Naciones Unidas A/78/310

Asamblea General

Distr. general
30 de agosto de 2023

Original: español

Septuagésimo octavo período de sesiones
Tema 73 b) del programa provisional*
Promoción y protección de los derechos humanos: cuestiones de derechos humanos, incluidos otros medios de mejorar el goce efectivo de los derechos humanos y las libertades fundamentales

Derecho a la privacidad

Nota del Secretario General

El Secretario General tiene el honor de transmitir a la Asamblea General el informe preparado por la Relatora Especial sobre el derecho a la privacidad, Ana Brian Nougrères, presentado de conformidad con la resolución 28/16 del Consejo.

* A/78/150.

Informe de la Relatora Especial sobre el derecho a la privacidad, Ana Brian Nougrères

Principios de transparencia y explicabilidad en el tratamiento de datos personales en la inteligencia artificial

Resumen

En el presente informe, la Relatora Especial sobre el derecho a la privacidad, Ana Brian Nougrères, recalca la importancia de los principios de transparencia y explicabilidad en el tratamiento de datos personales mediante la inteligencia artificial. La omnipresencia de la inteligencia artificial en todas las actividades y la toma de decisiones sobre las personas a partir del uso de esta obligan a analizar este tema y adoptar medidas para que el uso de la inteligencia artificial sea ético, responsable y respetuoso de los derechos humanos.

Lo anterior es relevante porque la transparencia y la explicabilidad no solo ayudan a generar confianza y fiabilidad en la inteligencia artificial, sino que contribuyen a proteger los derechos humanos. Mediante estos principios, de una parte, se informa de manera oportuna, completa, sencilla y clara a las personas sobre aspectos básicos respecto del uso de su información personal en procesos o proyectos de inteligencia artificial y sus consecuencias y, de otra parte, se exige que las personas afectadas por la inteligencia artificial conozcan los motivos concretos que dieron origen a dicha afectación. Con esto, la persona podrá ejercer sus derechos como, por ejemplo, al debido proceso y el derecho de defensa frente a las decisiones adoptadas mediante herramientas o tecnologías de inteligencia artificial.

I. Introducción

1. El Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial[1] de la Comisión Europea ha señalado que los principios de transparencia y explicabilidad son componentes relevantes para promover una inteligencia artificial fiable. Para ello, la inteligencia artificial debe ser lícita, ética y robusta, "tanto desde el punto de vista técnico como social, puesto que los sistemas de inteligencia artificial, incluso si las intenciones son buenas, pueden provocar daños accidentales"[2].

2. En la misma línea, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) señala que "[l]a transparencia y la explicabilidad están estrechamente relacionadas con las medidas adecuadas de responsabilidad y rendición de cuentas, así como con la fiabilidad de los sistemas de [inteligencia artificial]"[3]. Asimismo, establece que "[l]a transparencia y la explicabilidad de los sistemas de inteligencia artificial suelen ser condiciones previas fundamentales para garantizar el respeto, la protección y la promoción de los derechos humanos, las libertades fundamentales y los principios éticos"[4].

3. La inteligencia artificial está muy presente en la agenda global. A finales de diciembre de 2022, por ejemplo, la Organización de Cooperación y Desarrollo Económicos (OCDE) emitió una declaración sobre un futuro digital fiable, sostenible e inclusivo[5] mediante la cual se comprometió a trabajar para, entre otros aspectos, impulsar una transformación digital centrada en el ser humano y que promueva los derechos humanos, tanto en línea como sin conexión, así como una sólida protección de los datos personales, y de las leyes y normativas adecuadas a la era digital, y un uso fiable, seguro, responsable y sostenible de las tecnologías digitales emergentes y la inteligencia artificial[6]. Respecto de la inteligencia artificial, los Estados miembros de la OCDE solicitan a esta organización que respalde el desarrollo de marcos políticos y jurídicos con visión de futuro, coherentes y viables para regular la inteligencia artificial y gestionar sus riesgos de forma eficaz, y proporcione datos empíricos, previsiones, herramientas y una labor de seguimiento de incidentes para la planificación y ejecución de políticas públicas eficaces de cara a implementar una inteligencia artificial fiable de seguimiento[7].

4. El Parlamento Europeo, el Consejo y la Comisión, por su parte, aprobaron el 23 de enero de 2023 la Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital, por la que se comprometieron a:

a) promover sistemas de inteligencia artificial centrados en el ser humano, fiables y éticos a lo largo de todo su desarrollo, despliegue y uso, en consonancia con los valores de la [Unión Europea];

---

[1] Se trata de un grupo de expertos independientes constituido por la Comisión Europea en junio de 2018.

[2] Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, *Directrices éticas para una inteligencia artificial fiable* (2019), pág. 2. Disponible en: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[3] UNESCO, *Recomendación sobre la ética de la inteligencia artificial* (2021), pág. 23. Disponible en https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa.

[4] *Ibid.*, pág. 22.https://www.unesdoc.unesco.org/ark:/48223/pf0000381137_spa

[5] OCDE, *Declaration on a Trusted, Sustainable and Inclusive Digital Future* (2022). La declaración fue fruto de la reunión que se realizó en la isla de Gran Canaria (España) el 14 y el 15 diciembre de 2022. Disponible en: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0488.

[6] *Ibid*.

[7] *Ibid*.

b) velar por un nivel adecuado de transparencia en el uso de los algoritmos y la inteligencia artificial y por que las personas estén informadas y capacitadas para utilizarlos cuando interactúen con ellos;

c) velar por que los sistemas algorítmicos se basen en conjuntos de datos adecuados para evitar la discriminación y permitir la supervisión humana de todos los resultados que afecten a la seguridad y los derechos fundamentales de las personas;

d) garantizar que las tecnologías como la inteligencia artificial no se utilicen para anticiparse a las decisiones de las personas en ámbitos como, por ejemplo, la salud, la educación, el empleo y la vida privada;

e) proporcionar salvaguardias y adoptar las medidas adecuadas, en particular promoviendo normas fiables, para que la inteligencia artificial y los sistemas digitales sean seguros y se utilicen en todo momento con pleno respeto de los derechos fundamentales de las personas;

f) adoptar medidas para garantizar que la investigación en inteligencia artificial respete las normas éticas más estrictas y la legislación pertinente de la [Unión Europea][8].

5. En vista de lo anterior, a continuación se formulan algunas consideraciones sobre la inteligencia artificial y se hace referencia brevemente a los siguientes aspectos con miras a precisar en qué consisten los principios de transparencia y explicabilidad en el contexto del tratamiento de datos personales en procesos o proyectos de inteligencia artificial.

II. Inteligencia artificial y tratamiento de datos personales

6. La inteligencia artificial ha alcanzado un estado de omnipresencia en casi todos los aspectos de nuestra sociedad, desde los dispositivos móviles que usan permanentemente los ciudadanos hasta los sistemas de gestión empresarial más complejos. Esta creciente presencia de la inteligencia artificial ha abierto un amplio abanico de oportunidades en diversas actividades y sectores. Sin embargo, junto con estas oportunidades también surgen retos y peligros que se deben abordar de manera responsable con miras a que, entre otros aspectos, se aproveche todo el potencial de la inteligencia artificial de manera segura, ética y respetuosa de los derechos humanos.

7. No existe consenso sobre la definición de la inteligencia artificial, pero se han señalado algunas formas de clasificar por qué está constituida. En un texto de referencia sobre el tema, se propuso la siguiente taxonomía[9]:

- Sistemas que piensan como humanos (por ejemplo, arquitecturas cognitivas y redes neuronales).
- Sistemas que actúan como seres humanos (por ejemplo, razonamiento automatizado y aprendizaje).
- Sistemas que piensan racionalmente (por ejemplo, inferencias).

[8] Parlamento Europeo, Consejo y Comisión, “Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital”, *Diario Oficial de la Unión Europea*, 2023/C 23/01, 23 de enero de 2023. Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ%3AJOC_2023_023_R_0001.

[9] Stuart Russell y Peter Norvig, *Artificial Intelligence: A Modern Approach* (Essex, England: Pearson, 2009).

- Sistemas que actúan racionalmente (por ejemplo, agentes de *software* inteligentes y robots incorporados que logran objetivos mediante la percepción, la planificación, el razonamiento, el aprendizaje, la comunicación, la toma de decisiones y la actuación).

8. Todos esos sistemas procesan información para generar resultados, y esa información contiene, entre otros, datos personales. En este sentido, la Comisión Europea ha precisado lo siguiente:

> A los efectos del presente Libro Blanco, así como de todo posible debate sobre iniciativas políticas en el futuro, parece importante clarificar cuáles son los principales elementos que integran la inteligencia artificial, a saber: los «datos» y los «algoritmos». La inteligencia artificial puede incorporarse en los equipos informáticos. En lo que se refiere a las técnicas de aprendizaje automático, que constituyen un subapartado de la inteligencia artificial, los algoritmos son entrenados para inferir determinados modelos a partir de un conjunto de datos, a fin de determinar las acciones que se requieren para alcanzar un objetivo determinado[10].

9. En otras palabras, para desarrollar inteligencia artificial se recolectan, almacenan, analizan y procesan enormes cantidades de información, usada para generar diversos resultados, acciones o comportamientos por parte de las máquinas o de los usuarios de estas. No obstante, como lo exige la citada recomendación de la UNESCO, "[l]a privacidad, que constituye un derecho esencial para la protección de la dignidad, la autonomía y la capacidad de actuar de los seres humanos, debe ser respetada, protegida y promovida a lo largo del ciclo de vida de los sistemas de [inteligencia artificial]"[11].

10. El debido tratamiento de datos personales es imprescindible para evitar que, con el desarrollo de la inteligencia artificial, los derechos humanos se vean lesionados o amenazados, según el caso. Existen varias iniciativas y organizaciones que han trabajado para exigir un desarrollo de inteligencia artificial respetuoso de los derechos humanos y, a continuación, se presentan algunos ejemplos.

11. En primer lugar, la Asamblea Global de Privacidad adoptó, en el mes de octubre de 2020, la Resolución sobre la Rendición de Cuentas Responsable en el Desarrollo y la Utilización de Inteligencia Artificial[12]. Esa resolución, entre otras, insta a las organizaciones que desarrollan o utilizan sistemas de inteligencia artificial a considerar la implementación de las siguientes medidas:

- Evaluar el potencial impacto en los derechos humanos (incluidos los derechos de protección de datos y privacidad) previo al desarrollo y/o uso de la inteligencia artificial;
- Previo a su utilización, comprobar la solidez, credibilidad, exactitud y seguridad de los datos en la inteligencia artificial. Esto incluye la identificación y el tratamiento de los sesgos de los sistemas, así como de los datos que se utilizan, los cuales podrían conducir a resultados injustos;
- Implementar medidas de responsabilidad demostrada que sean apropiadas con respecto a los riesgos de interferencia con los derechos humanos.

[10] Comisión Europea, *Libro blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*, COM(2020) 65 final (2020). Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/?qid=1603192201335&uri=CELEX%3A52020DC0065.

[11] Véase https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa, pág. 22.

[12] Véase https://globalprivacyassembly.org/wp-content/uploads/2020/11/GPA-Resolution-on-Accountability-in-the-Development-and-Use-of-AI-ES.pdf, pág. 3.

12. En la misma línea, la UNESCO, en su recomendación, establece que:

> Los sistemas algorítmicos requieren evaluaciones adecuadas del impacto en la privacidad, que incluyan también consideraciones sociales y éticas de su utilización y un empleo innovador del enfoque de privacidad desde la etapa de concepción. Los actores de la [inteligencia artificial] deben asumir la responsabilidad de la concepción y la aplicación de los sistemas de [inteligencia artificial] de manera que se garantice la protección de la información personal durante todo el ciclo de vida del sistema de [inteligencia artificial][13].

13. Previo a ello, en junio de 2019, la Red Iberoamericana de Protección de Datos publicó las "Recomendaciones generales para el tratamiento de datos personales en la inteligencia artificial"[14]. En dicho documento se realizan algunas sugerencias a quienes desarrollan productos de inteligencia artificial con el fin de orientarlos para que desde el diseño del producto se tengan en cuenta las exigencias de las regulaciones sobre tratamiento de datos personales. Las recomendaciones son las siguientes:

- Cumplir las normas locales sobre tratamiento de datos personales;
- Efectuar estudios de impacto de privacidad;
- Incorporar la privacidad, la ética y la seguridad desde el diseño y por defecto;
- Materializar el principio de responsabilidad demostrada (accountability);
- Diseñar esquemas apropiados de gobernanza sobre tratamiento de datos personales en las organizaciones que desarrollan productos de inteligencia artificial;
- Adoptar medidas para garantizar los principios sobre tratamiento de datos personales en los proyectos de inteligencia artificial;
- Respetar los derechos de los titulares de los datos e implementar mecanismos efectivos para el ejercicio de los mismos;
- Asegurar la calidad de los datos personales;
- Utilizar herramientas de anonimización;
- Incrementar la confianza y la transparencia con los titulares de los datos personales.

14. Para conocer los detalles de la implementación de algunas de estas recomendaciones, la Red Iberoamericana de Protección de Datos ha elaborado unas directrices complementarias y más detalladas contenidas en el documento "Orientaciones específicas para el cumplimiento de los principios y derechos que rigen la protección de los datos personales en los proyectos de inteligencia artificial"[15]. En el presente informe se desarrolla el principio de transparencia al que se hará referencia posteriormente.

---

[13] Véase https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa, pág. 22.

[14] Red Iberoamericana de Protección de Datos, "Recomendaciones generales para el tratamiento de datos personales en la inteligencia artificial" (2019). Texto aprobado por las entidades integrantes de la Red en la sesión del 21 de junio de 2019, en Naucalpan de Juárez (México). Disponible en: https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf.

[15] Red Iberoamericana de Protección de Datos, "Orientaciones específicas para el cumplimiento de los principios y derechos que rigen la protección de los datos personales en los proyectos de inteligencia artificial" (2019), Disponible en: https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf.

III. Riesgos inherentes a la inteligencia artificial

15. La sociedad, y su transformación digital, está siendo influenciada por la inteligencia artificial, la cual se encuentra presente en diversos aspectos de nuestra vida diaria, la economía, la ciencia, la educación, la salud y muchas otras áreas o actividades.

16. Si bien son innegables las oportunidades y los beneficios de la inteligencia artificial en la sociedad en general, no debe perderse de vista que puede entrañar desafíos, peligros o amenazas que son intrínsecos a la inteligencia artificial. A título enunciativo, por ejemplo, dichos riesgos pueden incluir aspectos como la falta de ética en el desarrollo o uso de la inteligencia artificial, o la toma de decisiones sesgadas, no transparentes o incorrectas sobre seres humanos.

17. Los niveles de riesgos dependen de cada situación particular.

> La Comisión Europea considera que, en general, una aplicación de [inteligencia artificial] determinada debe considerarse de riesgo elevado en función de lo que esté en juego, y considerando si tanto el sector como el uso previsto suponen riesgos significativos, en especial desde la perspectiva de la protección de la seguridad, los derechos de los consumidores y los derechos fundamentales. De manera más específica, una aplicación de [inteligencia artificial] debe considerarse de riesgo elevado cuando presente la suma de los dos criterios siguientes:
>
> a) En primer lugar, que la aplicación de [inteligencia artificial] se emplee en un sector en el que, por las características o actividades que se llevan a cabo normalmente, es previsible que existan riesgos significativos. […]. Por ejemplo, la sanidad, el transporte, la energía y determinados ámbitos del sector público […];
>
> b) En segundo lugar, que la aplicación de [inteligencia artificial] en el sector en cuestión se use, además, de manera que puedan surgir riesgos significativos. […] La evaluación del nivel de riesgo de un uso determinado puede basarse en las repercusiones para las partes afectadas. Por ejemplo, el uso de aplicaciones de [inteligencia artificial] con efectos jurídicos o similares en los derechos de un particular o de una empresa; aplicaciones que presenten el riesgo de causar lesiones, la muerte, o daños materiales o inmateriales significativos; aplicaciones que produzcan efectos que las personas físicas o jurídicas no puedan evitar razonablemente[16].

18. La inteligencia artificial involucra diferentes tipos de riesgos. Entre las contingencias a tener en cuenta deben considerarse, entre otras, las inherentes a la operación de los algoritmos —sesgos humanos, fallas técnicas, vulnerabilidad de seguridad y fallas en la implementación—, y a su diseño. Sobre este punto se han identificado aspectos que inciden en la gestión de riesgos de los algoritmos que se ilustran en la siguiente gráfica[17]:

[16] Véase https://eur-lex.europa.eu/legal-content/ES/TXT/?qid=1603192201335&uri=CELEX%3A52020DC0065.

[17] Véase https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf, pág. 18.

19. Explica la doctrina que:

> los datos de entrada están afectados principalmente por dos variables: los sesgos (incorporación de datos parciales, insuficientes, no actualizados o manipulados) y la pertinencia (relevancia, inconsistencia o completitud de los datos); por su parte, el desarrollo del algoritmo se puede ver afectado por los patrones (sesgos de la lógica de programación, inclusión de funciones no previstas y fallas inherentes de las funciones utilizadas para su codificación) y los errores (condiciones de la operación que reflejan un funcionamiento diferente al previsto y que atentan contra las premisas del diseño planteado). Finalmente, los riesgos en las decisiones de salida están relacionados con la pertinencia y precisión del resultado de la ejecución del algoritmo y como respuesta al análisis de los datos de entrada[18].

IV. Principio de transparencia en el tratamiento de datos personales

20. La transparencia es un concepto usado en diversas disciplinas como las ciencias de la computación, el acceso a la información, el derecho y el tratamiento de datos personales. Para la UNESCO, “la transparencia tiene como objetivo proporcionar información adecuada a los respectivos destinatarios para permitir su comprensión y fomentar la confianza”[19].

21. No existe consenso sobre el alcance de este concepto en cada caso y, dentro de cada uno, presenta diferentes connotaciones. Por ejemplo, es diferente referirse al principio de transparencia en el tratamiento de datos personales en general que si se hace referencia a dicho principio en el contexto de la inteligencia artificial. En el presente informe se hace referencia a la transparencia en el tratamiento de datos personales en general y de forma específica en la inteligencia artificial.

[18] Alejandro Useche y Jeimy Cano, *Robo-Advisors: Asesoría automatizada en el mercado de valores*, Universidad del Rosario y Autorregulador del Mercado de Valores de Colombia (2019), págs. 9 y 10. Disponible en: https://www.researchgate.net/publication/331358231_Robo-Advisors_Asesoria_automatizada_en_el_mercado_de_valores.

[19] UNESCO, *Recomendación sobre la ética de la inteligencia artificial* (2021), pág. 22.

22. En múltiples documentos de organizaciones proveniente de diferentes partes del mundo se hace referencia a este principio[20]. La Relatora Especial ya se pronunció al respecto, en cuanto el principio de transparencia supone que todo responsable del tratamiento habrá de informar al titular del dato acerca de las condiciones de tratamiento a las que será sometida su información personal en todo el proceso de tratamiento del dato, esto es, desde el momento de la recogida, de forma tal que el titular de los datos posea la información necesaria para hacer un control efectivo[21].

23. En el informe mencionado se analizan los siguientes documentos internacionales sobre privacidad y tratamiento de datos personales con relación al principio de transparencia: a) Reglamento General de Protección de Datos de la Unión Europea; b) Convenio para la Protección de las Personas con respecto al Tratamiento Automatizado de Datos de Carácter Personal; c) Estándares de Protección de Datos Personales para los Estados Iberoamericanos, aprobados por la Red Iberoamericana de Protección de Datos; d) Recomendaciones del Consejo sobre las Directrices para la Protección de la Privacidad y los Flujos Transfronterizos de Datos Personales de la Organización para la Cooperación Económica y el Desarrollo; e) Marco de privacidad del Foro de Cooperación Económica de Asia y el Pacífico, y f) los Principios Actualizados sobre la Privacidad y la Protección de Datos Personales, con anotaciones, de la Organización de los Estados Americanos.

24. Del análisis de dichos documentos se concluye, en términos generales, que se debe informar lo siguiente:

- la identidad y domicilio del responsable o de su representante, así como los fines o propósitos del tratamiento, datos básicos iniciales para una actuación transparente;
- los derechos que le corresponden al titular de los datos y la forma de ejercerlos, los destinatarios de los datos o la categoría de estos;
- el fundamento o base jurídica que habilita el tratamiento, así como la existencia y las características principales del tratamiento;

[20] Organización de Cooperación y Desarrollo Económicos (OCDE), *Directrices sobre protección de la privacidad y flujos transfronterizos de datos personales*, 23 de septiembre de 1980 (no disponibles en español) y su actualización de julio de 2013; Consejo de Europa, Convenio para la Protección de las Personas con respecto al Tratamiento Automatizado de Datos de Carácter Personal, núm. 108, de 28 de enero de 1981; Naciones Unidas, Principios rectores sobre la reglamentación de los ficheros computadorizados de datos personales, de 14 de diciembre de 1990; Consejo de Europa, Protocolo Adicional al Convenio para la Protección de las Personas con respecto al Tratamiento Automatizado de Datos de Carácter Personal, sobre las Autoridades de Control y los Flujos Transfronterizos de Datos, de 8 de noviembre de 2001; Foro de Cooperación Económica de Asia y el Pacífico, *Marco de Privacidad del Foro de Cooperación Económica Asia-Pacífico*, de 2004; Agencia Española de Protección de Datos, *Propuesta Conjunta para la Redacción de Estándares Internacionales para la Protección de la Privacidad en relación con el Tratamiento de Datos de Carácter Personal*, Madrid, 5 de noviembre de 2009; Parlamento Europeo y Consejo de la Unión Europea, Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo relativo a la Protección de las Personas Físicas en lo que respecta al Tratamiento de Datos Personales y a la Libre Circulación de Estos Datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos), de 27 de abril de 2016; Red Iberoamericana de Protección de Datos, Directrices para la Armonización de la Protección de Datos en la Comunidad Iberoamericana, de 2017; Consejo de Europa, Protocolo de enmienda del Convenio para la Protección de las Personas con respecto al Tratamiento de Datos Personales, de octubre de 2018, y Organización de los Estados Americanos, Comité Jurídico Interamericano, *Principios Actualizados sobre la Privacidad y la Protección de Datos Personales*, de 2021.

[21] A/77/196, párr. 45.

- la categoría de los datos tratados y, cuando no se obtuvieron directamente del titular, el origen de estos datos.

25. Se entiende trascendente destacar que, para cumplir con el principio de transparencia, la información debe serle proporcionada al titular del dato en un lenguaje claro, sencillo, inteligible y de fácil acceso y comprensión. Si se trata de niños, niñas o adolescentes, igualmente debe cumplirse este mandato, con las adaptaciones que correspondan.

26. No todos los instrumentos normativos mencionados exigen que se informe sobre los mismos aspectos, pues algunos contemplan un listado más amplio de lo que se debe informar. En el caso particular del Reglamento General de Protección de Datos de la Unión Europea, se requiere, entre otras, la siguiente información[22]: los datos de contacto del delegado de protección de datos; el plazo de durante el cual se conservarán los datos personales o el criterio para determinarlo; si el responsable prevé realizar comunicaciones o transferencias y la normativa que lo autoriza; el derecho a presentar una reclamación ante la autoridad de control; si la comunicación es un requisito legal, contractual, o es necesaria para suscribir un contrato, y si el interesado está obligado a facilitar sus datos personales y las consecuencias que conllevaría que no los facilitara; la existencia de decisiones automatizadas, incluida la elaboración de perfiles, en cuyos casos se debe proporcionar información significativa sobre la lógica aplicada, la importancia y las consecuencias previstas de dicho tratamiento, y la información sobre el fin cuando el responsable proyecte un tratamiento ulterior para un fin que no sea aquel para el que se obtuvieron los datos.

V. Principio de transparencia en el tratamiento de datos personales en el ámbito de la inteligencia artificial

27. Es esencial garantizar la transparencia en la inteligencia artificial, pues su desconocimiento u omisión puede generar efectos negativos. En este sentido, se ha planteado que:

> La falta de transparencia (opacidad de la [inteligencia artificial]) hace difícil detectar y demostrar los posibles incumplimientos de la legislación, especialmente las disposiciones legales que protegen los derechos fundamentales, imputan responsabilidades y permiten reclamar una indemnización[23].

28. La eventual opacidad en la inteligencia artificial puede mitigarse exigiendo el cumplimiento de requisitos mínimos de transparencia. Por eso, se ha planteado que es necesario:

> Facilitar información clara con respecto de las capacidades y limitaciones del sistema de [inteligencia artificial] , en especial sobre el objetivo al que se destinan los sistemas, las condiciones en las que se espera que funcione según lo previsto y el nivel de exactitud esperado en la consecución del objetivo mencionado […]. Independientemente, debe informarse claramente a los ciudadanos de cuándo están interactuando con un sistema de [inteligencia artificial] y no con un ser humano […]. Es importante también que la información facilitada sea objetiva, concisa y fácilmente comprensible[24].

[22] Véase https://eur-lex.europa.eu/legal-content/ES/TXT/?qid=1532348683434&uri=CELEX%3A02016R0679-20160504.

[23] Comisión Europea *Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza* (2020), pág. 17.

[24] *Ibid*., págs. 23 y 24.

29. En la recomendación de la UNESCO se indica que:

> En el caso específico de los sistemas de [inteligencia artificial], la transparencia puede permitir a las personas comprender cómo se implementa cada etapa de un sistema de [inteligencia artificial], en función del contexto y la sensibilidad del sistema en cuestión. También puede proporcionar información sobre los factores que influyen en una predicción o decisión específicas, y sobre la existencia o no de garantías adecuadas (como medidas de seguridad o de equidad)[25].

30. El Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial ha puesto de presente que, para lograr una inteligencia artificial fiable, se deben garantizar algunos requisitos entre los cuales se encuentra la transparencia, que implica:

> Comunicar información a las partes interesadas, de un modo claro y proactivo, sobre las capacidades y limitaciones de los sistemas de [inteligencia artificial], posibilitando el establecimiento de expectativas realistas, así como sobre el modo en que se cumplen los requisitos[;] y [s]er transparentes acerca del hecho de que se está trabajando con un sistema de [inteligencia artificial]"[26].

En el mismo sentido, la mencionada recomendación de la UNESCO explicita que "los actores de la [inteligencia artificial] deberían informar a los usuarios cuando un producto o servicio se proporcione directamente o con la ayuda de sistemas de [inteligencia artificial] de manera adecuada y oportuna"[27].

31. Así pues, según recomendación de la UNESCO:

> [L]a explicabilidad está estrechamente relacionada con la transparencia, ya que los resultados y los subprocesos que conducen a ellos deberían aspirar a ser comprensibles y trazables, apropiados al contexto. Los actores de la [inteligencia artificial] deberían comprometerse a velar por que los algoritmos desarrollados sean explicables. En el caso de las aplicaciones de [inteligencia artificial] cuyo impacto en el usuario final no es temporal, fácilmente reversible o de bajo riesgo, debería garantizarse que se proporcione una explicación satisfactoria con toda decisión que haya dado lugar a la acción tomada, a fin de que el resultado se considere transparente[28].

32. El Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial explica que la transparencia "guarda una relación estrecha con el principio de explicabilidad e incluye la transparencia de los elementos pertinentes para un sistema de [inteligencia artificial]: los datos, el sistema y los modelos de negocio"[29]. Adicionalmente, señala la relevancia de la trazabilidad, la explicabilidad y comunicación en los siguientes términos:

- Trazabilidad: Los conjuntos de datos y los procesos que dan lugar a la decisión del sistema de [inteligencia artificial], incluidos los relativos a la recopilación y etiquetado de los datos, así como a los algoritmos utilizados, deberían documentarse con arreglo a la norma más rigurosa posible con el fin de posibilitar la trazabilidad y aumentar la transparencia. Esto también es aplicable a las decisiones que adopte el sistema de [inteligencia artificial]. Esto permitirá identificar los motivos de una decisión errónea por parte del sistema, lo que a

[25] Véase https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa, pág. 22.
[26] Véase https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, págs. 2 y 3.
[27] Véase https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa, pág. 22.
[28] *Ibid.*, pág. 23.
[29] Véase https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, pág. 22.

su vez podría ayudar a prevenir futuros errores. La trazabilidad, por tanto, facilita la auditabilidad y la explicabilidad.

- Explicabilidad: La explicabilidad concierne a la capacidad de explicar tanto los procesos técnicos de un sistema de [inteligencia artificial] como las decisiones humanas asociadas (por ejemplo, las áreas de aplicación de un sistema de [inteligencia artificial]). La explicabilidad técnica requiere que las decisiones que adopte un sistema de [inteligencia artificial] sean comprensibles para los seres humanos y estos tengan la posibilidad de rastrearlas. Además, puede que sea necesario buscar un equilibrio entre la mejora de la explicabilidad de un sistema (que puede reducir su precisión) o una mayor precisión de este (a costa de la explicabilidad). Cuando un sistema de [inteligencia artificial] tenga un impacto significativo en la vida de las personas, debería ser posible reclamar una explicación adecuada del proceso de toma de decisiones del sistema de [inteligencia artificial]. Dicha explicación debería ser oportuna y adaptarse al nivel de especialización de la parte interesada (que puede ser una persona no experta en la materia, un regulador o un investigador). Además, debería ser posible disponer de explicaciones sobre la medida en que el sistema de [inteligencia artificial] condiciona e influye en el proceso de toma de decisiones de la organización, sobre las decisiones de diseño del sistema y sobre la lógica subyacente a su despliegue (garantizando así la transparencia del modelo de negocio).

- Comunicación: Los sistemas de [inteligencia artificial] no deberían presentarse a sí mismos como humanos ante los usuarios; las personas tienen derecho a saber que están interactuando con un sistema de [inteligencia artificial]. Por lo tanto, los sistemas de [inteligencia artificial] deben ser identificables como tales. Además, cuando sea necesario, se debería ofrecer al usuario la posibilidad de decidir si prefiere interactuar con un sistema de [inteligencia artificial] o con otra persona, con el fin de garantizar el cumplimiento de los derechos fundamentales. Más allá de lo expuesto, se debería informar sobre las capacidades y limitaciones del sistema de [inteligencia artificial] a los profesionales o usuarios finales; dicha información debería proporcionarse de un modo adecuado según el caso de uso de que se trate y debería incluir información acerca del nivel de precisión del sistema de [inteligencia artificial], así como de sus limitaciones[30].

33. Al respecto, el Comité Europeo de Protección de Datos y el Supervisor Europeo de Protección de Datos en un dictamen conjunto establecen que:

> Siempre deberá informarse a las personas interesadas, cuando sus datos se utilicen para la formación o predicción en materia de [inteligencia artificial], de la base jurídica de dicho tratamiento, una explicación general de la lógica (procedimiento) y el alcance del sistema de [inteligencia artificial]. A este respecto, en esos casos deberá garantizarse siempre el derecho de las personas físicas a restringir el [...] tratamiento (artículo 18 del [Reglamento (UE) 2016/679, General de Protección de Datos] y artículo 20 del Reglamento relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión, y a la libre circulación de esos datos), así como a la supresión o el borrado de datos (artículo 16 del [Reglamento (UE) 2016/679, General de Protección de Datos] y artículo 19 del Reglamento relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión, y a la libre circulación de esos datos). Además, la

[30] *Ibid*.

> persona responsable del tratamiento deberá tener la obligación explícita de informar al interesado de los plazos aplicables para formular objeciones, limitaciones, supresión de datos, etc. El sistema de [inteligencia artificial] debe ser capaz de cumplir todos los requisitos de protección de datos mediante medidas técnicas y organizativas adecuadas. Un derecho a la explicación ofrecerá una mayor transparencia[31].

34. En el informe mencionado[32], se destaca que en el caso de que el interesado esté sometido a decisiones automatizadas, o a la elaboración de perfiles, el titular del dato puede entender cómo se produce el tratamiento al que será sometida la información que le concierne, por ejemplo, si se trata de un caso de inteligencia artificial se debe dar información significativa sobre la lógica aplicada y la importancia y las consecuencias previstas.

35. Sobre este punto, la Agencia Española de Protección de Datos recuerda que "la palabra 'significativa' […] se ha de interpretar como información que, proporcionada al interesado, le hace consciente del tipo de tratamiento que se está llevando a cabo con sus datos y le proporciona certeza y confianza sobre sus resultados"[33].

36. Asimismo, la Agencia destaca que:

> Cumplir con esta obligación ofreciendo una referencia técnica a la implementación del algoritmo puede ser opaco, confuso, e incluso conducir a la fatiga informativa. Debe facilitarse información que permita entender el comportamiento del tratamiento. Aunque dependerá del tipo de componente [de inteligencia artificial] utilizado, un ejemplo de información que podría tener relevancia de cara al interesado, sería:
>
> - El detalle de los datos empleados para la toma de decisión, más allá de la categoría, y en particular información sobre los plazos de uso de los datos (su antigüedad).
> - La importancia relativa que cada uno de ellos tiene en la toma de decisión.
> - La calidad de los datos de entrenamiento y el tipo de patrones utilizados.
> - Los perfilados realizados y sus implicaciones.
> - Valores de precisión o error según la métrica adecuada para medir la bondad de la inferencia.
> - La existencia o no de supervisión humana cualificada.
> - La referencia a auditorías, especialmente sobre las posibles desviaciones de los resultados de las inferencias, así como la certificación o certificaciones realizadas sobre el sistema de [inteligencia artificial]. En el caso de sistemas adaptativos o evolutivos, la última auditoría realizada.

[31] Comité Europeo de Protección de Datos y Supervisor Europeo de Protección de Datos, Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), 18 de junio de 2021, págs. 19 y 20. Disponible en https://edpb.europa.eu/system/files/2021-10/edpb-edps_joint_opinion_ai_regulation_es.pdf.

[32] A/77/196, párr. 55.

[33] Agencia Española de Protección de Datos, *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*, febrero de 2020, pág. 24. Disponible en: https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf.

- En el caso de que el sistema [de inteligencia artificial] contenga información de terceros identificables, la prohibición de tratar esa información sin legitimación y de las consecuencias de realizarlo[34].

37. El Supervisor Europeo de Protección de Datos sugiere en un dictamen que, si la Comisión presentara un nuevo marco normativo específico para la inteligencia artificial, deberían aplicarse cierto número de garantías razonables a todas las aplicaciones de inteligencia artificial, independientemente del nivel de riesgo, tales como contar con medidas técnicas y organizativas (incluida la documentación) siendo totalmente transparente sobre los objetivos, el uso y el diseño de los sistemas algorítmicos implementados; garantizar la solidez del sistema de inteligencia artificial, o implementar y ser transparente sobre los mecanismos disponibles de rendición de cuentas, reparación y supervisión independiente[35].

38. El Comité Europeo de Protección de Datos y el Supervisor Europeo de Protección de Datos, por su parte, enfatizan la necesidad de promover:

> [f]ormas nuevas, más proactivas y oportunas de informar a los usuarios de los sistemas de [inteligencia artificial] sobre la situación (de toma de decisiones) en que se encuentra el sistema en cualquier momento, proporcionando una alerta temprana de posibles resultados perjudiciales, de modo que las personas cuyos derechos y libertades puedan verse perjudicados por decisiones autónomas de las máquinas puedan reaccionar o corregir la decisión[36].

39. La Red Iberoamericana de Protección de Datos considera que, para materializar el principio de transparencia, se debe[37]:

- "Comunicar al titular las características principales del tratamiento al que será sometida su información personal";
- "Informar expresamente a los titulares que en el tratamiento de sus datos personales se utilizarán procesos de automatización";
- "Incluir en el medio que se seleccione por los responsables para dar cumplimiento al principio de transparencia todas las finalidades para las cuales serán tratados los datos de los titulares";
- "Informar el origen de los datos personales cuando estos se obtengan a través de una transferencia y, en el supuesto específico de que se pretendan utilizar para inteligencia artificial, validar que esta finalidad haya sido informada por el primer responsable que los obtuvo para poder hacer uso de los datos para dicha finalidad";
- Desarrollar formas innovadoras de dar a conocer a los titulares las características principales del tratamiento y el nivel de riesgo relacionado con el aumento o disminución de la expectativa de privacidad;
- "Salvaguardar el derecho a la autodeterminación informativa, al asegurarse que los titulares siempre estén informados de forma adecuada y oportuna de que

[34] *Ibid*.

[35] Supervisor Europeo de Protección de Datos, *Opinion 4/2020, EDPS Opinion on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence and trust*. 29 de junio de 2020, pág. 14. Disponible en: https://edps.europa.eu/sites/edp/files/publication/20-06-19_opinion_ai_white_paper_en.pdf.

[36] Comité Europeo de Protección de Datos y Supervisor Europeo de Protección de Datos, Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), 18 de junio de 2021, pág. 22.

[37] Véase https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf, págs. 17 a 19.

estarán interactuando directamente con un sistema de inteligencia artificial o cuando su información será tratada por la inteligencia artificial";

- "Proporcionar información significativa sobre la finalidad y los efectos de los sistemas de inteligencia artificial para verificar la alineación continua con la expectativa de privacidad de los titulares permitiendo que en todo momento puedan ejercer control sobre el tratamiento de sus datos personales";
- "Identificar y definir los términos comúnmente utilizados y crear una base de datos para que puedan ser reutilizados en diferentes contextos, con iconos estándar para dar a conocer información a los titulares";
- "Informar continuamente a los titulares de manera que puedan conocer la forma en la que las decisiones automatizadas pueden afectarlos y, en su caso, solicitar la intervención humana, con el objetivo de que puedan tomar una decisión informada respecto a si consiente o no el tratamiento".

40. Precisa la Red Iberoamericana de Protección de Datos que:

> La información que se proporcione respecto a la lógica del modelo de [inteligencia artificial] deberá incluir por lo menos aspectos básicos sobre su funcionamiento, así como la ponderación y correlación de los datos, redactados en un lenguaje claro, sencillo y de fácil comprensión, por lo que no será necesario proporcionar una explicación completa de los algoritmos utilizados o incluso incluirlos [38].

41. Plantea la Red Iberoamericana de Protección de Datos que los responsables del tratamiento de datos en inteligencia artificial deben ser innovadores para comunicar la información de manera concisa y simple. Señala que "existen varios enfoques innovadores para proporcionar avisos de privacidad, incluido el uso de videos, caricaturas e íconos estandarizados. El uso de una combinación de enfoques puede ayudar a que la información compleja sobre [inteligencia artificial] sea más fácil de comprender para los titulares de los datos personales" [39].

42. En las siguientes líneas, se presentan, a título enunciativo y no exhaustivo, ejemplos de algunos países que han abordado en su regulación local, de manera explícita o implícita, el principio de transparencia en el tratamiento de datos personales en la inteligencia artificial.

43. En el Ecuador se aprobó, en 2021, la Ley Orgánica de Protección de Datos, que establece en su artículo 12, numerales 14 y 17, el derecho a ser informado sobre la existencia del derecho a no ser objeto de una decisión basada únicamente en valoraciones automatizadas y la forma en que puede hacerse efectivo dicho derecho y la existencia de valoraciones y decisiones automatizadas, incluida la elaboración de perfiles.

44. En la misma Ley, se señala que en el caso de que los datos se obtengan directamente del titular, la información deberá serle comunicada de forma previa, es decir, en el momento mismo de la recogida del dato personal. Además, el artículo 12 añade:

> Cuando los datos personales no se obtuvieren de forma directa del titular o fueren obtenidos de una fuente accesible al público, el titular deberá ser informado dentro de los siguientes treinta (30) días o al momento de la primera comunicación con el titular, cualquiera de las dos circunstancias que ocurra

[38] Véase https://www.redipd.org/es/documentos/guia, págs. 17 a 19.
[39] *Ibid*.

primero. Se le deberá proporcionar información expresa, inequívoca, transparente, inteligible, concisa, precisa y sin barreras técnicas.

45. En el Perú, el Reglamento de la Ley núm. 29733, de Protección de Datos Personales, en su artículo 72, se refiere al derecho al tratamiento objetivo de datos personales, aclarando que:

> Para garantizar el ejercicio del derecho al tratamiento objetivo de conformidad con lo establecido en el artículo 23 de la Ley[40], cuando se traten datos personales como parte de un proceso de toma de decisiones sin participación del titular de los datos personales, el titular del banco de datos personales o responsable del tratamiento deberá informárselo a la brevedad posible, sin perjuicio de lo regulado para el ejercicio de los demás derechos en la Ley y el […] reglamento.

46. En Santo Tomé y Príncipe, por su parte, la Ley núm. 3/2016, de 2 de mayo de 2016, de Protección de Datos Personales de las Personas Físicas, tiene como particularidad que en su artículo 21 establece que el responsable, o su representante, deberá notificar, por escrito y dentro de los ocho días anteriores al inicio del tratamiento, a la Agencia Nacional de Protección de Datos Personales el inicio del tratamiento o conjunto de tratamientos, total o parcialmente automatizados, destinados a la consecución de uno o varios fines interrelacionados, con algunas excepciones. Por otra parte, el interesado, al ejercer el derecho de acceso, tiene derecho a ser informado por el responsable sobre las razones que subyacen al tratamiento automatizado de datos que le conciernen, de acuerdo con el artículo 11.

47. En el Uruguay, la Ley núm. 18831, de 11 de agosto de 2008, de Protección de Datos Personales, establece en su artículo 13 que los titulares de los datos tienen derecho a ser informados de forma expresa, precisa e inequívoca, previamente al momento de la recolección de los datos, sobre los criterios de valoración, los procesos aplicados y la solución tecnológica o el programa utilizado, en el caso de un tratamiento automatizado de datos destinado a evaluar determinados aspectos de su personalidad, como el rendimiento laboral, crédito, fiabilidad, conducta, entre otros, que tomen decisiones con efectos jurídicos que afecte de manera significativa al titular de los datos. Además, en la Ley se añade que "[c]uando los datos personales no sean recolectados directamente de sus titulares, la información […] les deberá ser proporcionada a estos en un plazo de cinco días hábiles de recibida la solicitud por parte de los responsables" del tratamiento.

VI. Principio de explicabilidad en el tratamiento de datos personales en proyectos de inteligencia artificial

48. Cada vez es más frecuente la creación de "perfiles virtuales" sobre las personas a partir de la información existente. Adicionalmente, a menudo se adoptan decisiones con respecto a ellas a partir del tratamiento automatizado de sus datos mediante diversas herramientas tecnológicas.

49. El ser humano puede verse afectado positiva o negativamente con las decisiones que se tomen a su respecto a partir del uso y tratamiento de datos en proyectos de

[40] "Artículo 23. Derecho al tratamiento objetivo. El titular de datos personales tiene derecho a no verse sometido a una decisión con efectos jurídicos sobre él o que le afecte de manera significativa, sustentada únicamente en un tratamiento de datos personales destinado a evaluar determinados aspectos de su personalidad o conducta, salvo que ello ocurra en el marco de la negociación, celebración o ejecución de un contrato o en los casos de evaluación con fines de incorporación a una entidad pública, de acuerdo a ley, sin perjuicio de la posibilidad de defender su punto de vista, para salvaguardar su legítimo interés."

inteligencia artificial. Preocupa la forma de proteger los derechos de las personas afectadas por las decisiones tomadas respecto a ellas con herramientas o tecnologías de inteligencia artificial. En el Libro Blanco sobre la inteligencia artificial, por ejemplo, se pone de presente que: "[c]omo sucede con toda nueva tecnología, el uso de la [inteligencia artificial] presenta tanto oportunidades como amenazas. Los ciudadanos temen quedarse indefensos a la hora de proteger sus derechos y su seguridad frente a los desequilibrios informativos de la toma de decisiones mediante algoritmos"[41].

50. Dado lo anterior, es necesario que las personas conozcan qué datos se utilizaron para tomar una decisión que les afecte y cuál fue la lógica utilizada para llegar a ella. Tener acceso a esta información, entre otros aspectos, le servirá a la persona afectada para saber si la decisión tomada a su respecto es correcta y, en caso de no serlo, poder defenderse. En otras palabras, esa información es necesaria para garantizar el debido proceso porque será prueba para debatir las posibles imprecisiones o injusticias generadas contra una persona con ocasión del tratamiento de sus datos personales en procesos de inteligencia artificial. En este sentido, el citado Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial ha sido enfático en recalcar que el principio de explicabilidad:

> es crucial para conseguir que los usuarios confíen en los sistemas de [inteligencia artificial] y para mantener dicha confianza. Esto significa que los procesos han de ser transparentes, que es preciso comunicar abiertamente las capacidades y la finalidad de los sistemas de [inteligencia artificial] y que las decisiones deben poder explicarse —en la medida de lo posible— a las partes que se vean afectadas por ellas de manera directa o indirecta. Sin esta información, no es posible impugnar adecuadamente una decisión […]. El grado de necesidad de explicabilidad depende en gran medida del contexto y la gravedad de las consecuencias derivadas de un resultado erróneo o inadecuado[42].

51. Todo ello explica por qué es importante la transparencia en la inteligencia artificial, dado que esta no debe ser oscura, secreta o engañosa. Por eso, en la citada Declaración Europea se puso de presente que:

> Toda persona debería estar empoderada para beneficiarse de las ventajas de los sistemas algorítmicos y de inteligencia artificial, especialmente a fin de tomar sus propias decisiones en el entorno digital con conocimiento de causa, así como estar protegida frente a los riesgos y daños a su salud, su seguridad y sus derechos fundamentales[43].

52. En línea con lo anterior, la Red Iberoamericana de Protección de Datos recomendó en 2019 que se incrementase la transparencia con los titulares de los datos personales[44].

53. Posteriormente, y también relacionado con lo anterior, en la citada resolución de 2020 de la Asamblea Global de Privacidad se insta a que las organizaciones que desarrollan o utilizan sistemas de inteligencia artificial deben tener en consideración las siguientes medidas: a) garantizar la transparencia y la apertura respecto a la divulgación del uso de la inteligencia artificial, los datos que se utilizan y la lógica implicada en la inteligencia artificial; b) asegurar que se identifique a un actor

[41] Véase https://eur-lex.europa.eu/legal-content/ES/TXT/?qid=1603192201335&uri=CELEX%3A52020DC0065, pág. 12.

[42] Véase https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, pág. 16.

[43] Véase https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ%3AJOC_2023_023_R_0001.

[44] Véase https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf, págs. 23 y 24.

humano responsable, a quien se le puedan plantear las preocupaciones relacionadas con las decisiones automatizadas y se puedan ejercer los derechos y, además, que pueda impulsar la evaluación del proceso de decisión y la intervención humana; c) proporcionar explicaciones, en un lenguaje claro y comprensible para los seres humanos, sobre las decisiones automatizadas que tome la inteligencia artificial, previa solicitud, y d) realizar una intervención humana en la decisión automatizada tomada por la inteligencia artificial, previa solicitud[45].

54. Todo lo anterior se corresponde parcialmente con lo previsto en el Reglamento General de Protección de Datos cuando, por ejemplo, establece que:

> Cuando los datos personales no se hayan obtenido del interesado, el responsable del tratamiento le facilitará la siguiente información: […] 2. […] g) la existencia de decisiones automatizadas, incluida la elaboración de perfiles, a que se refiere el artículo 22, apartados 1 y 4, y, al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado [46].

Adicionalmente, el interesado o titular del dato tiene derecho a:

> obtener del responsable del tratamiento confirmación de si se están tratando o no datos personales que le conciernen y, en tal caso, derecho de acceso a los datos personales y a la siguiente información: […] h) la existencia de decisiones automatizadas, incluida la elaboración de perfiles, a que se refiere el artículo 22, apartados 1 y 4, y, al menos en tales casos, información significativa sobre la lógica aplicada, así como la importancia y las consecuencias previstas de dicho tratamiento para el interesado[47].

55. En el siguiente gráfico, el Instituto Nacional de Normas y Tecnología resume el alcance de dicho principio[48]:

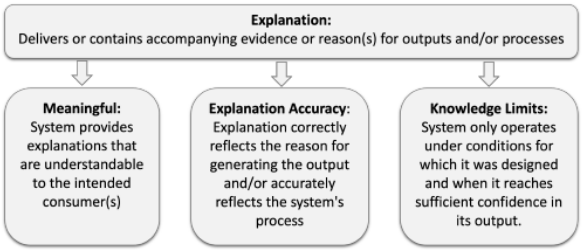

Fig. 1. Illustration of the four principles of explainable artificial intelligence. Arrows indicate that for a system to be explainable, it must provide an explanation. The remaining three principles are the fundamental properties of those explanations.

56. En el siguiente cuadro se explica lo más relevante de cada principio según el documento del Instituto Nacional de Normas y Tecnología[49]:

[45] Véase https://globalprivacyassembly.org/document-archive/adopted-resolutions/, pág. 3.
[46] Véase http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32016R0679, art. 14, párr. 2 g).
[47] *Ibid*., art. 15, párr. 1.
[48] Instituto Nacional de Normas y Tecnología, *Four Principles of Explainable Artificial Intelligence*, – NISTIR 8312 (2021), pág. 3. Disponible en: https://doi.org/10.6028/NIST.IR.8312.
[49] La explicación de la tabla es una adaptación, resumen y traducción del texto original en inglés citado y disponible en https://doi.org/10.6028/NIST.IR.8312.

Principio	*Significado o alcance*
Explanation (Explicación)	Se debe proporcionar evidencia, apoyo o razonamiento relacionado con un resultado o un proceso de un sistema de [inteligencia artificial].
Meaningful (Significado entendible)	Se debe explicar en términos que le permitan a la persona comprender la explicación. En otras palabras, este principio busca que la explicación sea inteligible para determinada audiencia. Muchos factores inciden en una buena explicación, razón por la cual se debe tener presente el público objetivo o la audiencia a la cual se dirige la explicación.
Explanation Accuracy (Explicación precisa)	Este principio exige rigurosidad, precisión y completitud de la explicación técnica.
Knowledge Limits (Límites del conocimiento)	Identificar y declarar los límites de conocimiento implica dejar claro que el sistema no es perfecto ni infalible porque la [inteligencia artificial] opera dentro de cierto límites y condicionamientos en los que ha sido programada. También depende, entre otras, de la calidad y la cantidad de información procesada.

57. Se ha planteado que la explicación debe: a) "ser comprensible y convincente para el usuario"; b) "reflejar de manera precisa el razonamiento del sistema; c) "ser completa", y d) "ser específica en el sentido de usuarios diferentes con diferentes circunstancias o diferentes resultados, deberían obtener explicaciones diferenciadas"[50]. Adicionalmente, se pone de presente que:

> la explicabilidad de la inteligencia artificial es una aspiración que se entiende desde un punto de vista ético e incluso legal, pero que tiene unas profundas dificultades técnicas que conviene conocer y, probablemente, gran parte de la solución sea también técnica, en la medida que se consiga rediseñar algoritmos o identificar otros nuevos que satisfagan las aspiraciones éticas y normativas[51].

La UNESCO, por su parte, indica que:

> La explicabilidad supone hacer inteligibles los resultados de los sistemas de [inteligencia artificial] y facilitar información sobre ellos. La explicabilidad de los sistemas de [inteligencia artificial] también se refiere a la inteligibilidad de la entrada, salida y funcionamiento de cada componente algorítmico y la forma en que contribuye a los resultados de los sistemas [52].

58. Para precisar el alcance del principio de explicabilidad es necesario tener presente su objetivo y, a partir de este, establecer lo necesario para alcanzarlo. En línea con lo anterior, se ha señalado que:

> si con el principio de explicabilidad se quiere que cualquier ser humano conozca por qué razón una decisión es tomada a partir del tratamiento de sus datos con herramientas de [inteligencia artificial], pues, por lo menos la explicación debería ser clara, sencilla, completa, veraz y fácilmente entendible por quien

[50] Gavilán, Ignacio, "Cuatro principios para una buena explicabilidad de los algoritmos" (2022). Disponible en: https://ignaciogavilan.com/cuatro-principios-para-una-buena-explicabilidad-de-los-algoritmos/.

[51] *Ibid*.

[52] Véase https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa, pág. 23.

> solicita la explicación. No basta que se informe sobre los datos utilizados como insumos para generar la decisión, sino la lógica o metodología empleada para llegar a la misma. El reto no es menor, pero es alcanzable si se tiene voluntad para explicarle fácilmente a la gente por qué razón se generó una decisión basada en el tratamiento de sus datos personales[53].

59. A continuación, se hace referencia a algunos ejemplos de la regulación local de unos países que han incorporado de manera tácita o explícita el principio de explicabilidad dentro de su marco legal.

60. La regulación de Colombia prohíbe el tratamiento de datos que "induzca a error"[54] y, en el caso concreto de las decisiones tomadas con respecto a solicitudes de crédito, exige que aquellos que rechacen dichas solicitudes informen por escrito a la persona afectada, si así lo requiere, "las razones objetivas del rechazo"[55] del crédito.

61. Por su parte, en el Ecuador, en el artículo 20 de su Ley Orgánica de Protección de Datos, se establece que el titular de los datos, ante una decisión basada única o parcialmente en valoraciones que sean producto de procesos automatizados, incluida la elaboración de perfiles, que produzcan efectos jurídicos en él o que atenten contra sus derechos y libertades fundamentales, puede exigir una explicación motivada de la decisión, obtener los criterios de valoración sobre el programa automatizado, presentar observaciones, solicitar información sobre los tipos de datos utilizados y la fuente de la cual se obtuvieron, e impugnar la decisión ante el responsable o encargado (con ciertas excepciones).

62. En el Uruguay, en el artículo 16 de la Ley núm. 18331, se establece que:

> las personas tienen derecho a no verse sometidas a una decisión con efectos jurídicos que les afecte de manera significativa, que se base en un tratamiento automatizado de datos destinado a evaluar determinados aspectos de su personalidad, como su rendimiento laboral, crédito, fiabilidad, conducta, entre otros. Quien se vea afectado tendrá derecho a obtener información del responsable de la base de datos tanto sobre los criterios de valoración como sobre el programa utilizado en el tratamiento que sirvió para adoptar la decisión manifestada en el acto.

VII. Conclusiones

63. **A partir de lo anterior se extraen las siguientes conclusiones**:

a) La transparencia y la explicabilidad contribuyen a generar confianza en la inteligencia artificial y a respetar los derechos humanos;

b) Quienes desarrollan inteligencia artificial deben ser transparentes con relación a cómo se tratan los datos (cómo se recopilan, almacenan y utilizan), así como también con relación a la forma en que se

[53] Nelson Remolina Angarita, "Del principio de explicabilidad en la inteligencia artificial (notas preliminares)", en *Protección de datos personales: doctrina y jurisprudencia*, Pablo Palazzi, ed., tomo III (Centro de Tecnologías y Sociedad de la Universidad de San Andrés, Buenos Aires, 2023).

[54] Ley Estatutaria 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales, art. 4 d).

[55] Ley Estatutaria 2157 de 2021, por medio de la cual se modifica y adiciona la Ley Estatutaria 1266 de 2008, y se dictan disposiciones generales del *habeas data* con relación a la información financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones, art. 5, párr. 1.

toman las decisiones basadas en la inteligencia artificial, la confiabilidad de estas y la seguridad de la información;

c) Las personas afectadas por las decisiones tomadas a partir de la inteligencia artificial merecen una explicación clara, sencilla, completa, veraz y comprensible de la motivación de esa decisión. En este sentido, el principio de explicabilidad es de cardinal importancia no solo porque se corresponde con el principio de transparencia, sino porque permitirá el derecho de defensa y el debido proceso de dichas personas;

d) La explicabilidad y la transparencia demandan claridad, completitud, veracidad, imparcialidad y publicidad de las decisiones adoptadas mediante inteligencia artificial y de la lógica, método o razonamiento para tomar decisiones sobre los seres humanos a partir de la información y, particularmente, los datos personales. La explicabilidad y la transparencia se oponen, desde luego, a la opacidad, la oscuridad, el engaño, la mentira y el abuso del poder informático, los cuales son algunos síntomas de un tratamiento de datos ilegal carente de ética y respeto por los seres humanos y su dignidad.

VIII. Recomendaciones

64. **Teniendo en cuenta lo anterior, la Relatora Especial exhorta a los Estados a**:

a) Promover la transparencia en la inteligencia artificial para mitigar los riesgos que la opacidad pueda generar en la sociedad y, especialmente, respecto de la protección de los derechos humanos;

b) Incorporar en sus regulaciones el principio de explicabilidad, no solo para que las personas comprendan cómo se adoptaron las decisiones que las afectan, sino para que puedan tener herramientas para defender sus derechos humanos frente a la inteligencia artificial;

c) Fomentar prácticas éticas que aseguren la transparencia y la explicabilidad en el tratamiento de datos personales en los proyectos o procesos de inteligencia artificial;

d) Impulsar, apoyar y facilitar la educación y la alfabetización digital para que los ciudadanos comprendan mejor los conceptos relacionados con la inteligencia artificial, la transparencia y la explicabilidad, de manera que puedan exigir el respeto de sus derechos.