



## **DÉCLARATION SUR L'ÉTHIQUE ET LA PROTECTION DES DONNÉES DANS LE SECTEUR DE L'INTELLIGENCE ARTIFICIELLE**

40e Conférence internationale des commissaires à la protection des données et de la vie privée

Mardi 23 octobre 2018, Bruxelles

### AUTEURS :

- Commission Nationale de l'Informatique et des Libertés (CNIL), France
- Contrôleur européen de la protection des données, Union européenne
- Garante per la protezione dei dati personali, Italie

### CO-SPONSORS :

- Agencia de Acceso a la Información Pública, Argentine
- Commission d'accès à l'information, Québec, Canada
- Datatilsynet (Inspection de l'informatique), Norvège
- Information Commissioner's Office (ICO), Royaume-Uni
- Préposé fédéral à la protection des données et à la transparence, Suisse
- Autorité chargée de la protection des données, Belgique
- Privacy Commissioner for Personal Data, Hong Kong
- Data Protection Commission, Irlande
- Bureau de la protection des données, Pologne
- INAI, Mexique
- Autorité nationale chargée de la protection des données et de la liberté d'information, Hongrie | Commissaire fédéral chargé de la protection des données et de la liberté d'information, Allemagne
- Commissariat à la protection de la vie privée du Canada

**La 40e Conférence internationale des commissaires à la protection des données et de la vie privée :**

*Prenant en considération* la discussion initiale tenue lors de la 38e Conférence internationale des commissaires à la protection des données et de la vie privée qui s'est déroulée à Marrakech sur les thèmes de l'intelligence artificielle, la robotique, la protection des données et la vie privée ;

*Reconnaissant* que les systèmes d'intelligence artificielle pourront peut-être avoir de grands avantages pour les utilisateurs et la société, y compris de la manière suivante : en accélérant la rapidité des processus et en soutenant la prise de décision ; en créant de nouveaux modes de participation aux processus démocratiques ; en améliorant l'efficacité du secteur public et de l'industrie ; en parvenant à une répartition plus équitable des ressources et des opportunités ; en proposant de nouvelles méthodes et solutions dans plusieurs domaines (par exemple : santé publique, soins médicaux, sécurité, développement durable, agriculture et transport) ; en offrant de nouvelles opportunités en matière de recherche scientifique et d'éducation ; et en proposant aux individus des services mieux personnalisés ;

*Tenant compte* des progrès importants réalisés dans certains domaines de l'intelligence artificielle, notamment en ce qui concerne le traitement de volumes importants d'informations, l'analyse et la prédiction des caractéristiques et du comportement humains, et dans des domaines connexes tels que la robotique, les systèmes autonomes et de vision artificielle, tous susceptibles de connaître un progrès fulgurant dans un avenir proche ;

*Mettant en avant* l'avancée rapide du big data et de l'intelligence artificielle, surtout l'apprentissage automatique, en particulier grâce au développement des technologies d'apprentissage profond, qui permettent aux algorithmes de résoudre des opérations complexes pouvant mener à des décisions, mais qui rendent toutefois de tels processus [potentiellement] moins transparents ;

*Affirmant* que le respect des droits à la vie privée et à la protection des données est de plus en plus remis en cause par le développement de l'intelligence artificielle, et que ce développement devrait être accompagné de réflexions d'ordre éthique et de droits de l'Homme ;

*Considérant* que les technologies d'apprentissage automatique en particulier, et les systèmes d'intelligence artificielle en général, peuvent dépendre, pour leur développement, du traitement de grands ensembles de données à caractère personnel, qui auront un impact potentiel sur la protection des données et la vie privée ; et en outre, tenant compte des risques potentiels inhérents à la tendance actuelle de concentration du marché dans le secteur de l'intelligence artificielle ;

Reconnaissant le lien entre collecte, utilisation et divulgation d'informations personnelles (le domaine traditionnel de la protection des données et de la vie privée) d'un côté et, de l'autre, les impacts directs sur les droits de l'homme de façon plus générale, surtout en ce qui concerne la discrimination et la liberté d'expression et d'information, et reconnaissant, par conséquent, que les autorités de protection des données et de la vie privée doivent réfléchir aux droits de l'homme de façon plus large et coopérer avec d'autres autorités pour défendre ces mêmes droits ;

*Indiquant* qu'il s'est avéré que certains ensembles de données servant à alimenter les systèmes basés sur l'apprentissage automatique et l'intelligence artificielle entraînent un biais inhérent, lequel mène à la prise de décisions qui peuvent faire preuve de discrimination à l'encontre de certains individus ou certains groupes, limitant potentiellement la disponibilité de certains services ou contenus et, par là-même, entravant certains droits individuels tels que la liberté d'expression ou entraînant l'exclusion de certaines personnes de plusieurs aspects de la vie personnelle, sociale ou professionnelle ;

Déclaration sur l'éthique et la protection des données dans le secteur de l'intelligence artificielle

*Insistant sur le fait* que les systèmes reposant sur l'intelligence artificielle, dont les décisions ne peuvent pas être expliquées, posent des questions fondamentales de transparence, non seulement dans le cadre de la législation en matière de protection des données et de la vie privée, mais aussi en termes de responsabilité en cas d'erreurs et de préjudices ;

*Notant* que de nombreuses parties prenantes dans le secteur de l'intelligence artificielle ont exprimé leurs inquiétudes en ce qui concerne les risques d'une utilisation malveillante de cette technologie, ainsi que les risques associés à la vie privée, la protection des données et la dignité humaine, montrant, par exemple, que le développement de l'intelligence artificielle, combiné à la surveillance de masse, fait naître des interrogations quant à leur utilisation possible pour restreindre les libertés et les droits fondamentaux ;

*Soulignant* que ces risques et ces défis peuvent avoir un impact sur les individus et la société, et que la portée et la nature de ces conséquences potentielles sont, à l'heure actuelle, mal connues ;

*Soulignant* l'importance de la confiance, car la mise en place de garde-fous robustes en matière de protection des données et de vie privée contribue à donner aux individus un sentiment de confiance sur la façon dont leurs données sont traitées. Cela encourage le partage des données et favorise ainsi l'innovation ;

*Étant de l'opinion* que les défis actuels, nés du développement des systèmes d'intelligence artificielle et d'apprentissage automatique, renforcent le besoin d'adopter une approche et des normes internationales afin de garantir la promotion et la protection des droits de l'homme dans tous les développements numériques à l'échelle mondiale ;

*Réaffirmant* l'engagement des autorités de protection des données et de la Conférence des commissaires à la protection des données et de la vie privée de faire respecter les principes de protection à cet égard en s'adaptant à cet environnement en pleine évolution, en particulier en allouant des ressources et en développant de nouvelles compétences afin de se préparer aux changements de demain.

La **40e Conférence internationale des commissaires à la protection des données et de la vie privée** juge que la création, le développement et l'utilisation de systèmes d'intelligence artificielle doivent respecter entièrement les droits de l'homme, particulièrement ceux relatifs à la protection des données à caractère personnel et de la vie privée, ainsi que la dignité humaine, la non-discrimination et les valeurs fondamentales ; elle fournira des solutions permettant aux individus de garder le contrôle sur les systèmes d'intelligence artificielle et de continuer à en comprendre le contenu.

**Par conséquent, la Conférence soutient les principes directeurs suivants, les adoptant comme valeurs fondamentales pour protéger les droits de l'homme dans le contexte du développement de l'intelligence artificielle :**

1. Les technologies d'intelligence artificielle et d'apprentissage automatique doivent être conçues, développées et utilisées dans le respect des droits fondamentaux de l'homme et conformément au **principe de loyauté**, en particulier en procédant ainsi :

- a. En tenant compte des attentes raisonnables des individus, en s'assurant à cette fin que l'utilisation des systèmes d'intelligence artificielle reste fidèle aux objectifs d'origine et que l'utilisation des données est compatible avec l'objectif premier de leur collecte ;

## Déclaration sur l'éthique et la protection des données dans le secteur de l'intelligence artificielle

- b. En tenant compte non seulement de l'impact de l'utilisation de l'intelligence artificielle sur l'individu, mais aussi de son impact collectif sur les groupes et la société dans son ensemble ;
- c. En s'assurant que les systèmes d'intelligence artificielle sont développés de façon à faciliter l'épanouissement de l'individu, sans l'entraver ni le mettre en danger, reconnaissant ainsi le besoin de définir et de délimiter certaines utilisations.

2. Il est nécessaire de **continuer à faire preuve d'attention et de vigilance**, ainsi que de transparence, en ce qui concerne les effets et les conséquences des systèmes d'intelligence artificielle, en particulier en procédant ainsi :

- a. En encourageant toutes les parties prenantes pertinentes à appliquer le principe de transparence au niveau des individus, des autorités de contrôle et d'autres tiers le cas échéant, y compris grâce à la réalisation d'audits, de suivi permanent et d'évaluation de l'impact des systèmes d'intelligence artificielle, ainsi que par un examen périodique des mécanismes de surveillance ;
- b. En encourageant la responsabilité collective et conjointe, impliquant l'ensemble des acteurs et des parties prenantes, par exemple pour le développement de normes collaboratives et le partage des bonnes pratiques ;
- c. En investissant dans une meilleure sensibilisation, l'éducation, la recherche et la formation afin de garantir un niveau satisfaisant d'information et de compréhension dans le domaine de l'intelligence artificielle et de ses effets potentiels sur la société ; et
- d. En mettant en place, pour tous les acteurs concernés, des processus de gouvernance dont on peut apporter la preuve, par exemple en s'appuyant sur des tiers dignes de confiance ou en créant des comités d'éthique indépendants.

3. Il convient d'améliorer la **transparence et l'intelligibilité** des systèmes d'intelligence artificielle, l'objectif étant de permettre une mise en œuvre efficace, en particulier en procédant ainsi :

- a. En investissant dans la recherche scientifique publique et privée portant sur l'intelligence artificielle explicable ;
- b. En promouvant la transparence, l'intelligibilité et l'accessibilité, par exemple en développant des modes de communication innovants, en tenant compte des différents niveaux de transparence et d'information requis en fonction de chaque audience concernée ;
- c. En rendant les pratiques des organisations plus transparentes, en particulier en mettant l'accent sur la transparence des algorithmes et la vérifiabilité des systèmes, tout en garantissant le sérieux des informations fournies ; et
- d. En garantissant la liberté des individus de maîtriser les informations les concernant, en particulier en s'assurant qu'ils sont toujours informés de façon appropriée lorsqu'ils interagissent directement avec un système d'intelligence artificielle ou qu'ils fournissent des données à caractère personnel qui seront traitées par de tels systèmes ;
- e. En fournissant des informations adéquates sur les objectifs et les effets de l'intelligence artificielle, 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.

4. Dans le cadre d'une approche globale basée sur l'« Ethics by design » (éthique intégrée), les systèmes d'intelligence artificielle doivent être **conçus et développés de manière responsable**, en appliquant les principes de **protection de la vie privée par défaut et protection intégrée de la vie privée**, en particulier en procédant ainsi :

- a. En mettant en œuvre des mesures et procédures technico-organisationnelles, en fonction du type de système développé, pour s'assurer du respect de la vie privée des personnes concernées et de leurs données à caractère personnel, aussi bien au moment de déterminer les méthodes de traitement des données que lors du traitement lui-même ;
- b. En évaluant et en décrivant les impacts attendus sur les individus et la société, au début d'un projet reposant sur l'intelligence artificielle et au cours des développements pertinents durant tout son cycle de vie ; et
- c. En identifiant les besoins spécifiques en vue d'une utilisation éthique et loyale des systèmes, et pour respecter les droits de l'homme, dans le cadre du développement et de l'exploitation de tout système d'intelligence artificielle.

5. Il convient de **donner davantage de pouvoirs à chaque personne** et d'encourager l'exercice des droits individuels, tout en créant des opportunités de participation publique, en particulier en procédant ainsi :

- a. En respectant les droits en matière de vie privée et de protection des données, y compris, le cas échéant, le droit à l'information, le droit d'accès, le droit de s'opposer au traitement des données et le droit à l'effacement, ainsi qu'en promouvant ces droits grâce aux campagnes d'éducation et de sensibilisation ;
- b. En respectant les droits connexes, y compris la liberté d'expression et d'information ainsi que la non-discrimination ;
- c. En reconnaissant que le droit d'opposition ou de recours s'applique aux technologies qui ont une influence sur les opinions ou le développement personnels et en garantissant, le cas échéant, le droit de chaque individu de ne pas faire l'objet d'une décision du simple fait d'un traitement automatique, si cette décision a un impact significatif sur le plan personnel et, si ce n'est pas le cas, en garantissant le droit de chaque individu de contester une telle décision;
- d. En utilisant les capacités des systèmes d'intelligence artificielle à doter chaque individu de pouvoirs égaux et à améliorer la participation publique, via des interfaces adaptables et des outils accessibles, par exemple.

6. Il convient de réduire et d'atténuer les **préjugés ou les discriminations** illicites pouvant résulter de l'utilisation de données présentes dans les systèmes d'intelligence artificielle, y compris en procédant ainsi :

- a. En assurant le respect des instruments juridiques internationaux en matière de droits de l'Homme et de non-discrimination ;
- b. En investissant dans la recherche portant sur les moyens techniques d'identifier, de traiter et de limiter les préjugés ;
- c. En prenant des mesures raisonnables pour s'assurer que les données et les informations à caractère personnel utilisées dans les prises de décision automatiques sont exactes, actuelles et aussi exhaustives que possible ; et
- d. En rédigeant des conseils et des principes spécifiques destinés au traitement des préjugés et de la discrimination, et en sensibilisant davantage les individus et les parties prenantes.

Tenant compte des principes ci-dessus, la 40e Conférence internationale des commissaires à la protection des données et de la vie privée appelle à l'établissement de **principes communs de gouvernance en matière d'intelligence artificielle**, encourageant les efforts internationaux concertés dans ce secteur, afin de veiller à ce que son développement et son utilisation se fassent dans le respect de l'éthique, et de la dignité et des valeurs humaines. Ces principes communs de gouvernance doivent pouvoir confronter les défis posés par les évolutions rapides des technologies d'intelligence artificielle, sur la base d'une approche impliquant plusieurs parties prenantes, afin de pouvoir aborder l'ensemble des questions intersectorielles en jeu. L'envergure doit être internationale, car le développement de l'intelligence artificielle est un phénomène qui va au-delà des frontières et qui a le potentiel de toucher toute l'humanité. La Conférence doit s'impliquer dans cet effort international, coopérer avec les autorités générales et sectorielles, et les soutenir, dans d'autres secteurs tels que la concurrence, la réglementation des marchés et les règles de protection des consommateurs.

Par conséquent, pour participer à la création d'une gouvernance commune future au niveau international, et mieux définir l'orientation qui accompagnera ces principes d'éthique et de protection des données en matière d'intelligence artificielle, la 40e Conférence internationale des commissaires à la protection des données et de la vie privée établit un **groupe de travail permanent** dont la mission sera de relever les défis associés au développement de l'intelligence artificielle. Ce **groupe de travail sur l'éthique et la protection des données en matière d'intelligence artificielle** sera chargé de favoriser la compréhension et le respect des principes de la présente résolution par toutes les parties concernées impliquées dans le développement des systèmes d'intelligence artificielle, y compris les gouvernements et les autorités publiques, les organismes de normalisation, les concepteurs de systèmes d'intelligence artificielle, les prestataires et les chercheurs, les entreprises, les citoyens et les utilisateurs finaux de ces systèmes. Le groupe de travail sur l'éthique et la protection des données en matière d'intelligence artificielle tiendra compte du travail effectué par d'autres groupes de travail de la Conférence et présentera ses activités à cette dernière de façon régulière. La Conférence s'efforce ainsi de soutenir de façon proactive un débat public dynamique sur l'éthique numérique, dans le but de créer une culture éthique solide et de sensibiliser chacun dans ce domaine.

- *La présente déclaration sera ouverte à une consultation publique* -