliciter les commentaires du public, comme cela s'imposait.

Les recommandations qui suivent ont pour but de clarifier et de renforcer la politique et d'aider la Commission à améliorer sa capacité à régir l'utilisation de l'IA d'une manière qui protège les droits en matière de vie privée et d'accès à l'information ainsi que d'autres droits fondamentaux. Comme la politique est conçue pour s'appliquer à un large éventail d'applications et de technologies actuelles et éventuelles de l'IA, nos observations sont de portée générale et se concentrent sur les questions clés, les principes essentiels et les prochaines étapes recommandées. Ces observations reflètent l'évolution constante de l'approche de mon bureau en matière d'IA, compte tenu de la nature dynamique de ce domaine.

Le CIPVP continuera à collaborer avec les parties prenantes pour promouvoir la gouvernance responsable de l'IA dans le contexte du maintien de l'ordre, dans le cadre de notre priorité stratégique *La nouvelle génération des forces de l'ordre*. Nous sommes disposés à collaborer avec la Commission et le Service de police de Toronto relativement à la politique et aux procédures connexes.

2 Bloor Street East
Suite 1400
Toronto, Ontario
Canada M4W 1A8

2, rue Bloor Est
Bureau 1400
Toronto (Ontario)
Canada M4W 1A8

Tel/Tél : (416) 326-3333
1 (800) 387-0073
TTY/ATS : (416) 325-7539
Web : www.ipc.on.ca

Par souci de transparence envers la population ontarienne, la présente lettre et les pièces jointes seront publiées sur notre site Web.

Veuillez agréer, Monsieur, mes sincères salutations.

Patricia Kosseim
Commissaire

Résumé des recommandations du CIPVP à la Commission

Clarifier l'objet de la politique

La politique serait plus efficace si elle était assortie de définitions et d'un champ d'application clairs. Cette clarté aiderait le Service de police de Toronto (le Service) à interpréter, à appliquer et à respecter la politique de manière uniforme, et le public à comprendre l'étendue des mesures de précaution que prévoit la politique.

1. Donner une définition brève et claire de l'IA en langage simple.

2. Clarifier la définition de « biais » afin qu'elle se concentre sur les résultats plutôt que sur la cause d'un extrant biaisé.

3. Inclure une section relative au champ d'application, avec des critères et des exemples concernant les types de technologies et d'utilisations de l'IA qui font partie de ce champ et ceux qui n'en font pas partie, et étendre explicitement l'application de la politique à l'utilisation des technologies d'IA (y compris les essais gratuits) par les membres du Service dans le cadre de leur travail.

4. Inclure explicitement, dans le champ d'application de la politique, des technologies d'IA qui s'appuient sur des renseignements anonymisés ou des renseignements concernant des groupes ou des communautés.

5. Exiger que le Service évalue les technologies d'IA qu'il utilise déjà en regard du cadre d'évaluation des risques de la politique.

6. Au sujet du cadre d'évaluation des risques :

   a) préciser que les catégories de risques à inclure dans le système de classification des risques doivent être fondées sur des catégories spécifiques de préjudices, leur gravité et leur probabilité, plutôt que sur des applications;

   b) exiger que les évaluations des risques soient axées sur le contexte dans lequel les technologies d'IA seront utilisées, y compris la façon dont de multiples systèmes d'IA peuvent être utilisés par le Service d'une manière qui, cumulativement, peut donner lieu à de nouveaux risques;

   c) exiger que les parties prenantes participant à l'élaboration du système de classification des risques comprennent des membres des communautés touchées.

Renforcer la transparence, la responsabilisation et la surveillance

La politique comprend des mécanismes conçus pour assurer la transparence, la responsabilisation et la surveillance de l'utilisation de l'IA par le Service. Toutefois, nous recommandons l'adoption de mesures supplémentaires.

7. Prévoir un mécanisme de dénonciation permettant aux membres du Service de signaler à la Commission, de manière anonyme et sécurisée, les infractions à la politique.

8. Prévoir dans la politique une description claire des rôles et des responsabilités en matière de réalisation des évaluations des risques et de participation à ces évaluations, et des modalités prévoyant que le Service est tenu de disposer d'une expertise interdisciplinaire et de perspectives diverses, ainsi que des ressources et du temps nécessaire pour un processus d'évaluation solide.

9. Établir des exigences en matière de tenue de registres pour l'IA prévoyant la conservation des renseignements personnels utilisés par les technologies d'IA et l'obligation de documenter les activités du Service faisant intervenir l'IA.

10. Élargir les renseignements divulgués sur le site Web public concernant l'IA conformément aux recommandations de la Commission du droit de l'Ontario dans *Réglementer l'intelligence artificielle : Enjeux et choix essentiels* et du gouvernement du Royaume-Uni dans sa *norme sur la transparence algorithmique*. Il devrait notamment y avoir divulgation proactive des évaluations de l'incidence algorithmique, sous réserve d'exceptions légitimes en vertu de la *Loi sur l'accès à l'information municipale et la protection de la vie privée*, ou du moins de résumés de ces évaluations.

Assurer la mise en œuvre de la politique

La politique prévoit que le Service doit élaborer un grand nombre de processus et de procédures à l'appui du cadre de gestion des risques qu'elle contient. En ce qui concerne la surveillance de l'intervention humaine et l'explicabilité de l'IA, nos recommandations sont les suivantes :

11. Prévoir des exigences relatives aux processus nécessitant une intervention humaine comprenant ce qui suit :

    a) Mettre au point un processus pour déterminer les aspects de la prise de décision qui devraient être automatisés et ceux qui devraient recourir à l'intervention humaine. Ce processus devrait évaluer :

        i. le rôle de la latitude dans la prise de décision;

        ii. la précision, le rendement et la robustesse du système d'IA envisagé;

        iii. les conséquences d'une mauvaise décision.

    b) S'assurer que les personnes chargées de l'intervention humaine ont le temps, les ressources, les renseignements et les capacités nécessaires pour bien examiner les décisions automatisées ou les prédictions.

    c) Prévoir des exigences de consignation et de tenue de registres concernant les activités humaines d'examen et d'intervention.

    d) Inclure des mesures sur la surveillance humaine parmi les indicateurs dont il faut rendre compte à la Commission.

12. Considérer l'explicabilité comme un facteur atténuant et l'absence d'explicabilité comme un facteur aggravant lors de l'évaluation des préjudices éventuels dans le cadre du processus d'évaluation des risques, et fixer des exigences de base en matière d'explicabilité qui doivent être respectées sans égard au niveau de risque.

Étapes subséquentes

Si la politique est un premier pas important vers une gouvernance efficace de l'IA, de nombreuses autres étapes doivent suivre. Un engagement continu avec le public et le CIPVP sera essentiel.

13. Continuer de consulter un large éventail de parties prenantes de la collectivité, y compris le CIPVP, dans le cadre des travaux de la Commission aux fins de la mise au point de sa politique.

14. S'assurer que le Service fait également appel à un large éventail de parties prenantes, y compris le CIPVP, aux fins de l'élaboration de processus et de procédures de mise en œuvre de la politique.

15. Prévoir l'élaboration de critères afin de déterminer les circonstances où une consultation publique devrait faire partie du processus d'évaluation des risques préalable à la mise en œuvre.

OBSERVATIONS DU CIPVP SUR LA *POLITIQUE CONCERNANT L'UTILISATION DES NOUVELLES TECHNOLOGIES D'INTELLIGENCE ARTIFICIELLE* DE LA COMMISSION DE SERVICES POLICIERS DE TORONTO

Le Bureau du commissaire à l'information et à la protection de la vie privée de l'Ontario (CIPVP) a le plaisir de faire part de ses observations à la Commission de services policiers de Toronto (la « Commission ») sur son projet de Politique concernant l'utilisation des nouvelles technologies d'intelligence artificielle (la « politique »).

Le CIPVP est chargé de la surveillance indépendante d'organismes assujettis à la *Loi sur l'accès à l'information municipale et la protection de la vie privée* (*LAIMPVP*). En vertu de cette loi, le CIPVP fait enquête sur des plaintes, règle des appels, examine des pratiques de protection de la vie privée et de gestion de l'information, mène des recherches et fournit des conseils et des observations sur les initiatives de palier municipal concernant l'accès à l'information, la protection de la vie privée et, par extension, la transparence et la reddition de comptes en matière de renseignements.

Les présentes observations s'appuient sur nos travaux antérieurs relatifs à l'intelligence artificielle (IA), notamment les suivants :

a) *Observations du CIPVP dans le cadre des consultations du gouvernement de l'Ontario sur le cadre de l'intelligence artificielle (IA) de confiance de l'Ontario* (observations sur le cadre de l'IA de confiance);

b) Observations fournies à la Commission par des membres du personnel sur une ébauche antérieure de la politique, en juin 2021;

c) Contributions au *Document d'orientation préliminaire sur la protection de la vie privée à l'intention des services de police relativement au recours à la reconnaissance faciale* (document d'orientation sur la reconnaissance faciale), publié conjointement par les autorités de la protection de la vie privée du Canada;

d) *Modèle de cadre de gouvernance pour les programmes de caméras d'intervention des services de police de l'Ontario*, éclairé par des consultations approfondies avec la Commission et le Service;

e) Commentaires du CIPVP sur le livre blanc du gouvernement de l'Ontario intitulé *Modernisation de la protection de la vie privée en Ontario* (commentaires sur la protection de la vie privée dans le secteur privé).

Contexte général

L'utilisation de l'IA par la police pourrait améliorer la sécurité publique et favoriser l'innovation axée sur les données dans les services de police. Cependant, elle soulève également d'importantes questions liées à la protection de la vie privée, à la transparence et à la responsabilisation, dont bon nombre sont décrites dans nos commentaires ci-après.

La consultation de la Commission sur sa politique intervient à un moment critique pour la confiance du public tant dans l'IA que dans les services de police. Un sondage mondial réalisé par Edelman en 2021 a révélé que seulement 39 % des Canadiens croient que l'on peut faire confiance aux entreprises de technologie d'intelligence artificielle, soit une baisse de 5 % depuis 2020[1]. En outre, une récente enquête du gouvernement du Canada a révélé que les Canadiens craignent que l'IA ait des conséquences négatives dans le secteur de l'application de la loi, plus que tout autre secteur[2]. Une étude fondée sur des entrevues menées en 2019 concernant l'opinion de la population canadienne de l'IA, menée par Deloitte, a permis d'observer que les biais fondés sur la race et le genre « influent largement sur l'opinion publique »[3].

Des sentiments de méfiance et des expériences de discrimination associés au maintien de l'ordre ont également été communiqués à la Commission dans le cadre d'enquêtes et de réunions publiques organisées ces dernières années[4]. Il est indéniable que les enjeux sont de taille pour la Commission, le Service de police de Toronto (le Service) et le public. Les systèmes d'IA doivent être utilisés de manière à respecter la vie privée, à favoriser la transparence et à promouvoir

---

[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, Sciences et Développement économique Canada (mai 2021). « Opinions des Canadiens sur l'intelligence artificielle : Rapport final », https://ised-isde.canada.ca/site/recherche-opinion-publique/fr/opinions-canadiens-lintelligence-artificielle-rapport-final

[3] Deloitte (2019). *Impératif de l'IA au Canada - Surmonter les risques, instaurer la confiance*, https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/deloitte-analytics/ca-overcoming-risks-building-trust-aoda-fr-updated.pdf

[4] Voir notamment Commission de services policiers de Toronto (août 2020). « '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; Commission de services policiers de Toronto (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

l'équité. Ils doivent servir à réduire, et non à exacerber, les formes actuelles de discrimination, notamment dans le contexte des interventions policières insuffisantes ou excessives.

Nous saluons les efforts de la Commission en vue d'interdire l'utilisation des technologies d'IA présentant des risques qu'elle juge inacceptables, et à élaborer un cadre et un modèle de gouvernance axés sur les risques. Il s'agit là d'une approche semblable à celle de la législation sur l'intelligence artificielle proposée dans l'Union européenne, qui associe les exigences relatives à la gouvernance au niveau de risque de l'IA[5]. Le concept de base de la politique, qui consiste à assujettir les utilisations de l'IA présentant un risque accru à des mesures de précaution plus strictes, constitue un fondement conceptuel raisonnable pour la gouvernance de ce domaine en évolution rapide. Cependant, comme nous l'indiquons dans nos recommandations, il reste beaucoup à faire pour que la politique remplisse son objectif.

Les commentaires qui suivent sont divisés en quatre catégories :

1. Clarifier l'objet de la politique
2. Renforcer la transparence, la responsabilisation et la surveillance
3. Mettre en œuvre la politique
4. Étapes subséquentes

1. Clarifier l'objet de la politique

Comme il est indiqué dans nos observations sur le cadre de l'IA de confiance, l'efficacité d'un modèle de gouvernance de l'IA repose sur des définitions et une portée claires. Cette clarté aidera le Service à interpréter, à appliquer et à observer la politique de façon uniforme. Elle aidera aussi le public et les organismes de réglementation à comprendre les exigences de la politique et éclairera l'adoption de normes et d'attentes raisonnables.

1.1. Clarifier les définitions des principaux termes

Définir l'« intelligence artificielle » (ou des termes similaires, comme la prise de décision automatisée) est une tâche difficile. Or, la définition adoptée par la Commission est inutilement compliquée, car elle conjugue des caractéristiques techniques des technologies avec des déclarations sur le champ d'application de la politique. La définition inclut également [traduction] « tout bien ou service dont l'acquisition, le déploiement ou l'utilisation exigent la réalisation d'une évaluation de l'incidence sur la vie privée », ce qui pourrait faire entrer un large éventail de biens et de services non liés à l'intelligence artificielle dans le champ d'application de la politique.

Nous recommandons à la Commission de modifier la politique afin de donner une définition brève et claire de l'IA en langage simple.

La définition de « biais » de la politique pourrait être simplifiée en se concentrant sur le résultat plutôt que sur la cause du biais. Il n'est pas toujours facile d'établir si un extrant déficient est attribuable à un défaut de conception d'une technologie d'IA ou aux données d'apprentissage. Par

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

exemple, des données d'apprentissage peuvent constituer le reflet fidèle de données générées dans une situation inéquitable. Ces données d'apprentissage, et les extrants qui sont produits au moyen de ces données, sont-ils donc « déficients » en vertu de cette définition? En axant la définition sur le résultat, par exemple, « extrant qui identifie systématiquement de façon incorrecte certains types de sujets ou leur attribue des caractéristiques qui leur sont défavorables pour des motifs illégitimes (p. ex., des motifs protégés par le *Code* [*des droits de la personne* de l'Ontario]) », les personnes qui appliquent la politique ou cherchent à la comprendre pourront soulever l'existence d'un biais qu'ils puissent ou non en démontrer la cause.

Nous recommandons à la Commission de clarifier la définition de « biais » afin qu'elle se concentre sur les résultats plutôt que sur la cause d'un extrant biaisé.

1.2. Clarifier à quoi et à qui s'applique la politique dans une section distincte sur son champ d'application

La définition de « technologie d'IA » restreint le champ d'application de la politique en limitant ce terme aux biens et services « qui recueillent ou utilisent des renseignements sur les membres du public » et à l'égard desquels des décisions sont prises « concernant les renseignements ou les membres du public auxquels ils se rapportent ». Étant donné que le champ d'application de la politique repose sur le terme « membres du public », la politique doit indiquer clairement qui est inclus dans le sens de ce terme et qui en est exclu. Il ne devrait y avoir aucun doute sur la question de savoir si les « membres du public » comprennent les victimes d'actes criminels, les plaignants, les personnes d'intérêt, les suspects, les personnes en détention, les membres du Service eux-mêmes ou les candidats à un emploi au sein du Service.

L'ambiguïté quant au champ d'application de la politique pourrait conduire à une incertitude quant à savoir si des outils tels que ceux permettant de détecter des empreintes digitales altérées[6], de découvrir les mots de passe d'appareils mobiles[7], de surveiller le niveau de stress des employés du Service[8] ou de rationaliser les processus de recrutement du Service seraient assujettis ou non à la politique.

La longue définition du terme « nouvelle technologie d'IA » énonce cinq circonstances différentes dans lesquelles une technologie d'IA serait considérée comme étant « nouvelle ». Il pourrait s'agir de technologies jamais utilisées, de technologies que le Service utilise déjà et dont on envisage une nouvelle utilisation, de technologies augmentées par l'IA ou améliorées au moyen d'ensembles de données supplémentaires, ou de technologies d'IA pour lesquelles de nouveaux ensembles de données sont constitués. Nous saluons les efforts de la Commission en vue de décrire le large éventail de situations dans lesquelles une nouvelle technologie d'IA pourrait être utilisée.

---

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

[7] O'Kane, J. (5 nov. 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. (8 mai 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/

Cependant, afin de communiquer plus clairement les utilisations de l'IA qui sont visées par la politique, il serait préférable d'ajouter à celle-ci une section décrivant son champ d'application.

Un champ d'application clair est essentiel pour permettre à la Commission, au Service et aux membres du public de comprendre les limites de ce cadre de gouvernance. En l'absence d'un champ d'application clair, les membres du Service pourraient considérer à tort que des technologies telles que des logiciels offerts par des entreprises pour la tenue d'essais gratuits en sont exclues[9]. De même, le public pourrait avoir la surprise de constater que certaines initiatives importantes d'IA ne font pas l'objet d'une évaluation[10].

Nous recommandons d'inclure dans la politique une section qui en définit le champ d'application. Cette section devrait comprendre des critères et des exemples concernant les types de technologies et d'utilisations de l'IA qui font partie de ce champ et ceux qui n'en font pas partie, et étendre explicitement l'application de la politique à l'utilisation des technologies d'IA (y compris les essais gratuits) par les membres du Service dans le cadre de leur travail.

1.3. Veiller à ce que les renseignements anonymisés et la vie privée des groupes fassent partie du champ d'application de la politique

Les systèmes d'IA qui ne traitent pas de renseignements personnels peuvent néanmoins avoir une incidence sur les valeurs que les mesures de protection de la vie privée cherchent à protéger. Par exemple, l'IA permet de généraliser à des communautés ou à des groupes entiers des déductions faites sur un petit échantillon de particuliers. Des décisions basées sur les résultats de l'IA peuvent être prises au niveau du groupe et avoir une incidence sur l'autonomie, l'épanouissement personnel et d'autres libertés des particuliers, le tout sans s'appuyer sur des renseignements identificatoires sur une personne en particulier[11].

Par exemple, les systèmes de détection de coups de feu ou les statistiques anonymisées sur la criminalité qui alimentent des systèmes d'IA n'ont pas à s'appuyer sur des renseignements personnels ou à prendre des décisions sur des particuliers pouvant être identifiés. Cependant, ces systèmes déclencheront des alertes qui pourraient multiplier les interventions policières dans les communautés où ils sont situés, ce qui pourrait influer sur les décisions quant aux itinéraires de patrouille et aux niveaux de risque des communautés, et assujettir certains quartiers à une surveillance policière plus intense. Au moyen de cette surveillance accrue, la police pourrait mettre au jour un plus grand nombre d'actes criminels dans les communautés marginalisées que dans les quartiers qui ne font pas l'objet d'une surveillance aussi active, ou trop se concentrer sur certains

---

[9] Voir p. ex. Mac, R. et coll. (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] Voir p. ex. Cardoso, T., et B. Curry (7 févr. 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] Voir p. ex. Kammourieh, L. et coll. (2017). « Group Privacy in the Age of Big Data », *Group Privacy: New Challenges of Data Technologies*, Taylor, L., Floridi, L. et B. van der Sloot, éd. Philosophical Studies Series, Springer, p. 37-66.

types d'actes criminels au détriment d'autres, comme les cyberattaques ou les crimes économiques.

La déclaration d'intention de la politique établit l'objectif de préserver la vie privée, les droits et la dignité des particuliers et des communautés, et elle fait mention de la vie privée des groupes. Cependant, sa définition de « technologie d'AI » ne précise pas si les renseignements anonymisés font partie de son champ d'application. Il en va de même des renseignements sur un groupe de particuliers qui sont analysés afin de prendre des décisions concernant un particulier, un groupe de particuliers ou l'ensemble d'une communauté.

Nous recommandons d'inclure explicitement, dans le champ d'application de la politique, des technologies d'IA qui s'appuient sur des renseignements anonymisés ou des renseignements concernant des groupes ou des communautés.

1.4. Exiger explicitement une évaluation rétroactive des technologies d'IA que le Service utilise actuellement

Nous saluons l'engagement de la Commission d'obliger le Service à cesser d'utiliser des technologies d'IA considérées comme posant un risque extrême dès que ces technologies seront identifiées (ou au plus tard en 2024), et à lui rendre compte des utilisations actuelles de technologies d'IA à risque élevé ou modéré.

Or, pour attribuer un niveau de risque à une technologie, il faut l'évaluer. Pour garantir au public que les évaluations nécessaires seront effectuées, la politique devrait exiger explicitement du chef de police qu'il veille à ce que les technologies, les biens et les services existants soient examinés pour déterminer s'ils comportent une composante d'IA et, le cas échéant, à ce que leurs risques soient évalués. Exiger explicitement une telle évaluation rétroactive aidera la Service à comprendre ses obligations et à mieux garantir au public que la politique gérera les risques associés à l'IA en usage actuellement.

Nous recommandons que la politique exige plus clairement que le Service évalue les technologies d'IA qu'il utilise déjà en regard du cadre d'évaluation des risques de la politique.

1.5. Revoir le système de classification des risques

Dans nos commentaires sur la protection de la vie privée dans le secteur privé, nous avons soutenu que les systèmes de prise de décision automatisée – un terme qui englobe généralement l'IA – devraient faire l'objet d'une évaluation de l'incidence algorithmique. Nous avons également proposé d'associer les exigences concernant la responsabilisation et la surveillance aux niveaux de risque établis. Nous sommes heureux de constater que l'on a tenté d'intégrer cette approche dans la politique.

La politique comprend un cadre d'évaluation des risques assorti de niveaux pour les technologies d'IA présentant un risque minime, faible, modéré, élevé ou extrême. Cependant, ces niveaux de risque ne sont pas définis; ils sont plutôt illustrés par des exemples de circonstances où la Commission attribuerait un niveau de risque particulier. Ces exemples se divisent en plusieurs

catégories, qu'il s'agisse des préjudices, des déficiences dans les données, des contraintes sur le plan de la surveillance, ou encore d'applications ou d'utilisations précises des technologies. Ces différents niveaux de risque s'appliquant à diverses catégories font en sorte qu'il est difficile de procéder à une évaluation de façon pratique et méthodique.

Par exemple, la politique considère la « surveillance de masse, définie comme la surveillance secrète et sans discernement d'une population ou d'une composante importante de la population » comme étant une technologie qui présente un risque extrême. Si le CIPVP convient qu'il ne devrait pas y avoir de surveillance de masse secrète, la surveillance de masse pratiquée ouvertement présente également des risques majeurs. La politique n'indique pas clairement si une telle surveillance serait autorisée ou quel niveau de risque elle poserait.

Nous reconnaissons que la politique de la Commission n'établit pas de niveau de risque, mais elle prévoit plutôt que le chef de police doit établir des catégories de risques pour les technologies d'AI en collaboration avec des experts et des parties prenantes. Cependant, à notre avis, cela pourrait aboutir à la création d'un cadre d'évaluation des risques qui est difficile à appliquer pour comparer les risques de différents produits d'IA destiné aux mêmes fins, ou la mesure dans laquelle différentes mesures d'atténuation parviennent à réduire un risque.

Au lieu de s'appuyer uniquement sur des exemples d'applications de l'IA, le système proposé de classification des risques devrait se concentrer sur des catégories précises de préjudices, leur gravité et leur probabilité. Les préjudices causés aux particuliers et aux groupes pourraient comprendre des mauvais traitements, de la discrimination, une perte d'autonomie, la perturbation des relations personnelles et des atteintes à la vie privée, des préjudices psychologiques ou une perte de confiance dans les forces de l'ordre et l'administration de la justice. Les exemples pourraient alors illustrer en quoi la gravité et l'incidence d'un préjudice varient selon le niveau de risque.

Le Service devrait aussi évaluer le risque en contexte, au lieu d'évaluer chaque technologie en vase clos. Une telle évaluation contextuelle porterait sur l'utilisation possible de la technologie en parallèle avec d'autres technologies et pratiques du Service, et déterminerait les préjudices cumulatifs éventuels en conséquence. Soulignons également que les communautés qui sont susceptibles de faire l'objet de l'utilisation de l'IA (c.-à-d. qui seront surveillées ou soumises à une prise de décision fondée sur l'IA) pourraient avoir à partager, aux fins de l'élaboration d'un système pertinent d'évaluation des risques, des perspectives importantes pouvant éclairer ce que sont les préjudices et leurs conséquences[12].

Nous recommandons donc à la Commission :

a) de préciser que les catégories de risques à inclure dans le système de classification des risques doivent être fondées sur des catégories spécifiques de préjudices, leur gravité et leur probabilité, plutôt que sur des applications;

[12] Moss, E. et coll. (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/

b) d'exiger que les évaluations des risques soient axées sur le contexte dans lequel les technologies d'IA seront utilisées, y compris la façon dont de multiples systèmes d'IA peuvent être utilisés par le Service d'une manière qui, cumulativement, peut donner lieu à de nouveaux risques;

c) d'exiger que les parties prenantes participant à l'élaboration du système de classification des risques comprennent des membres des communautés touchées.

2. Renforcer la transparence, la responsabilisation et la surveillance

Bien que la politique comprenne de nombreuses dispositions prévoyant d'informer la Commission de l'utilisation que le Service fait de l'IA, nous recommandons de prendre des mesures supplémentaires pour faire en sorte que la Commission et le public aient un accès élargi à des renseignements pertinents sur les utilisations de l'IA et leurs risques.

2.1. Élargir les dispositions relatives à la surveillance interne et à la responsabilisation

En 2020, l'Assemblée mondiale de la protection de la vie privée (GPA) a adopté une résolution coparrainée par le CIPVP énonçant 12 mesures pour favoriser la responsabilisation dans le développement et l'utilisation de l'IA[13]. Dans nos observations sur le cadre de l'IA de confiance, nous mentionnons cette résolution de la GPA et recommandons au gouvernement de l'Ontario d'envisager différents mécanismes pour assurer une surveillance adéquate de l'IA, dont les suivants :

- Élaborer des procédures pour garantir qu'un système d'IA n'est pas utilisé à de nouvelles fins sans réévaluation;
- Mettre en place des pratiques solides de documentation et de tenue de dossiers;
- Établir des mécanismes de dénonciation permettant de signaler sans crainte de représailles les infractions à la politique;
- Mettre en place le financement, la formation et les structures institutionnelles nécessaires à la surveillance et à l'intervention humaines;
- Assurer une surveillance indépendante par des organismes de surveillance existants, dont le CIPVP, la Commission ontarienne des droits de la personne et l'Ombudsman de l'Ontario.

Nous sommes heureux de constater que la politique comprend déjà de nombreuses dispositions de surveillance conçues pour promouvoir l'utilisation responsable de l'IA; ainsi :

- les membres du Service devront suivre une formation afin de pouvoir identifier les technologies d'IA;

[13] Assemblée mondiale de la protection de la vie privée (octobre 2020). « Résolution sur la responsabilisation dans le développement et l'utilisation de l'intelligence artificielle adoptée lors de la 42e Assemblée mondiale de la protection de la vie privée », https://globalprivacyassembly.org/wp-content/uploads/2021/01/GPA-Resolution-on-Accountability-in-the-Development-and-Use-of-AI-FR.pdf

- des rapports devront être fournis à la Commission concernant les technologies d'IA présentant un risque élevé ou modéré afin qu'elle les approuve avant leur utilisation;
- un avis devra être fourni à la Commission lorsque le Service souhaite utiliser une technologie d'IA à faible risque;
- des indicateurs sont prévus pour évaluer l'efficacité et les conséquences imprévues des technologies d'IA à risque élevé ou modéré, aux fins d'un rapport à la Commission faisant suite à leur mise en œuvre;
- un site public Web sera mis sur pied et recensera les technologies d'IA à risque élevé, modéré et faible que le Service utilise;
- les technologies d'IA à risque élevé, modéré ou faible seront évaluées selon un cycle quinquennal pour déterminer si elles ont été utilisées à de nouvelles fins qui pourraient modifier considérablement les données recueillies ou utilisées, et il en sera rendu compte à la Commission;
- un mécanisme est prévu pour permettre aux membres du public de communiquer leurs préoccupations à la Commission sur les technologies d'IA qui sont utilisées; le directeur exécutif devra évaluer ces préoccupations et les mesures de suivi requises et en rendre compte à la Commission (p. ex., en recommandant que les technologies que le Service considère comme étant à risque faible ou minime soient réévaluées).

Ces dispositions aideront la Commission à superviser l'implantation et l'utilisation de l'IA au sein du Service. Cependant, la politique ne prévoit pas de mécanisme interne de dénonciation. Un tel mécanisme pourrait constituer pour la Commission une source d'information supplémentaire qui ne repose pas sur la structure hiérarchique officielle.

En outre, une étude récente concernant les évaluations de l'incidence algorithmique a permis de constater que ces évaluations, en règle générale, nécessitent le concours de différents acteurs spécialisés au sein des institutions[14]. De plus, comme l'a déclaré récemment le National Institute of Standards and Technology :

> [Traduction]
>
> « Des études ont montré que l'un des déterminants les plus importants de la capacité d'une équipe à faire face aux préjugés nuisibles dans l'IA est la diversité de l'équipe. Les équipes moins diversifiées ont plus de mal à atténuer les préjugés involontaires dans leurs modèles d'apprentissage automatique que les équipes composées de membres issus d'un large éventail de genres, de groupes ethniques et de milieux. En outre, les équipes qui ne représentent pas le point de vue des communautés

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

> touchées par les systèmes d'IA ont plus de mal à prévoir et à atténuer les préjudices que le système peut causer à ces communautés. »[15]

Il est donc important de disposer de ressources suffisantes pour mener les évaluations de l'incidence algorithmique, et d'avoir accès pour les exécuter à des experts interdisciplinaires, en tenant compte d'un éventail de perspectives et d'expériences[16]. La politique n'exige pas, à l'heure actuelle, que le Service définisse des rôles et des responsabilités clairs ou qu'il assure la diversité sur le plan de l'expertise, des perspectives et des expériences qui sont prises en compte dans ses évaluations. Elle n'exige pas non plus que du temps et des ressources raisonnables soient consacrés au processus.

Enfin, nous constatons que la politique ne prévoit pas d'obligations spécifiques de tenue de registres pour les technologies d'IA, à part le contenu requis des rapports à la Commission et l'inventaire public des technologies d'IA. Les renseignements personnels utilisés par les systèmes d'IA doivent être conservés conformément à la *LAIMPVP* afin de permettre aux particuliers de faire valoir leur droit d'accéder aux renseignements personnels qui les concernent.

Outre les renseignements personnels, la transparence et la responsabilisation d'une institution s'accompagnent de l'obligation générale de documenter et de conserver les renseignements utilisés pour prendre des décisions, élaborer des politiques et assurer l'exécution des programmes. En ce qui concerne l'IA, ces renseignements doivent comprendre ce qui suit :

- de la documentation sur la conception du système;
- la démarche suivie pour assurer l'apprentissage du système, y compris des renseignements sur les données d'apprentissage;
- les renseignements utilisés comme intrants;
- des explications sur le processus qui a mené une technologie d'IA à faire une prédiction ou à prendre une décision donnée;
- les renseignements qui ont été remis aux personnes chargées de l'intervention humaine, et les mesures que celles-ci ont prises lorsqu'elles ont examiné ou annulé une décision automatisée[17];
- les incidents où le système a présenté un biais ou un manque de robustesse;
- les résultats des essais techniques ou de sécurité dont a fait l'objet le système;
- les journaux du système et des événements de sécurité.

---

[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] Pour un aperçu des types d'intervenants qu'il y aurait lieu de faire participer à l'évaluation des risques de technologies d'IA, voir p. ex. CIO Strategy Council (2020). « Intelligence artificielle : Conception éthique et utilisation de systèmes de décision automatisés », *CAN/CIOSC 101:2019*, https://ciostrategycouncil.com/normes/conception-ethique/?lang=fr

[17] Les exigences relatives à l'intervention humaine et à l'explicabilité sont décrites en détail aux sections 3.1 et 3.2.

Nous recommandons que la politique prévoie un mécanisme de dénonciation permettant aux membres du Service de signaler à la Commission, de manière anonyme et sécurisée, les infractions à la politique.

Nous recommandons également que la Commission ajoute à la politique une description claire des rôles et des responsabilités en matière de réalisation des évaluations des risques et de participation à ces évaluations, et des modalités prévoyant que le Service est tenu de disposer d'une expertise interdisciplinaire et de perspectives diverses, ainsi que des ressources et du temps nécessaire pour un processus d'évaluation solide.

Nous recommandons aussi que la Commission établisse des exigences en matière de tenue de registres pour l'IA prévoyant la conservation des renseignements personnels utilisés par les technologies d'IA et l'obligation de documenter les activités du Service faisant intervenir l'IA.

2.2. Publier des renseignements supplémentaires sur les technologies d'IA

L'un des défis les plus importants que pose l'utilisation de l'IA a trait à la transparence. Les modèles d'IA sont souvent si complexes que même les développeurs et les utilisateurs de ces modèles ont parfois du mal à décrire le raisonnement ou les étapes spécifiques que suit un système pour prendre des décisions particulières. Ce défi revêt une grande importance dans les contextes où l'équité procédurale est essentielle, comme dans le domaine de l'application de la loi.

D'autres problèmes de transparence liés à l'IA concernent les données utilisées pour l'apprentissage du modèle d'IA. Il est particulièrement important de savoir comment les données ont été obtenues, si elles sont représentatives du contexte réel dans lequel le modèle d'IA sera utilisé, et si elles ont été générées par des pratiques modernes ou de longue date qui présentent des préjugés à l'égard d'un ou de plusieurs groupes. La transparence peut aider les utilisateurs des systèmes d'IA et le public à déterminer si des mesures ont été prises pour garantir que les données d'apprentissage ont été recueillies de manière légale et éthique, et pour y réduire les biais statistiques[18].

Un autre défi majeur consiste à faire en sorte que les personnes qui interagissent avec les technologies d'IA ou qui y sont exposées de quelque autre manière sachent que l'IA est utilisée pour faire des prédictions, pour effectuer des classifications ou pour prendre d'autres décisions les concernant, et soient informées de la nature des renseignements utilisés[19].

---

[18] Voir p. ex. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D. et K. Crawford (2021). « Datasheets for datasets », *Communications of the ACM*, vol. 64, n° 12, p. 86-92, https://arxiv.org/pdf/1803.09010.pdf

[19] P. ex., comme il est décrit dans le principe « Aucune activité d'intelligence artificielle secrète » du cadre de l'intelligence artificielle de confiance proposé par le gouvernement de l'Ontario. https://www.ontario.ca/fr/page/consultations-sur-le-cadre-de-lintelligence-artificielle-ia-de-confiance-de-lontario

Pour relever certains de ces défis, la plupart des cadres et des politiques d'éthique de l'IA doivent être transparents[20]. Parmi les politiques proposées pour renforcer la transparence figurent les registres publics d'IA, qui sont des sites Web publics contenant des renseignements sur les systèmes d'IA dont se servent les pouvoirs publics. La Commission du droit de l'Ontario a publié un exposé exhaustif précisant ce que devraient contenir de tels registres[21]. Par ailleurs, le gouvernement du Royaume-Uni a publié une norme de transparence algorithmique qui décrit les composantes obligatoires et facultatives à inclure dans les renseignements publics sur un outil algorithmique[22]. Le gouvernement de l'Ontario et le gouvernement du Canada se sont également engagés à publier respectivement des inventaires de systèmes d'IA et de systèmes algorithmiques[23] et les résultats d'évaluations de l'incidence algorithmique[24].

Nous reconnaissons que la politique comprend des dispositions visant à divulguer des informations sur l'utilisation de technologies d'IA par le Service. En particulier, elle exige la création d'un site Web public faisant office de registre de l'IA. Ce site Web public documenterait les technologies à risque élevé, modéré et faible qui sont utilisées. En ce qui concerne plus particulièrement les technologies à risque élevé et modéré, ce registre comprendrait des informations sur la vocation de la technologie, la manière dont elle est utilisée, les renseignements recueillis et le contexte d'utilisation prévu.

Bien qu'un site Web public constitue une mesure de transparence essentielle, nous convenons avec la Commission du droit de l'Ontario que les évaluations de l'incidence algorithmique des technologies d'IA devraient également être révélées publiquement et de manière proactive[25]. Nous recommandons à la Commission de divulguer de façon proactive les évaluations de l'incidence algorithmique, sous réserve des exemptions existantes en matière d'accès à l'information en vertu de la *LAIMPVP*, que la Commission peut invoquer, au besoin, pour protéger les renseignements confidentiels. À tout le moins, la politique devrait exiger la préparation et la publication d'un résumé de ces évaluations, qui comprendrait suffisamment d'informations pour faire connaître au public le niveau de risque, la nature des risques encourus et les stratégies d'atténuation à employer si la technologie d'IA est mise en œuvre.

---

[20] Une étude qui a examiné les sujets abordés dans 112 cadres éthiques de l'IA provenant d'organisations publiques, privées et non gouvernementales du monde entier a révélé que la transparence était le deuxième sujet le plus fréquemment abordé après la responsabilité sociale générale. Voir D. Schiff, J. et coll., « AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection », *IEEE Transactions on Technology and Society*, vol. 2, n° 1, p. 31-42, mars 2021, préimpression : https://doi.org/10.36227/techrxiv.14109482.v1.

[21] Thomas, N., Chocla, E., et S. Lindsay (2021). « Réglementer l'intelligence artificielle – Enjeux et choix essentiels », *Commission du droit de l'Ontario*, https://www.lco-cdo.org/wp-content/uploads/2021/04/LCO-Regulating-AI-Critical-Issues-and-Choices-Toronto-April-2021-1.pdf (rapport intégral en anglais); https://www.lco-cdo.org/wp-content/uploads/2021/04/CDO-Reglementer-lintelligence-artificielle-Enjeux-et-choix-essentiels-Sommaire-Toronto-avril-2021.pdf (résumé en français)

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

[23] Gouvernement de l'Ontario. *Catalogue de données : Artificial Intelligence and Algorithms*, https://data.ontario.ca/fr/group/artificial-intelligence-and-algorithms

[24] Voir p. ex. Gouvernement du Canada (2021). *Évaluation de l'incidence algorithmique - Reconnaissance de preuve de vaccination d'ArriveCAN*, https://open.canada.ca/data/fr/dataset/afc17416-3781-422d-a4a9-cc55e3a053c8

[25] Thomas, N. (sept. 2021). « Letter to TPSB re: AI Policy », *Commission du droit de l'Ontario*, https://tpsb.ca/images/consultations/AI/LCO_Letter_to_TPSB_re_AI_Policy.pdf

Nous recommandons que la politique élargisse les renseignements divulgués sur le site Web public concernant l'IA conformément aux recommandations de la Commission du droit de l'Ontario dans *Réglementer l'intelligence artificielle* et du gouvernement du Royaume-Uni dans sa norme sur la transparence algorithmique. Il devrait notamment y avoir divulgation proactive des évaluations de l'incidence algorithmique, sous réserve d'exceptions légitimes en vertu de la *LAIMPVP*, ou du moins de résumés de ces évaluations.

3. Assurer la mise en œuvre de la politique

La politique prévoit que le Service doit élaborer un grand nombre de processus et de procédures à l'appui du cadre de gestion des risques qu'elle contient. Dans la présente section, nous soulignons les exigences à établir et présentons des recommandations préliminaires sur ce qu'elles devraient aborder.

3.1. Établir des exigences pour une intervention humaine efficace

Le rôle de l'intervention humaine est souvent considéré comme étant une mesure très importante pour que les décisions concernant des particuliers ne soient pas prises uniquement au moyen d'un système algorithmique. Dans nos observations sur le cadre de l'IA de confiance, nous avons souligné l'importance d'adopter une démarche axée sur les risques pour établir les exigences relatives à la surveillance et à l'intervention humaines. Nous avons abordé cette question en détail dans nos commentaires sur la protection de la vie privée dans le secteur privé, soulignant que la surveillance humaine n'est pas une panacée qui permet de remédier à tous les préjudices causés par les algorithmes.

Dans le contexte de l'application de la loi, si des systèmes prédictifs font des prédictions fondées sur des données d'apprentissage concernant le maintien de l'ordre dans une communauté, ils peuvent reproduire les biais associés aux pratiques historiques de maintien de l'ordre[26]. Prévoir une intervention humaine ne permet pas en soi de résoudre ce problème. La personne responsable de cette intervention devra justifier de l'expérience ou de la formation nécessaire pour déceler ces biais et les atténuer. Elle aura également besoin de soutien pour éviter de s'en remettre au jugement de systèmes d'IA qu'elle pourrait considérer comme étant plus objectifs qu'elle – un phénomène que l'on appelle « biais d'automatisation »[27]. De même, cette personne devra être consciente du fait que les décisions et biais humains ont causé les situations inéquitables qui ont conduit à ces données, et qu'ils pourraient entraîner ou perpétuer des biais dans les systèmes d'IA.

La politique devrait prévoir des objectifs clairs concernant la surveillance et l'intervention humaines. Ces objectifs devraient être assortis de critères et d'une méthodologie d'évaluation de l'efficacité du processus d'intervention humaine en vue de les atteindre. Ces objectifs, critères et méthodologies peuvent alors éclairer le contenu de la formation des membres du Service qui seront

---

[26] Richardson, R., Schultz, J., et K. Crawford (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] Pour un exposé sur les limites de l'intervention humaine, voir p. ex. Green, B., et A. Kak (15 juin 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

chargés de l'intervention humaine et les exigences en matière d'approvisionnement, et contribueront à déterminer comment il serait possible de tester, de mesurer ou de surveiller l'intervention et l'examen humains. Ces critères doivent accorder une attention particulière au rôle du pouvoir discrétionnaire dans le processus décisionnel que l'on envisage de faire automatiser par le système d'IA. Dans des circonstances où un degré important de latitude est requis, il pourrait ne pas être approprié que l'IA fasse ne serait-ce que la recommandation initiale, alors que dans d'autres cas, cela pourrait être acceptable. Il faut également veiller à ce que les personnes chargées d'exercer une surveillance humaine disposent du temps, des ressources, des renseignements et des capacités nécessaires pour bien examiner les décisions. Faute de tels soutiens, la surveillance humaine ne présente que des avantages limités[28].

Dans la politique, la surveillance humaine n'est pas présentée comme un mécanisme de contrôle utilisé pour atténuer d'autres risques. Elle est plutôt abordée dans le contexte de la définition des niveaux de risque, de sorte que l'absence de surveillance humaine ou la difficulté à l'exercer est traitée comme s'il s'agissait d'un préjudice en soi. Selon nous, l'intervention humaine devrait plutôt être considérée comme une mesure d'atténuation requise ou recommandée, en fonction de l'évaluation des risques.

Nous recommandons à la Commission de prévoir des exigences relatives aux processus d'intervention humaine comprenant ce qui suit :

a) Mettre au point un processus pour déterminer les aspects de la prise de décision qui devraient être automatisés et ceux qui devraient recourir à l'intervention humaine. Ce processus devrait évaluer :

i. le rôle de la latitude dans la prise de décision;

ii. la précision, le rendement et la robustesse du système d'IA envisagé;

iii. les conséquences d'une mauvaise décision.

b) S'assurer que les personnes chargées de l'intervention humaine ont le temps, les ressources, les renseignements et les capacités nécessaires pour bien examiner les décisions automatisées ou les prédictions.

c) Prévoir des exigences de consignation et de tenue de registres concernant les activités humaines d'examen et d'intervention.

d) Inclure des mesures sur la surveillance humaine parmi les indicateurs dont il faut rendre compte à la Commission.

3.2. Établir des exigences relatives à l'explicabilité

Une autre mesure importante de responsabilisation en matière d'IA est l'« explicabilité ». Il s'agit de la mesure dans laquelle un système d'IA peut révéler son fonctionnement interne pour décrire les renseignements, les critères et le raisonnement qu'il a utilisés pour prendre une décision ou arriver à une conclusion. L'explicabilité peut appuyer l'intervention humaine en fournissant aux

[28] Voir p. ex. Green, B. (2021). « The Flaws of Policies Requiring Human Oversight of Government Algorithms », *Working paper*, http://dx.doi.org/10.2139/ssrn.3921216

personnes chargées de cette intervention les renseignements dont elles ont besoin pour examiner les décisions prises par le système.

Il n'existe pas encore de norme précise sur ce qui constitue une explication pertinente[29]. Cependant, l'objectif du système d'IA, la catégorie de personnes susceptibles de demander une explication et à quelles fins (c.-à-d. l'auditoire cible) et les risques associés à l'impossibilité d'expliquer un produit pourraient constituer un bon point de départ pour établir les exigences relatives à l'explicabilité pour un système donné[30]. Cela va dans le sens du document d'orientation préliminaire sur les technologies de reconnaissance faciale, selon laquelle il faut pouvoir expliquer la performance d'une technologie de reconnaissance faciale, c'est-à-dire que cette technologie devrait être soumise à des essais pour déterminer les taux de faux positifs et de faux négatifs, au lieu de fournir une explication exhaustive de la *manière* dont la technologie a permis d'identifier un particulier.

La politique indique qu'une technologie d'IA qui n'est pas entièrement explicable devrait être considérée comme posant un risque extrême. Cependant, comme dans notre exposé sur l'intervention humaine, il serait préférable de considérer l'explicabilité comme une mesure d'atténuation de certains risques, et l'absence d'explicabilité comme un facteur aggravant pour d'autres. Il peut être raisonnable de s'attendre à ce qu'une technologie à risque extrême soit explicable de façon pertinente, mais il ne s'ensuit pas que tout système d'IA présente un risque extrême s'il ne peut être ainsi expliqué. Inversement, une certaine forme d'explicabilité est tout aussi importante pour de nombreuses technologies d'IA à faible risque. Un certain niveau d'explicabilité permet de répondre de manière pertinente aux demandes d'accès à l'information, y compris à des renseignements personnels, et protège des risques liés à l'équité procédurale et à la responsabilisation. C'est pourquoi l'utilisation d'une technologie d'IA qui diminue la transparence globale du Service ou qui fait en sorte qu'il est plus difficile pour ce dernier de répondre aux demandes d'accès à l'information serait préjudiciable.

Nous recommandons que la politique considère l'explicabilité comme un facteur atténuant et l'absence d'explicabilité comme un facteur aggravant lors de l'évaluation des préjudices éventuels dans le cadre du processus d'évaluation des risques, et qu'elle établisse des exigences de base en matière d'explicabilité qui doivent être respectées sans égard au niveau de risque.

4. Étapes subséquentes

En 2020, la Commission s'est engagée à renforcer la confiance du public et à lutter contre le racisme systémique[31], notamment en consultant des experts et des communautés, en rehaussant la

---

[29] Voir p. ex. Arrieta A.B. et coll. (2020). « Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI », *Information Fusion (58)*. Préimpression : https://arxiv.org/abs/1910.10045

[30] Voir p. ex. Amarasinghe, K. et coll. (2021). « Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions ». Document de travail : https://arxiv.org/abs/2010.14374

[31] Commission de services policiers de Toronto (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

transparence et en assurant une vérification indépendante et une meilleure responsabilisation du Service. Ces engagements se reflètent dans la politique, dont les principes directeurs comprennent la nécessité de préserver la dignité des particuliers et des communautés de même que leur droit à la vie privée et leurs autres droits.

Le CIPVP salue l'engagement de la Commission d'élaborer un cadre de gouvernance de l'IA de façon transparente en invitant les commentaires du public. Cependant, comme nous l'indiquons dans les présentes observations, certains aspects nécessitent des modifications, et des travaux approfondis s'imposent afin d'élaborer les procédures et processus opérationnels requis pour mettre en œuvre la politique.

4.1. Continuer de consulter le public sur l'élaboration des procédures

La politique exige que le Service élabore, en consultation avec des experts et des intervenants, des procédures et des processus pour l'évaluation de nouvelles technologies d'IA. Nous constatons que les procédures et processus qui restent à élaborer sont nombreux, notamment :

- lignes directrices aux fins de l'identification des technologies d'IA à l'intention des membres du Service;
- exigences relatives à la tenue de registres;
- exigences à respecter pour une intervention humaine efficace;
- exigences relatives à l'explicabilité;
- une ou plusieurs méthodologies d'évaluation des risques;
- mesures d'atténuation requises dans le cas de différents risques;
- lignes directrices sur l'identification et la surveillance des conséquences indésirables;
- critères d'évaluation de la qualité des données.

Nous recommandons à la Commission de continuer de consulter un large éventail de parties prenantes de la collectivité, y compris le CIPVP, dans le cadre de ses travaux aux fins de la mise au point de sa politique.

Nous recommandons aussi à la Commission de faire en sorte que le Service fasse également appel à un large éventail de parties prenantes, y compris le CIPVP, aux fins de l'élaboration de processus et de procédures de mise en œuvre de la politique.

4.2. Consulter le CIPVP et le public sur la méthodologie d'évaluation des risques

Une étude mondiale des politiques de responsabilisation algorithmique dans le secteur public a révélé que peu de politiques prévoient une participation significative du public et des parties prenantes externes[32]. Il est essentiel d'assurer la participation concrète du public afin d'intégrer dans les initiatives d'IA les perspectives des personnes qui seront touchées par ces technologies

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

ainsi que d'experts ayant des compétences techniques et en droit, pour relever et élaborer des stratégies d'atténuation des risques.

Dans nos commentaires sur la protection de la vie privée dans le secteur privé, nous avons recommandé d'envisager des consultations avec le CIPVP et le public dans le cas des systèmes de prise de décision automatisée pouvant poser un risque élevé.

Nous sommes conscients du fait que la politique confère un rôle au CIPVP dans le cadre du processus d'évaluation des risques. Les modalités du rapport à présenter à la Commission sur les technologies à risque élevé et modéré font état de consultations avec le CIPVP et d'analyses que ce dernier exigerait[33]. Nous saurions gré au Service de nous consulter sur les nouvelles initiatives d'IA (ou les évaluations de systèmes existants) dans les plus brefs délais, notamment en ce qui concerne l'examen des évaluations de l'incidence sur la vie privée et de l'incidence algorithmique.

La politique exige la tenue de consultations publiques sur les technologies d'IA à risque élevé après leur mise en œuvre, et l'inclusion de leurs résultats dans un rapport de suivi présenté à la Commission. Bien que de telles consultations puissent être utiles pour déterminer les répercussions de la technologie d'IA sur la confiance du public et mettre au jour des conséquences involontaires éventuelles, la consultation du public devrait également jouer un rôle clé au cours du processus d'évaluation lui-même.

Toutes les technologies d'IA ne devraient pas nécessairement être soumises à une consultation publique; cependant, **nous recommandons que la politique prévoie l'élaboration de critères afin de déterminer les circonstances où une consultation publique devrait faire partie du processus d'évaluation des risques préalable à la mise en œuvre.**

[33] Alinéas 5 g) et 5 j) de la politique, respectivement.