



Organización  
de las Naciones Unidas  
para la Educación,  
la Ciencia y la Cultura

## Grupo Especial de Expertos (GEE) encargado de elaborar un proyecto de recomendación sobre la ética de la inteligencia artificial

Distribución: limitada

SHS/BIO/AHEG-AI/2020/4  
París, 7 de mayo de 2020  
Original: inglés

### DOCUMENTO FINAL

#### PRIMERA VERSIÓN DEL PROYECTO DE RECOMENDACIÓN SOBRE LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

De conformidad con la decisión adoptada por la Conferencia General de la UNESCO en su 40ª reunión ([resolución 40 C/37](#)), la Directora General constituyó, en marzo de 2020, el Grupo Especial de Expertos (GEE) encargado de elaborar un proyecto de recomendación sobre la ética de la inteligencia artificial.

Adaptándose a la difícil situación causada por la pandemia de COVID-19, el GEE trabajó virtualmente desde finales de marzo hasta principios de mayo de 2020 y elaboró la primera versión de un proyecto de recomendación sobre la ética de la inteligencia artificial, que figura en el presente documento.

Hay que subrayar que el GEE seguirá revisando esta primera versión del proyecto de recomendación hasta principios de septiembre de 2020, teniendo en cuenta las observaciones recibidas durante el proceso de consulta entre múltiples interesados que se celebrará de junio a julio de 2020.

El presente documento no pretende ser exhaustivo ni representa necesariamente las opiniones de los Estados Miembros de la UNESCO.

## PRIMERA VERSIÓN DEL PROYECTO DE RECOMENDACIÓN SOBRE LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

### PREÁMBULO

La Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), reunida en París del xxx al xxx, en su xx reunión,

**Recordando** que, con arreglo a su Constitución, la UNESCO aspira a erigir los baluartes de la paz en la mente de los seres humanos y tiene por objeto estrechar, mediante la educación, la ciencia, la cultura y la comunicación y la información, la colaboración entre las naciones, a fin de asegurar el respeto universal a la justicia, a la ley, a los derechos humanos y a las libertades fundamentales que se reconocen a todos los pueblos del mundo,

**Reflexionando** sobre la profunda influencia que la inteligencia artificial (IA) puede tener en las sociedades, los ecosistemas y las vidas humanas, en particular en la mente humana, debido en parte a las nuevas formas en que influye en el pensamiento humano y en la adopción de decisiones y afecta a la educación, la ciencia, la cultura y la comunicación y la información,

**Considerando** que los sistemas de IA pueden ser de gran utilidad para la humanidad, pero también suscitan preocupaciones éticas fundamentales, por ejemplo, en relación con los sesgos que pueden incorporar y exacerbar, lo que puede llegar a provocar desigualdad, exclusión y una amenaza para la diversidad cultural y social y la igualdad de género; la necesidad de transparencia e inteligibilidad del funcionamiento de los algoritmos y los datos con los que han sido formados; y su posible impacto en la privacidad, la libertad de expresión, los procesos sociales, económicos y políticos y el medio ambiente,

**Reconociendo** que el desarrollo de la IA puede profundizar las divisiones y desigualdades existentes en el mundo y que no se debería dejar a nadie atrás contra su voluntad, ya sea en el disfrute de los beneficios de la IA o en la protección contra sus consecuencias negativas, reconociendo al mismo tiempo las diferentes circunstancias de los distintos países,

**Consciente** de que los países de ingreso mediano bajo, incluidos, entre otros, los de África, América Latina y el Caribe y Asia Central, así como los pequeños Estados insulares en desarrollo, se enfrentan a una aceleración del uso de las tecnologías de la información y la IA, y de que la economía digital presenta importantes desafíos sociales y ofrece oportunidades para las sociedades creativas, que requieren el reconocimiento de las culturas, los valores y los conocimientos endógenos a fin de desarrollar las economías,

**Reconociendo** que la IA puede ser beneficiosa para el medio ambiente, a través de las funciones que desempeña en la investigación ecológica y climática, la gestión del riesgo de desastres y la agricultura, pero que para que esos beneficios se materialicen es necesario asegurar un acceso justo a la tecnología y encontrar un equilibrio entre los posibles beneficios y el impacto ambiental de todo el ciclo de producción de la IA y la tecnología de la información,

**Observando** que el hecho de tener en cuenta los riesgos y las preocupaciones éticas no debería obstaculizar la innovación, sino más bien estimular nuevas prácticas de investigación e innovación responsables en las que la investigación, el diseño, el desarrollo, el despliegue y la utilización de la IA se fundamenten en valores morales y en la reflexión ética,

**Recordando** que en noviembre de 2019 aprobó, en su 40ª reunión, la resolución 40 C/37, en la que encargó a la Directora General que “[preparara] un instrumento normativo internacional sobre la ética de la inteligencia artificial (IA) en forma de recomendación”, que se le presentaría en su 41ª reunión en 2021,

**Convencida** de que el instrumento normativo que aquí se presenta, basado en un enfoque normativo mundial y centrado en la dignidad humana y los derechos humanos, incluidas la diversidad, la interconexión, la inclusión y la equidad, puede orientar la investigación, el diseño, el desarrollo, el despliegue y la utilización de la IA en una dirección responsable,

**Observando** que el marco normativo sobre la IA y sus implicaciones sociales se encuentra en la intersección de la ética, los derechos humanos, los marcos jurídicos internacionales y nacionales, la libertad de investigación e innovación y el bienestar humano,

**Reconociendo** que los valores y principios éticos no son necesariamente normas jurídicas en sí mismos, pero pueden influir poderosamente en la elaboración y aplicación de medidas en materia de políticas y de normas jurídicas, proporcionando orientación cuando el ámbito de las normas no esté claro o cuando esas normas no se hayan aún establecido debido a la rapidez del desarrollo tecnológico combinada con el ritmo relativamente más lento de las respuestas políticas,

**Convencida** de que las normas éticas aceptadas mundialmente pueden desempeñar una función útil en la armonización de las normas jurídicas relacionadas con la IA en todo el mundo, así como en la aplicación responsable del derecho internacional existente, si esa aplicación se ajusta a los marcos éticos y no causa perjuicios a nivel local,

**Reconociendo** la Declaración Universal de Derechos Humanos (1948), en particular su artículo 27, que hace hincapié en el derecho a participar en el progreso científico y en los beneficios que de él resulten; los instrumentos del marco internacional de derechos humanos, entre ellos la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (Naciones Unidas, 1979), la Convención sobre los Derechos del Niño (Naciones Unidas, 1989) y la Convención sobre los Derechos de las Personas con Discapacidad (Naciones Unidas, 2006); y la Convención sobre la Protección y Promoción de la Diversidad de las Expresiones Culturales (UNESCO, 2005),

**Tomando nota** de la Declaración sobre las Responsabilidades de las Generaciones Actuales para con las Generaciones Futuras (UNESCO, 1997); de la Declaración de las Naciones Unidas sobre los Derechos de los Pueblos Indígenas (2007); del informe de 2011 del Secretario General de las Naciones Unidas sobre el seguimiento de la Segunda Asamblea Mundial sobre el Envejecimiento (A/66/173), centrado en la situación de los derechos humanos de las personas de edad; del informe de 2011 del Representante Especial del Secretario General para la cuestión de los derechos humanos y las empresas transnacionales y otras empresas (A/HRC/17/31), en el que se esbozan los “Principios Rectores sobre las empresas y los derechos humanos: puesta en práctica del marco de las Naciones Unidas para ‘proteger, respetar y remediar’”; de la resolución del Consejo de Derechos Humanos sobre “El derecho a la privacidad en la era digital” (A/HRC/RES/42/15), aprobada el 26 de septiembre de 2019; de la Recomendación sobre la Ciencia y los Investigadores Científicos (UNESCO, 2017); de los indicadores de la UNESCO sobre la universalidad de Internet (2019), incluidos los principios ROAM; del informe del Panel de Alto Nivel del Secretario General de las Naciones Unidas sobre la Cooperación Digital titulado “La era de la interdependencia digital” (2019); y de los resultados e informes de las cumbres mundiales de la UIT sobre la “IA para el bien”,

**Tomando nota también** de los marcos existentes relacionados con la ética de la IA elaborados por otras organizaciones intergubernamentales, como los instrumentos pertinentes de derechos humanos y otros instrumentos jurídicos aprobados por el Consejo de Europa, y de la labor de su Comité Especial sobre IA (CAHAI); de los trabajos de la Unión Europea relacionados con la IA y del Grupo de expertos de alto nivel sobre IA de la Comisión Europea, incluidas las Directrices Éticas para una IA Fiable; de la labor del Grupo de expertos sobre IA (AIGO) de la Organización de Cooperación y Desarrollo Económicos (OCDE) y la

Recomendación del Consejo de la OCDE sobre IA; de los Principios relativos a la IA del G20, inspirados en los trabajos de la OCDE y esbozados en la Declaración Ministerial del G20 sobre Comercio y Economía Digital; de la Visión Común de Charlevoix sobre el Futuro de la IA, aprobada por el G7; de la labor del Grupo de Trabajo de la Unión Africana sobre la IA; y de la labor del Grupo de Trabajo de la Liga de los Estados Árabes sobre la IA,

**Haciendo hincapié** en que debe prestarse atención específica a los países de ingreso mediano bajo, incluidos, entre otros, los de África, América Latina y el Caribe y Asia Central, así como los pequeños Estados insulares en desarrollo, ya que han estado insuficientemente representados en el debate sobre la ética de la IA, lo que suscita la preocupación de que se descuiden los conocimientos locales, el pluralismo cultural y ético, los sistemas de valores y las exigencias de equidad mundial,

**Consciente** de los numerosos marcos nacionales relacionados con la ética y la reglamentación de la IA,

**Consciente también** de las numerosas iniciativas y marcos relacionados con la ética de la IA desarrollados por el sector privado, las organizaciones profesionales y las organizaciones no gubernamentales, como la Iniciativa Mundial sobre la Ética de los Sistemas Inteligentes y Autónomos del Instituto de Ingenieros Electricistas y Electrónicos (IIEE) y sus trabajos sobre el diseño conforme a la ética; la “Gobernanza Mundial de la Tecnología: un enfoque de múltiples interesados” del Foro Económico Mundial; los “Diez Principios Fundamentales de la Ética de la Inteligencia Artificial” de UNI Global Union; la Declaración de Montreal para el Desarrollo Responsable de la IA; los Principios de la Inteligencia Artificial Armoniosa (HAIP); y los Principios de la Asociación sobre la IA,

**Convencida** de que la IA puede aportar importantes beneficios, pero que su consecución también puede ir acompañada de una deuda de la innovación, un acceso asimétrico a los conocimientos, barreras al derecho a la información, lagunas en la capacidad de creatividad de los ciclos de desarrollo y las capacidades humanas e institucionales, obstáculos al acceso a la innovación tecnológica y una falta de infraestructura y reglamentación adecuadas en materia de datos,

**Reconociendo** que se produce una competencia económica dentro de los Estados y entre ellos, así como entre empresas multinacionales, lo que puede hacer que las estrategias y los marcos reguladores de la IA se centren en los intereses nacionales y comerciales, mientras que se necesita una cooperación mundial para hacer frente a los desafíos que conlleva la IA en culturas y sistemas éticos diversos e interconectados y para reducir la posibilidad de usos indebidos,

**Teniendo plenamente en cuenta** que el rápido desarrollo de los sistemas de IA encuentra obstáculos para comprender y aplicar la IA, debido a la diversidad de orientaciones éticas y culturas en todo el mundo, la falta de agilidad de la ley en relación con la tecnología y la sociedad de la información, y el riesgo de que las normas y valores éticos locales y regionales se vean perturbados por la IA,

1. **Aprueba** la presente Recomendación sobre la Ética de la Inteligencia Artificial;
2. **Recomienda** que los Estados Miembros apliquen las disposiciones de la presente Recomendación adoptando las medidas adecuadas, en particular las medidas legislativas o de otra índole que puedan ser necesarias, de conformidad con la práctica constitucional y las estructuras de gobierno de cada Estado, con el fin de dar efecto en sus respectivas jurisdicciones a los principios y normas enunciados en la Recomendación;

3. **Recomienda también** que los Estados Miembros señalen la Recomendación a la atención de las autoridades, órganos, instituciones y organizaciones de los sectores público, comercial y no comercial que participan en la investigación, el diseño, el desarrollo, el despliegue y la utilización de sistemas de IA.

## I. ÁMBITO DE APLICACIÓN

1. En la presente Recomendación se tratan las cuestiones éticas relacionadas con la IA. Se aborda la ética de la IA como un marco integral de valores, principios y acciones interdependientes que puede guiar a las sociedades en el ciclo de vida de los sistemas de IA, tomando la dignidad y el bienestar humanos como brújula para afrontar de manera responsable los efectos conocidos y desconocidos de los sistemas de IA en sus interacciones con los seres humanos y su entorno. El ciclo de vida de un sistema de IA abarca la investigación, el diseño, el desarrollo, el despliegue y la utilización de ese sistema, y se puede entender que la utilización de un sistema de IA incluye el mantenimiento, funcionamiento, uso final y desmontaje de dicho sistema. El presente instrumento no pretende proporcionar una única definición de la IA, ya que tal definición tendría que cambiar con el tiempo en función de los avances tecnológicos. Su objetivo es más bien abordar las características de los sistemas de IA que tienen una importancia ética central y sobre las que existe un gran consenso internacional. A los efectos de la presente Recomendación, los sistemas de IA pueden considerarse sistemas tecnológicos que tienen la capacidad de procesar información de una manera que se asemeja a un comportamiento inteligente, y suelen incluir funciones de aprendizaje, percepción, predicción, planificación o control. La presente Recomendación aborda los sistemas de IA de acuerdo con las siguientes líneas:

- a) en primer lugar, los sistemas de IA representan modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, como formular recomendaciones y tomar decisiones en entornos reales y virtuales. Los sistemas de IA están diseñados para funcionar con diversos niveles de autonomía mediante la modelización y la representación del conocimiento y la explotación de datos y el cálculo de correlaciones. Pueden incluir varios enfoques y tecnologías, como, por ejemplo, aunque no exclusivamente:
  - i) el aprendizaje automático, incluido el aprendizaje profundo y el aprendizaje de refuerzo;
  - ii) el razonamiento automático, incluidas la planificación, la programación, la representación del conocimiento, la búsqueda y la optimización;
  - iii) los sistemas ciberfísicos, incluidas la Internet de las cosas y la robótica, que comportan el control, la percepción, el procesamiento de los datos recogidos por sensores y el funcionamiento de los actuadores en el entorno en que funcionan los sistemas de IA;
- b) en segundo lugar, además de plantear cuestiones éticas similares a las de cualquier otra tecnología, los sistemas de IA también plantean nuevos tipos de cuestiones. Algunas de ellas tienen que ver con el hecho de que los sistemas de IA son capaces de hacer cosas que antes solo podían hacer los seres vivos y que, en algunos casos, incluso se limitaban solo a los seres humanos. Estas características otorgan a los sistemas de IA una función nueva y determinante en las prácticas y la sociedad humanas. Yendo aún más lejos, a largo plazo, los sistemas de IA podrían desafiar el sentido especial de la experiencia y la conciencia propio del ser humano, planteando cuestiones adicionales sobre la autonomía, el valor y la dignidad del ser humano, pero este no es todavía el caso;

- c) en tercer lugar, aunque las cuestiones éticas relativas a la IA tienen que ver generalmente con el impacto concreto de los sistemas de IA en los seres y las sociedades humanas, otro conjunto de cuestiones éticas apunta a las interacciones entre los sistemas de IA y los seres humanos y sus implicaciones para nuestra comprensión tanto de los seres humanos como de las tecnologías. En la presente Recomendación se reconoce que ambos tipos de cuestiones están estrechamente relacionados y son elementos necesarios de un enfoque ético de la IA.

2. En la presente Recomendación se presta especial atención a las repercusiones éticas más amplias de la IA en relación con las principales esferas de competencia de la UNESCO –la educación, la ciencia, la cultura y la comunicación y la información–, examinadas en el estudio preliminar sobre la ética de la inteligencia artificial elaborado en 2019 por la Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST) de la UNESCO:

- a) los sistemas de IA están vinculados a la educación de múltiples maneras: cuestionan la función de la educación en la sociedad debido a sus implicaciones para el mercado laboral y la empleabilidad, podrían tener repercusiones en las prácticas educativas y exigen que en la formación de los ingenieros en IA y los informáticos se incorpore la sensibilización sobre las implicaciones sociales y éticas de la IA;
- b) en todos los ámbitos de las ciencias y las ciencias sociales y humanas, la IA influye en nuestros conceptos de comprensión y explicación científicas, así como en las formas en que el conocimiento científico puede aplicarse como base para la adopción de decisiones;
- c) la IA tiene implicaciones para la identidad y la diversidad culturales. Puede influir positivamente en las industrias culturales y creativas, pero también puede dar lugar a una mayor concentración de la oferta de contenidos, los datos y los ingresos de la cultura en manos de unos pocos actores, lo que puede tener consecuencias negativas para la diversidad de las expresiones culturales y la igualdad;
- d) en el ámbito de la comunicación y la información, es probable que la traducción automática de idiomas desempeñe una función cada vez más importante. Esto podría tener un impacto sustancial en el lenguaje y la expresión humana, en todas las dimensiones de la vida, lo que conlleva la responsabilidad de actuar con cautela respecto a los idiomas humanos y su diversidad. Además, la IA cuestiona las prácticas del periodismo y la función social de los periodistas, los trabajadores de los medios de comunicación y los productores de medios sociales que se dedican a actividades periodísticas, y está relacionada tanto con la difusión como con la detección de información falsa o interpretaciones erróneas.

3. La presente Recomendación se dirige a los Estados. Según proceda y sea pertinente, también proporciona orientación para las decisiones o prácticas de particulares, grupos, comunidades, instituciones y empresas, públicas y privadas, en particular los actores de la IA, entendidos como aquellos que desempeñan una función activa en el ciclo de vida de los sistemas de IA, incluidas las organizaciones y las personas que investigan, diseñan, desarrollan, despliegan o utilizan la IA.

## II. FINES Y OBJETIVOS

4. La presente Recomendación tiene por objeto la formulación de valores éticos, principios y recomendaciones de política para la investigación, el diseño, el desarrollo, el

despliegue y la utilización de la IA, a fin de que los sistemas de IA estén al servicio de la humanidad, las personas, las sociedades y el medio ambiente.

5. La complejidad de las cuestiones éticas que rodean a la IA exige respuestas igualmente complejas que precisan de la cooperación de múltiples partes interesadas en los diversos niveles y sectores de las comunidades internacionales, regionales y nacionales.

6. Si bien esta Recomendación se dirige principalmente a los encargados de formular políticas en los Estados Miembros de la UNESCO y en otros Estados, también tiene por objeto proporcionar un marco para las organizaciones internacionales, las empresas nacionales y transnacionales, los ingenieros y los científicos, en particular en ciencias humanas, naturales y sociales, las organizaciones no gubernamentales, las organizaciones religiosas y la sociedad civil, estimulando un enfoque de múltiples interesados basado en un marco ético aceptado mundialmente que permita a los interesados colaborar y asumir su responsabilidad común sobre la base de un diálogo mundial e intercultural.

### **III. VALORES Y PRINCIPIOS**

7. Los valores y principios no son necesariamente normas jurídicas en sí mismos, como se indica en el preámbulo de la presente Recomendación. Desempeñan una importante función en la configuración de las medidas de política y las normas jurídicas, porque los valores engloban las expectativas acordadas internacionalmente sobre lo que es bueno y lo que debe preservarse. Así, los valores apuntalan los principios.

8. Los valores inspiran, por tanto, un buen comportamiento moral en consonancia con el entendimiento que la comunidad internacional tiene de ese comportamiento y son los cimientos de los principios. Los principios, por su parte, revelan los valores subyacentes de manera más concreta, de modo que los valores pueden actualizarse más fácilmente en las declaraciones de política y las acciones.

#### **III.1 VALORES**

##### **Dignidad humana**

9. La investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA deberían respetar y preservar la dignidad humana. La dignidad de toda persona humana es un valor que constituye la base de todos los derechos humanos y libertades fundamentales y es esencial para el desarrollo y la adaptación de los sistemas de IA. La dignidad humana se relaciona con el reconocimiento del valor intrínseco de cada ser humano y, por lo tanto, no está vinculada al origen nacional, la condición jurídica, la posición socioeconómica, el género y la orientación sexual, la religión, el idioma, el origen étnico, la ideología política o las opiniones.

10. Este valor debería, en primer lugar, ser respetado por todos los actores que participan en la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA y, en segundo lugar, ser promovido mediante nueva legislación, iniciativas de gobernanza, buenos ejemplos de desarrollo y utilización de la IA en colaboración, o directrices técnicas y metodológicas nacionales e internacionales publicadas por los gobiernos a medida que avanzan las tecnologías de la IA.

##### **Derechos humanos y libertades fundamentales**

11. El valor del respeto, la protección y la promoción de los derechos humanos y las libertades fundamentales en el contexto de la IA significa que la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA deberían ser coherentes y

conformes con el derecho, los principios y las normas internacionales de los derechos humanos.

### **No dejar a nadie atrás**

12. Es fundamental velar por que en la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA se respete a todos los grupos de la humanidad y se fomente la creatividad en toda su diversidad. La discriminación y los prejuicios, las brechas digitales y de conocimientos y las desigualdades mundiales deben abordarse a lo largo del ciclo de vida de los sistemas de IA.

13. Así pues, la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA deben ser compatibles con el empoderamiento de todos los seres humanos, tomando en consideración las necesidades específicas de los diferentes grupos de edad, los sistemas culturales, las personas con discapacidad, las mujeres y las niñas y las poblaciones desfavorecidas, marginadas y vulnerables, y no deberían utilizarse para restringir las elecciones de estilo de vida o el alcance de las experiencias personales, incluida la utilización opcional de los sistemas de IA. Además, habría que esforzarse para paliar la falta de infraestructura, educación y competencias tecnológicas necesarias, así como de marcos jurídicos, en particular en los países de ingreso bajo y mediano.

### **Vivir en armonía**

14. El valor de vivir en armonía supone que la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA reconocen la interconexión de todos los seres humanos. La noción de estar interconectado se basa en el conocimiento de que cada ser humano pertenece a un todo más amplio, que se debilita cuando los demás se ven mermados de alguna manera.

15. Este valor exige que la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA eviten el conflicto y la violencia y no segreguen ni cosifiquen a los seres humanos ni socaven su seguridad, no dividan y enfrenten entre sí a las personas y los grupos y no amenacen la coexistencia armoniosa entre los seres humanos y el medio natural, ya que esto tendría un impacto negativo en la humanidad como colectivo. El propósito de este valor es reconocer la función facilitadora que los actores de la IA deberían desempeñar para lograr el objetivo de vivir en armonía, que consiste en asegurar un futuro para el bien común.

### **Fiabilidad**

16. Los sistemas de IA deberían ser fiables. La fiabilidad es un concepto sociotécnico que implica que la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA deberían inspirar, en lugar de vulnerar, la confianza entre las personas y en los sistemas de IA.

17. La confianza tiene que ganarse en cada contexto de utilización y, en términos más generales, es un indicador de referencia de la aceptación social de los sistemas de IA. Por consiguiente, las personas deberían tener buenas razones para confiar en que la tecnología de la IA aporta beneficios, al tiempo que se adoptan medidas adecuadas para mitigar los riesgos.

### **Protección del medio ambiente**

18. El objetivo de este valor es asegurar que la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA reconozcan la promoción del bienestar



ambiental. Todos los actores que participan en el ciclo de vida de los sistemas de IA deberían seguir las leyes internacionales y nacionales pertinentes en el ámbito de la protección del medio ambiente y el desarrollo sostenible para garantizar la minimización de los factores de riesgo asociados al cambio climático, en particular la emisión de carbono de los sistemas de IA, y evitar la explotación y el agotamiento de los recursos naturales, que contribuyen al deterioro del medio ambiente.

19. Al mismo tiempo, los sistemas de IA deberían utilizarse para aportar soluciones encaminadas a proteger el medio ambiente y preservar el planeta, apoyando enfoques como la economía circular.

### **III.2 PRINCIPIOS**

20. Teniendo en cuenta que todo sistema de IA tiene una serie de características circunstanciales esenciales y en evolución que dependen de factores humanos y tecnológicos, los principios se presentan en dos grupos.

21. El primer grupo está integrado por aquellos principios que reflejan las características asociadas a la interfaz humano-tecnológica, es decir, la interacción entre los seres humanos y los sistemas de IA. Cabe señalar que la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA influyen en los agentes humanos de dos maneras: en primer lugar, amplían el margen de autonomía y toma de decisiones de las máquinas y, en segundo lugar, influyen en la calidad de la acción humana tanto de manera positiva como negativa.

22. El segundo grupo reúne los principios que reflejan las propiedades de los propios sistemas de IA que son pertinentes para que la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA se realicen de conformidad con las expectativas de comportamiento ético aceptadas internacionalmente.

#### **GRUPO 1**

##### **Al servicio de la humanidad y la prosperidad**

23. Los sistemas de IA deberían ser investigados, diseñados, desarrollados, desplegados y utilizados para que los humanos y el medio ambiente en el que viven prosperen. A lo largo de su ciclo de vida, los sistemas de IA deberían mejorar la calidad de vida y favorecer el disfrute de todos los derechos humanos de cada ser humano, dejando a las personas o los grupos la tarea de definir el concepto de “calidad de vida”, siempre que ningún ser humano resulte dañado física o mentalmente o que su dignidad no disminuya como resultado de esa definición.

24. Los sistemas de IA pueden investigarse, diseñarse, desarrollarse, desplegarse o utilizarse para contribuir a las interacciones en las que participen personas vulnerables, incluidos, entre otros, niños, ancianos o enfermos, pero nunca deberían cosificar a los seres humanos ni socavar la dignidad humana, así como tampoco violar los derechos humanos o abusar de ellos.

##### **Proporcionalidad**

25. La investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA no podrán superar el nivel necesario para lograr propósitos u objetivos legítimos y deberían ser adecuados al contexto.

26. La elección de un método de IA debería justificarse de las siguientes maneras: a) el método de IA elegido debería ser conveniente y proporcional para lograr un objetivo determinado; b) el método de IA elegido no debería tener una excesiva repercusión negativa en los valores fundamentales enunciados en la presente Recomendación; c) el método de IA debería ser adecuado al contexto.

### **Supervisión y decisión humanas**

27. Siempre debería ser posible atribuir la responsabilidad ética y jurídica de la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA a una persona física o a una entidad jurídica existente. La supervisión humana se refiere, por tanto, no solo a la supervisión humana individual, sino también a la supervisión pública.

28. Puede ocurrir que, en algunos casos, los seres humanos tengan que compartir el control con los sistemas de IA por razones de eficacia, pero la decisión de ceder el control en contextos limitados seguirá recayendo en los seres humanos, ya que los sistemas de IA deberían ser investigados, diseñados, desarrollados, desplegados y utilizados para ayudar a los seres humanos en la toma de decisiones y en la actuación, pero nunca para reemplazar la responsabilidad humana final.

### **Sostenibilidad**

29. En el contexto de la promoción del desarrollo de sociedades sostenibles, los actores de la IA deberían respetar las dimensiones sociales, económicas y ambientales del desarrollo sostenible de toda la humanidad y del medio ambiente. Los sistemas de IA deberían ser investigados, diseñados, desarrollados, desplegados y utilizados para promover el logro de la sostenibilidad en relación con marcos mundialmente aceptados como los Objetivos de Desarrollo Sostenible.

### **Diversidad e inclusión**

30. La investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA deberían respetar y fomentar la diversidad y la inclusión, ajustándose, como mínimo, al derecho, las normas y los principios internacionales de los derechos humanos, en particular en lo que respecta a la diversidad y la inclusión demográficas, culturales y sociales.

### **Privacidad**

31. La investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA deberían respetar, proteger y promover la privacidad, que constituye un derecho esencial para la protección de la dignidad y la acción humanas. Deberían garantizarse mecanismos adecuados de gobernanza de los datos a lo largo del ciclo de vida de los sistemas de IA, en particular en relación con la reunión y el control de la utilización de datos, mediante el consentimiento informado y autorizaciones y divulgaciones de la aplicación y la utilización de los datos, garantizando los derechos personales sobre los datos y el acceso a ellos.

### **Sensibilización y alfabetización**

32. La sensibilización del público y su comprensión de las tecnologías de IA y del valor de los datos deberían promoverse mediante la educación, campañas públicas y capacitación, a fin de asegurar una participación pública eficaz, de modo que los ciudadanos puedan adoptar decisiones informadas sobre su utilización de los sistemas de IA. Los niños deberían ser protegidos de los daños razonablemente previsibles que se deriven de los sistemas de IA y

tener acceso a esos sistemas mediante la educación y la capacitación, y sus capacidades no deberían verse mermadas por su interacción con dichos sistemas.

### **Gobernanza adaptable y de múltiples interesados**

33. La gobernanza de la IA debería responder a los cambios en la tecnología y los modelos empresariales conexos, ser inclusiva (con la participación de múltiples interesados), estar potencialmente distribuida en diferentes niveles y garantizar, mediante un enfoque sistémico intersectorial, respuestas de gobernanza adecuadas a los fines perseguidos.

34. La gobernanza debería considerar una serie de respuestas que irían desde la gobernanza “blanda”, pasando por procesos de autorregulación y certificación, hasta la gobernanza “dura”, con leyes nacionales y, cuando sea posible y necesario, instrumentos internacionales. A fin de evitar consecuencias negativas y daños involuntarios, la gobernanza debería incluir aspectos de previsión, protección, seguimiento de los efectos, aplicación y reparación.

## **GRUPO 2**

### **Equidad**

35. Durante todo el ciclo de vida de los sistemas de IA, los actores de la IA deberían respetar la equidad y la inclusión, así como hacer todo lo posible por minimizar y evitar reforzar o perpetuar los prejuicios sociotécnicos, en particular los culturales y los relacionados con la raza, el origen étnico, el género y la edad.

### **Transparencia y explicabilidad**

36. Si bien, en principio, hay que hacer todo lo posible por aumentar la transparencia y la explicabilidad de los sistemas de IA para garantizar la confianza de los seres humanos, el nivel de transparencia y explicabilidad debería ser siempre adecuado al contexto de utilización, ya que es preciso encontrar el equilibrio justo entre la transparencia y la explicabilidad y otros principios como la seguridad y la protección.

37. La transparencia significa permitir que las personas comprendan cómo se investigan, diseñan, desarrollan, despliegan y utilizan los sistemas de IA, de manera adecuada al contexto de utilización y a la sensibilidad del sistema en cuestión. También puede incluir la comprensión de los factores que influyen en una predicción o decisión específicas, pero no suele incluir el intercambio de códigos o conjuntos de datos específicos. En este sentido, la transparencia es una cuestión sociotécnica, cuyo objetivo es ganar la confianza de los seres humanos en los sistemas de IA.

38. La explicabilidad se refiere a hacer inteligibles y facilitar la comprensión de los resultados de los sistemas de IA. La explicabilidad de los modelos de IA también se refiere a la inteligibilidad de la entrada, salida y comportamiento de cada componente del algoritmo y la forma en que contribuye a los resultados de los modelos. Así pues, la explicabilidad está estrechamente relacionada con la transparencia, ya que los resultados y los subprocesos que conducen a ellos deberían ser comprensibles y trazables, apropiados al contexto de utilización.

### **Seguridad y protección**

39. En la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA se deberían evitar los daños involuntarios (riesgos de seguridad) y las

vulnerabilidades a los ataques (tareas de protección), a fin de garantizar la seguridad y la protección durante todo el ciclo de vida de los sistemas de IA.

40. Los gobiernos deberían desempeñar una función de liderazgo para garantizar la seguridad y la protección de los sistemas de IA, incluso mediante el establecimiento de normas nacionales e internacionales acordes con el derecho, las normas y los principios internacionales aplicables de los derechos humanos. La investigación estratégica sobre los posibles riesgos de seguridad y protección asociados a los diferentes enfoques para hacer efectiva la IA a largo plazo debería recibir un apoyo constante para evitar daños catastróficos.

### **Responsabilidad y rendición de cuentas**

41. Los actores de la IA deberían asumir la responsabilidad moral y jurídica de conformidad con el derecho internacional de los derechos humanos vigente y las directrices éticas establecidas durante todo el ciclo de vida de los sistemas de IA. La responsabilidad y la obligación de rendir cuentas de las decisiones y las acciones basadas de alguna manera en un sistema de IA siempre deberían ser atribuibles, en última instancia, a los actores de la IA.

42. Deberían elaborarse mecanismos adecuados para garantizar la rendición de cuentas respecto de los sistemas de IA y sus resultados. Deberían considerarse tanto dispositivos técnicos como institucionales para garantizar la auditabilidad y la trazabilidad (del funcionamiento) de los sistemas de IA.

## **IV. ÁMBITOS DE ACCIÓN POLÍTICA**

### **OBJETIVO DE ACCIÓN I: ADMINISTRACIÓN ÉTICA**

43. Velar por la adecuación de la investigación, el diseño, el desarrollo, el despliegue y la utilización de la IA a los valores éticos fundamentales, como los derechos humanos, la diversidad y la inclusión, etc.

#### **Acción política 1: promover la diversidad y la inclusión**

44. Los Estados Miembros deberían colaborar con las organizaciones internacionales para garantizar la participación activa de todos los Estados Miembros, en especial de los países de ingreso mediano bajo, en los debates internacionales sobre la IA. Esto puede hacerse mediante la provisión de fondos, velando por la participación regional en igualdad de condiciones o mediante cualquier otro mecanismo.

45. Los Estados Miembros deberían exigir a los actores de la IA que revelen y combatan todo estereotipo cultural y social en el funcionamiento de los sistemas de IA, ya sea deliberado o por negligencia, y procuren que los conjuntos de datos de capacitación para los sistemas de IA no fomenten las desigualdades culturales y sociales. Deberían adoptarse mecanismos que permitan a los usuarios finales informar sobre esas desigualdades, sesgos y estereotipos.

46. Los Estados Miembros deberían velar por que los actores de la IA tengan en cuenta y respeten las diversidades culturales y sociales existentes, incluidas las costumbres locales y las tradiciones religiosas, en la investigación, el diseño, el desarrollo, el despliegue y la utilización de los sistemas de IA, en consonancia con las normas internacionales de derechos humanos.

47. Los Estados Miembros deberían esforzarse por subsanar las deficiencias en materia de diversidad que se observan actualmente en el desarrollo de los sistemas de IA, en particular la diversidad en los conjuntos de datos de capacitación y en los propios actores de la IA. Los Estados Miembros deberían colaborar con todos los sectores, las organizaciones internacionales y regionales y otras entidades a fin de empoderar a las mujeres y las niñas

para que participen en todas las etapas del ciclo de vida de los sistemas de IA, ofreciéndoles incentivos, acceso a mentores y modelos de conducta, así como protección contra el acoso. También deberían procurar que el ámbito de la IA sea más accesible a personas de diversos orígenes étnicos, así como a personas con discapacidad. Además, debería promoverse la igualdad de acceso a los beneficios de los sistemas de IA, en particular para los grupos marginados.

48. Los Estados Miembros deberían colaborar con las organizaciones internacionales para generalizar la ética de la IA, incluyendo debates sobre cuestiones éticas relacionadas con la IA en los foros internacionales, intergubernamentales y de múltiples interesados pertinentes.

## **OBJETIVO DE ACCIÓN II: EVALUACIÓN DEL IMPACTO**

49. Crear capacidades de observación y anticipación para responder a tiempo a las consecuencias negativas u otras consecuencias imprevistas derivadas de los sistemas de IA.

### **Acción política 2: abordar los cambios en el mercado laboral**

50. Los Estados Miembros deberían procurar evaluar y abordar el impacto de la IA en el mercado laboral y sus consecuencias en las necesidades educativas. Para ello puede ser preciso introducir una gama más amplia de “competencias básicas” en todos los niveles educativos a fin de dar a las nuevas generaciones una oportunidad justa de encontrar empleo en un mercado en rápida evolución y para asegurar que sean conscientes de los aspectos éticos de la IA. Junto a las competencias técnicas especializadas deberían enseñarse competencias como “aprender a aprender”, comunicación, trabajo en equipo, empatía y la capacidad de transferir los conocimientos propios a diversos ámbitos. Es fundamental ser transparente respecto de las competencias de las que existe demanda y actualizar los programas de estudios en torno a ellas.

51. Los Estados Miembros deberían colaborar con entidades privadas, organizaciones no gubernamentales y otras partes interesadas a fin de garantizar una transición justa para los empleados en situación de riesgo. Esto supone poner en marcha programas de perfeccionamiento y reconversión, encontrar formas creativas de retener a los empleados durante esos periodos de transición y explorar programas de “cobertura social” para aquellos que no puedan reconvertirse.

52. Los Estados Miembros deberían alentar a los investigadores a analizar el impacto de la IA en el mercado laboral local con miras a anticipar las tendencias y los desafíos futuros. Estos estudios deberían arrojar luz sobre los sectores económicos, sociales y geográficos que se verán más afectados por la incorporación masiva de la IA.

53. Los Estados Miembros deberían elaborar políticas relativas a la fuerza de trabajo destinadas a apoyar a las mujeres y las poblaciones infrarrepresentadas para asegurarse de que nadie se quede fuera de la economía digital impulsada por la IA. Debería estudiarse la posibilidad de hacer inversiones especiales, y hacerlas si es factible, para ofrecer programas específicos destinados a aumentar la preparación, la empleabilidad, el desarrollo de las perspectivas de carrera y el crecimiento profesional de las mujeres y las poblaciones infrarrepresentadas.

### **Acción política 3: afrontar el impacto social y económico de la IA**

54. Los Estados Miembros deberían idear mecanismos destinados a impedir la monopolización de la IA y las desigualdades resultantes, ya se trate de monopolios de datos, investigación, tecnología, mercados o de otro tipo.

55. Los Estados Miembros deberían colaborar con las organizaciones internacionales y entidades privadas y no gubernamentales para impartir a la población conocimientos adecuados en materia de IA, especialmente en los países de ingreso mediano bajo, a fin de reducir la brecha digital y las desigualdades en el acceso a la tecnología digital resultantes de la adopción a gran escala de sistemas de IA.

56. Los Estados Miembros deberían establecer mecanismos de supervisión y evaluación de las iniciativas y políticas relacionadas con la ética de la IA. Entre los posibles mecanismos cabe citar los siguientes: un repositorio que englobe las iniciativas de cumplimiento ético en las esferas de competencia de la UNESCO, un mecanismo de intercambio de experiencias para que los Estados Miembros soliciten información a otros Estados Miembros sobre sus políticas e iniciativas, y una guía para que los desarrolladores de sistemas de IA evalúen su grado de cumplimiento de las recomendaciones de política mencionadas en el presente documento.

57. Se alienta a los Estados Miembros a que consideren la posibilidad de establecer un mecanismo de certificación para los sistemas de IA similar al utilizado para los dispositivos médicos. Podría haber diferentes clases de certificación según la sensibilidad del ámbito de aplicación y el impacto previsto en las vidas humanas, el medio ambiente y consideraciones éticas como la igualdad, la diversidad y los valores culturales, entre otras. Ese mecanismo podría incluir diferentes niveles de auditoría de los sistemas, los datos y el cumplimiento ético. Al mismo tiempo, no debería obstaculizar la innovación ni perjudicar a las pequeñas empresas o las empresas incipientes exigiendo numerosos trámites. Ese mecanismo también incluiría un componente de seguimiento periódico para garantizar la solidez y la integridad y el cumplimiento continuos de los sistemas de IA durante todo su ciclo de vida, exigiendo una nueva certificación si fuera necesario.

58. Los Estados Miembros deberían alentar a las empresas privadas a que incorporen a diferentes partes interesadas a su gestión de la IA y consideren la posibilidad de crear una función de responsable de ética de la IA o algún otro mecanismo para supervisar la evaluación del impacto, las auditorías y las actividades de seguimiento continuo, así como para garantizar el cumplimiento ético de los sistemas de IA.

59. Los Estados Miembros deberían procurar elaborar estrategias de gobernanza de datos que garanticen la evaluación continua de la calidad de los datos de capacitación para los sistemas de IA, en particular la idoneidad de los procesos de reunión y selección de datos, y establezcan medidas adecuadas de seguridad y protección de los datos, así como mecanismos de retroinformación para aprender de los errores y compartir las mejores prácticas entre todos los actores de la IA. Una de las prioridades de esas estrategias debería ser el logro de un equilibrio entre los metadatos y la protección de la privacidad de los usuarios.

#### **Acción política 4: impacto en la cultura y el medio ambiente**

60. Se alienta a los Estados Miembros a que incorporen sistemas de IA, cuando proceda, a la preservación, el enriquecimiento y la comprensión del patrimonio cultural, tanto material como inmaterial, incluidos los idiomas poco comunes, por ejemplo, introduciendo o actualizando programas educativos relacionados con la aplicación de los sistemas de IA en esas esferas, dirigidos a las instituciones y al público.

61. Se alienta a los Estados Miembros a que examinen y aborden las repercusiones de los sistemas de IA, especialmente las aplicaciones de procesamiento del lenguaje natural, como la traducción automática y los asistentes de voz, en los matices del lenguaje humano. Ese examen puede incluir la maximización de los beneficios de esos sistemas mediante la reducción de las diferencias culturales y el aumento del entendimiento humano, así como repercusiones negativas, como la menor penetración de los idiomas poco comunes, los

dialectos locales y las variaciones tonales y culturales asociadas con el lenguaje y el habla humanos.

62. Los Estados Miembros deberían alentar y promover la investigación colaborativa de los efectos de la interacción a largo plazo de las personas con los sistemas de IA. Para ello deberían utilizarse múltiples normas, principios, protocolos, enfoques disciplinarios y un análisis de la modificación de los hábitos, así como una cuidadosa evaluación de los impactos culturales y sociales posteriores.

63. Los Estados Miembros deberían promover la educación de los artistas y los profesionales creativos en materia de IA a fin de evaluar la idoneidad de la IA para utilizarla en su profesión, ya que la IA se emplea para crear, producir, distribuir y difundir una enorme variedad de bienes y servicios culturales, teniendo en cuenta la importancia de preservar el patrimonio y la diversidad culturales.

64. Los Estados Miembros deberían promover el conocimiento y la evaluación de los instrumentos de IA entre las industrias culturales locales y las empresas incipientes que trabajan en el ámbito de la cultura, a fin de evitar el riesgo de una mayor concentración en el mercado cultural.

65. Los Estados Miembros deberían procurar evaluar y reducir el impacto ambiental de los sistemas de IA, incluida, entre otras cosas, su huella de carbono. También deberían introducir incentivos para promover soluciones ambientales éticas impulsadas por la IA y facilitar su adopción en diferentes contextos. Algunos ejemplos incluyen la utilización de la IA para:

- a) acelerar la protección, el seguimiento y la gestión de los recursos naturales;
- b) apoyar la prevención, el control y la gestión de los problemas relacionados con el clima;
- c) apoyar un ecosistema alimentario más eficiente y sostenible;
- d) acelerar el acceso a la energía verde y su adopción a gran escala.

### **OBJETIVO DE ACCIÓN III: CREACIÓN DE CAPACIDADES PARA LA ÉTICA DE LA IA**

66. Desarrollar las capacidades humanas e institucionales para posibilitar la evaluación del impacto, la supervisión y la gobernanza éticas.

#### **Acción política 5: promover la educación y la sensibilización en materia de ética de la IA**

67. Los Estados Miembros deberían alentar, de conformidad con sus programas de educación y tradiciones nacionales, la incorporación de la ética de la IA en los planes de estudios escolares y universitarios de todos los niveles y promover la colaboración cruzada entre las competencias técnicas y las ciencias sociales y humanas. Deberían elaborarse cursos en línea y recursos digitales en los idiomas locales y en formatos accesibles para las personas con discapacidad.

68. Los Estados Miembros deberían promover la adquisición de “competencias previas” para la educación en materia de IA, como la alfabetización básica, la aritmética elemental y las competencias de codificación, especialmente en los países en los que existen lagunas notables en la enseñanza de esas competencias.

69. Los Estados Miembros deberían flexibilizar los planes de estudios universitarios y facilitar su actualización, habida cuenta del ritmo acelerado de las innovaciones en los sistemas de IA. Además, debería considerarse la integración de la educación en línea y la

educación permanente y la acumulación de credenciales para facilitar planes de estudios ágiles y actualizados.

70. Los Estados Miembros deberían promover programas generales de sensibilización sobre la IA y el acceso inclusivo a los conocimientos sobre las oportunidades que ofrece y los retos que plantea la IA. Estos conocimientos deberían ser accesibles a los grupos técnicos y no técnicos, prestando especial atención a las poblaciones infrarrepresentadas.

71. Los Estados Miembros deberían alentar las iniciativas de investigación sobre la utilización de la IA en la enseñanza, la formación de docentes y el aprendizaje electrónico, entre otras materias, de manera que se potencien las oportunidades y se atenúen los problemas y los riesgos asociados a estas tecnologías. Esto siempre debería ir acompañado de una evaluación adecuada del impacto de la calidad de la educación y de las repercusiones que la utilización de la IA tiene para los educandos y los docentes, y garantizar que la IA confiera más autonomía a ambos grupos y mejore su experiencia.

72. Los Estados Miembros deberían apoyar los acuerdos de colaboración entre las instituciones académicas y la industria para salvar la brecha de las competencias exigidas y promover la colaboración entre los sectores de la industria, el mundo académico, la sociedad civil y el gobierno para adecuar los programas y estrategias de capacitación que ofrecen las instituciones educativas a las necesidades de la industria. Deberían promoverse enfoques de aprendizaje de la IA basados en proyectos, facilitando las asociaciones de colaboración entre empresas, universidades y centros de investigación.

73. Los Estados Miembros deberían promover en particular la participación de las mujeres, las personas de raza y culturas diversas y las personas con discapacidad en los programas de educación en materia de IA, desde la enseñanza básica hasta la educación superior, así como promover el seguimiento y el intercambio de las mejores prácticas con otros Estados Miembros.

#### **Acción política 6: promover la investigación sobre la ética de la IA**

74. Los Estados Miembros deberían promover la investigación sobre la ética de la IA, ya sea mediante inversiones directas o creando incentivos para que los sectores público y privado inviertan en este ámbito.

75. Los Estados Miembros deberían velar por que los investigadores en IA reciban formación en ética de la investigación y exigirles que incluyan consideraciones éticas en el diseño de sus investigaciones y en los productos finales, en particular análisis de los conjuntos de datos que utilizan, la forma en que se anotan y la calidad y el alcance de los resultados.

76. Cuando sea posible, los Estados Miembros y las empresas privadas deberían facilitar a la comunidad científica nacional el acceso a los datos para la investigación, a fin de promover la capacidad de esta comunidad, en particular en los países en desarrollo. Este acceso no debería realizarse a expensas de la privacidad de los ciudadanos.

77. Los Estados Miembros deberían promover la diversidad de género en la investigación en IA en el mundo académico y la industria ofreciendo incentivos a las mujeres para que se incorporen a este campo, establecer mecanismos para luchar contra los estereotipos de género y el acoso en la comunidad de investigadores en IA y alentar a las entidades académicas y privadas a que compartan las mejores prácticas sobre la forma de promover la diversidad.

78. Los Estados Miembros y los organismos de financiación deberían promover la investigación interdisciplinaria en IA incluyendo disciplinas distintas de la ciencia, la



tecnología, la ingeniería y las matemáticas (STEM), por ejemplo, derecho, relaciones internacionales, ciencias políticas, educación, filosofía, cultura y estudios lingüísticos, para garantizar un enfoque crítico de la investigación en IA y un seguimiento adecuado de los posibles usos indebidos o efectos adversos.

#### **OBJETIVO DE ACCIÓN IV: DESARROLLO Y COOPERACIÓN INTERNACIONAL**

79. Velar por un enfoque cooperativo y ético de la utilización de la IA en las aplicaciones de desarrollo, habida cuenta de la gran oportunidad que ofrece esta tecnología para la aceleración de los esfuerzos de desarrollo.

##### **Acción política 7: promover la utilización ética de la IA en el desarrollo**

80. Los Estados Miembros deberían alentar la utilización ética de la IA en esferas de desarrollo como la atención de la salud, la agricultura/el suministro de alimentos, la educación, la cultura, el medio ambiente, la ordenación de los recursos hídricos, la gestión de infraestructuras, la planificación y el crecimiento económicos y otras.

81. Los Estados Miembros y las organizaciones internacionales deberían esforzarse por proporcionar plataformas de cooperación internacional en el ámbito de la IA para el desarrollo, entre otras cosas, aportando competencias técnicas, financiación, datos, conocimientos del sector e infraestructura y facilitando la celebración de talleres entre expertos técnicos y empresariales para hacer frente a los difíciles problemas de desarrollo, especialmente para los países de ingreso mediano bajo y los países menos adelantados.

82. Los Estados Miembros deberían procurar promover la colaboración internacional en materia de investigación en IA, incluidos centros y redes de investigación que promuevan una mayor participación de los investigadores de países de ingreso mediano bajo y otras zonas geográficas emergentes.

##### **Acción política 8: promover la cooperación internacional en materia de ética de la IA**

83. Los Estados Miembros deberían trabajar en el marco de organizaciones internacionales e instituciones de investigación para llevar a cabo investigaciones sobre la ética de la IA. Tanto las entidades públicas como las privadas deberían velar por que los algoritmos y los datos utilizados en una amplia variedad de esferas de la IA –desde la vigilancia policial y la justicia penal hasta el empleo, la salud y la educación– se apliquen de manera equitativa y justa, lo que incluye investigar qué tipos de igualdad y justicia son adecuados en diferentes culturas y contextos, y estudiar la manera de hacerlos coincidir con soluciones técnicamente viables.

84. Los Estados Miembros deberían alentar la cooperación internacional en materia de desarrollo y despliegue de la IA para salvar las divisiones geotecnológicas. Esto requiere un esfuerzo de múltiples interesados en los planos nacional, regional e internacional. Deberían realizarse intercambios/consultas de carácter tecnológico entre los Estados Miembros y su población, entre los sectores público y privado y entre los Estados Miembros.

#### **OBJETIVO DE ACCIÓN V: GOBERNANZA AL SERVICIO DE LA ÉTICA DE LA IA**

85. Promover y orientar la inclusión de consideraciones éticas en la gobernanza de los sistemas de IA.

##### **Acción política 9: establecer mecanismos de gobernanza al servicio de la ética de la IA**

86. Los Estados Miembros deberían velar por que todo mecanismo de gobernanza de la IA sea:

- a) inclusivo: invite y aliente la participación de representantes de comunidades indígenas, mujeres, jóvenes y ancianos, personas con discapacidad y otros grupos minoritarios e insuficientemente representados;
- b) transparente: acepte la supervisión de las estructuras nacionales competentes o de terceras partes de confianza. En el caso de los medios de comunicación, podría tratarse de un grupo de trabajo intersectorial que verifique las fuentes; en el de las empresas de tecnología, podrían ser auditorías externas de los procesos de diseño, despliegue y auditoría interna; en los Estados Miembros, podrían ser exámenes realizados por foros de derechos humanos;
- c) multidisciplinario: todas las cuestiones deberían considerarse de manera global y no solo desde el punto de vista tecnológico;
- d) multilateral: deberían establecerse acuerdos internacionales para atenuar y reparar cualquier daño que pueda aparecer en un país causado por una empresa o usuario situado en otro país. Esto no impide que los diferentes países y regiones elaboren sus propias normas de acuerdo con su cultura.

87. Los Estados Miembros deberían fomentar el desarrollo de un ecosistema digital para la IA ética y el acceso a él. Ese ecosistema incluiría, en particular, tecnologías e infraestructuras digitales y mecanismos para compartir los conocimientos en materia de IA, según proceda. A este respecto, los Estados Miembros deberían considerar la posibilidad de revisar sus políticas y marcos reguladores, incluso en lo que respecta al acceso a la información y el gobierno abierto, para reflejar los requisitos específicos de la IA y promover mecanismos, como los fideicomisos de datos, a fin de apoyar el intercambio seguro, justo, legal y ético de datos, entre otros.

88. Los Estados Miembros deberían alentar la elaboración y utilización de directrices comparables en materia de IA, incluidos los aspectos éticos en los planos mundial y regional, y reunir las pruebas necesarias para evaluar, seguir y controlar la progresión de la aplicación ética de los sistemas de IA.

89. Los Estados Miembros deberían considerar la posibilidad de elaborar y aplicar un marco jurídico internacional para alentar la cooperación internacional entre los Estados y otros interesados.

#### **Acción política 10: garantizar la fiabilidad de los sistemas de IA**

90. Los Estados Miembros y las empresas privadas deberían aplicar medidas adecuadas para vigilar todas las fases del ciclo de vida de los sistemas de IA, incluido el comportamiento de los algoritmos encargados de la adopción de decisiones, los datos y los actores de la IA que participan en el proceso, especialmente en los servicios públicos y en los casos en que se necesita una interacción directa con el usuario final.

91. Los Estados Miembros deberían procurar establecer requisitos claros en relación con la transparencia y la explicabilidad de los sistemas de IA, a saber:

- a) ámbito de aplicación: es probable que algunos sectores, como el de la aplicación de la ley, la seguridad, la educación y la atención de la salud, tengan una mayor necesidad de transparencia y explicabilidad que otros;
- b) destinatarios: el nivel de información sobre los algoritmos y resultados de un sistema de IA y la forma de explicación requerida pueden variar en función de

quién solicite la explicación, por ejemplo, usuarios, expertos en esos ámbitos, desarrolladores, etc.;

- c) viabilidad: muchos algoritmos de IA aún no son explicables; para otros, la explicabilidad añade importantes gastos generales de aplicación. Hasta que la plena explicabilidad sea técnicamente posible con un impacto mínimo en la funcionalidad, habrá que encontrar un compromiso entre la precisión/calidad de un sistema y su nivel de explicabilidad.

92. Los Estados Miembros deberían alentar la investigación sobre la transparencia y la explicabilidad destinando fondos adicionales a esas esferas para diferentes ámbitos y en diferentes niveles (técnico, de lenguaje natural, etc.).

93. Los Estados Miembros y las organizaciones internacionales deberían considerar la posibilidad de elaborar normas internacionales que describan niveles de transparencia mensurables y comprobables, de modo que se puedan evaluar objetivamente los sistemas y determinar los niveles de cumplimiento.

### **Acción política 11: garantizar la responsabilidad, la rendición de cuentas y la privacidad**

94. Los Estados Miembros deberían examinar y adaptar, según proceda, los marcos reguladores y jurídicos para lograr la rendición de cuentas y la responsabilidad por el contenido y los resultados de los sistemas de IA en las diferentes fases de su ciclo de vida. Los gobiernos deberían introducir marcos de responsabilidad o aclarar la interpretación de los marcos existentes para poder atribuir la responsabilidad de las decisiones y el comportamiento de los sistemas de IA. Al elaborar los marcos reguladores, los gobiernos deberían tener en cuenta, en particular, que la responsabilidad y la rendición de cuentas deberían recaer siempre en una persona física o jurídica; no se debería delegar la responsabilidad en un sistema de IA, ni se debería otorgar personalidad jurídica a un sistema de IA.

95. Se alienta a los Estados Miembros a que establezcan evaluaciones del impacto para determinar y analizar los beneficios y los riesgos de los sistemas de IA, así como medidas de prevención, atenuación y seguimiento de los riesgos. En la evaluación de los riesgos se deberían determinar las repercusiones en los derechos humanos y el medio ambiente, así como las consecuencias éticas y sociales, de conformidad con los principios establecidos en la presente Recomendación. Los gobiernos deberían adoptar un marco regulador que establezca un procedimiento para que las autoridades públicas lleven a cabo evaluaciones del impacto de los sistemas de IA que hayan adquirido, desarrollado o desplegado a fin de anticipar las repercusiones, atenuar los riesgos, evitar las consecuencias perjudiciales, facilitar la participación de los ciudadanos y hacer frente a los problemas de la sociedad. Como parte de esa evaluación del impacto, se debería exigir a las autoridades públicas que realicen una autoevaluación de los sistemas de IA existentes y propuestos, en la que se debería analizar, en particular, si la utilización de sistemas de IA en una esfera determinada del sector público es apropiada y cuál es el método adecuado. La evaluación también debería establecer mecanismos de supervisión adecuados, como la auditabilidad, la trazabilidad y la explicabilidad que permite evaluar los algoritmos, los datos y los procesos de diseño, así como incluir un examen externo de los sistemas de IA. Esa evaluación también debería ser multidisciplinaria, multicultural, pluralista e inclusiva y contar con múltiples interesados.

96. Los Estados Miembros deberían hacer participar a todos los actores del ecosistema de la IA (incluidos, entre otros, representantes de la sociedad civil, organismos encargados de hacer cumplir la ley, aseguradores, inversores, fabricantes, ingenieros, abogados y usuarios) en un proceso destinado a establecer normas cuando estas no existan. Las normas pueden llegar a convertirse en mejores prácticas y leyes. Se alienta además a los Estados

Miembros a que utilicen mecanismos como los entornos de pruebas reguladores para acelerar la formulación de leyes y políticas acordes con el rápido desarrollo de las nuevas tecnologías y garantizar que las leyes se puedan poner a prueba en un entorno seguro antes de su aprobación oficial.

97. Los Estados Miembros deberían velar por que los daños causados a los usuarios mediante sistemas de IA puedan investigarse, castigarse y repararse, entre otras cosas alentando a las empresas del sector privado a que proporcionen mecanismos de reparación. Con ese fin, debería promoverse la auditabilidad y la trazabilidad de los sistemas de IA, especialmente de los autónomos.

98. Los Estados Miembros deberían aplicar salvaguardias adecuadas del derecho fundamental de las personas a la privacidad, incluso mediante la adopción o la aplicación de marcos legislativos que proporcionen una protección adecuada, conforme al derecho internacional. De no existir esa legislación, los Estados Miembros deberían alentar enérgicamente a todos los actores de la IA, incluidas las empresas privadas que desarrollan y explotan sistemas de IA, a que apliquen el principio de protección de la privacidad desde el diseño de sus sistemas.

99. Los Estados Miembros deberían velar por que las personas puedan supervisar el uso que se hace de su información o sus datos privados, en particular por que conserven el derecho a acceder a sus datos personales y “el derecho al olvido”.

100. Los Estados Miembros deberían garantizar una mayor seguridad de los datos identificables de las personas o de los datos que, de ser divulgados, puedan causar daños, lesiones o dificultades excepcionales a una persona. Cabe citar como ejemplos los datos relativos a los delitos, las actuaciones penales y las condenas, así como las medidas de seguridad conexas; los datos biométricos; y los datos personales relativos al origen “racial” o étnico, las opiniones políticas, la pertenencia a sindicatos, las creencias religiosas o de otro tipo, la salud o la vida sexual.

101. Los Estados Miembros deberían procurar adoptar un enfoque común de los datos a fin de promover la interoperabilidad de los conjuntos de datos, asegurando al mismo tiempo su solidez y ejerciendo una vigilancia extrema en la supervisión de su reunión y utilización. Esto podría suponer, cuando sea posible y factible, la inversión en la creación de conjuntos de datos de excelencia, incluidos conjuntos de datos abiertos y fiables, que sean diversos, establecidos con el consentimiento de los interesados, cuando la ley exija el consentimiento, y fomentar las prácticas éticas en la tecnología, apoyadas por el intercambio de datos de calidad en un espacio común de datos fiables y seguros.

## **V. SEGUIMIENTO Y EVALUACIÓN**

102. Los Estados Miembros, de acuerdo con sus circunstancias, estructuras de gobierno y disposiciones constitucionales específicas, deberían hacer el seguimiento y evaluar las políticas, los programas y los mecanismos relativos a la ética de la IA mediante una combinación de enfoques cuantitativos y cualitativos, según proceda. Se alienta a los Estados Miembros a que estudien la posibilidad de:

- a) establecer mecanismos de investigación adecuados para medir la eficacia y la eficiencia de las políticas y los incentivos relacionados con la ética de la IA con respecto a objetivos definidos;
- b) recopilar y difundir información relativa a los avances, las buenas prácticas, las innovaciones y la investigación sobre la ética de la IA y sus repercusiones, con el apoyo de la UNESCO y las comunidades internacionales de ética de la IA.

103. Entre los posibles mecanismos de seguimiento y evaluación cabe mencionar los siguientes: un observatorio de la IA que abarque el cumplimiento ético en todas las esferas de competencia de la UNESCO, un mecanismo de intercambio de experiencias para que los Estados Miembros proporcionen información sobre sus iniciativas a otros Estados Miembros, y un “medidor de cumplimiento” para que los desarrolladores de sistemas de IA determinen la medida en que cumplen las recomendaciones de política mencionadas en el presente documento.

104. Deberían elaborarse instrumentos e indicadores adecuados para medir la eficacia y la eficiencia de las políticas relativas a la ética de la IA en función de las normas, prioridades y objetivos acordados, incluidos objetivos específicos para los grupos desfavorecidos y vulnerables. Ello podría comportar evaluaciones de instituciones públicas y privadas, proveedores y programas, incluidas autoevaluaciones, así como estudios de seguimiento y la elaboración de conjuntos de indicadores. La reunión y el procesamiento de datos deberían realizarse de conformidad con la legislación en materia de protección de datos.

105. Los procesos de seguimiento y evaluación deberían asegurar una amplia participación de las partes interesadas pertinentes, entre ellas, aunque no exclusivamente, personas de diferentes grupos de edad, personas con discapacidad, mujeres y niñas y poblaciones desfavorecidas, marginadas y vulnerables, y respetar la diversidad social y cultural, con miras a mejorar los procesos de aprendizaje y fortalecer los nexos entre las conclusiones, la toma de decisiones, la transparencia y la rendición de cuentas sobre los resultados.

## **VI. UTILIZACIÓN Y APLICACIÓN DE LA PRESENTE RECOMENDACIÓN**

106. Los Estados Miembros deberían esforzarse por ampliar y complementar su propia acción en lo que respecta a la presente Recomendación, cooperando con todas las organizaciones gubernamentales y no gubernamentales nacionales e internacionales cuyas actividades correspondan al alcance y a los objetivos de la presente Recomendación.

107. Los Estados Miembros y las partes interesadas que se indican en la presente Recomendación deberían adoptar todas las medidas viables para aplicar las disposiciones enunciadas anteriormente a fin de hacer efectivos los valores fundamentales, los principios y las medidas establecidos en la presente Recomendación.

## **VII. PROMOCIÓN DE LA PRESENTE RECOMENDACIÓN**

108. La UNESCO tiene la vocación de ser el principal organismo de las Naciones Unidas encargado de promover y difundir la presente Recomendación y, en consecuencia, trabajará en colaboración con otras entidades de las Naciones Unidas, entre ellas, aunque no exclusivamente, el Panel de Alto Nivel del Secretario General de las Naciones Unidas sobre la Cooperación Digital, la Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST), el Comité Internacional de Bioética (CIB), el Comité Intergubernamental de Bioética (CIGB), la Unión Internacional de Telecomunicaciones (UIT) y otras entidades pertinentes de las Naciones Unidas que se ocupan de la ética de la IA.

109. La UNESCO trabajará también en colaboración con otras organizaciones internacionales, entre ellas, aunque no exclusivamente, la Unión Africana (UA), la Asociación de Naciones de Asia Sudoriental (ASEAN), el Consejo de Europa (CE), la Unión Económica Euroasiática (UEE), la Unión Europea (UE), la Organización de Cooperación y Desarrollo Económicos (OCDE) y la Organización para la Seguridad y la Cooperación en Europa (OSCE), así como el Instituto de Ingenieros Electricistas y Electrónicos (IIEE) y la Organización Internacional de Normalización (ISO).

## **VIII. DISPOSICIONES FINALES**

110. La presente Recomendación debe entenderse como un todo, y los valores fundamentales y los principios deben considerarse complementarios y relacionados entre sí. Cada principio debe considerarse en el contexto de los valores fundamentales.

111. Ninguna disposición de la presente Recomendación podrá interpretarse como si autorizara a un Estado, otro agente social, grupo o individuo a emprender actividades o realizar actos que vayan en contra de los derechos humanos, las libertades fundamentales, la dignidad humana y el respeto de la vida en la Tierra y más allá.