

INTELIGENCIA ARTIFICIAL BASADA EN LAS PERSONAS: APUESTA EUROPEA POR UNA TECNOLOGÍA ÉTICA¹

*PEOPLE-BASED ARTIFICIAL INTELLIGENCE: EUROPEAN
COMMITMENT TO AN ETHICAL TECHNOLOGY*

Álvaro A. Sánchez Bravo¹ 

¹Facultad de Derecho de la Universidad de Sevilla, Sevilla, España. Doutor em Direito. E-mail: elialva@us.es

Cada tres años se dispone de más información nueva que la creada en toda la historia de la humanidad. El único modo de gestionar esa información es mediante el uso de tecnologías digitales.

El aumento de la capacidad computacional hizo que fuera posible la implementación de algoritmos cada vez más complejos, potentes y flexibles. Al mismo tiempo, la amplia disponibilidad de datos dio lugar a grandes avances en el campo de la inteligencia artificial (IA). Los datos están, por tanto, en el centro de esta transformación. Pero la forma en que se recojan y utilicen los datos debe situar los intereses de las personas en primer lugar, conforme los valores, derechos fundamentales y las normas jurídicas propias de Estados democráticos de Derecho.

Uno de los más relevantes objetivos de los sistemas modernos de IA es distinguir y extraer patrones de datos sin procesar para construir su propio conocimiento. Frente a los sistemas expertos, la solución actual no es trabajar con una base de datos de conocimiento, sino aprender conocimiento. Esa capacidad de la IA para aprender se conoce como aprendizaje de máquina y permiten que las computadoras resuelvan problemas que requieren cierta comprensión del mundo real y tomen decisiones situacionales y subjetivas.

DOI: <http://dx.doi.org/10.20912/rdc.v16i39.441>

Autor convidado

¹ Texto publicado en *Derecho, Inteligencia Artificial y Nuevos Entornos Digitales*, edic. a cargo de A. Sánchez Bravo, Punto Rojo Libros, Sevilla, 2020.



El aprendizaje supone que las máquinas puedan encontrar patrones diferentes de los generalmente asimilados por los cerebros humanos.^{1 2}

El crecimiento de la capacidad informática y la disponibilidad de datos, así como los avances de los algoritmos hacen de la inteligencia artificial una de las tecnologías más estratégicas del siglo XXI.

La Unión Europea (UE), consciente de la competencia feroz que se desarrollan en los mercados, empezó sus trabajos en orden a establecer un enfoque sólido y conjunto, y abordar los retos que presenta la IA. La idea era triple: avanzar en el conocimiento científico; ahondar en el liderazgo tecnológico, y, sobre todo, garantizar que las nuevas tecnologías de IA estén al servicio de todos los europeos, de una manera inclusiva y garantizando sus derechos.

1 Comisión Económica para América Latina y el Caribe (CEPAL), Datos, algoritmos y políticas: la redefinición del mundo digital (LC/CMSI.6/4), Santiago, 2018, p. 169-176.

2 La IA puede utilizarse para el desarrollo económico y social, basado en los Objetivos de Desarrollo del Milenio (ODM), identificándose cuatro elementos que enmarcan los efectos de la IA en el desarrollo:

Transferencias de inteligencia.

1) La inteligencia a distancia hace referencia a que las modernas redes de telecomunicaciones permiten aplicar a distancia sistemas de inteligencia artificial altamente entrenados.

La inteligencia a distancia es la capacidad de las tecnologías de inteligencia artificial (IA), en combinación con las telecomunicaciones, para remediar la carencia de recursos en campos que no cuentan con personal suficiente o han sido poco investigados. Esto es especialmente importante si se considera que el aprendizaje multitarea y de transferencia permite reutilizar la inteligencia generada u obtenida en otro lugar. Una de las aplicaciones pioneras es el uso de la IA en los sectores de educación y salud, como en el caso de la educación a distancia automatizada y los diagnósticos a distancia para tratar una serie de enfermedades (cataratas congénitas, tuberculosis y cáncer de mama, entre otras).

En el ámbito de la educación, las soluciones de IA permiten automatizar los sistemas de educación y tutoría, proceso que a su vez permite soluciones de bajo costo a gran escala. Se pueden automatizar actividades especialmente estructuradas, como el aprendizaje de idiomas, la programación de software o las habilidades analíticas cuantitativas. Los sistemas de IA de aprendizaje posibilitan la masificación de una experiencia de educación individualizada para un curso estructurado.

La inteligencia a distancia puede revolucionar la industria de la salud al incrementar la eficiencia y la cobertura.

2) La inteligencia local se refiere al hecho de que los sistemas IA se pueden aplicar de forma autónoma localmente, adaptándose al contexto y requisitos locales.

Casos emblemáticos son los relacionados con el cambio climático, igualdad de género en el lugar de trabajo y en el aula y ciudades inteligentes.

Manipulación de la realidad.

3) La realidad aumentada, virtual y duplicada se refiere al hecho de que la IA permite crear los llamados *gemelos digitales* de aspectos de la realidad que luego puedan usarse para mejorar nuestra comprensión de la realidad o duplicar aspectos ésta.

En muchas aplicaciones prácticas, la inteligencia a distancia y local se combina cada vez más con el uso de la realidad virtual y aumentada. Así los vehículos autónomos, por ejemplo, pueden usar mapas tridimensionales para tomar decisiones en tiempo real. Las realidades virtuales guiadas por la IA también se utilizan para fomentar la educación y la igualdad de género.

Más allá de las realidades aumentadas y virtuales, la IA también se está utilizando para duplicar el diseño de átomos del mundo real y objetos moleculares, como los alimentos. Se trata de duplicar la estructura de un determinado artículo para desarrollar una versión más sostenible de este. La duplicación se podría utilizar para combatir el hambre.

4) La realidad de grano fino hace referencia al hecho de que la huella digital proporciona mapas cada vez más detallados de la realidad y el aprendizaje de máquina permite explotar la información resultante para impulsar el logro de los objetivos de desarrollo.

Una de las formas en que la IA puede proporcionar información más detallada sobre áreas específicas en materia de desarrollo económico y social es refinando nuestra comprensión de la realidad mediante una nueva manera de recopilar datos con mayor granularidad. El aprendizaje automatizado de representación permite transformar detalles recién obtenidos en características útiles.

Ibídem.

Pero para ello, el punto de partida debía ser, dado el enorme impacto de la IA en la vida de las personas, generar confianza en su uso, y para ello, debe cimentarse en los valores y derechos fundamentales de la UE³, como la dignidad humana y la protección de la intimidad.

En materia de datos, ya desde 2014, la Comisión adoptó una serie de medidas. Así el Reglamento General de Protección de Datos⁴, el Reglamento relativo a la libre circulación de datos no personales⁵, el “Reglamento sobre Ciberseguridad”⁶, y la Directiva sobre Datos Abiertos⁷. Merece destacarse también la Directiva sobre contenidos digitales⁸, que introduce derechos contractuales cuando se prestan servicios digitales a consumidores que facilitan el acceso a sus datos.

En el campo de la IA, y para abordar los retos y aprovechar al máximo las oportunidades, en abril de 2028 la Comisión Europea, inicia las actividades comunitarias con la publicación de una Estrategia europea⁹. Esta primera iniciativa europea destaca por colocar a la persona en el centro del desarrollo de la IA, adoptando un planteamiento triple para potenciar la capacidad tecnológica e industrial de la UE e impulsar la IA en todos los ámbitos económicos, prepararse para las transformaciones socioeconómicas y garantizar la aprobación de un marco ético y jurídico adecuado.

Un año después la Comisión Europea elaboró otra estrategia, cuyo título no deja dudas acerca de la opción de cómo debe ser la IA en la UE: *generar confianza en la inteligencia artificial centrada en el ser humano*¹⁰.

3 Carta de los Derechos Fundamentales de la Unión Europea, DOUE C 83. 30.03.2010.

4 Reglamento (UE) 2016/679 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. DOUE L 119. 04.05.2016.

DIRECTIVA (UE) 2016/680 DEL PARLAMENTO EUROPEO Y DEL CONSEJO, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, y a la libre circulación de dichos datos y por la que se deroga la Decisión Marco 2008/977/JAI del Consejo. DOUE L 119. 04.05.2016.

Vid. Sánchez Bravo, A. “Nuevo marco europeo de protección de datos personales”, en la obra colectiva, edic. a cargo de Álvaro Sánchez Bravo, *Derechos humanos y protección de datos personales en el Siglo XXI: homenaje a Cinta Castillo Jiménez*, Punto Rojo Libros, Sevilla, 2014, pp. 255-288; y, “Hacia un nuevo marco europeo de protección de datos personales: empoderamiento de los ciudadanos en la sociedad tecnológica”, en la obra colectiva, coord. a cargo de Yarina Amoroso Fernández, *Sociocibernética e Infoética: contribución a una nueva cultura y praxis jurídica*, Editorial UNIJURIS, La Habana, 2015. pp. 108-133.

5 REGLAMENTO (UE) 2018/1807 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 14 de noviembre de 2018 relativo a un marco para la libre circulación de datos no personales en la Unión Europea. DOUE L 303. 28.11.2018.

6 REGLAMENTO (UE) 2019/881 DEL PARLAMENTO EUROPEO Y DEL CONSEJO, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) n° 526/2013 («Reglamento sobre la Ciberseguridad»). DOUE L 151. 07.06.2019.

7 DIRECTIVA (UE) 2019/1024 DEL PARLAMENTO EUROPEO Y DEL CONSEJO, de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público (versión refundida). DOUE L 172. 26.06.2019.

8 Directiva (UE) 2019/770 del Parlamento Europeo y del Consejo, de 20 de mayo de 2019, relativa a determinados aspectos de los contratos de suministro de contenidos y servicios digitales. DOUE L 136. 22.05.2019.

9 COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Inteligencia artificial para Europa. COM (2018) 237. Bruselas. 25.04.2018.

10 COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES.

La confianza es un requisito imprescindible y previo para garantizar que la IA esté garantizada y los ciudadanos confíen en ella.

Pero la IA presenta nuevos desafíos, ya que permite a las máquinas “aprender” y tomar decisiones y llevarlas a término sin acción humana. Ahora bien, las decisiones adaptadas “autónomamente” pueden dar como resultado datos incompletos, y por tanto, no fiables, que puedan ser manipulados por ciberataques, puedan ser datos sesgados o simplemente estar equivocados. Aplicar la tecnología de forma irreflexiva o poco cuidadosa generará resultados problemáticos, y ataques lesivos a los derechos de los ciudadanos, lo que supondrá su rechazo.

Por tanto, todas las acciones de IA deben ajustarse no sólo a la ley, sino también respetar unos principios éticos y garantizar que su implementación no genera daños involuntarios. En cada una de las fases de desarrollo de la IA deben garantizarse los derechos de los ciudadanos, y especialmente, la diversidad en cuanto al género, origen racial o étnico, religión o creencias, discapacidad y edad.

Debe quedar claro que la IA debe mejorar y apoyar la capacidad de las personas, pero no sustituirla.

Es por ello que se imponía la elaboración de unas directrices éticas aplicables a desarrolladores, proveedores y usuarios de IA en el UE estableciendo unas condiciones de competencia ética en todos los Estados miembros.

Para dar cumplimiento a este mandato, el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial¹¹, elaboró en abril de 2019 unas *Directrices Éticas para una IA fiable*¹².

La fiabilidad de la IA se apoya en tres componentes: que debe cumplirse a lo largo de todo el ciclo de vida del sistema. Estos tres componentes son imprescindibles cada uno de ellos, pero por sí mismos no son suficientes para garantizar una IA fiable. Estos componentes son:

a) la IA debe ser lícita; es decir, cumplir todas las leyes y reglamentos aplicables. Las Directrices no profundizan en este apartado, remitiéndose al Derecho primario y secundario de la UE y las normas y acuerdos internacionales aplicables.

b) la IA ha de ser ética. Deben desarrollarse, desplegarse y utilizarse los sistemas de inteligencia artificial respetando los principios éticos de respeto de la autonomía humana¹³,

Generar confianza en la inteligencia artificial centrada en el ser humano. COM (2019) 168. Bruselas. 08.04.2019.

11 El Grupo de Expertos de alto nivel sobre IA es un grupo de expertos independiente constituido por la Comisión Europea en junio de 2018.

<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

12 Directrices Éticas para una IA fiable.

<https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es/format-PDF/source-121796438>

13 Los derechos fundamentales en los que se apoya la UE van dirigidos a garantizar el respeto de la libertad y la autonomía de los seres humanos. Las personas que interactúen con sistemas de IA deben poder mantener una autonomía plena y efectiva sobre sí mismas y ser capaces de participar en el proceso democrático. Los sistemas de IA no deberían subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de manera injustificada. En lugar de ello, los sistemas de IA deberían diseñarse de forma que aumenten, complementen y potencien las aptitudes cognitivas, sociales y culturales de las personas. La distribución de funciones entre los seres humanos y los sistemas de IA debería seguir principios de diseño centrados en las personas, y dejar amplias oportunidades para la elección humana. Esto implica garantizar la supervisión y el control humanos sobre los procesos de trabajo de los sistemas de IA. Ídem, p.15.

prevención del daño¹⁴, equidad¹⁵ y explicabilidad¹⁶, reconociendo y desarrollando mecanismos de resolución de conflictos que puedan surgir entre estos principios¹⁷.

Deberá prestarse especial atención a las situaciones que afecten a los grupos de población más vulnerables¹⁸, así como a aquellas situaciones donde se de una manifiesta situación de desequilibrio de poder o información.

- 14 Los sistemas de IA no deberían provocar daños (o agravar los existentes) ni perjudicar de cualquier otro modo a los seres humanos. Esto conlleva la protección de la dignidad humana, así como de la integridad física y mental. Todos los sistemas y entornos de IA en los que operan estos deben ser seguros. También deberán ser robustos desde el punto de vista técnico, y debería garantizarse que no puedan destinarse a usos malintencionados. Las personas vulnerables deberían recibir mayor atención y participar en el desarrollo y despliegue de los sistemas de IA. Se deberá prestar también una atención particular a las situaciones en las que los sistemas de IA puedan provocar efectos adversos (o agravar los existentes) debido a asimetrías de poder o de información, por ejemplo entre empresarios y trabajadores, entre empresas y consumidores o entre gobiernos y ciudadanos. La prevención del daño implica asimismo tener en cuenta el entorno natural y a todos los seres vivos. *Íbidem*.
- 15 El desarrollo, despliegue y utilización de sistemas de IA debe ser equitativo. Pese a que reconocemos que existen muchas interpretaciones diferentes de la equidad, creemos que esta tiene tanto una dimensión sustantiva como procedimental. La dimensión sustantiva implica un compromiso de: garantizar una distribución justa e igualitaria de los beneficios y costes, y asegurar que las personas y grupos no sufran sesgos injustos, discriminación ni estigmatización. Si se pueden evitar los sesgos injustos, los sistemas de IA podrían incluso aumentar la equidad social. También se debería fomentar la igualdad de oportunidades en términos de acceso a la educación, los bienes los servicios y la tecnología. Además, el uso de sistemas de IA no debería conducir jamás a que se engañe a los usuarios (finales) ni se limite su libertad de elección. Asimismo, la equidad implica que los profesionales de la IA deberían respetar el principio de proporcionalidad entre medios y fines, y estudiar cuidadosamente cómo alcanzar un equilibrio entre los diferentes intereses y objetivos contrapuestos. La dimensión procedimental de la equidad conlleva la capacidad de oponerse a las decisiones adoptadas por los sistemas de IA y por las personas que los manejan, así como de tratar de obtener compensaciones adecuadas frente a ellas³². Con este fin, se debe poder identificar a la entidad responsable de la decisión y explicar los procesos de adopción de decisiones. *Ídem*, pp. 15-16.
- 16 Esto significa que los procesos han de ser transparentes, que es preciso comunicar abiertamente las capacidades y la finalidad de los sistemas de IA 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. No siempre resulta posible explicar por qué un modelo ha generado un resultado o una decisión particular (ni qué combinación de factores contribuyeron a ello). Esos casos, que se denominan algoritmos de «caja negra», requieren especial atención. En tales circunstancias, puede ser necesario adoptar otras medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema), siempre y cuando el sistema en su conjunto respete los derechos fundamentales. 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. *Ídem*, p.16
- 17 Cabe la posibilidad de que surjan tensiones entre los principios anteriores, y no existe una solución establecida para resolverlas. En consonancia con el compromiso fundamental de la UE con la participación democrática, el respeto de las garantías procesales y la participación abierta en la esfera política, deberían establecerse métodos que posibiliten un debate responsable sobre dichas tensiones. Además, los beneficios globales de los sistemas de IA deberían ser sustancialmente superiores a los riesgos individuales previsibles.
- 18 Debido a su heterogeneidad, no existe una definición generalmente aceptada ni que cuente con un consenso amplio del concepto de «personas vulnerables». Lo que se considera una persona o grupo vulnerable suele depender del contexto. Los sucesos vitales de carácter temporal (como la infancia o la enfermedad), los factores de mercado (como la asimetría de información o el poder de mercado), los factores económicos (como la pobreza), los vinculados a la identidad de las personas (como el género, la religión o la cultura) y otros pueden desempeñar un papel en ese sentido. La Carta de los Derechos Fundamentales de la Unión Europea recoge en su artículo 21, relativo a la no discriminación, los motivos de discriminación siguientes, que pueden servir como punto de referencia, entre otros: el sexo, la raza, el color, los orígenes étnicos o sociales, las características genéticas, la lengua, la religión o las convicciones, las opiniones políticas o de cualquier otro tipo, la pertenencia a una minoría nacional, el patrimonio, el nacimiento, la discapacidad, la edad o la orientación sexual. En las disposiciones de otras leyes se abordan los derechos de determinados grupos, además de los enumerados anteriormente. Este tipo de listas nunca pueden ser exhaustivas, y pueden cambiar a lo largo del tiempo. Un grupo vulnerable es un grupo de personas que comparten una o varias características de vulnerabilidad. *Ídem*,

c) la IA debe ser robusta; es decir, sólida desde el punto de vista técnico (adecuada en un contexto determinado, como el ámbito de aplicación o la fase de ciclo de vida) como social (que garantice que el sistema de IA tenga adecuadamente el contexto y el entorno en que opera)¹⁹.

Los principios indicados deben traducirse en requisitos que permitan de manera efectiva conseguir una IA fiable. Los mismos deben ser aplicables a todas las partes que intervienen en algún momento en el ciclo de vida de sistemas de IA: desarrolladores, responsables de despliegue y usuarios finales, así como a la sociedad en su conjunto²⁰.

Así, cada una de esas partes desempeñarán diversos papeles para garantizar el cumplimiento de los requisitos: a) los desarrolladores deben incluir y aplicar los principios en la fase de diseño y desarrollo; b) los responsables del despliegue deben asegurarse de que los sistemas utilizados y los productos y servicios ofertados cumplen los requisitos imprescindibles; y, c) los usuarios y la sociedad deben estar informados sobre los requisitos exigidos a los anteriores, y disponer de mecanismos para exigir su cumplimiento.

Siete son los requisitos esenciales que debe cumplir y respetar cualquier aplicación de IA para ser fiable:

- Intervención y supervisión humana. Los sistemas de IA deben ayudar a las personas a escoger mejor y con más conocimiento de sus opciones, en función de sus objetivos. Con ello se garantiza que un sistema de IA no menoscaba la autonomía humana, ni causa efectos adversos. Por ello deberán incrementarse medidas estrictas de control, incluida la adaptabilidad, la exactitud y la explicabilidad. Al respecto las autoridades públicas deben establecer funciones de supervisión mediante sistemas de gobernanza²¹, que determinen, entre otras cosas, que cuanto más difícil sea para un ciudadano controlar un sistema de IA, más estricta será la gobernanza.

- Solidez y seguridad técnicas. Los algoritmos deben ser seguros, fiables y sólidos para resolver los errores o incoherencias durante todo el ciclo vital del sistema de IA, solucionando adecuadamente los errores que se produzcan. Al respecto, deberán elaborarse mecanismos de seguridad, incluyendo la seguridad desde el diseño, para verificar que son objetivamente verificables en cada fase del proceso, teniendo muy en cuenta la seguridad física y psicológica de

p.51.

19 Ídem, pp. 17-30.

20 El término «desarrolladores» hace referencia a las personas dedicadas a la investigación, el diseño o el desarrollo de sistemas de IA. Por «responsables del despliegue» se entienden las organizaciones públicas o privadas que utilizan sistemas de IA en sus procesos internos y para ofrecer productos y servicios a otros agentes. Los «usuarios finales» son aquellos que interactúan con el sistema de IA, ya sea de forma directa o indirecta. Por último, la «sociedad en su conjunto» engloba el resto de agentes, personas y entidades afectados de manera directa o indirecta por los sistemas de IA. *Ibíd.*

21 Entre los sistemas de gobernanza pueden indicarse, Human-in-the-loop (HITL) se refiere a la intervención humana en cada ciclo de decisión del sistema, lo que en muchos casos no es posible ni deseable. Human-on-the-loop (HOTL) se refiere a la capacidad de la intervención humana durante el ciclo de diseño del sistema y a la supervisión del funcionamiento del sistema. Human-in-command (HIC) se refiere a la capacidad de supervisar la actividad global del sistema de IA (incluido su impacto más amplio económico, social, jurídico y ético) y a la capacidad de decidir cuándo y cómo utilizar el sistema en cada situación determinada. Esto puede incluir la decisión de no utilizar un sistema de IA en una situación concreta, establecer niveles de discreción humana durante el uso del sistema o garantizar la capacidad de imponerse a una decisión tomada por el sistema. COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Generar confianza en la inteligencia artificial centrada en el ser humano, cit. p.5.

los usuarios. Igualmente, debe garantizarse la minimización, y cuando sea posible, la reversibilidad de consecuencias no deseadas o errores causados por el sistema²².

- Privacidad y gestión de datos. La protección de los datos personales debe garantizarse en todas las fases del ciclo vital del sistema de IA. Esto incluye los datos inicialmente suministrados por el usuario, así como la información generada sobre el usuario en el contexto de su utilización del sistema. Como sabemos, los registros digitales del comportamiento humano, pueden producir no sólo el suministro de preferencias al usuario, sino también que la IA infiera su orientación sexual, edad, género u opiniones políticas o religiosas. Es por ello, que es necesario garantizar que los datos recogidos no se utilizarán para discriminar a los usuarios de forma injusta o ilegal.

Será igualmente necesario garantizar la calidad e integridad de los datos, así como el acceso a los mismos²³.

- Transparencia. Debe garantizarse la trazabilidad de los sistemas de IA, registrar y documentar las decisiones tomadas por los sistemas como la totalidad del proceso que dio lugar a la decisión. Igualmente debe aportarse la explicabilidad del proceso algorítmico de tomas de decisiones, pero adaptada al nivel de las personas afectadas.

Deberá comunicarse adecuadamente las capacidades y limitaciones del sistema de IA a las diferentes partes interesadas afectadas conforme al caso de que se trate²⁴.

- Diversidad, no discriminación y equidad. Los conjuntos de datos que alimentan los sistemas de inteligencia artificial pueden verse afectados por la inclusión de sesgos históricos involuntarios, por no ser completos o por no haberse implementado modelos de gobernanza adecuados.

Puede ayudar a resolver estos problemas establecer equipos de diseño diversificados y crear mecanismos que garanticen la participación, en particular de los ciudadanos, en el desarrollo de la IA. Los sistemas de IA deberían tener en cuenta toda la gama de capacidades, habilidades y necesidades humanas y garantizar la accesibilidad mediante un enfoque de diseño universal para tratar de lograr la igualdad de acceso para las personas con discapacidades²⁵.

- Bienestar social y medio ambiental. Los sistemas de inteligencia artificial deben utilizarse para mejorar el cambio social positivo y aumentar la sostenibilidad y la responsabilidad ecológicas.

- Rendición de cuentas. Deben implantarse mecanismos que garanticen la responsabilidad y la rendición de cuentas de los sistemas de inteligencia artificial y de sus resultados.

El planteamiento de la UE en materia de inteligencia artificial (IA), fundado en la confianza y la excelencia, ofrecerá a los ciudadanos la seguridad necesaria para adoptar estas tecnologías, animando al mismo tiempo a las empresas a desarrollarlas²⁶.

22 Ibidem.

23 Directrices Éticas para una IA fiable, cit., p.22.

24 COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Generar confianza en la inteligencia artificial centrada en el ser humano, cit. p.6.

25 Ibidem.

26 https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_es#%C3%BAltimas-novedades

En febrero de 2020, la Comisión Europea presentó sus ideas y acciones para una transformación digital que funcione para todos, reflejando lo mejor de Europa: abierto, justo, diverso, democrático y confiado²⁷. Presenta una sociedad europea impulsada por soluciones digitales que priorizan a las personas, abre nuevas oportunidades para las empresas y fomenta el desarrollo de tecnología confiable para fomentar una sociedad abierta y democrática y una economía vibrante y sostenible. Digital es un facilitador clave para combatir el cambio climático y lograr la transición verde.

La Estrategia de Datos europea²⁸ y las opciones de política para garantizar el desarrollo centrado en el ser humano de la Inteligencia Artificial (IA)²⁹ presentada simultáneamente, son los primeros pasos para lograr estos objetivos.

Como señaló la Presidenta de la Comisión Europea, Úrsula von der Leyen: “*presentamos nuestra ambición de dar forma al futuro digital de Europa. Cubre todo, desde la ciberseguridad hasta las infraestructuras críticas, la educación digital hasta las habilidades, la democracia hasta los medios. Quiero que la Europa digital refleje lo mejor de Europa: abierta, justa, diversa, democrática y segura*”³⁰.

El objetivo es convertir a Europa en líder digital de confianza. Las tecnologías digitales, si se usan con un propósito, beneficiarán a los ciudadanos y las empresas de muchas maneras. Durante los próximos cinco años, la Comisión se centrará en tres objetivos clave en digital:

- *Tecnología que funciona para las personas;*
- *Una economía justa y competitiva; y*
- *Una sociedad abierta, democrática y sostenible.*

Europa se basará en su larga historia de tecnología, investigación, innovación e ingenio, y en su fuerte protección de los derechos y valores fundamentales. Las nuevas políticas y marcos permitirán a Europa desplegar tecnologías digitales de vanguardia y fortalecer sus capacidades de ciberseguridad. Europa continuará preservando su sociedad abierta, democrática y sostenible, y las herramientas digitales pueden apoyar estos principios. Desarrollará y seguirá su propio camino para convertirse en una economía y sociedad digital globalmente competitiva, basada en el valor e inclusiva, a la vez que seguirá siendo un mercado abierto pero basado en normas y trabajará en estrecha colaboración con sus socios internacionales.

Por su parte, el Libro Blanco sobre Inteligencia Artificial, como señala y resume certeramente la propia Comisión, prevé un marco para la Inteligencia Artificial confiable, basado en la **excelencia** y la **confianza**. En asociación con el sector público y privado, el objetivo es movilizar recursos a lo largo de toda la cadena de valor y crear los incentivos correctos para acelerar el despliegue de IA, incluso por parte de las pequeñas y medianas empresas. Esto incluye trabajar con los Estados miembros y la comunidad investigadora para atraer y conservar el talento. Como

27 COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Configurar el futuro digital de Europa. COM (2020) 67. Bruselas. 19.02.2020.

28 COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Una Estrategia Europea de Datos. COM (2020) 66. Bruselas. 19.02.2020.

29 LIBRO BLANCO sobre la inteligencia artificial – un enfoque europeo orientado en la excelencia y la confianza. COM (2020) 65. Bruselas. 19.02.2020.

30 https://ec.europa.eu/info/files/communication-shaping-europes-digital-future_en

los sistemas de IA pueden ser complejos y conllevar riesgos significativos en ciertos contextos, es esencial generar confianza. Se necesitan reglas claras para abordar los sistemas de IA de alto riesgo sin poner demasiada carga en los menos riesgosos. Se siguen aplicando normas estrictas de la UE para la protección del consumidor, para abordar las prácticas comerciales desleales y para proteger los datos personales y la privacidad.

Para casos de alto riesgo, como en salud, vigilancia o transporte, los sistemas de IA deben ser transparentes, trazables y garantizar la supervisión humana. Las autoridades deberían poder probar y certificar los datos utilizados por los algoritmos a medida que revisan cosméticos, automóviles o juguetes. Se necesitan datos imparciales para capacitar a los sistemas de alto riesgo para que funcionen adecuadamente y para garantizar el respeto de los derechos fundamentales, en particular la no discriminación. Si bien hoy en día, el uso del reconocimiento facial para la identificación biométrica remota está generalmente prohibido y solo puede usarse en casos excepcionales, debidamente justificados y proporcionados, sujetos a salvaguardas y basados en la legislación nacional o de la UE, la Comisión desea iniciar un amplio debate sobre qué circunstancias, si las hay, podrían justificar tales excepciones.

Para aplicaciones de IA de bajo riesgo, la Comisión prevé un esquema de etiquetado voluntario si aplican estándares más altos.

Todas las aplicaciones de IA son bienvenidas en el mercado europeo siempre que cumplan con las normas de la UE³¹.

En lo tocante a la estrategia europea de datos, la cantidad de datos generados por empresas y organismos públicos está en constante crecimiento. La próxima ola de datos industriales transformará profundamente la forma en que producimos, consumimos y vivimos. Pero la mayor parte de su potencial sigue sin cumplirse. El objetivo de la estrategia europea de datos es garantizar que la UE se convierta en un modelo a seguir y un líder para una sociedad potenciada por los datos. Para esto, tiene como objetivo establecer un verdadero espacio de datos europeo, un mercado único de datos, para desbloquear los datos no utilizados, lo que le permite fluir libremente dentro de la Unión Europea y en todos los sectores en beneficio de las empresas, los investigadores y las administraciones públicas. Los ciudadanos, las empresas y las organizaciones deben estar facultados para tomar mejores decisiones basadas en los conocimientos extraídos de los datos no personales. Esos datos deberían estar disponibles para todos, ya sean públicos o privados, de nueva creación o gigantes.

Para lograr esto, la Comisión propondrá primero establecer el marco regulatorio adecuado con respecto a la gobernanza, el acceso y la reutilización de datos entre empresas, entre empresas y gobierno, y dentro de las administraciones. Esto implica crear incentivos para compartir datos, establecer reglas prácticas, justas y claras sobre el acceso y uso de datos, que cumplan con los valores y derechos europeos, como la protección de datos personales, la protección del consumidor y las reglas de competencia. También significa hacer que los datos del sector público estén más ampliamente disponibles mediante la apertura de conjuntos de datos de alto valor en toda la UE y permitir su reutilización para innovar en la parte superior.

En segundo lugar, la Comisión tiene como objetivo apoyar el desarrollo de los sistemas tecnológicos y la próxima generación de infraestructuras, lo que permitirá a la UE y a todos

31 https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

los actores aprovechar las oportunidades de la economía de datos. Contribuirá a inversiones en proyectos europeos de alto impacto en espacios de datos europeos e infraestructuras de nube confiables y energéticamente eficientes.

Finalmente, lanzará acciones sectoriales específicas, para construir espacios de datos europeos en, por ejemplo, fabricación industrial, el acuerdo verde, movilidad o salud.

La Comisión también trabajará para reducir aún más la brecha de habilidades digitales entre los europeos y explorar cómo dar a los ciudadanos un mejor control sobre quién puede acceder a sus datos generados por máquina.

En los próximos meses la Comisión seguirá desarrollando nuevas iniciativas en estos campos, para establecer un marco europeo de vanguardia en IA. Pero no debe olvidarse en ningún momento, los riesgos que la propia IA incorpora si nos alejamos de marcos de gobernanza basados en criterios democráticos que tengan en su frontispicio los derechos fundamentales de los ciudadanos. Es muy importante la rentabilidad y la competitividad, la liderazgo en los mercados y el avance en la investigación y la innovación, pero nada servirá para hacer de Europa un referente mundial en IA, si ésta no se desarrolla por y para las personas.