

# Ética, igualdad e inteligencia artificial

Sofia B. Nunes<sup>1</sup>, Rui Nunes<sup>1</sup>

1. Faculdade de Medicina da Universidade do Porto, Porto, Portugal.

## Resumen

La inteligencia artificial conlleva profundos cambios en la sociedad, ya sean sociales, económicos o culturales, con importantes repercusiones en principios éticos universalmente compartidos, como la libertad, la privacidad, la igualdad y la no discriminación, especialmente en el sector salud. Si, por un lado, su desarrollo puede generar crecimiento económico, por otro, puede agravar la vulnerabilidad de diversos grupos, como las mujeres, las personas mayores o las personas con discapacidad, a consecuencia de la desinformación o los sesgos algorítmicos. Este artículo analiza los instrumentos legales más relevantes y que enmarcan la inteligencia artificial como herramienta para promover la igualdad y la inclusión general, y concluye que es necesario promover estrategias holísticas para anteponer las oportunidades que genera a los riesgos.

**Palabras clave:** Bioética. Ética. Inteligencia artificial. Derechos humanos. Políticas públicas antidiscriminación.

## Resumo

### Ética, igualdade e inteligência artificial

A inteligência artificial traz profundas alterações à sociedade, sejam sociais, econômicas ou culturais, com importantes repercussões em princípios éticos universalmente partilhados, como liberdade, privacidade, igualdade e não discriminação, em especial no setor da saúde. Se por um lado seu desenvolvimento pode gerar crescimento econômico, por outro pode agravar a vulnerabilidade de diversos grupos, como mulheres, idosos ou pessoas com deficiência, em decorrência de desinformação ou de vieses algorítmicos. O artigo analisa os mais relevantes instrumentos legais existentes que enquadram a inteligência artificial como ferramenta de fomento de igualdade e de inclusão geral e conclui que é preciso promover estratégias holísticas para colocar as oportunidades geradas por ela à frente dos riscos.

**Palavras-chave:** Bioética. Ética. Inteligência artificial. Direitos humanos. Políticas públicas antidiscriminatórias.

## Abstract

### Ethics, equality and artificial intelligence

Artificial intelligence brings profound changes to society, whether social, economic, or cultural, with important repercussions on universally shared ethical principles such as freedom, privacy, equality, and non-discrimination, especially in the health sector. While its development can generate economic growth, it can also exacerbate the vulnerability of various groups, such as women, older adults, or people with disabilities, as a result of misinformation or algorithmic biases. This article analyzes the most relevant existing legal instruments that frame artificial intelligence as a tool for promoting equality and general inclusion and concludes that holistic strategies must be promoted to put the opportunities generated by it ahead of the risks.

**Keywords:** Bioethics. Ethics. Artificial intelligence. Human Rights. Public nondiscrimination policies.

Los autores declaran que no existe ningún conflicto de interés.

La inteligencia artificial (IA) ha revolucionado la vida cotidiana al automatizar tareas, aumentando la productividad, creando nuevas formas de trabajo, pero también eliminando ciertos puestos de trabajo, con un tremendo potencial de crecimiento económico. En la medicina y en las demás profesiones de la salud también se prevé esta evolución. Esto no es diferente de lo que ocurrió en las revoluciones tecnológicas anteriores, desde la Primera Revolución Industrial, alrededor del siglo XVIII, con la introducción de máquinas para sustituir el trabajo manual; pasando por la Segunda Revolución Industrial, en el siglo XIX, con avances científicos como el telégrafo y la electrificación; la Tercera Revolución Industrial, en el siglo XX, con la transición de las industrias a las tecnologías de la información, en particular las computadoras e internet; hasta, finalmente, la Cuarta Revolución Industrial, que estamos viviendo actualmente, con rápidos avances tecnológicos. Toda revolución de este tipo ha provocado profundas transformaciones en las sociedades y ha exigido que las personas y entidades se adapten a una nueva realidad.

Sin embargo, se necesitan medidas específicas para contrarrestar los posibles efectos negativos de la IA en la sociedad, ya que, dependiendo de las medidas adoptadas, podrá crear o intensificar desigualdades o fomentar la inclusión. Es fundamental hacer de los sistemas de IA herramientas para la inclusión, lo que se logra trabajando del lado creador, desde la creación de las bases del sistema hasta su implementación, y del lado de los ciudadanos, desde los más jóvenes hasta los mayores, que necesitan entender cómo funciona y cómo se implementa la IA. Por lo tanto, es fundamental invertir en educación y formación para las nuevas tecnologías y, de esta forma, combatir los riesgos para la población, como la desinformación. La alfabetización en inteligencia artificial, sobre todo en inteligencia artificial generativa (IAGen), es esencial.

De hecho, la desinformación se considera la mayor amenaza a corto plazo para cualquier sociedad y puede poner en peligro el proceso democrático de un país al generar desconfianza y polarización y socavar derechos humanos fundamentales, como el derecho a la salud. Por ejemplo, en la Unión Europea, se observó un aumento de casi el 1.000% en los casos de sarampión entre el 2023 y el 2024 debido a la desinformación sobre los beneficios de la vacunación. En la medicina, la desinformación

puede tener un efecto dramático al inducir a los pacientes a diagnósticos y tratamientos no probados científicamente. Aunque no es nuevo, se trata de un fenómeno agravado por el rápido intercambio de información, como resultado de las nuevas tecnologías, en concreto las redes sociales. Si la IA puede generar contenido falso sin que nadie se dé cuenta de que es falso, ¿cómo podemos confiar en la información transmitida por los medios de comunicación?

Por ello, la legislación nacional e internacional se vuelve esencial para establecer los requisitos mínimos y las directrices éticas que deben seguir los sistemas de IA para asegurar que la población esté protegida. Sin embargo, las leyes deben implementarse y supervisarse de manera adecuada para detectar más fácilmente cualquier posible violación. El rápido avance de la tecnología también exige que la legislación se mantenga al día, lo que puede ser particularmente desafiante; sin embargo, debe haber adaptaciones constantes y se deben reevaluar las sanciones por cibercrimes para salvaguardar a las personas.

Este artículo comienza enumerando los efectos negativos del avance de la inteligencia artificial; pasa a sus potencialidades, incluido su papel como motor de inclusión a la luz de los principios bioéticos universales; y concluye con la legislación creada en este campo tanto en el ámbito nacional como internacional, centrándose en Portugal y Brasil, y en los principios éticos específicos de la IA, que siempre deben prevalecer.

## La inteligencia artificial generativa como motor de inclusión

Las revoluciones tecnológicas generan transformaciones en las sociedades en muy diversos niveles, con repercusiones tanto positivas como negativas cuyo impacto depende de las medidas implementadas. Si bien por una parte puede haber un agravamiento de las desigualdades y de la exclusión social, por otra, la igualdad puede generarse en varios frentes. El caso de las nuevas tecnologías digitales, en concreto la inteligencia artificial, no es diferente, pero trae consigo nuevos efectos negativos que es necesario prevenir, que son visibles en el ámbito social, económico, político y cultural y afectan a toda la población. Newfield<sup>1</sup> menciona seis problemas potenciales asociados a los sistemas de IA:

1. prejuicios sociales (especialmente vinculados a la etnicidad y al género);
2. opacidad (los usuarios no saben cómo se generan los resultados);
3. coerción (los algoritmos y los datos son controlados por los creadores y las plataformas, no por los usuarios);
4. violaciones sistémicas de la privacidad (la IA depende del acceso a grandes cantidades de datos, a menudo de diversas fuentes y que pueden contener información personal);
5. falta de libertad académica respecto a la investigación en las empresas (los objetivos comerciales pueden influir en los resultados);
6. el término “inteligencia” distorsiona el estado real y los efectos de estos sistemas (existe potencial de uso indebido, y los propios sistemas indican que tienen fallas).

Es necesario, por lo tanto, tener en cuenta los cambios socioeconómicos que pueden experimentar los más diversos segmentos poblacionales. La automatización que impulsa la IA conducirá al reemplazo de ciertas tareas y aumentará la productividad, lo que predice un crecimiento económico significativo. Se reconoce que la IA hará que muchos empleos se vuelvan obsoletos y revolucionará la mayoría de los sectores económicos, si no todos, como ocurrió en revoluciones tecnológicas anteriores. Se espera que, en las economías más avanzadas, haya un impacto en alrededor del 60% de los empleos y, en los mercados emergentes y en las economías menos avanzadas, el impacto sea menor, del 40% y del 26%, respectivamente<sup>2</sup>. También se estima que las mujeres se verán más afectadas que los hombres —alrededor del 21% de las mujeres—, dado que ocupan empleos más susceptibles a tales impactos, como los de los sectores de salud y educación, o el trabajo administrativo<sup>3</sup>.

No solo eso, sino que las capacidades cognitivas podrán verse afectadas, dado que, *aunque la inteligencia artificial generativa puede mejorar la eficiencia de los trabajadores, puede inhibir la visión crítica respecto al trabajo y potencialmente llevar a una dependencia excesiva a largo plazo de la herramienta y a la disminución de las habilidades para la resolución independiente de problemas. Una mayor confianza en la capacidad de la IA gen para realizar una tarea está relacionada con un menor esfuerzo de pensamiento crítico*<sup>4</sup>. Y también debemos hablar de la potencial exclusión social de la población de

mayor edad, ya que, por regla general, las generaciones más jóvenes demuestran una mayor aptitud para seguir el ritmo y adaptarse a la rápida evolución tecnológica. Ya sea en las interacciones sociales, ya sea en el acceso a los servicios, puede haber una mayor dificultad, e incluso segregación, de las personas mayores a medida que las interacciones se vuelven más tecnológicas, especialmente para aquellos que viven en zonas sin internet o en las que el servicio de internet no funciona bien.

La más reciente transformación digital ha traído consigo otro gran riesgo para la sociedad: la rápida propagación de información falsa, también conocida como desinformación y misinformación. Es necesario distinguir la desinformación de la misinformación (traducidos del inglés *desinformation* y *misinformation*, respectivamente). Ambas se refieren a información falsa, pero la primera no necesariamente conlleva la intención de engañar, mientras que la segunda tiene el objetivo de manipular. Así, a lo largo del texto, se utilizará la expresión “desinformación” para referirse a una información falsa, dado que se desconocen los motivos detrás de su transmisión.

De hecho, actualmente ha habido una cantidad creciente de desinformación, generada por el consumo y la difusión inmediata de información en los medios de comunicación, particularmente las redes sociales, con el potencial de socavar los sistemas democráticos. Esto no es nuevo, ya que las campañas de desinformación fueron una técnica ampliamente utilizada por los Estados Unidos de América y la Unión Soviética durante la Guerra Fría<sup>5</sup>. Aun así, la fácil difusión de contenido falso, intencional o no, hace que el proceso de manipulación sea más peligroso que nunca. Y la comunicación social es extremadamente importante en el sistema democrático para responsabilizar y criticar el sistema político, un poder que conlleva el deber de desmitificar información potencialmente falsa y ayudar a la población a discernir la verdad. La tecnología no es mala en sí misma; lo que puede causar daño es la forma en que la gente la utiliza.

El Foro Económico Mundial, en el *Informe de Riesgo Global 2024*<sup>6</sup>, identificó la desinformación y la misinformación como el mayor riesgo a corto plazo (dos años) para la humanidad, destacando tres aspectos:

1. En los próximos dos años, la misinformación y la desinformación podrían perturbar radicalmente los procesos electorales en varias economías.

2. Una creciente desconfianza en la información, así como en los medios de comunicación y en los gobiernos como fuentes, profundizará la polarización de las opiniones, un círculo vicioso que podrá desencadenar disturbios civiles y, posiblemente, confrontaciones.
3. Existe un riesgo de represión y erosión de los derechos a medida que las autoridades repriman la proliferación de información falsa, así como riesgos derivados de la inacción.

Sin embargo, las potencialidades de las nuevas tecnologías de la información y comunicación (TIC), impulsadas por la IA, también son vastas y pueden constituir un medio de inclusión, especialmente para las mujeres<sup>7-9</sup>. De hecho, el acceso a la educación se ha vuelto más sencillo gracias a la disponibilidad de una gama de recursos para expandir el conocimiento, desde cursos en línea hasta programas de enseñanza a distancia<sup>10</sup>. Ha aumentado también la posibilidad de empoderamiento económico mediante la creación y el desarrollo de negocios en plataformas digitales y mercados en línea, que generan nuevas oportunidades de empleo y facilitan la conciliación entre la vida profesional y personal, especialmente para las mujeres, que tradicionalmente asumen mayor responsabilidad en el cuidado del hogar y de los hijos. La flexibilidad en el trabajo —por ejemplo, con el trabajo remoto— les proporcionó un mayor equilibrio.

Además, las nuevas herramientas digitales permiten un mayor compromiso y participación política, especialmente en las plataformas de redes sociales, que tienen la capacidad de crear contenido viral, es decir, contenido que se difunde de forma masiva. En lo que se refiere a la salud y al bienestar, las TIC pueden contribuir significativamente a mejorar el acceso de las mujeres a la atención sanitaria y a brindarles un mayor cuidado para su bienestar.

Por supuesto, para que esto suceda, es necesario mitigar los efectos nocivos, ya sea del lado del usuario, ya sea del lado del propio creador de la tecnología. Mirando primero al destinatario del contenido, es necesario no dejar fuera a ninguna generación, desde los más jóvenes hasta los mayores. Es esencial educar sobre la IA, sus riesgos y potencialidades. Esto debería hacerse en la escuela, para concienciar a los jóvenes sobre la nueva realidad, pero también en formaciones dirigidas a las personas mayores, que pueden no tener tantas habilidades para manejar las nuevas tecnologías. La formación

a lo largo de la vida es esencial para proporcionar nuevas competencias y facilitar, así, la adaptación a la nueva realidad. Existen, por ejemplo, juegos en línea que estimulan de forma didáctica el uso de IA, como “bad news”, que *expone las tácticas y técnicas de manipulación que se utilizan para engañar a las personas y construir seguidores (...)* jugar crea resistencia cognitiva contra formas comunes de manipulación que el usuario puede encontrar en línea<sup>11</sup> y que es utilizado por algunos países en la Unión Europea para enseñar a los más jóvenes a detectar noticias falsas, las llamadas *fake news*.

En Portugal, se creó un proyecto piloto en las escuelas para promover la enseñanza de programación, robótica y otros temas relacionados con la inteligencia artificial<sup>12</sup>. En Brasil, ya hay escuelas que utilizan con éxito la IA en diversas actividades pedagógicas, pero las disparidades en el acceso a la tecnología entre escuelas con más y menos recursos (por ejemplo, en cuanto a infraestructura tecnológica) dificultan la proliferación de las potencialidades de la IA entre los estudiantes<sup>13</sup>.

En el caso específico de las mujeres, sigue existiendo una gran brecha en términos de existencia y participación en las áreas STEM (*science, technology, engineering and mathematics*), es decir, en los campos más relacionados con las tecnologías. Dado que ya existen desigualdades en la creación original de nuevas tecnologías, se espera que después los propios *softwares* presenten sesgos de género y no sean tan sensibles en este aspecto como podrían serlo si más mujeres participaran en su creación y desarrollo. Por lo tanto, es necesario atraer más mujeres a estas áreas científicas. De hecho, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco) ha estado advirtiendo sobre este hecho y ha elaborado una serie de recomendaciones para incorporar la perspectiva de género en la IA, especialmente para *establecer una comprensión básica de los imperativos de la igualdad de género y de cómo la IA, como sistema sociotécnico, puede reforzar o desafiar la desigualdad. (...) Hay que asegurar que la igualdad de género se entienda en términos de abordar los daños, aumentando la visibilidad de las implicaciones de género de la IA y fomentando las aplicaciones positivas de la IA y los impactos beneficiosos en el empoderamiento de las mujeres*<sup>14</sup>.

La IA generativa puede ser un instrumento dinámico para incluir a las mujeres en los campos Stem

debido a su aplicación interdisciplinaria, facilidad de uso y beneficios sociales; según las Naciones Unidas, *mediante políticas e iniciativas específicas destinadas a mejorar la participación y el liderazgo de las mujeres en la IA generativa, tales como programas de mentoría, financiación equitativa y promoción de reformas políticas inclusivas de género, se puede lograr un progreso significativo hacia el logro de los Objetivos de Desarrollo Sostenible 5 (ODS)*<sup>15</sup>, es decir, lograr la igualdad de género y empoderar a todas las mujeres y niñas<sup>16</sup>.

Aunque aquí nos centramos en las mujeres, hay que mencionar que otros grupos minoritarios y especialmente vulnerables pueden sufrir desigualdades en este ámbito. Como se ha demostrado anteriormente, todos los ciudadanos, independientemente de determinadas características, como el género, la etnia, la discapacidad, la edad o el lugar de origen, pueden contribuir de forma válida a la sociedad si se integran adecuadamente en el tejido social y productivo<sup>17</sup>. De hecho, esto está en consonancia con la doctrina internacional de derechos humanos que ha adoptado la mayoría de los países civilizados<sup>18,19</sup>.

Se destaca la enorme interdependencia entre los principios clásicos de la bioética —respeto a la autonomía, beneficencia, no maleficencia y justicia— y la implementación de una inteligencia artificial inclusiva y confiable. La bioética respeta la vida y el amor al fenómeno vital. Si la dignidad humana es el pilar de la bioética moderna, también debe enfatizarse la protección de otras formas de vida, incluido el medio ambiente y la biodiversidad. El concepto mismo de inclusión debe revisarse a la luz de estos principios, porque una sociedad solo puede ser inclusiva en el futuro si es económica y ambientalmente sostenible.

Por otra parte, la verdadera inclusión, es decir, la inclusión de todos en una sociedad moderna y avanzada, presupone la libertad individual. En la salud, el ejercicio de la autonomía se manifiesta, por ejemplo, en el consentimiento informado. La inclusión sin autonomía no tiene sentido, y la autonomía sin inclusión es inconsecuente. A su vez, la beneficencia y la correlativa no maleficencia son la base moral de la acción humana. La plena inclusión de las mujeres, de las personas mayores, de las personas con discapacidad, entre otros, en la sociedad debe surgir de buenas intenciones y motivaciones humanas y solidarias, y no solo de un contrato social. Ya sea que esto se

establezca por ley o por convención, siempre es de cariz más utilitario que benevolente.

Por su parte, el complejo principio bioético de justicia encuentra su expresión más amplia en la inclusión<sup>20</sup>. La ética de la justicia es la ética de la igualdad de oportunidades. Es decir, la igualdad en bioética implica asegurar a todos oportunidades sociales y económicas efectivas para desarrollar sus talentos individuales. De ello se desprende que la verdadera igualdad de oportunidades solo existe si el acceso a la salud, educación, formación profesional o protección social en la vejez y en la enfermedad va acompañado, en la sociedad digital contemporánea, de una auténtica inclusión digital. Desde un punto de vista bioético, la igualdad de oportunidades es un instrumento para la autorrealización personal; o, en la visión de Amartya Sen<sup>21</sup>, para el desarrollo de todo el potencial de la persona humana, de sus capacidades más avanzadas, que solo pueden expresarse en un entorno económico y social que promueva la inclusión.

Cuando se afirma que los sistemas de IA deben ser más inclusivos, se hace referencia a la inclusión de todas las personas, independientemente de género, clase social, etnia, tipo de discapacidad, entre otros, porque cualquier sistema de IA que no sea inclusivo puede ser bastante perjudicial, y *los impactos más graves incluyen daños materiales a los usuarios de estos sistemas, ya sea en forma de malas clasificaciones de crédito injustificadas, disminución de oportunidades de educación y empleo, diagnósticos médicos imprecisos o detenciones criminales injustificadas*<sup>22</sup>.

De hecho, se ha constatado que ciertos sistemas de IA, al basarse en datos que conllevan sesgos inherentes, aunque no intencionales, perpetúan e incluso intensifican las desigualdades existentes. Esto es lo que ocurrió, por ejemplo, en la empresa Amazon, que utilizó un sistema de IA para reclutar nuevos empleados con el fin de automatizar el proceso y preseleccionarlos, y terminó discriminando a las candidatas mujeres, porque la herramienta:

*fue creada con base en los estándares de currículum enviados a la empresa durante los últimos diez años. La gran mayoría de estos currículos era de hombres, como es el caso en la mayor parte de la industria de tecnología, por lo que los candidatos hombres eran considerados naturalmente más aptos para las vacantes. Una simple mención*

del término mujer en el currículum era penalizada por la herramienta y reducía las posibilidades de que las profesionales lograran el trabajo<sup>23</sup>.

En otras palabras, a medida que el sistema aprendió a hacer distinciones básicas sin considerar el por qué, la discriminación sentida en el pasado se magnificó en el presente. Por ello, el ser humano debe estar en el centro de los principios éticos de la IA e integrado en todas las etapas del proceso de creación y desarrollo de estos sistemas, para que exista un verdadero respeto a la persona humana, independientemente de sus características específicas.

## Aparición de legislación relativa a la inteligencia artificial

Mirando ahora al lado de los creadores de la tecnología en sí o de contenido —por ejemplo, de videos e imágenes generados por IA o de *chatbots* de conversación en la salud—, es necesario que sigan las directrices establecidas en los reglamentos nacionales e internacionales para que haya mayor transparencia y responsabilidad. A nivel de Unión Europea (UE), se creó la Regulación de la Inteligencia Artificial en el 2024 —la primera regulación internacional en este ámbito—, que se basa en el riesgo que determinado sistema de IA genera a las personas y a la sociedad. En otras palabras, si va en contra de los valores de la UE, este sistema no se permitirá. La nueva ley trajo consigo la obligación para las personas:

*que utilizan un sistema de IA para generar o manipular contenido de imagen, audio o video cuya similitud sustancial con personas, objetos, lugares, entidades o eventos reales podría llevar a una persona a creer, erróneamente, que son auténticos o verdaderos (falsificaciones profundas/deep fakes), también deberán revelar de forma clara y perceptible que el contenido ha sido creado o manipulado artificialmente, identificando los resultados de la IA como tales y revelando su origen artificial<sup>24</sup>.*

Tal regla es bastante relevante, dada la proliferación de *deep fakes*, con efectos increíblemente negativos en la vida de las personas por quienes se hacen pasar. Hay varias personas famosas, como artistas y políticos, que han sufrido este

tipo de ataques, y a menudo es sumamente difícil determinar si el contenido es real o ficticio.

Por eso, es importante que las propias plataformas tengan formas de detectar si el contenido ha sido creado de forma artificial, porque, si hay mala intención por parte del creador, no respetará ningún principio. Entonces, se está desarrollando un *software* que detecta contenido falso, pero también será necesario ir al ritmo de los avances que se están dando en el aspecto de creadores de contenido, es decir, las tecnologías de ambos lados tienen que ir paso a paso. Sorprendentemente, algunas falsificaciones profundas/*deep fakes* ocurren en la medicina; hay casos documentados de alteración deliberada de imágenes de TAC o de resonancia magnética con un impacto real en la salud y en la vida de las personas.

En Brasil, se aprobó en el 2024 el Proyecto de Ley 2.338 del 2023<sup>25</sup>, que regula la IA y también prevé una evaluación basada en riesgos de los sistemas de IA. Este documento establece que:

*el ser humano como principio “de todas las cosas”, en la centralidad de las decisiones (e) prohíbe la evaluación de rasgos de personalidad y características de conducta para predecir la comisión de delitos. E impide la clasificación de individuos en función de su conducta social o personalidad para determinar, de forma ilegítima y desproporcionada, el acceso a bienes, servicios y políticas públicas<sup>26</sup>.*

Respecto del contenido generado artificialmente, es bastante claro respecto a la no instigación a hacer daño y a la verificación de su autenticidad y procedencia.

Las consecuencias de no regular sistemas de IA son inmensas, y los ciberdelitos que han ocurrido en los últimos tiempos son prueba de ello, especialmente aquellos contra mujeres, con una mayor incidencia de *violencia sexual basada en imágenes, coloquialmente conocida como pornografía de venganza*<sup>27</sup>. Este es el caso del *deep fake* de cariz sexual, que, dependiendo del país y de las relaciones de género que allí existan (por ejemplo, si son más o menos conservadoras), puede tener repercusiones desastrosas para las mujeres, como la pérdida del empleo, el rechazo de la familia y posteriores problemas de salud mental.

La Directiva de la Unión Europea (UE) 1.385/2024<sup>28</sup>, relativa a la lucha contra la violencia hacia las mujeres y la violencia doméstica,

criminaliza precisamente este tipo de delito digital. En Brasil, en el 2024, se aprobó un proyecto de ley para sancionar la violencia digital contra las mujeres, que quedará incluida en la Ley Maria da Penha —Ley 11.340/2006<sup>29</sup>—, que tiene como objetivo eliminar la violencia doméstica y familiar contra la mujer<sup>30</sup>. Los avances en materia legislativa demuestran la importancia de mantenerse al día con el progreso tecnológico. Por lo tanto, es necesario que las leyes se implementen y se supervisen de manera adecuada para identificar posibles violaciones.

Para intentar densificar los principios éticos aplicados específicamente en la IA fiable, la Comisión Europea creó, en el 2019, el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial<sup>31</sup>, que establece siete criterios éticos esenciales:

1. agencia y supervisión humanas;
2. robustez técnica y seguridad;
3. privacidad y buena gobernanza de datos;
4. transparencia;
5. diversidad, no discriminación y justicia;
6. bienestar social y ambiental;
7. *accountability* (responsabilidad).

Los criterios, que son de aplicación general, pretenden asegurar que la IA se diseñe y se aplique de manera que promueva la diversidad, la no discriminación y la distribución justa de las cargas y los beneficios de su uso. Además, la Organización Mundial de la Salud (OMS) ha abordado la importancia de una gobernanza ética e inclusiva de la inteligencia artificial<sup>32</sup>, ahora específicamente en la salud, destacando la necesidad de:

1. proteger la autonomía;
2. promover el bienestar humano, la seguridad humana y el interés público;
3. asegurar la transparencia, la explicabilidad y la inteligibilidad;
4. promover la responsabilidad y la rendición de cuentas;
5. asegurar la inclusión y la equidad;
6. promover una inteligencia artificial responsiva y sostenible.

En el 2024 se creó el primer tratado internacional a nivel del Consejo de Europa: el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, que es jurídicamente vinculante y está abierto a países fuera de Europa. Establece los

principios éticos que todos los sistemas de IA deben seguir a lo largo de su ciclo de vida<sup>33</sup>:

1. dignidad humana y autonomía individual;
2. igualdad y no discriminación;
3. respeto a la privacidad y protección de datos personales;
4. transparencia y supervisión;
5. responsabilidad y rendición de cuentas;
6. confiabilidad;
7. innovación segura.

Se percibe que existe una fundada preocupación de diferentes organismos internacionales por asegurar la fiabilidad de la IA y el control y supervisión humanos, así como por asegurar la diversidad y la inclusión promoviendo el acceso y la alfabetización digital de grupos vulnerables, tradicionalmente excluidos, o al menos desatendidos, ahora con el agravante del factor generacional. En otras palabras, es importante combatir el “dividendo generacional” desde ahora incluyendo a todas las generaciones y evitando las interseccionalidades que agravan la exclusión.

Sin embargo, cada vez está más extendida la idea de que la simple enunciación de estos y otros principios no es suficiente<sup>34</sup>. Es necesario, por una parte, promover campañas intensas de educación y formación en IA y otras competencias digitales en todos los rangos de edad y estratos sociales y, por otra parte, implementar regulación específica, nacional e internacional, incluida legislación que proporcione un marco ético para la inteligencia artificial. Por ejemplo, garantizando la “igualdad epistémica”, es decir, que los sistemas de IA no se desarrollarán de manera que privilegien ciertas perspectivas éticas o sistemas de conocimiento en detrimento de otros.

También debe haber una evolución constante de la legislación y de las políticas públicas para seguir el ritmo de los avances tecnológicos y asegurar que no haya discrepancias entre el progreso alcanzado, con los peligros potenciales que de él se derivan, y las medidas que protegen a los seres humanos. Sancionar los actos discriminatorios y los delitos digitales es esencial para frenar la conducta inhumana y excesiva resultante del uso de sistemas de IA.

Estas tecnologías también deben verse desde una perspectiva positiva, es decir, por su contribución a la lucha contra la desigualdad y la exclusión. Por ejemplo, las personas con discapacidad pueden y deben encontrar en este conjunto ampliado de

innovaciones tecnológicas —que incluye la IA o la síntesis biomecánica— las herramientas necesarias para integrarse en el *mainstream* social<sup>35</sup>. De hecho, las tecnologías siempre han sido un recurso esencial para superar una serie de deficiencias y permitir así una mayor y mejor expresión de la identidad personal y de la autonomía individual. Los dispositivos mecánicos o electrónicos permiten superar discapacidades orgánicas o funcionales con expresión interna o externa al organismo humano, pero con gran valor terapéutico y enorme inclusión social<sup>36</sup>.

En un contexto médico o de rehabilitación, la IA asociada a este tipo de tecnología podrá ser un factor promotor de la inclusión y la igualdad<sup>37</sup>. No se trata necesariamente de alcanzar un futuro poshumano desde hace mucho deseado por la humanidad, sino de generalizar y democratizar el acceso a evoluciones tecnológicas modernas, como asistentes digitales, *software* de voz a texto, avatares de lenguaje de señas, miembros protéticos avanzados, exoesqueletos sofisticados o apoyo a la salud mental<sup>38</sup>. Todas estas evoluciones promueven la igualdad y la inclusión mediante una simbiosis única y un diseño innovador, y permiten una nueva *performance* individual al mitigar diferentes discapacidades.

Sin embargo, cabe señalar que, al hablar de inclusión e igualdad, existen importantes riesgos interseccionales que no solo agravan, sino que también potencian las desigualdades<sup>39</sup>, que puede ser potenciados por la IA si no se detectan y se corrigen. Es decir, la combinación de desigualdades ya existentes basadas en el género, la etnia o la discapacidad puede intensificar vulnerabilidades previamente existentes<sup>40</sup>. En síntesis, y como señala Francis Fukuyama, *los seres humanos han sido preparados por el proceso evolutivo para ser criaturas sociales que buscan naturalmente integrarse en una atmósfera de relaciones comunitarias*<sup>41</sup>.

La IA debería servir para profundizar la solidaridad y la inclusión social, no lo contrario. En la medicina y en la salud, donde la IA es especialmente sensible, se deben tener en cuenta, de forma sistemática, seis principios éticos adicionales para que la inteligencia artificial se considere fiable y, sobre todo, para que esté bajo supervisión humana y en línea con los Objetivos de Desarrollo Sostenible: justicia, universalidad, trazabilidad, usabilidad, robustez y explicabilidad<sup>42,43</sup>. Solo así se podrá garantizar el legítimo derecho de todos a un futuro abierto<sup>44</sup>.

## Consideraciones finales

Las potencialidades de la inteligencia artificial son inmensas, pero, para aprovecharlas en beneficio de la humanidad y evitar que se convierta en riesgos previsible, se necesita una estrategia concertada para atacar múltiples frentes, desde el lado del diseño hasta el lado del usuario, y hacer de la IA un verdadero motor de la inclusión general en la sociedad digital y del acceso a la salud. Sin embargo, el desafío de crear mecanismos específicos es complejo y requiere la colaboración de varios agentes, sobre todo en la medicina, desde entidades gubernamentales hasta empresas privadas. Los programas de formación, el acceso a recursos digitales para todas las personas, la implementación y el seguimiento de la legislación son ejemplos de medidas que deben implementarse para evitar exacerbar la desigualdad en el nuevo mundo digital, incluido el acceso a la salud.

De hecho, la IA tiene el potencial de promover el desarrollo económico y el bienestar social y de traducirse en el pleno desarrollo humano de todos los ciudadanos, en todas las sociedades. Para ello, deben respetarse los principios universales de la bioética, especialmente en el diseño e implementación de los diversos sistemas de inteligencia artificial. La presencia humana (*human in the loop*) en todos los niveles de implementación de la IA es la garantía del respeto a la dignidad humana como pilar central de la bioética contemporánea, así como de los principios irrenunciables de igualdad, no discriminación, responsabilidad y respeto a la privacidad, al igual que una autonomía que contribuya a la autorrealización de las personas, y no a su alienación. La igualdad significa que todos deben tener acceso a la inclusión digital, sin discriminación basada en edad, género, ingresos, discapacidad u otra condición. La autonomía implica que la IA refuerce la libertad individual, especialmente en la salud, permitiendo tomar decisiones mejor informadas.

En materia de salud, el ejercicio del derecho universal al acceso a servicios de calidad requiere que las principales amenazas éticas y sociales identificadas por el uso generalizado de la IA —como sesgos algorítmicos, exclusión digital, riesgos interseccionales, desinformación y violaciones sistemáticas de la privacidad— se detecten instantáneamente para asegurar que el derecho a la salud se haga realidad y

que la IA contribuya a su disfrute efectivo. El uso de inteligencia artificial para optimizar la gestión hospitalaria y, así, traducirse en ganancias de eficiencia en la asignación de recursos podría ser otra forma de mejorar el acceso a la salud y contribuir a la inclusión.

Sin embargo, todas estas virtualidades de IA carecen de un marco regulatorio nacional e internacional para armonizar las prácticas y contribuir al bienestar de los ciudadanos. La Unión Europea, mediante el Reglamento de Inteligencia Artificial, aprobado en el 2024, pretende asegurar que no se vulneren los principios éticos y bioéticos. Para ello, el riesgo en el uso de la IA se estratifica en niveles de acción, y la privacidad individual y la protección de datos personales merecen especial énfasis en este instrumento legal. Brasil sigue la misma línea de acción por medio del Proyecto de Ley 2.338/2023, actualmente en análisis en el Congreso, que establece directrices éticas para el desarrollo y aplicación de la IA. Estos y otros instrumentos jurídicos tienen como objetivo

salvaguardar no solo los principios bioéticos tradicionales, sino también principios específicos de una IA confiable, como la justicia, la universalidad, la trazabilidad, la usabilidad, la robustez, la responsabilidad y la necesaria explicabilidad de cómo la IA toma decisiones y llega a conclusiones.

En resumen, para que la IA se considere confiable y un instrumento para el pleno desarrollo humano, se deben abordar varios ejes. Por ejemplo, debe considerarse que la IA tiene efectos en todos los dominios de la actividad humana e influye en todos los Objetivos de Desarrollo Sostenible, que están estrechamente relacionados con los principios bioéticos; o que solo con la colaboración de todos los agentes implicados la IA puede ser realmente una herramienta de ayuda a los seres humanos; y, finalmente, que la igualdad como valor bioético intrínseco puede y debe alcanzarse, de manera que cada individuo, utilizando o no la IA, desarrolle sus capacidades y alcance su máximo potencial, logrando así la máxima autorrealización personal.

## Referencias

1. Newfield C. How to make "AI" intelligent; or, the question of epistemic equality. *Critical AI* [Internet]. 2023 [acceso 11 ago 2025];1(1-2). DOI: 10.1215/2834703X-10734076
2. Quase 40% dos empregos em todo o mundo serão afetados pela IA, alerta diretora do FMI. *Euronews* [Internet]. Trabalho; 15 jan 2024 [acceso 11 ago 2025]. Disponível: <https://bit.ly/42h1UIZ>
3. Will generative AI disproportionately affect the jobs of women?. *Kenan Institute of Private Enterprise* [Internet]. Kenan Insight; 18 abr 2023 [acceso 11 ago 2025]. Disponível: <https://bit.ly/4mJyW5z>
4. Lee H, Sarkar A, Tankelevitch L, Drosos I, Rintel S, Banks R, Wilson N. The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers [Internet]. In: CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems; 26 abr-1 maio; Yokohama, Japão. New York: Association for Computing Machinery; 2025 [acceso 11 ago 2025]. Tradução livre. DOI: 10.1145/3706598.3713778
5. Ward M, Pierson S, Beyer J. Formative battles: Cold War disinformation campaigns and mitigation strategies [Internet]. Washington, DC: Wilson Center; 2019 [acceso 11 ago 2025]. Disponível: <https://bit.ly/4mCi1BL>
6. World Economic Forum. The Global Risks Report 2024 [Internet]. 19ª ed. Geneva: WEF; 2024 [acceso 11 ago 2025]. Disponível: <https://bit.ly/46uQe1w>
7. World Economic Forum. Global Gender Gap Report 2022 [Internet]. Geneva: WEF; 2022 [acceso 11 ago 2025]. Disponível: <https://bit.ly/4mKgWYT>
8. Estratégia de igualdade de gênero [Internet]. Brussels: European Commission; 2020 [acceso 11 ago 2025]. Disponível: <https://bit.ly/4nVKPX0>
9. Unesco. Unesco Priority Gender Equality Action Plan: 2014-2021 [Internet]. Paris: Unesco; 2014 [acceso 11 ago 2025]. Disponível: <https://bit.ly/3Il2RTC>
10. Nunes S. A União Europeia como guardiã dos direitos das mulheres. In: Nunes R, Rego G, Melo H, Nunes S, Duarte I, coordenadores. *Perspectives in gender equality: making education the core of public policies*. Basel: MDPI Books; 2025.

11. Bad News [Internet]. Groningen: Tilt; 2021 [acesso 11 ago 2025]. Tradução livre. Disponível: <https://www.getbadnews.com/en>
12. Governo vai lançar projeto-piloto para ensino de inteligência artificial e programação. Jornal de Notícias [Internet]. Inovação; 12 dez 2024 [acesso 11 ago 2025]. Disponível: <https://bit.ly/42iwsDX>
13. Costa AG. Escolas do Brasil já usam inteligência artificial em aulas, mas há desafios de inclusão. Terra [Internet]. Byte; 30 maio 2024 [acesso 11 ago 2025]. Disponível: <https://bit.ly/42jDkkm>
14. Unesco. Artificial intelligence and gender equality: key findings of Unesco's Global Dialogue [Internet]. Paris: Unesco; 2020 [acesso 11 ago 2025]. Tradução livre. Disponível: <https://bit.ly/4nE0f2c>
15. Nedungadi P, Ramesh M, Govindaraju V, Rao B, Berbeglia P, Raman R. Emerging leaders or persistent gaps? Generative AI research may foster women in STEM. *Int J Inf Manage* [Internet]. 2024 [acesso 11 ago 2025];77. Tradução livre. DOI: 10.1016/j.ijinfomgt.2024.102785
16. United Nations. Convention on the Elimination of All Forms of Discrimination against Women New York, 18 December 1979 [Internet]. Geneva: OHCHR; 1979 [acesso 11 ago 2025]. Disponível: <https://bit.ly/42S23MP>
17. Nunes S. Migration and economic development: the unintended consequences for technological innovation and, thus, the development of societies arising from barriers to high-skilled immigrants [dissertação]. London: University College London; 2018.
18. United Nations. United Nations Charter [Internet]. Geneva: United Nations; 1945 [acesso 11 ago 2025]. Disponível: <https://bit.ly/3Koz9xD>
19. United Nations. Universal Declaration of Human Rights [Internet]. Geneva: United Nations; 1948 [acesso 11 ago 2025]. Disponível: <https://bit.ly/46Z1QtE>
20. Nunes R. Healthcare as a universal human right: sustainability in global health. New York: Routledge; 2022.
21. Sen A. The idea of justice. Cambridge: Harvard University Press; 2009.
22. Zowghi D, Rimini F. Diversity and inclusion in artificial intelligence In: Lu Q, Zhu L, Whittle J, Xu X, editores. Responsible AI: best practices for creating trustworthy AI systems. Boston: Addison-Wesley Professional; 2023. p. 216.
23. Valeri J. Inteligência artificial utiliza base de dados que refletem preconceitos e desigualdades. Jornal da USP [Internet]. Atualidades; 7 jul 2023 [acesso 11 ago 2025]. Disponível: <https://bit.ly/4mI8hWC>
24. União Europeia. Regulamento 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024. Cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento da Inteligência Artificial). Jornal Oficial da União Europeia [Internet]. Bruxelas, 12 jul 2024 [acesso 11 ago 2025]. Disponível: <https://bit.ly/4n0qT4B>
25. Projeto de Lei n° 2.338, de 2023. Dispõe sobre o uso da inteligência artificial [Internet]. Brasília: Senado Federal; 2023 [acesso 11 ago 2025]. Disponível: <https://bit.ly/474LDn3>
26. Senado aprova regulamentação da inteligência artificial; texto vai à Câmara. Senado Notícias [Internet]. 2024 [acesso 11 ago 2025]. Disponível: <https://bit.ly/4gKQ6ya>
27. Venema AE. Deepfakes as a security issue: why gender matters. *Women in International Security* [Internet]. 2025 [acesso 11 ago 2025]. Disponível: <https://bit.ly/4nlu73o>
28. União Europeia. Diretiva 2024/1385, de 14 de maio de 2024. Relativa ao combate à violência contra as mulheres e à violência doméstica. Jornal Oficial da União Europeia [Internet]. Bruxelas, 24 maio 2024 [acesso 11 ago 2025]. Disponível: <https://bit.ly/3lwYXXL>
29. Brasil. Lei n° 11.340, de 7 de agosto de 2006. Lei Maria da Penha. Diário Oficial da União [Internet]. Brasília, 8 ago 2006 [acesso 11 ago 2025]. Disponível: <https://bit.ly/3IMcNWg>
30. Projeto aprovado pune uso de inteligência artificial para criar material prejudicial a mulheres. Portal da Câmara dos Deputados [Internet]. 2024 [acesso 11 ago 2025]. Disponível: <https://bit.ly/3VDwx1c>
31. High-Level Expert Group on AI. Ethics guidelines for trustworthy AI [Internet]. Brussels: European Commission; 2019 [acesso 11 ago 2025]. Disponível: <https://bit.ly/3KLgJqM>

32. World Health Organization. Ethics & Governance of Artificial Intelligence for Health [Internet]. Geneva: WHO; 2021 [acceso 15 ago 2025]. Disponible: <https://bit.ly/3lwZ8Cp>
33. Council of Europe. Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law [Internet]. Strasbourg: Council of Europe; 2025 [acceso 11 ago 2025]. Disponible: <https://bit.ly/4nWaWx5>
34. Mittelstadt BD. Principles alone cannot guarantee ethical AI. *Nat Mach Intell* [Internet]. 2019 [acceso 11 ago 2025];1(11):501-7. Disponible: <https://bit.ly/42kUZrP>
35. Romanska M. The bionic body: disability, technology and posthumanism. *Body Space Technol* [Internet]. 2024 [acceso 11 ago 2025];23(1). DOI: 10.16995/bst.11480
36. Murray S. Disability and the posthuman: bodies, technology, and cultural futures. Liverpool: Liverpool University Press; 2020.
37. Shildrick M. Disability, digital technologies and the ambivalent allure of posthumanist/transhumanist futures. *Med Humanit* [Internet]. 2024 [acceso 11 ago 2025];50:685-93. DOI: 10.1136/medhum-2024-013020
38. Hayles NK. How we became posthuman: virtual bodies in cybernetics, literature, and informatics. Chicago: University of Chicago Press; 1999.
39. Crenshaw K. On intersectionality: essential writings. New York: New Press; 2019.
40. Verloo M. Multiple inequalities, intersectionality and the European Union. *Eur J Womens Stud* [Internet]. 2006 [acceso 11 ago 2025];13(3):193-210. DOI: 10.1177/1350506806065753
41. Fukuyama F. O nosso futuro pós-humano: consequências da revolução biotecnológica. Lisboa: Quetzal; 2002. p. 193.
42. Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas C, Langlotz CP *et al*. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* [Internet]. 2025 [acceso 11 ago 2025];388:e081554. DOI: 10.1136/bmj-2024-081554
43. Nunes R, Nunes S. Reliable artificial intelligence: the 18th sustainable development goal. *J Ethics Leg Technol* [Internet]. 2024 [acceso 11 ago 2025];6(2):1-19. Disponible: <https://bit.ly/4nXw9qp>
44. Feinberg J. The child's right to an open future. In: Aiken W, LaFollette H, editores. *Whose child?: children's rights, parental authority and state power*. Totowa: Littlefield, Adams & Co; 1980.

Sofia B. Nunes – Estudiante de doctorado – [asnunes@med.up.pt](mailto:asnunes@med.up.pt)

 0000-0002-7182-9890

Rui Nunes – Doctor – [ruinunes@med.up.pt](mailto:ruinunes@med.up.pt)

 0000-0002-1377-9899

#### Correspondencia

Sofia B. Nunes – Al. Prof. Hernâni Monteiro. CEP 4200-319. Porto, Portugal.

#### Contribución de los autores (CRediT)

Ambos autores contribuyeron por igual a la concepción, elaboración, redacción y revisión crítica del manuscrito, aprobaron la versión final y asumen la responsabilidad de la integridad del contenido.

**Disponibilidad de los datos:** Todos los datos utilizados o generados en la investigación se describen y presentan íntegramente en el cuerpo del artículo.

**Editora responsable:** Dilza Teresinha Ambrós Ribeiro

**Recibido:** 20.3.2025

**Revisado:** 18.8.2025

**Aprobado:** 19.8.2025