



# Assemblée générale

Distr. générale  
30 août 2023  
Français  
Original : espagnol

---

## Soixante-dix-huitième session

Point 73 b) de l'ordre du jour provisoire\*

**Promotion et protection des droits humains : questions relatives aux droits humains, y compris les divers moyens de mieux assurer l'exercice effectif des droits humains et des libertés fondamentales**

## Droit à la vie privée

### Note du Secrétaire général

Le Secrétaire général a l'honneur de transmettre à l'Assemblée générale le rapport établi par la Rapporteuse spéciale sur le droit à la vie privée, Ana Brian Nougrères, en application de la résolution [28/16](#) du Conseil des droits de l'homme.

---

\* [A/78/150](#)



## **Rapport de la Rapporteuse spéciale sur le droit à la vie privée, Ana Brian Nougrères**

### **Principes de transparence et d'explicabilité dans le traitement de données personnelles par l'intelligence artificielle**

#### *Résumé*

Dans le présent rapport, la Rapporteuse spéciale sur le droit à la vie privée, Ana Brian Nougrères, souligne l'importance des principes de transparence et d'explicabilité dans le traitement de données à caractère personnel par l'intelligence artificielle. L'omniprésence de l'intelligence artificielle dans toutes les activités et son utilisation dans la prise de décisions qui concernent les personnes imposent d'analyser la question et de mettre en place des mesures pour faire en sorte que le recours à l'intelligence artificielle soit éthique, responsable et respectueux des droits humains.

En effet, la transparence et l'explicabilité contribuent non seulement à instaurer la confiance dans l'intelligence artificielle, mais aussi à protéger les droits humains. En vertu de ces principes, d'une part, les personnes disposent en temps utile d'une information complète, simple et claire sur des aspects fondamentaux relatifs à l'utilisation de leurs informations personnelles dans les processus ou projets d'intelligence artificielle et sur les conséquences de ladite utilisation. D'autre part, les personnes affectées par l'intelligence artificielle doivent impérativement être informées des raisons précises à l'origine du préjudice subi. Ainsi, elles seront en mesure d'exercer leurs droits, par exemple le droit à une procédure régulière et le droit à la défense face aux décisions prises par les outils ou les technologies de l'intelligence artificielle.

## I. Introduction

1. Selon le Groupe d'experts de haut niveau sur l'intelligence artificielle<sup>1</sup> de la Commission européenne, les principes de transparence et d'explicabilité sont des éléments essentiels pour promouvoir une intelligence artificielle digne de confiance. Pour cela, l'intelligence artificielle doit être licite, éthique et robuste, « sur le plan tant technique que social car, même avec de bonnes intentions, les systèmes d'intelligence artificielle peuvent causer des préjudices involontaires »<sup>2</sup>.

2. Dans le même esprit, l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) indique que « [l]a transparence et l'explicabilité sont étroitement liées à des mesures adéquates en matière de responsabilité et d'obligation de rendre compte, ainsi qu'à la fiabilité des systèmes d'[intelligence artificielle] »<sup>3</sup>. Elle précise également que « [l]a transparence et l'explicabilité des systèmes d'intelligence artificielle sont souvent des conditions préalables essentielles pour garantir le respect, la protection et la promotion des droits humains, des libertés fondamentales et des principes éthiques »<sup>4</sup>.

3. L'intelligence artificielle est l'une des grandes questions prioritaires mondiales. Par exemple, fin décembre 2022, l'Organisation de coopération et de développement économiques (OCDE) a adopté une déclaration sur un avenir numérique de confiance, durable et inclusif<sup>5</sup> par laquelle elle entend œuvrer, entre autres aspects, pour favoriser une transformation numérique centrée sur l'humain et fondée sur les droits, qui aille de pair avec la promotion du respect des droits humains au sein comme en dehors de l'environnement numérique, de mécanismes solides de protection des données à caractère personnel, de lois et réglementations adaptées à l'ère numérique, et d'une utilisation fiable, sécurisée, responsable et durable des technologies numériques émergentes et de l'intelligence artificielle<sup>6</sup>. En ce qui concerne l'intelligence artificielle, les États membres de l'OCDE appellent cette organisation à appuyer l'élaboration de cadres cohérents, applicables et tournés vers l'avenir pour assurer la gouvernance de l'intelligence artificielle et gérer efficacement les risques connexes ; et fournir des données probantes, éléments prospectifs, outils et mécanismes de surveillance des incidents à l'appui d'une planification et d'une exécution efficaces des politiques pour tendre vers une intelligence artificielle digne de confiance<sup>7</sup>.

4. Pour leur part, le 23 janvier 2023, le Parlement européen, le Conseil et la Commission ont approuvé la Déclaration européenne sur les droits et principes numériques pour la décennie numérique, par laquelle les trois institutions se sont engagées à :

---

<sup>1</sup> Il s'agit d'un groupe d'experts indépendants constitué par la Commission européenne en juin 2018.

<sup>2</sup> Groupe d'experts de haut niveau sur l'intelligence artificielle, *Lignes directrices en matière d'éthique pour une intelligence artificielle digne de confiance* (2019), p. 2. Disponible sur : <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

<sup>3</sup> UNESCO, *Recommandation sur l'éthique de l'intelligence artificielle* (2021), p. 23. Disponible sur : [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre).

<sup>4</sup> Ibid, p. 22.

<sup>5</sup> OCDE, *Déclaration sur un avenir numérique de confiance, durable et inclusif* (2022). La déclaration a été adoptée à l'occasion de la réunion tenue sur l'île de la Grande Canarie (Espagne) les 14 et 15 décembre 2022. Disponible sur : <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0488>.

<sup>6</sup> Ibid.

<sup>7</sup> Ibid.

a) promouvoir des systèmes d'intelligence artificielle axés sur l'humain, fiables et éthiques tout au long de leur mise au point, de leur déploiement et de leur utilisation, conformément aux valeurs de l'[Union européenne] ;

b) assurer un niveau de transparence adéquat quant à l'utilisation des algorithmes et de l'intelligence artificielle, et à faire en sorte que les citoyens soient formés à les utiliser et qu'ils soient informés lorsqu'ils interagissent avec ces technologies ;

c) veiller à ce que les systèmes algorithmiques reposent sur des ensembles de données appropriés, afin d'éviter toute discrimination et de permettre une surveillance humaine de tous les résultats qui affectent la sécurité et les droits fondamentaux des citoyens ;

d) veiller à ce que les technologies telles que l'intelligence artificielle ne soient pas utilisées pour préjuger des choix des personnes, par exemple en matière de santé, d'éducation, d'emploi et de vie privée ;

e) prévoir des garanties et prendre des mesures appropriées, y compris en promouvant des normes fiables, pour que l'intelligence artificielle et les systèmes numériques soient, en permanence, sûrs et utilisés dans le plein respect des droits fondamentaux ;

f) prendre des mesures pour faire en sorte que la recherche en matière d'intelligence artificielle respecte les normes éthiques les plus élevées et la législation pertinente de l'[Union européenne]<sup>8</sup>.

5. À la lumière de ce qui précède, la suite du présent rapport expose quelques considérations sur l'intelligence artificielle puis aborde les grandes lignes des aspects qui suivent, dans le but de préciser en quoi consistent les principes de transparence et d'explicabilité dans le cadre du traitement des données à caractère personnel dans les processus ou les projets d'intelligence artificielle.

## II. Intelligence artificielle et traitement de données personnelles

6. L'intelligence artificielle est désormais omniprésente dans presque tous les aspects de notre société, depuis les appareils mobiles que les particuliers utilisent au quotidien jusqu'aux systèmes de gestion d'entreprise les plus complexes. Cette présence croissante de l'intelligence artificielle a ouvert un large éventail de possibilités dans divers secteurs et activités. Toutefois, ces possibilités nouvelles vont aussi de pair avec des défis et des risques qu'il convient de prendre en compte de manière responsable afin, notamment, d'exploiter pleinement le potentiel de l'intelligence artificielle dans le respect de la sécurité, de l'éthique et des droits humains.

7. Il n'y a pas de consensus sur la définition de l'intelligence artificielle, mais certaines idées ont été avancées pour classifier ce qui la constitue. Un ouvrage de référence sur le sujet a ainsi proposé la taxonomie suivante<sup>9</sup> :

- Systèmes qui pensent comme un humain (par exemple les architectures cognitives et les réseaux neuronaux).

<sup>8</sup> Parlement européen, Conseil et Commission, « Déclaration européenne sur les droits et principes numériques pour la décennie numérique », *Journal officiel de l'Union européenne*, 2023/C 23/01, 23 janvier 2023. Disponible sur : [https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ%3AJOC\\_2023\\_023\\_R\\_0001](https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ%3AJOC_2023_023_R_0001).

<sup>9</sup> Stuart Russell et Peter Norvig, *Artificial Intelligence: A Modern Approach* (Essex, England: Pearson, 2009).

- Systèmes qui agissent comme un humain (par exemple le raisonnement automatisé et l'apprentissage).
  - Systèmes qui pensent rationnellement (par exemple les inférences).
  - Systèmes qui agissent rationnellement (par exemple les agents logiciels intelligents et les robots intégrés qui atteignent des objectifs par la perception, la planification, le raisonnement, l'apprentissage, la communication, la prise de décision et l'action).
8. Tous ces systèmes traitent des informations pour produire des résultats, et ces informations contiennent notamment des données à caractère personnel. À cet égard, la Commission européenne a apporté les précisions suivantes :

Aux fins du présent Livre blanc ainsi que des débats qui pourront avoir lieu sur d'autres initiatives éventuelles, il semble important de préciser les principaux éléments dont se compose l'intelligence artificielle, à savoir les « données » et les « algorithmes ». L'intelligence artificielle peut être intégrée dans des dispositifs matériels. Dans le cas des techniques d'apprentissage automatique, qui constituent l'une des branches de l'intelligence artificielle, les algorithmes sont entraînés à reconnaître des structures à partir d'un ensemble de données afin de déterminer les actions à prendre pour atteindre un objectif donné<sup>10</sup>.

9. En d'autres termes, pour développer l'intelligence artificielle, d'énormes quantités d'informations sont collectées, stockées, analysées, traitées et utilisées pour générer divers résultats, actions ou comportements de la part des machines ou de leurs utilisateurs. Cependant, comme l'exige la recommandation précitée de l'UNESCO, « [l]a vie privée, qui constitue un droit essentiel pour la protection de la dignité, de l'autonomie et de la capacité d'action des êtres humains, doit être respectée, protégée et promue tout au long du cycle de vie des systèmes d'intelligence artificielle »<sup>11</sup>.

10. Un traitement approprié des données personnelles est indispensable pour éviter que les droits humains soient lésés ou menacés, le cas échéant, par le développement de l'intelligence artificielle. Plusieurs initiatives et organisations, dont voici quelques exemples, ont œuvré pour exiger un développement de l'intelligence artificielle qui respecte les droits humains.

11. L'Assemblée mondiale pour la protection de la vie privée a adopté en octobre 2020 la Résolution sur la responsabilisation dans le développement et l'utilisation de l'intelligence artificielle<sup>12</sup>. Cette résolution, entre autres, exhorte les organisations qui développent ou utilisent des systèmes d'intelligence artificielle à envisager de mettre en œuvre les mesures suivantes :

- Évaluer l'impact potentiel sur les droits de l'Homme (y compris la protection des données personnelles et le droit à la vie privée) avant le développement et/ou l'utilisation de l'intelligence artificielle ;
- Tester la robustesse, la fiabilité, l'exactitude et la sécurité des données de l'intelligence artificielle avant de la mettre en œuvre, y compris en identifiant et en traitant les biais dans les systèmes et les données qu'ils utilisent qui peuvent conduire à des résultats inévitables ;

<sup>10</sup> Commission européenne, *Livre blanc sur l'intelligence artificielle – Une approche européenne axée sur l'excellence et la confiance*, COM(2020) 65 final (2020). Disponible sur : <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1603192201335&uri=CELEX%3A52020DC0065>.

<sup>11</sup> Voir [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre), p. 22.

<sup>12</sup> Voir <https://globalprivacyassembly.org/wp-content/uploads/2020/11/GPA-Resolution-on-Accountability-in-the-Development-and-Use-of-AI-FR.pdf>, p. 3 et 4.

- Mettre en œuvre des mesures de responsabilisation appropriées concernant les risques d'interférence avec les droits de l'Homme.

12. Dans le même ordre d'idées, dans sa recommandation, l'UNESCO déclare :

Les systèmes algorithmiques nécessitent des évaluations adéquates de l'impact sur la vie privée, qui incluent également des considérations sociétales et éthiques de leur utilisation et une approche innovante consistant à appliquer le respect de la vie privée dès la conception des systèmes. Les acteurs de l'[intelligence artificielle] doivent assumer la responsabilité de la conception et de la mise en œuvre des systèmes d'[intelligence artificielle] de manière à garantir que les informations personnelles sont protégées tout au long du cycle de vie du système d'[intelligence artificielle]<sup>13</sup>.

13. Auparavant, en juin 2019, le Réseau ibéro-américain de protection des données avait publié les « Recommandations générales pour le traitement des données personnelles dans le cadre de l'intelligence artificielle »<sup>14</sup>. Ce document contient les recommandations suivantes destinées aux personnes qui développent des produits d'intelligence artificielle, afin de les guider pour que les exigences de la réglementation sur le traitement des données personnelles soient prises en compte dès la phase de conception :

- respecter les règles locales relatives au traitement des données personnelles ;
- réaliser des évaluations d'impact sur la vie privée ;
- prendre en compte le respect de la vie privée, l'éthique et la sécurité dès la conception, et par défaut ;
- démontrer la mise en œuvre du principe de responsabilisation (*accountability*) ;
- concevoir des structures de gouvernance appropriées pour le traitement des données personnelles dans les organisations développant des produits d'intelligence artificielle ;
- prendre des mesures pour garantir le respect des principes du traitement des données personnelles dans les projets d'intelligence artificielle ;
- respecter les droits des titulaires de données et mettre en œuvre des mécanismes efficaces pour l'exercice de ces droits ;
- garantir la qualité des données à caractère personnel ;
- utiliser des outils d'anonymisation ;
- renforcer la confiance et la transparence à l'égard des titulaires des données personnelles.

14. Afin de préciser les modalités de mise en œuvre de certaines de ces recommandations, le Réseau ibéro-américain de protection des données a élaboré des lignes directrices complémentaires et plus détaillées, qui font l'objet du document « Orientations spécifiques pour le respect des principes et des droits régissant la protection des données personnelles dans les projets d'intelligence artificielle »<sup>15</sup>. Le présent rapport détaille le principe de transparence qui sera abordé plus avant.

<sup>13</sup> Voir [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre), p. 22.

<sup>14</sup> Réseau ibéro-américain de protection des données, *Recomendaciones generales para el tratamiento de datos personales en la inteligencia artificial* (2019). Texte approuvé par les entités membres du Réseau lors de la session du 21 juin 2019, tenue à Naucalpan de Juárez (Mexique). Disponible sur : <https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf>.

<sup>15</sup> Réseau ibéro-américain de protection des données, *Orientaciones específicas para el*

### III. Risques inhérents à l'intelligence artificielle

15. La société, de même que sa transformation numérique, est influencée par l'intelligence artificielle qui est présente dans différents aspects de notre vie quotidienne, de l'économie, de la science, de l'éducation, de la santé et de bien d'autres domaines ou activités.

16. Si les possibilités et les avantages de l'intelligence artificielle pour la société en général sont indéniables, il ne faut pas perdre de vue qu'elle peut être source de défis, de dangers ou de risques qui lui sont inhérents. Parmi ces risques, on peut par exemple citer le manque d'éthique dans le développement ou l'utilisation de l'intelligence artificielle, ou encore la prise de décisions biaisées, non transparentes ou inappropriées concernant les êtres humains.

17. Les degrés de risque dépendent de chaque situation.

La Commission européenne considère qu'une application d'[intelligence artificielle] devrait généralement être considérée comme étant à haut risque en fonction de ce qui est en jeu, en examinant si des risques importants sont associés à la fois au secteur et à l'utilisation envisagée, notamment du point de vue de la protection de la sécurité, des droits des consommateurs et des droits fondamentaux. En particulier, une application d'[intelligence artificielle] devrait être considérée comme étant à haut risque si elle remplit cumulativement les deux critères suivants :

a) Premièrement, l'application d'[intelligence artificielle] est employée dans un secteur où, compte tenu des caractéristiques des activités normalement menées, des risques importants sont à prévoir. [...] Par exemple, les soins de santé ; les transports ; l'énergie et certains pans du secteur public [...]

b) Deuxièmement, l'application d'[intelligence artificielle] dans le secteur en question est, de surcroît, utilisée de façon telle que des risques importants sont susceptibles d'apparaître. [...] L'appréciation du niveau de risque d'une utilisation donnée pourrait reposer sur ses conséquences pour les parties concernées. Par exemple, les utilisations d'applications d'[intelligence artificielle] qui produisent des effets juridiques sur les droits d'une personne physique ou d'une entreprise, ou l'affectent de manière significative de façon similaire ; qui occasionnent un risque de blessure, de décès ou de dommage matériel ou immatériel important ; dont les effets ne peuvent être évités par les personnes physiques ou morales<sup>16</sup>.

18. L'intelligence artificielle implique différents types de risques. Parmi les éventualités à prendre en compte figurent notamment celles inhérentes au fonctionnement des algorithmes – biais humains, défaillances techniques, vulnérabilités en matière de sécurité, défaillances de mise en œuvre –, ainsi qu'à leur conception. Le schéma suivant représente les éléments considérés comme ayant une incidence sur la gestion des risques liés aux algorithmes<sup>17</sup> :

---

*cumplimiento de los principios y derechos que rigen la protección de los datos personales en los proyectos de inteligencia artificial* (2019), Disponible sur : <https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf>.

<sup>16</sup> Voir <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1603192201335&uri=CELEX%3A52020DC0065>.

<sup>17</sup> Voir <https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf>, p. 18.

## Gestión de riesgos de los algoritmos



## Factores de riesgo inherentes

19. Selon les auteurs de ce schéma :

les intrants sont principalement affectés par deux variables : les biais (incorporation de données partielles, insuffisantes, obsolètes ou manipulées) et la pertinence (données inadaptées, incohérentes ou incomplètes) ; par ailleurs, la mise en œuvre de l'algorithme peut être affectée par les modèles (biais dans la logique de programmation, inclusion de fonctions imprévues et défaillances inhérentes aux fonctions utilisées pour son codage) et par les erreurs (conditions opérationnelles qui reflètent un fonctionnement différent de celui prévu et qui vont à l'encontre des postulats de la conception envisagée). Enfin, les risques liés aux décisions de sortie sont liés à la pertinence et à la précision du résultat de l'exécution de l'algorithme et à la réponse à l'analyse des intrants)<sup>18</sup>.

## IV. Principe de transparence dans le traitement des données personnelles

20. La transparence est un concept utilisé dans différents domaines tels que l'informatique, l'accès à l'information, le droit et le traitement des données à caractère personnel. Selon l'UNESCO, « la transparence vise à fournir des informations appropriées aux destinataires respectifs afin de permettre la compréhension et de favoriser la confiance »<sup>19</sup>.

21. Il n'y a pas de consensus sur la portée de ce concept dans chacun de ces domaines et, même au sein de ceux-ci, il revêt des connotations différentes. Par exemple, il ne revient pas au même de se référer au principe de transparence dans le traitement des données personnelles en général que de s'y référer dans le contexte de l'intelligence artificielle. Le présent document traite de la transparence dans le traitement des données à caractère personnel en général et, plus spécifiquement dans le cadre de l'intelligence artificielle.

<sup>18</sup> Alejandro Useche et Jeimy Cano, *Robo-Advisors: Asesoría automatizada en el mercado de valores*, Universidad del Rosario y Autorregulador del Mercado de Valores de Colombia (2019), p. 9 et 10. Disponible sur : [https://www.researchgate.net/publication/331358231\\_Robo-Advisors\\_Asesoria\\_automatizada\\_en\\_el\\_mercado\\_de\\_valores](https://www.researchgate.net/publication/331358231_Robo-Advisors_Asesoria_automatizada_en_el_mercado_de_valores).

<sup>19</sup> UNESCO, *Recommandation sur l'éthique de l'intelligence artificielle* (2021), p. 22.



22. De très nombreux documents émanant d'organisations de différentes parties du monde mentionnent ce principe<sup>20</sup>. La Rapporteuse spéciale s'est déjà exprimée sur ce sujet, en indiquant que le principe de transparence suppose que le responsable du traitement informe la personne concernée des modalités de traitement auxquelles ses informations personnelles seront soumises, dès leur collecte, de manière à ce que la personne concernée soit en mesure d'exercer le pouvoir de contrôle qui lui est dévolu sur ces données<sup>21</sup>.

23. Le rapport cité ci-dessus analyse les documents internationaux suivants relatifs à la vie privée et au traitement des données personnelles, en ce qui concerne le principe de transparence : a) règlement général sur la protection des données de l'Union européenne ; b) Convention pour la protection des personnes à l'égard du traitement des données à caractère personnel ; c) Normes de protection des données personnelles pour les États ibéro-américains, approuvées par le Réseau ibéro-américain de protection des données ; d) recommandations du Conseil concernant les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel de l'Organisation de coopération et de développement économiques ; e) Cadre sur la vie privée de l'Association de coopération économique Asie-Pacifique ; f) Principes actualisés sur la vie privée et la protection des données personnelles, avec annotations, de l'Organisation des États américains.

24. De l'analyse de ces documents, il ressort, de façon générale, que des informations doivent être données sur les éléments suivants :

- l'identité et les coordonnées du responsable du traitement des données ou de son représentant, ainsi que les finalités ou les objectifs du traitement. Ces informations constituent les principaux fondements de la transparence ;
- les droits de la personne concernée ainsi que la manière de les exercer, ainsi que les destinataires ou la catégorie de destinataires des données ;
- le fondement ou la base juridique autorisant le traitement, ainsi que l'existence et les principales caractéristiques du traitement ;
- la catégorie de données traitées et leur origine lorsqu'elles n'ont pas été obtenues directement auprès de la personne concernée.

<sup>20</sup> Organisation de coopération et de développement économiques (OCDE), *Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel*, 23 septembre 1980 et amendement de juillet 2013 ; Conseil de l'Europe, Convention 108 pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, 28 janvier 1981 ; Nations Unies, Principes directeurs pour la réglementation des fichiers personnels informatisés, 14 décembre 1990 ; Conseil de l'Europe, Protocole additionnel à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel concernant les autorités de contrôle et les flux transfrontières de données, 8 novembre 2001 ; Association de coopération économique Asie-Pacifique, *Cadre sur la vie privée de l'Association de coopération économique Asie-Pacifique*, 2004 ; Agence espagnole de protection des données, *Proposition conjointe pour un projet de normes internationales sur la protection de la vie privée au regard du traitement de données personnelles*, Madrid, 5 novembre 2009 ; Parlement européen et Conseil de l'Union européenne, Règlement (UE) 2016/679 du Parlement européen et du Conseil relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) du 27 avril 2016 ; Réseau ibéro-américain de protection des données, *Directrices para la Armonización de la Protección de Datos en la Comunidad Iberoamericana*, 2017 ; Conseil de l'Europe, Protocole d'amendement à la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel, octobre 2018 ; Organisation des États américains, Comité juridique interaméricain, *Principes actualisés relatifs à la vie privée et à la protection des données personnelles*, 2021.

<sup>21</sup> [A/77/196](#), par. 45.

25. Il est important de souligner que, pour respecter le principe de transparence, l'information dont doit disposer la personne concernée doit être communiquée dans un langage clair, simple et facile à comprendre. Cette règle doit également s'appliquer s'il s'agit d'enfants ou d'adolescents, en apportant les adaptations correspondantes.

26. Tous les instruments normatifs n'exigent pas la fourniture des mêmes informations, d'ailleurs certains d'entre eux prévoient une liste plus importante de ce qui doit être divulgué. Ainsi, le règlement général de l'Union européenne sur la protection des données prescrit notamment la communication des éléments suivant<sup>22</sup> : les coordonnées du délégué à la protection des données ; la durée de conservation des données à caractère personnel ou les critères utilisés pour déterminer cette durée ; si le responsable prévoit de communiquer ou de transférer des données, la réglementation qui l'y autorise ; le droit d'introduire une réclamation auprès d'une autorité de contrôle ; les informations sur l'exigence de fourniture de données si elle a un caractère réglementaire ou contractuel ou si elle conditionne la conclusion d'un contrat et si la personne concernée est tenue de fournir ses données personnelles, et sur les conséquences de leur non-fourniture ; l'existence d'une prise de décision automatisée, y compris un profilage, auquel cas les informations utiles concernant la logique sous-jacente, l'importance et les conséquences prévues de ce traitement doivent être communiquées ; les informations sur la finalité poursuivie si le responsable prévoit un traitement ultérieur pour une finalité autre que celle pour laquelle les données ont été obtenues.

## V. Principe de transparence dans le traitement des données personnelles dans le cadre de l'intelligence artificielle

27. Il est essentiel de garantir la transparence de l'intelligence artificielle, car le fait de la négliger ou de l'omettre peut avoir des effets négatifs. À cet égard, il a été souligné que :

En raison du manque de transparence (opacité de l'[intelligence artificielle]), il est difficile de déceler et de prouver d'éventuelles infractions à la législation, notamment aux dispositions juridiques qui protègent les droits fondamentaux, mais aussi d'imputer la responsabilité et de remplir les conditions requises pour prétendre à une indemnisation<sup>23</sup>.

28. La possible opacité de l'intelligence artificielle peut être atténuée en imposant le respect d'exigences minimales de transparence. Il a donc été suggéré de :

Veiller à ce que des informations claires soient fournies en ce qui concerne les capacités et les limites du système d'[intelligence artificielle], en particulier l'objectif qu'il poursuit, les conditions dans lesquelles il devrait fonctionner comme prévu et le niveau de précision attendu dans la réalisation de l'objectif spécifié [...]. Par ailleurs, des informations devraient être clairement fournies aux citoyens lorsqu'ils interagissent avec un système d'[intelligence artificielle] et non avec un être humain [...]. Il importe en outre que les informations fournies soient objectives, concises et facilement compréhensibles<sup>24</sup>.

<sup>22</sup> Voir <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1532348683434&uri=CELEX%3A02016R0679-20160504>.

<sup>23</sup> Commission européenne, *Livre blanc sur l'intelligence artificielle – Une approche européenne axée sur l'excellence et la confiance* (2020), p. 16.

<sup>24</sup> Ibid., p. 23 et 24.

29. La recommandation de l'UNESCO indique que :

Concernant spécifiquement les systèmes d'[intelligence artificielle], la transparence peut permettre aux individus de comprendre comment chaque étape d'un système d'[intelligence artificielle] est mise en place, en fonction du contexte et du degré de sensibilité du système concerné. Il est possible également de fournir des informations sur les facteurs qui influencent une prévision ou une décision particulière, et sur l'existence ou non des garanties appropriées (telles que des mesures de sécurité ou d'équité)<sup>25</sup>.

30. Le Groupe d'experts de haut niveau sur l'intelligence artificielle a mis en évidence que, pour parvenir à une intelligence artificielle digne de confiance, un certain nombre d'exigences doivent être respectées, en particulier la transparence, ce qui implique de :

Fournir de façon proactive des informations claires aux parties prenantes sur les capacités et les limites des systèmes d'[intelligence artificielle], afin de leur permettre de formuler des attentes réalistes, ainsi que sur la manière dont les exigences sont mises en œuvre [;] et [f]aire preuve de transparence sur le fait qu'elles interagissent avec un système d'[intelligence artificielle]<sup>26</sup>.

De même, la recommandation précitée de l'UNESCO explique que « les acteurs de l'[intelligence artificielle] devraient informer les utilisateurs lorsqu'un produit ou un service est fourni directement ou par le biais de systèmes d'intelligence artificielle de manière appropriée et en temps utile »<sup>27</sup>.

31. Ainsi, selon ladite recommandation de l'UNESCO :

[L]'explicabilité est donc étroitement liée à la transparence, puisqu'il convient de rendre les résultats et les sous-processus qui y conduisent intelligibles et traçables, en fonction du contexte. Les acteurs de l'[intelligence artificielle] devraient s'engager à assurer l'explicabilité des algorithmes mis au point. Dans le cas d'applications de l'[intelligence artificielle] dont l'impact sur l'utilisateur final n'est pas temporaire, facilement réversible ou à faible risque, il convient de s'assurer qu'une explication valable est fournie avec toute décision ayant entraîné l'action entreprise, afin que les résultats soient considérés comme transparents<sup>28</sup>.

32. Le Groupe d'experts de haut niveau sur l'intelligence artificielle explique que la transparence « est étroitement liée au principe de l'explicabilité et comprend la transparence des éléments pertinents d'un système d'[intelligence artificielle] : les données, le système et les modèles économiques »<sup>29</sup>. Le Groupe souligne en outre dans les termes suivants l'importance de la traçabilité, de l'explicabilité et de la communication :

- Traçabilité : Les ensembles de données et les processus permettant au système d'[intelligence artificielle] de rendre une décision, y compris les processus de collecte et d'étiquetage de données, ainsi que les algorithmes utilisés, devraient être documentés selon les normes les plus strictes afin de permettre la traçabilité ainsi qu'une amélioration de la transparence. Ce principe s'applique également aux décisions rendues par le système d'[intelligence artificielle]. Cela permet de déterminer les raisons pour lesquelles une décision d'[intelligence

<sup>25</sup> Voir [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre), p. 22 et 23.

<sup>26</sup> Voir <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, p. 3.

<sup>27</sup> Voir [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre), p. 22.

<sup>28</sup> Ibid., p. 23.

<sup>29</sup> Voir <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, p. 22.

artificielle] était erronée ce qui, en retour, pourrait contribuer à éviter de futures erreurs. La traçabilité facilite donc l'auditabilité et l'explicabilité.

- **Explicabilité** : L'explicabilité concerne la capacité d'expliquer à la fois les processus techniques d'un système d'[intelligence artificielle] et les décisions humaines qui s'y rapportent (par exemple, domaines d'application d'un système d'[intelligence artificielle]). L'explicabilité technique suppose que les décisions prises par un système d'[intelligence artificielle] puissent être comprises et retracées par des êtres humains. Par ailleurs, des arbitrages peuvent s'avérer nécessaires entre le renforcement de l'explicabilité d'un système (qui pourrait réduire sa précision) et l'amélioration de sa précision (au détriment de l'explicabilité). Dès qu'un système d'[intelligence artificielle] a une incidence importante sur la vie des personnes, il devrait être possible d'exiger une explication appropriée du processus de décision du système d'[intelligence artificielle]. Ces explications devraient être présentées en temps opportun et adaptées à l'expertise de la partie prenante concernée (par exemple, non-spécialiste, autorité de réglementation ou chercheur). Des explications devraient également être fournies sur la mesure dans laquelle un système d'[intelligence artificielle] influence et façonne le processus de prise de décisions organisationnel, les choix opérés dans la conception du système, et la justification de son déploiement (de manière à assurer la transparence du modèle économique).
- **Communication** : Les systèmes d'[intelligence artificielle] ne devraient pas se présenter comme des êtres humains auprès des utilisateurs ; lorsqu'ils interagissent avec un système d'[intelligence artificielle], les êtres humains ont le droit d'en être informés. Cet aspect implique que les systèmes d'[intelligence artificielle] doivent être identifiables en tant que tels. Qui plus est, la possibilité de s'opposer à cette interaction au profit d'une interaction humaine devrait être proposée le cas échéant afin de garantir le respect des droits fondamentaux. Outre cet aspect, il convient de communiquer aux professionnels de l'[intelligence artificielle] ou aux utilisateurs finaux des informations appropriées sur les capacités et les limites du système d'[intelligence artificielle], selon des modalités adaptées au contexte d'utilisation concerné. Ces informations pourraient comprendre le degré de précision du système d'[intelligence artificielle], ainsi que ses limites<sup>30</sup>.

33. À ce propos, le Comité européen de la protection des données et le Contrôleur européen de la protection des données déclarent dans un avis commun que :

Les personnes concernées devraient toujours être informées lorsque leurs données sont utilisées à des fins de formation et/ou de prévision en matière d'[intelligence artificielle], de la base juridique de ce traitement, de l'explication générale de la logique (procédure) et de la portée du système d'[intelligence artificielle]. À cet égard, le droit des personnes à limiter le traitement (article 18 du [règlement (UE) 2016/679, général sur la protection des données] et article 20 du règlement relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les institutions, organes et organismes de l'Union et à la libre circulation de ces données) ainsi qu'à supprimer/effacer des données (article 16 du [règlement (UE) 2016/679, général sur la protection des données] et article 19 du règlement relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les institutions, organes et organismes de l'Union et à la libre circulation de ces données) devrait toujours être garanti

<sup>30</sup> Ibid.

dans ces cas. En outre, le responsable du traitement devrait être explicitement tenu d'informer la personne concernée des délais d'objection, de limitation, de suppression des données, etc. Le système d'[intelligence artificielle] doit être en mesure de satisfaire à toutes les exigences en matière de protection des données au moyen de mesures techniques et organisationnelles adéquates. Un droit à l'explication devrait permettre une plus grande transparence<sup>31</sup>.

34. Dans son rapport précité<sup>32</sup>, la Rapporteuse spéciale relève que, si la personne concernée est exposée à des prises de décisions automatisées ou à un profilage, elle doit pouvoir comprendre les modalités du traitement auquel seront soumises ses données, et par exemple s'il s'agit d'intelligence artificielle, il faut lui fournir des informations utiles concernant la logique sous-jacente et l'importance et les conséquences prévues.

35. Sur ce point, l'Agence espagnole de protection des données rappelle que « le terme « utile » [...] doit être interprété comme une information qui, fournie à la personne concernée, lui permet de connaître le type de traitement qui est effectué sur ses données, lui procure des certitudes et lui donne confiance quant aux résultats »<sup>33</sup>.

36. Par ailleurs, l'Agence souligne que :

S'acquitter de cette obligation en fournissant des renseignements techniques sur la mise en œuvre de l'algorithme peut s'avérer opaque, déroutant et même entraîner une certaine fatigue informationnelle. Il convient donc de donner des informations qui permettent de comprendre le comportement du traitement. Bien que cela dépende du type de composant [d'intelligence artificielle] utilisé, voici des exemples d'informations qui pourraient être pertinentes pour la personne concernée :

- le détail des données utilisées pour la prise de décision, au-delà de leur catégorie, et en particulier des informations sur la durée d'utilisation des données (leur ancienneté) ;
- l'importance relative de chacune d'entre elles dans le processus décisionnel ;
- la qualité des données d'entraînement et le type de modèles utilisés ;
- les profilages effectués et leurs implications ;
- les valeurs de précision ou d'erreur selon la formule appropriée pour mesurer la qualité de l'inférence ;
- l'existence ou non de supervision humaine qualifiée ;
- la référence aux audits, notamment en ce qui concerne les possibles distorsions des résultats des inférences, ainsi que la ou les certifications effectuées sur le système [d'intelligence artificielle]. Dans le cas de systèmes adaptatifs ou évolutifs, le dernier audit effectué ;

<sup>31</sup> Comité européen de la protection des données et Contrôleur européen de la protection des données, *Avis conjoint 5/2021 sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle)*, 18 juin 2021, p. 20. Disponible sur :

[https://edpb.europa.eu/system/files/2021-10/edpb-edps\\_joint\\_opinion\\_ai\\_regulation\\_fr.pdf](https://edpb.europa.eu/system/files/2021-10/edpb-edps_joint_opinion_ai_regulation_fr.pdf).

<sup>32</sup> A/77/196, par. 55.

<sup>33</sup> Agence espagnole de protection des données, *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*, février 2020, p. 24. Disponible sur : <https://www.aepd.es/sites/default/files/2020-02/adequacion-rgpd-ia.pdf>.

- si le système [d'intelligence artificielle] contient des informations sur des tiers identifiables, l'interdiction de traiter ces informations sans motif légitime et les conséquences d'un tel traitement<sup>34</sup>.

37. Dans l'un de ses avis, le Contrôleur européen de la protection des données suggère que, si la Commission devait présenter un nouveau cadre réglementaire spécifique à l'intelligence artificielle, un certain nombre de garanties raisonnables devraient s'appliquer à toutes les applications d'intelligence artificielle, indépendamment du niveau de risque, comme la mise en place de mesures techniques et organisationnelles (y compris la documentation), la transparence complète concernant les objectifs, l'utilisation et la conception des systèmes algorithmiques mis en œuvre, la solidité du système d'intelligence artificielle ou la mise en œuvre et la transparence des mécanismes de responsabilité, de recours et de contrôle indépendants disponibles<sup>35</sup>.

38. Le Comité européen de la protection des données et le Contrôleur européen de la protection des données, pour leur part, insistent sur la nécessité de promouvoir :

de nouveaux moyens, plus proactifs et opportuns, d'informer les utilisateurs des systèmes d'[intelligence artificielle] sur le statut (décisionnel) du système à tout moment, en les avertissant rapidement des éventuels effets préjudiciables, de manière à ce que les personnes dont les droits et libertés peuvent être lésés par des décisions autonomes de la machine puissent réagir ou remédier à la décision<sup>36</sup>.

39. Le Réseau ibéro-américain de protection des données considère que, pour concrétiser le principe de transparence, il convient de<sup>37</sup> :

- « Communiquer à la personne concernée les principales caractéristiques du traitement auquel ses données personnelles seront soumises » ;
- « Informer expressément les personnes concernées que des processus d'automatisation seront utilisés dans le traitement de leurs données personnelles » ;
- « Indiquer dans les supports choisis par les responsables du traitement pour respecter le principe de transparence toutes les finalités pour lesquelles les données des personnes concernées seront traitées » ;
- « Communiquer l'origine des données personnelles lorsqu'elles sont obtenues au moyen d'un transfert et, dans le cas particulier où elles sont destinées à être utilisées dans le cadre de l'intelligence artificielle, s'assurer que cette finalité a été communiquée par le premier responsable du traitement qui les a obtenues afin de pouvoir utiliser les données à cette fin » ;
- Mettre au point des méthodes innovantes pour sensibiliser les personnes concernées aux principales caractéristiques du traitement et au degré de risque lié à la hausse ou à la baisse des attentes en matière de respect de la vie privée ;

<sup>34</sup> Ibid.

<sup>35</sup> Contrôleur européen de la protection des données, *Avis 4/2020 du CEPD sur le Livre blanc de la Commission européenne intitulé « Intelligence artificielle – Une approche européenne axée sur l'excellence et la confiance »*, 29 juin 2020, p. 15. Disponible sur : [https://edps.europa.eu/sites/default/files/publication/20-06-19\\_ai\\_white\\_paper\\_fr.pdf](https://edps.europa.eu/sites/default/files/publication/20-06-19_ai_white_paper_fr.pdf).

<sup>36</sup> Comité européen de la protection des données et Contrôleur européen de la protection des données, *Avis conjoint 5/2021 sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle)*, 18 juin 2021, p. 22 et 23.

<sup>37</sup> Voir <https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf>, p. 17 à 19.

- « Préserver le droit à l'autodétermination informationnelle, en veillant à ce que les personnes concernées soient toujours informées de manière adéquate et en temps utile qu'elles vont interagir directement avec un système d'intelligence artificielle ou que leurs informations vont être traitées par l'intelligence artificielle » ;
- « Fournir des informations utiles sur la finalité et les effets des systèmes d'intelligence artificielle afin de vérifier la conformité permanente par rapport aux attentes des personnes concernées en matière de respect de la vie privée, en leur permettant à tout moment d'exercer un contrôle sur le traitement de leurs données personnelles » ;
- « Recenser et définir les termes couramment utilisés et créer une base de données afin qu'ils puissent être réutilisés dans différents contextes, avec des icônes normalisées pour fournir des informations aux personnes concernées » ;
- « Informer en permanence les personnes concernées afin qu'elles puissent savoir comment les décisions automatisées peuvent les affecter et, le cas échéant, demander une intervention humaine, dans le but de leur permettre de décider en connaissance de cause de consentir ou non au traitement ».

40. Le Réseau ibéro-américain de protection des données précise que :

Les informations fournies sur la logique du modèle [d'intelligence artificielle] doivent comprendre au moins les aspects fondamentaux de son fonctionnement, ainsi que la pondération et la corrélation des données, et elles doivent être rédigées dans un langage clair, simple et facilement compréhensible, de sorte qu'il ne soit pas nécessaire de fournir des explications détaillées sur les algorithmes utilisés, ni même de les évoquer<sup>38</sup>.

41. Le Réseau ibéro-américain de protection des données incite les responsables du traitement des données dans le domaine de l'intelligence artificielle à faire preuve d'innovation en communiquant des informations de manière concise et simple. Il remarque « qu'il existe plusieurs méthodes innovantes pour fournir les déclarations de confidentialité, notamment l'utilisation de vidéos, de petits dessins animés et d'icônes normalisées. Combiner de telles méthodes peut contribuer à rendre les informations complexes sur l'[intelligence artificielle] plus faciles à comprendre pour les titulaires des données personnelles »<sup>39</sup>.

42. Les paragraphes suivants présentent, à titre indicatif et non exhaustif, des exemples de pays qui ont explicitement ou implicitement abordé dans leur réglementation interne le principe de transparence dans le traitement des données à caractère personnel par l'intelligence artificielle.

43. En Équateur, la loi organique sur la protection des données, approuvée en 2021, établit dans son article 12, paragraphes 14 et 17, le droit d'être informé de l'existence du droit de ne pas faire l'objet d'une décision fondée uniquement sur des évaluations automatisées et de la manière d'exercer ce droit, ainsi que sur l'existence d'évaluations et de décisions automatisées, y compris le profilage.

44. La même loi précise que, si les données sont collectées directement auprès de la personne concernée, l'information doit lui être communiquée au préalable, c'est-à-dire au moment du recueil de ces données. En outre, l'article 12 ajoute :

Lorsque les données à caractère personnel ne sont pas collectées directement auprès de la personne concernée ou qu'elles le sont auprès d'une source

<sup>38</sup> Voir <https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf>, p. 18.

<sup>39</sup> Ibid.

accessible au public, la personne concernée doit être informée dans un délai de trente (30) jours ou lors de la première communication avec la personne concernée, selon ce qui se produit en premier. La personne concernée doit recevoir des informations expresses, non équivoques, transparentes, intelligibles, concises, précises et sans difficultés techniques.

45. Au Pérou, l'article 72 du règlement de la loi n° 29733 sur la protection des données personnelles fait référence au droit au traitement objectif de telles données, en précisant que :

Afin de garantir l'exercice du droit au traitement objectif conformément aux dispositions de l'article 23 de la loi<sup>40</sup>, lorsque des données personnelles sont traitées dans le cadre d'un processus décisionnel sans la participation de la personne concernée, le propriétaire de la banque de données personnelles ou le responsable du traitement doit informer dans les meilleurs délais la personne concernée, sans préjudice de ce qui est prévu pour l'exercice des autres droits dans la loi et le [...] règlement.

46. À Sao Tomé-et-Principe, la loi n° 3/2016 du 2 mai 2016, relative à la protection des données personnelles des personnes physiques, contient une disposition particulière : son article 21 prévoit en effet que le responsable du traitement, ou son représentant, doit informer l'Agence nationale de protection des données personnelles, par écrit et dans les huit jours précédant le traitement, de la date de début du traitement ou de l'ensemble de traitements, en tout ou partie automatisés, destinés à réaliser une ou plusieurs finalités interdépendantes, sous réserve de certaines exceptions. En outre, la personne concernée, lorsqu'elle exerce son droit d'accès, a le droit d'être informée par le responsable des motifs qui sous-tendent le traitement automatisé des données la concernant, conformément à l'article 11.

47. En Uruguay, la loi n° 18831 du 11 août 2008, relative à la protection des données personnelles, prévoit dans son article 13 que les personnes concernées ont le droit d'être informées de manière expresse, précise et non équivoque, et préalablement à la collecte des données, des critères d'évaluation, des processus appliqués et de la solution technologique ou du programme utilisé – dans le cas d'un traitement automatisé de données visant à évaluer certains aspects de leur personnalité, tels que la performance professionnelle, la solvabilité, la fiabilité, le comportement, entre autres – dans le but de prendre des décisions avec des effets juridiques susceptibles d'affecter de manière significative la personne concernée. La loi ajoute en outre que « [l]orsque les données personnelles ne sont pas collectées directement auprès des personnes concernées, les informations [...] doivent leur être fournies dans un délai de cinq jours ouvrables à compter de la réception de la demande par les responsables » du traitement.

## **VI. Principe d'explicabilité dans le traitement de données personnelles dans les projets d'intelligence artificielle**

48. La création de « profils virtuels » d'individus à partir d'informations existantes est de plus en plus courante. En outre, les décisions concernant ces personnes sont

<sup>40</sup> « Article 23. Droit au traitement objectif. Le titulaire des données personnelles a le droit de ne pas faire l'objet d'une décision produisant des effets juridiques à son égard ou l'affectant de manière significative, sur la seule base d'un traitement de données personnelles destiné à évaluer certains aspects de sa personnalité ou de son comportement, à moins que cela ne se produise dans le cadre de la négociation, de la conclusion ou de l'exécution d'un contrat ou dans les cas d'évaluation aux fins d'intégration dans un organisme public, conformément à la loi, sans préjudice de la possibilité de défendre son point de vue, afin de préserver son intérêt légitime. »



souvent prises sur la base d'un traitement automatisé de leurs données à l'aide de divers outils technologiques.

49. Les êtres humains peuvent être affectés positivement ou négativement par les décisions prises à leur sujet à partir de l'utilisation et du traitement des données dans les projets d'intelligence artificielle. La protection des droits des personnes concernées par les décisions prises à leur sujet à l'aide d'outils ou de technologies d'intelligence artificielle suscite des inquiétudes. Le Livre blanc sur l'intelligence artificielle, par exemple, relève que : « [à] l'instar de toute technologie nouvelle, l'utilisation de l'[intelligence artificielle] crée à la fois des possibilités et des risques. Les citoyens craignent d'être impuissants à défendre leurs droits et leur sécurité lorsqu'ils sont confrontés à l'asymétrie de l'information en matière de prise de décision algorithmique »<sup>41</sup>.

50. Par conséquent, il est nécessaire que les personnes sachent quelles données ont été utilisées pour prendre une décision qui les concerne et quelle a été la logique utilisée pour y parvenir. L'accès à ces informations, entre autres, aidera la personne concernée à savoir si la décision prise à son égard est correcte et, si elle ne l'est pas, à se défendre. En d'autres termes, ces informations sont indispensables pour garantir une procédure régulière, car ce seront les preuves sur lesquelles se fonder pour examiner les éventuelles inexactitudes ou injustices commises à l'égard d'une personne à l'occasion du traitement de ses données personnelles dans le cadre de processus d'intelligence artificielle. À cet égard, le Groupe d'experts de haut niveau sur l'intelligence artificielle a insisté sur le fait que l'explicabilité :

est essentielle pour renforcer et conserver la confiance des utilisateurs envers les systèmes d'[intelligence artificielle]. Cela signifie que les processus doivent être transparents, que les capacités et la finalité des systèmes d'[intelligence artificielle] doivent être communiquées ouvertement, et que les décisions – dans la mesure du possible – doivent pouvoir être expliquées aux personnes directement et indirectement concernées. Sans ces informations, une décision ne peut être dûment contestée. [...]. La mesure dans laquelle l'explicabilité est nécessaire dépend fortement du contexte et de la gravité des conséquences si ce résultat est erroné ou imprécis d'une autre manière<sup>42</sup>.

51. Pour toutes ces raisons, la transparence dans le domaine de l'intelligence artificielle est primordiale, sachant que l'intelligence artificielle ne doit être ni obscure, ni secrète, ni trompeuse. C'est pourquoi la Déclaration européenne précitée indique que :

Toute personne devrait être en mesure de bénéficier des avantages qu'offrent les systèmes algorithmiques et d'intelligence artificielle, y compris en faisant des choix libres et éclairés dans l'environnement numérique, tout en étant protégée contre les risques et les atteintes à sa santé, à sa sécurité et à ses droits fondamentaux<sup>43</sup>.

52. Dans l'esprit de ce qui précède, le Réseau ibéro-américain de protection des données a recommandé en 2019 d'accroître la transparence à l'égard des titulaires des données à caractère personnel<sup>44</sup>.

<sup>41</sup> Voir <https://eur-lex.europa.eu/legal-content/FR/TXT/?qid=1603192201335&uri=CELEX%3A52020DC0065>, p. 10.

<sup>42</sup> Voir <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>, p. 16.

<sup>43</sup> Voir [https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ%3AJOC\\_2023\\_023\\_R\\_0001](https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=OJ%3AJOC_2023_023_R_0001).

<sup>44</sup> Voir <https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf>, p. 23 et 24.

53. Par la suite, et également dans le prolongement de ce qui précède, la résolution de 2020 de l'Assemblée mondiale de la protection de la vie privée, citée plus haut, exhorte les organisations qui développent ou utilisent des systèmes d'intelligence artificielle à envisager de mettre en œuvre les mesures suivantes : a) garantir la transparence et l'ouverture en divulguant l'utilisation de l'[intelligence artificielle], les données utilisées et la logique utilisée par l'[intelligence artificielle] ; b) garantir l'identification d'un acteur humain responsable auprès duquel les préoccupations liées aux décisions automatisées peuvent être soulevées et les droits peuvent être exercés, et qui peut déclencher l'évaluation du processus de décision et l'intervention humaine ; c) fournir sur demande des explications dans un langage clair et compréhensible pour les décisions automatisées prises par l'[intelligence artificielle] ; d) effectuer sur demande une intervention humaine sur la décision automatisée prise par l'[intelligence artificielle]<sup>45</sup>.

54. Tout ce qui précède correspond en partie aux dispositions du règlement général européen sur la protection des données qui dispose, par exemple, que :

Lorsque les données à caractère personnel n'ont pas été collectées auprès de la personne concernée, le responsable du traitement fournit à celle-ci toutes les informations suivantes : [...] 2. [...] g) l'existence d'une prise de décision automatisée, y compris un profilage, visée à l'article 22, paragraphes 1 et 4, et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée<sup>46</sup>.

En outre, la personne concernée a le droit :

d'obtenir du responsable du traitement la confirmation que des données à caractère personnel la concernant sont ou ne sont pas traitées et, lorsqu'elles le sont, l'accès aux dites données à caractère personnel ainsi que les informations suivantes : [...] h) l'existence d'une prise de décision automatisée, y compris un profilage, visée à l'article 22, paragraphes 1 et 4, et, au moins en pareils cas, des informations utiles concernant la logique sous-jacente, ainsi que l'importance et les conséquences prévues de ce traitement pour la personne concernée<sup>47</sup>.

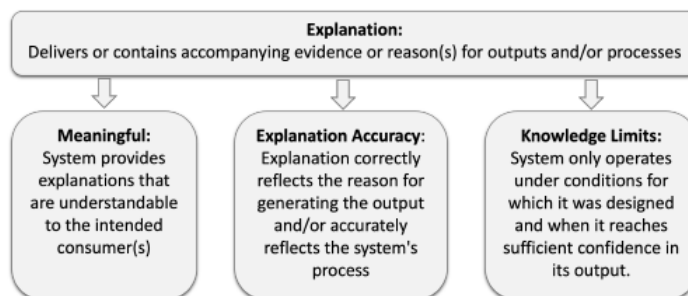
55. Dans le schéma ci-dessous, le National Institute of Standards and Technology synthétise la portée de ce principe<sup>48</sup> :

<sup>45</sup> Voir <https://globalprivacyassembly.org/document-archive/adopted-resolutions/>, p. 3.

<sup>46</sup> Voir <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:32016R0679>, art. 14, par. 2 g).

<sup>47</sup> *Ibid.*, art. 15, par. 1.

<sup>48</sup> National Institute of Standards and Technology, *Four Principles of Explainable Artificial Intelligence*, – NISTIR 8312 (2021), p. 3. Disponible sur : <https://doi.org/10.6028/NIST.IR.8312>.



**Fig. 1.** Illustration of the four principles of explainable artificial intelligence. Arrows indicate that for a system to be explainable, it must provide an explanation. The remaining three principles are the fundamental properties of those explanations.

56. Le tableau suivant expose les grandes lignes de chaque principe, selon le document du National Institute of Standards and Technology<sup>49</sup> :

<i>Principe</i>	<i>Signification ou portée</i>
Explanation (Explication)	Les éléments de preuve, les arguments ou le raisonnement relatifs à un résultat ou à un processus d'un système [d'intelligence artificielle] doivent être fournis.
Meaningful (Compréhension)	L'explication doit être donnée dans des termes qui permettent à la personne de la comprendre. Autrement dit, ce principe vise à rendre l'explication intelligible pour un public donné.  De nombreux facteurs influencent une bonne explication, c'est pourquoi le public cible ou le public auquel l'explication s'adresse doit être pris en compte.
Explanation Accuracy (Précision de l'explication)	Ce principe exige que l'explication technique soit rigoureuse, précise et complète.
Knowledge Limits (Limites de la connaissance)	Cerner et énoncer les limites de la connaissance implique de préciser que le système n'est ni parfait ni infaillible, car [l'intelligence artificielle] fonctionne selon certaines limites et conditions dans lesquelles elle a été programmée. Elle dépend aussi, notamment, de la qualité et de la quantité des informations traitées.

57. Il a été avancé que l'explication doit : a) « être compréhensible et convaincante pour l'utilisateur » ; b) « refléter fidèlement le raisonnement du système » ; c) « être complète » ; d) « être spécifique, en ce sens que différents utilisateurs dans des circonstances diverses ou avec des résultats distincts devraient recevoir des explications différentes »<sup>50</sup>. Il est en outre précisé que :

l'explicabilité de l'intelligence artificielle est une aspiration compréhensible d'un point de vue éthique et même juridique, mais qui se heurte à des difficultés

<sup>49</sup> Le contenu du tableau est une adaptation, un résumé et une traduction du texte original en anglais, disponible sur : <https://doi.org/10.6028/NIST.IR.8312>.

<sup>50</sup> Ignacio Gavilán, « Cuatro principios para una buena explicabilidad de los algoritmos » (2022). Disponible sur : <https://ignaciogavilan.com/cuatro-principios-para-una-buena-explicabilidad-de-los-algoritmos/>.

techniques profondes qu'il faut connaître ; et il est probable qu'une grande partie de la solution sera elle aussi technique, étant donné que les algorithmes peuvent être repensés ou que de nouveaux algorithmes peuvent être définis pour répondre aux aspirations éthiques et réglementaires<sup>51</sup>.

Pour sa part, l'UNESCO indique que :

L'explicabilité implique de rendre les résultats des systèmes d'[intelligence artificielle] intelligibles et de fournir des renseignements à leur sujet. L'explicabilité des systèmes d'[intelligence artificielle] renvoie également à l'intelligibilité des intrants, des extrants, du fonctionnement des différents modules algorithmiques et de leur contribution aux résultats des systèmes<sup>52</sup>.

58. Afin de préciser la portée du principe d'explicabilité, il convient de garder à l'esprit son objectif et, sur cette base, de définir ce qui est nécessaire pour le réaliser. Il a ainsi été souligné que :

si le principe d'explicabilité vise à permettre à tout être humain de savoir pourquoi une décision est prise sur la base du traitement de ses données personnelles à l'aide d'outils [d'intelligence artificielle], l'explication doit au moins être claire, simple, complète, véridique et facilement compréhensible par la personne qui la demande. Donner des informations sur les données utilisées comme intrants pour générer la décision ne suffit pas, il faut aussi en fournir sur la logique ou la méthodologie utilisée pour parvenir à la décision. Ce n'est pas là tâche facile, mais elle est réalisable avec la volonté d'expliquer simplement aux personnes pourquoi une décision a été prise sur la base du traitement de leurs données personnelles<sup>53</sup>.

59. Les points suivants fournissent quelques exemples de réglementations locales dans des pays qui ont implicitement ou explicitement intégré le principe d'explicabilité dans leur cadre juridique.

60. La réglementation colombienne interdit le traitement de données qui « induisent en erreur »<sup>54</sup> et, dans le cas particulier des décisions prises en ce qui concerne les demandes de crédit, impose aux personnes qui les rejettent d'informer par écrit la personne concernée, sur sa demande, des « raisons objectives du refus »<sup>55</sup> du crédit.

61. En Équateur, l'article 20 de la loi organique sur la protection des données personnelles dispose que la personne concernée, lorsqu'elle est face à une décision fondée exclusivement ou partiellement sur des évaluations résultant de procédés automatisés, y compris le profilage, et qui produisent des effets juridiques sur elle ou qui portent atteinte à ses droits et libertés fondamentaux, peut exiger une explication motivée de la décision, obtenir les critères d'évaluation du programme automatisé, présenter des observations, demander des informations sur les types de données utilisées et la source à partir de laquelle elles ont été obtenues, et contester la décision devant le responsable du traitement ou le sous-traitant (avec certaines exceptions).

62. En Uruguay, l'article 16 de la loi n° 18331 dispose que :

<sup>51</sup> Ibid.

<sup>52</sup> Voir [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_fre](https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre), p. 23.

<sup>53</sup> Nelson Remolina Angarita, « Del principio de explicabilidad en la inteligencia artificial (notas preliminares) », in *Protección de datos personales: doctrina y jurisprudencia*, Pablo Palazzi, ed., tome III (Centro de Tecnologías y Sociedad de la Universidad de San Andrés, Buenos Aires, 2023).

<sup>54</sup> Loi statutaire 1581 de 2012, qui établit des dispositions générales pour la protection des données personnelles, art. 4 d).

<sup>55</sup> Loi statutaire 2157 de 2021, qui modifie et complète la loi statutaire 1266 de 2008 et établit des dispositions générales d'*habeas data* relatif aux informations financières, de crédit, commerciales, de service et provenant de pays tiers, ainsi que d'autres dispositions, art. 5, par. 1.

les personnes ont le droit de ne pas être soumises à une décision produisant des effets juridiques qui les affectent de manière significative, prise sur la base d'un traitement automatisé de données destiné à évaluer certains aspects de leur personnalité, tels que leur performance professionnelle, leur solvabilité, leur fiabilité, leur comportement, entre autres. La personne affectée a le droit d'obtenir du responsable de la base de données des informations sur les critères d'évaluation comme sur le programme utilisé dans le traitement et qui a servi à prendre la décision notifiée.

## VII. Conclusions

63. Les conclusions suivantes peuvent être tirées de ce qui précède :

a) la transparence et l'explicabilité contribuent à renforcer la confiance dans l'intelligence artificielle et à respecter les droits humains ;

b) les développeurs d'intelligence artificielle doivent faire preuve de transparence sur la manière dont les données sont traitées (modalités de collecte, de stockage et d'utilisation), ainsi que sur la manière dont sont prises les décisions fondées sur l'intelligence artificielle, sur la fiabilité de ces décisions et sur la sécurité de l'information ;

c) les personnes affectées par des décisions prises sur la base de l'intelligence artificielle méritent de recevoir une explication claire, simple, complète, véridique et compréhensible de la motivation de ces décisions. En ce sens, le principe d'explicabilité est d'une importance capitale, non seulement parce qu'il répond au principe de transparence, mais aussi parce qu'il permet aux personnes affectées d'exercer leur droit à la défense et de bénéficier d'une procédure régulière ;

d) l'explicabilité et la transparence exigent la clarté, l'exhaustivité, la véracité, l'impartialité et la divulgation des décisions prises par l'intelligence artificielle, ainsi que de la logique, de la méthode ou du raisonnement permettant de prendre des décisions concernant des êtres humains à partir d'informations et, en particulier, de données à caractère personnel. L'explicabilité et la transparence s'opposent bien entendu à l'opacité, à la dissimulation, à la tromperie, au mensonge et à l'abus de la puissance de calcul, qui sont autant d'indices d'un traitement des données illicite et contraire à l'éthique et d'un manque de respect pour les êtres humains et leur dignité.

## VIII. Recommandations

64. À la lumière de tout ce qui a été exposé, la Rapporteuse spéciale exhorte les États à :

a) promouvoir la transparence dans le domaine de l'intelligence artificielle afin d'atténuer les risques que l'opacité peut engendrer pour la société et, en particulier, en ce qui concerne la protection des droits humains ;

b) inscrire le principe d'explicabilité dans leur réglementation, non seulement pour que les personnes comprennent comment les décisions qui les concernent ont été prises, mais aussi pour qu'elles puissent disposer des outils pour défendre leurs droits humains face à l'intelligence artificielle ;

**c) encourager les pratiques éthiques qui garantissent la transparence et l'explicabilité du traitement des données à caractère personnel dans le cadre de projets ou de processus d'intelligence artificielle ;**

**d) favoriser, soutenir et faciliter l'éducation et la culture numériques afin d'améliorer la compréhension par les citoyennes et les citoyens des concepts liés à l'intelligence artificielle, à la transparence et à l'explicabilité, de manière à ce que toute personne puisse exiger le respect de ses droits.**

---