47e séance à huis clos de l'Assemblée mondiale pour la protection de la vie privée

Septembre 2025

Résolution sur la surveillance humaine véritable des décisions prises à l'aide de systèmes d'IA

Cette résolution est soumise par le Commissariat à la protection de la vie privée du Canada au nom du Groupe de travail sur l'éthique et la protection des données dans le secteur de l'intelligence artificielle.

PARRAINEURS :

- Commissariat à la protection de la vie privée du Canada

COPARRAINEURS :

- Agence pour l'accès à l'information publique (Agencia de Acceso a la Información Pública), Argentine
- Autorité de protection des données, Belgique
- Commissariat à l'information et à la protection de la vie privée de la Colombie-Britannique
- Commission pour la Protection de Données à caractère personnel, Bulgarie
- Contrôleur européen de la protection des données
- Commission nationale de l'informatique et des libertés, France
- Autorité de protection des données, bailliage de Guernesey
- Commissariat à la protection de la vie privée et des données personnelles, Hong Kong, Chine
- Commission de la protection des données de l'Irlande
- Commissaire à l'information de l'île de Man
- Commission de protection des informations personnelles (PIPC), République de Corée
- Commissariat à l'information et à la protection de la vie privée de l'Ontario, Canada
- Préposé fédéral à la protection des données et à la transparence (PFPDT), Suisse

À l'occasion de la 47e Assemblée mondiale pour la protection de la vie privée qui a lieu en 2025 :

Reconnaissant que les systèmes d'intelligence artificielle (IA) sont de plus en plus intégrés aux processus décisionnels;

Ayant conscience qu'une approche centrée sur l'être humain à l'égard de l'IA pourrait procurer d'importants avantages pour l'économie, la société et l'intérêt public, notamment en favorisant la prospérité et en permettant de s'attaquer aux défis mondiaux urgents;

Insistant sur l'importance de veiller à ce que l'utilisation des systèmes d'IA respecte les droits de la personne, ne permette pas la discrimination, favorise l'équité et la justice, et prévienne les biais;

Reconnaissant que les processus décisionnels ayant recours à des systèmes d'IA pourraient avoir des répercussions négatives importantes sur les droits et libertés fondamentaux des individus, en particulier lorsque les décisions ne font pas l'objet d'une surveillance humaine véritable effectuée par des personnes capables de surveiller adéquatement les systèmes d'IA pertinents ou lorsqu'il n'y a aucun recours efficace ou possibilité pour un individu touché par la décision (l'« individu touché ») de réellement contester la décision;

Reconnaissant l'importance de veiller à ce que, lorsque des décisions automatisées pourraient avoir une incidence importante sur les droits et libertés fondamentaux des individus, ceux-ci puissent demander que ces décisions soient examinées par un être humain en temps opportun;

Reconnaissant les avantages d'une surveillance humaine véritable de la prise de décisions automatisée pour accroître la responsabilité des institutions dans le développement et l'utilisation des systèmes d'IA; améliorer leur fiabilité; déceler et atténuer proactivement les biais potentiels dans les données et les algorithmes; améliorer la transparence, la capacité de contestation et l'explicabilité; et contribuer à l'amélioration des systèmes d'IA et leur adaptabilité aux environnements réels en constante évolution afin de favoriser la confiance des personnes concernées par les données;

Ayant conscience que, si la conception et le développement des systèmes de décisions automatisées ne sont pas examinés attentivement, les décisions refléteront les tendances que l'on trouve dans les données d'entraînement d'un système et reproduiront ou renforceront donc les biais passés, ou des décisions seront prises en supposant que les conditions des décisions passées demeurent vraies pour le présent et l'avenir;

Rappelant que le plan de travail du Groupe de travail sur l'éthique et la protection des données dans le secteur de l'IA comprend l'établissement d'une compréhension commune des facteurs qui constituent une surveillance humaine « véritable » des processus décisionnels ayant recours aux systèmes d'IA;

Faisant une distinction entre les termes « surveillance humaine », un concept qui a lieu tout au long du processus décisionnel, et « examen humain », un concept qui a lieu dans le cadre d'un examen postérieur à une décision où un individu touché est en mesure de justifier son point de

vue, et notant que l'examen humain est une activité qui fait partie du processus plus large de surveillance humaine;

Reconnaissant que les organisations sont responsables d'établir leurs propres processus internes de surveillance humaine;

Soulignant que certaines lois sur la protection des données et des renseignements personnels établissent le droit pour les individus de ne pas être soumis à certains processus décisionnels entièrement automatisés, et qu'une décision est considérée comme entièrement automatisée en l'absence d'une surveillance humaine suffisante et véritable du processus décisionnel, ce qui fournit à la personne concernée par les données le droit d'obtenir une intervention humaine, d'exprimer son point de vue et de contester les décisions découlant de tels processus;

Soulignant que les normes intergouvernementales sur l'IA, comme la recommandation de l'Organisation de coopération et de développement économiques (OCDE) sur l'IA (OECD/LEGAL/0449), reconnaissent que les organisations qui déploient ou exploitent des systèmes d'IA devraient mettre en œuvre des mécanismes et des mesures de protection, notamment la surveillance humaine, pour gérer les risques découlant d'utilisations à des fins autres que celles prévues ou d'une mauvaise utilisation délibérée ou involontaire des systèmes d'IA;

Reconnaissant que le type de surveillance humaine requise pour un processus décisionnel dans lequel un système d'IA est utilisé dépend généralement du contexte et des répercussions potentielles de la décision, et qu'une surveillance et des mesures de protection accrues seront nécessaires lorsque les systèmes d'IA traitent certaines catégories de renseignements personnels ou de données liées à des condamnations criminelles et à des infractions;

Rappelant que la 40e Assemblée mondiale pour la protection de la vie privée a approuvé le principe selon lequel la transparence et l'intelligibilité des systèmes d'IA devraient être améliorées, notamment « [e]n fournissant des informations adéquates sur les objectifs et les effets de l'[IA], afin de vérifier que cette dernière s'aligne toujours sur les attentes des individus et que ces derniers peuvent exercer un contrôle global sur ces systèmes »;

Rappelant que la 42e Assemblée mondiale pour la protection de la vie privée a exhorté les organisations qui développent ou utilisent des systèmes d'IA à envisager de mettre en œuvre des mesures de responsabilité, notamment « garantir l'identification d'un acteur humain responsable a) auprès duquel les préoccupations liées aux décisions automatisées peuvent être soulevées et les droits peuvent être exercés, et b) qui peut déclencher l'évaluation du processus de décision et l'intervention humaine »;

Rappelant qu'en ce qui concerne l'IA et l'emploi, la 45e Assemblée mondiale pour la protection de la vie privée a souligné l'importance de permettre à une personne concernée par des données qui subit des répercussions causées par le système d'IA d'un employeur d'obtenir un examen humain consigné et véritable des décisions d'emploi, et l'importance de former les utilisateurs des outils d'IA, notamment ceux qui assurent la surveillance humaine des processus décisionnels ayant recours à un système d'IA;

Notant que certaines lois et orientations réglementaires sur l'IA indiquent que la surveillance humaine est nécessaire pour atténuer efficacement les risques pour la santé, la sécurité et les droits fondamentaux posés par certains systèmes d'IA ou leurs applications;

Reconnaissant que le processus de surveillance humaine des décisions automatisées doit tenir compte de considérations comme le biais d'automatisation, dans lequel les personnes effectuant la surveillance font trop confiance aux décisions prises par les systèmes d'IA;

Insistant sur le fait que la surveillance humaine ne sera pas un remède adéquat à un système d'IA mal conçu ou développé, mal appliqué ou autrement fondamentalement imparfait, et qu'il demeure essentiel que les organisations qui ont recours à des systèmes d'IA déterminent à la fois s'il est approprié d'utiliser un système d'IA dans un contexte donné et, le cas échéant, si le système d'IA qu'elles ont choisi sera utile dans ledit contexte;

Soulignant que pour la surveillance humaine des processus décisionnels ayant recours à un système d'IA, il faut que la personne qui effectue la surveillance ait accès à l'ensemble des renseignements pertinents à la décision, que ces renseignements soient présentés de façon appropriée selon le contexte du processus de surveillance, mais que ceux-ci soient limités à ce qui est nécessaire dans le contexte de la surveillance particulière qui est effectuée (c'est-à-dire que, pour ce qui est des renseignements personnels, la personne qui effectue la surveillance ne devrait avoir accès qu'aux renseignements pertinents au processus décisionnel);

Reconnaissant que, dans de nombreux cas, le processus par lequel un système d'IA parvient à une décision n'est pas toujours évident à comprendre pour un être humain et que des mesures doivent être prises pour concevoir le système d'IA d'une manière qui permet ou augmente l'explicabilité, en mettant l'accent sur le fait qu'il faut veiller à ce que la personne qui effectue la surveillance ainsi que les individus touchés – des personnes qui n'ont peut-être pas d'expertise technique – puissent comprendre la décision et demander que celle-ci soit examinée;

Reconnaissant également que, pour que des individus puissent demander un examen de décisions prises à l'aide de systèmes d'IA ou exercer d'autres droits ou capacités en ce sens, il faut que l'utilisation de ces systèmes soit transparente;

Réalisant que, dans certaines circonstances, il se peut qu'une surveillance humaine véritable ne soit pas possible, par exemple lorsque les processus décisionnels donnent lieu à des décisions à une échelle et dans un délai qui rendent impossible la surveillance de chaque cas, et que, dans de telles situations, les organisations devraient envisager d'autres moyens de surveillance (comme l'analyse d'échantillons de décisions pour veiller à ce que le processus décisionnel fonctionne comme prévu);

Reconnaissant qu'une partie ou l'ensemble des considérations ci-dessous peuvent avoir une incidence sur le caractère véritable de la surveillance humaine d'une décision prise à l'aide d'un système d'IA :

- **Faculté d'agir :** L'organisation devrait concevoir le processus de surveillance de manière à ce que la personne qui effectue la surveillance dispose d'une autonomie et d'un contrôle adéquats qui lui permettent de prendre des décisions et d'agir de façon indépendante. Elle devrait notamment s'assurer que la personne qui effectue la surveillance se sente à l'aise d'exercer son rôle sans crainte

de répercussions et ait les moyens de le faire. Par exemple, il pourrait s'agir d'établir une procédure de dénonciation;

- **Clarté du rôle :** L'organisation devrait veiller à ce que la personne qui effectue la surveillance sache clairement si son rôle consiste à évaluer une décision prise par un système d'IA; à accepter, rejeter ou modifier une recommandation faite par un système d'IA; ou à considérer les résultats obtenus au moyen d'un système d'IA comme un facteur parmi tant d'autres dans son processus décisionnel, en notant que la responsabilité finale de la décision revient toujours à l'organisation;
- **Connaissances et expertise :** L'organisation devrait s'assurer que la personne qui effectue la surveillance possède les connaissances et l'expertise nécessaires pour évaluer la décision du système d'IA, notamment sa pertinence, son exactitude et les répercussions possibles de la décision sur l'individu touché. De plus, la personne qui effectue la surveillance doit être suffisamment formée pour bien comprendre les opérations et les limites du système d'IA, afin d'être en mesure de cerner les situations ou les circonstances dans lesquelles les résultats obtenus au moyen du système d'IA peuvent nécessiter un examen plus approfondi, et bien comprendre des facteurs comme le biais d'automatisation qui peuvent avoir une incidence sur ses propres actions;
- **Ressources :** L'organisation devrait fournir à la personne qui effectue la surveillance les ressources nécessaires pour qu'une décision puisse faire l'objet d'une surveillance adéquate. La personne devrait entre autres avoir suffisamment de temps pour effectuer la surveillance et avoir une charge de travail raisonnable, des renseignements sur la façon dont le système a été entraîné (notamment sur la nature des données d'entraînement) et sur la logique de la prise de décision, des données pertinentes dans un format interprétable et de l'information contextuelle appropriée pour appuyer la surveillance, ou accès à des collègues, à des experts ou à d'autres ressources avec qui la personne qui effectue la surveillance peut discuter. S'il y a lieu, cela peut également comprendre le fait de permettre à la personne qui effectue la surveillance de s'entretenir avec l'individu touché pour lui poser des questions et obtenir des précisions;
- **Moment et efficacité :** L'organisation devrait veiller à ce que la surveillance ait lieu à un moment et d'une manière qui permet à la personne qui effectue la surveillance d'accepter, de contester ou d'atténuer les répercussions potentielles de la décision prise par le système d'IA; plus précisément, il est peu probable que la surveillance humaine d'un système d'IA soit véritable ou efficace si les recours ne sont disponibles qu'une fois qu'un individu a subi les répercussions d'une décision;
- **Évaluation et responsabilisation :** L'organisation devrait évaluer les personnes qui effectuent la surveillance en fonction de la diligence avec laquelle ils ont accompli leur tâche prescrite, et non du résultat de la décision. La responsabilité de la décision définitive et de ses répercussions reviendra toujours à l'organisation.

Reconnaissant que les organisations peuvent prendre des mesures pour assurer une surveillance humaine véritable des décisions prises à l'aide de systèmes d'IA, notamment les mesures ci-dessous :

- **Clarification de l'intention et de la valeur de la surveillance :** Pour éliminer de possibles biais d'automatisation et mettre l'accent sur la valeur du rôle de surveillance, les organisations devraient indiquer clairement aux personnes qui effectuent la surveillance à quelles connaissances ou expériences elles devraient faire appel lorsqu'elles examinent des décisions (comme une expertise dans un domaine donné, une expérience de vie générale, etc.);
- **Formation :** En plus de l'expertise du domaine et de la connaissance des utilisations prévues et des limites du système d'IA, les organisations devraient veiller à ce que les personnes qui effectuent la surveillance reçoivent une formation sur certains concepts, comme la façon de repérer et d'atténuer les biais, y compris ceux qui ont une incidence sur les individus touchés par une décision prise par le

système d'IA et ceux qui ont une incidence sur la personne qui effectue la surveillance (comme le biais d'automatisation). Cette formation devrait être suivie avant que la personne assume son rôle de surveillance et être régulièrement revue;

- **Conception du processus de surveillance :** Les organisations devraient concevoir le processus de surveillance en s'assurant qu'il est convivial, notamment que les renseignements pertinents sont présentés d'une manière qui sera utile et compréhensible dans la pratique. Le processus devrait être réexaminé et, au besoin, amélioré à mesure que les personnes qui effectuent la surveillance acquièrent de l'expérience pratique, et celles-ci peuvent contribuer au processus de conception. Les organisations devraient également chercher à établir si les personnes qui effectuent la surveillance sont touchées par des problèmes connus, comme l'« effet d'ancrage », soit des cas où le jugement d'un individu est trop influencé par un point de référence donné (comme le premier élément d'information qui lui est présenté);
- **Transmission des cas à l'échelon supérieur :** Les organisations devraient mettre en place des mesures pour transmettre les cas de prise de décisions à l'échelon supérieur, notamment dans les situations ou les circonstances déterminées par la personne qui effectue la surveillance. Ces mesures pourraient également comprendre la conception d'un processus de prise de décision automatisée dans lequel au-delà d'un certain degré de risque, les décisions feraient automatiquement l'objet d'un signalement menant à une intervention obligatoire avant que des mesures ne soient prises;
- **Documentation :** Les organisations devraient exiger des personnes qui effectuent la surveillance qu'elles documentent leurs décisions, particulièrement dans le cas où elles rejettent la décision d'un système d'IA. Ainsi, il sera possible à la fois de cerner les tendances des mauvaises décisions prises par le système d'IA et de démontrer le fait que des mécanismes de surveillance humaine mal appliqués peuvent introduire un biais;
- **Évaluations :** Les organisations devraient inclure la nature et l'étendue de la surveillance humaine dans une analyse d'impact relative à la protection des données et toute autre évaluation (comme une évaluation de l'incidence algorithmique ou une évaluation de l'incidence sur les droits fondamentaux) de l'utilisation proposée d'un système d'IA;
- **Évaluation et mise à l'essai du processus de surveillance :** Les organisations devraient régulièrement mettre à l'essai l'efficacité de leur processus de surveillance. Elles pourraient le faire au moyen de « fausses décisions » (soit en soumettant délibérément des décisions erronées à une surveillance pour déterminer si elles sont repérées);
- **Évaluation des résultats :** Les organisations devraient évaluer régulièrement s'il y a des tendances dans les décisions du système d'IA qui sont rejetées ou renversées par les personnes qui effectuent la surveillance, ou des décisions incorrectes prises même après une surveillance humaine. Ces tendances pourraient indiquer des problèmes avec le système d'IA ou avec le processus de surveillance.

La 47e Assemblée mondiale pour la protection de la vie privée décide donc de ce qui suit :

1. Favoriser une compréhension commune de la notion de surveillance humaine véritable des décisions prises à l'aide de systèmes d'IA, ce qui comprend notamment les considérations énoncées dans la présente résolution.

2. Exhorter les organisations qui ont recours à des systèmes d'IA dans les processus décisionnels à désigner des personnes pour effectuer la surveillance. Ces personnes devront avoir la compétence, la formation, les ressources et les connaissances de toute information contextuelle sur le processus

décisionnel nécessaires et comprendre les capacités, les limites, les modes de défaillance possibles et les risques connexes propres au système d'IA, notamment les biais potentiels. De plus, les exhorter à adopter des interfaces personne-machine et des outils d'interprétation appropriés pour l'adoption de technologies et de processus qui permettent une surveillance humaine véritable, en particulier lorsque ces décisions peuvent avoir des répercussions importantes sur les libertés et les droits fondamentaux des individus.

3. Par l'intermédiaire du Groupe de travail sur l'éthique et la protection des données dans le secteur de l'IA de l'Assemblée mondiale pour la protection de la vie privée, partager les connaissances et les pratiques exemplaires relatives à la mise en œuvre pratique de la notion de surveillance humaine véritable dans les cadres juridiques applicables respectifs et mettre au point des ressources à l'intention des autorités de protection des données et de la vie privée qui appuient les efforts de ces dernières visant à favoriser l'adoption des pratiques décrites dans la présente résolution par les contrôleurs de leur juridiction respective.

4. Continuer d'encourager le développement de technologies ou l'élaboration de processus qui contribuent à l'explicabilité des systèmes d'IA, en reconnaissant qu'il s'agira d'un élément important pour assurer une surveillance humaine véritable des décisions prises à l'aide de systèmes d'IA.

La Commission fédérale du commerce des États-Unis s'abstient d'adopter cette résolution.