

## Recomendación sobre las cuestiones éticas relativas a la IA

### **PREÁMBULO**

La UNESCO, en su Constitución, tiene como misión "contribuir a erigir los baluartes de paz en las mentes de las mujeres y los hombres, además de estrechar, mediante la educación, la ciencia y la cultura, la colaboración entre las naciones"(...)

Tras **reflexionar** sobre la influencia profunda que la IA podría tener sobre la sociedad; **considerando** que si bien la IA pueda prestar grandes servicios a la humanidad, también implica preocupaciones éticas fundamentales; **teniendo en cuenta** que el desarrollo de la IA podría acentuar las brechas y desigualdades existentes en el mundo; **estando consciente** de que, en los países de escasos y medianos recursos, la economía digital supone tanto retos como posibilidades enormes para las sociedades creativas, lo que a su vez requiere considerar las culturas, valores y conocimientos endógenos; **reconociendo** que la IA podría tener efectos beneficiosos en el medio ambiente, pero que la materialización de dichas ventajas supone garantizar el acceso equitativo a la tecnología, así como sopesar las ventajas potenciales del impacto medioambiental tanto de la IA como de las tecnologías de información; **sabiendo** que la consideración de los riesgos y preocupaciones éticas no debería frenar la innovación; **recordando** que la conferencia general de la UNESCO estableció la directriz general de "elaborar un instrumento normativo internacional sobre la ética de la inteligencia artificial, mediante una recomendación"; **convencida** de que el instrumento normativo presentado, basado en un enfoque normativo global puede brindar una orientación responsable en cuanto al uso de la IA; **constatando** que el marco normativo aplicable tanto a la IA como a sus implicaciones sociales se encuentra en una encrucijada de la ética, los derechos humanos, los marcos jurídicos internacionales y nacionales, la libertad de investigación e innovación, y el bienestar humano; **reconociendo** la Declaración Universal de los Derechos Humanos (1948), en particular el artículo 27, donde se cita el derecho de participar en el progreso científico; **examinando** el contenido de una serie de actividades y documentos de la ONU; **destacando** que es necesario prestar una atención particular a los países de escasos y medianos recursos; **consciente** de los numerosos marcos éticos nacionales; **consciente también** de que existen muchas iniciativas y marcos relativos a la IA derivados de sectores ajenos al sector público; **convencida** de que la IA puede traer ventajas importantes, pero que su respectiva materialización podría suponer un endeudamiento por innovación; **reconociendo** que una competencia económica podría surgir al interior de los países y entre naciones, lo cual podría ocasionar una orientación de las estrategias y marcos reguladores relativos a la IA hacia intereses nacionales y comerciales; **teniendo plenamente en cuenta** el hecho de que el rápido desarrollo de los sistemas de IA hace frente a obstáculos en relación con la comprensión y puesta en marcha de la IA, a raíz de la diversidad de enfoques a nivel mundial,

La Conferencia General de la UNESCO: **Adopta** la presente Recomendación sobre la Ética de la Inteligencia Artificial, **Recomienda** a los Estados miembros aplicarla tomando las medidas apropiadas para que entre en vigencia en su respectiva jurisdicción y **Recomienda** a los Estados miembros transmitir la Recomendación a todos los actores involucrados en la IA.

### **I. CAMPO DE APLICACIÓN**

*Se considera la ética de la IA en el presente texto como (1) un marco global de valores, principios y acciones interdependientes. El presente documento no provee una definición*

*única de la IA, dado que ésta debe evolucionar con el paso del tiempo. Por otro lado, aborda las características de la IA que sean de importancia ética central.*

La IA puede considerarse como el conjunto de sistemas tecnológicos con la capacidad de analizar información a modo de asemejarse a un comportamiento inteligente, y que además suele abarcar aspectos de aprendizaje, percepción, predicción, planificación o control. La IA encarna (1a) modelos y algoritmos que generan una capacidad tanto de aprendizaje como de ejecución de tareas cognitivas. La IA puede abarcar varios enfoques y tecnologías, entre ellas, 1) el aprendizaje automático, incluyendo el aprendizaje profundo y el aprendizaje por refuerzo; 2) el razonamiento automático, incluyendo la planificación, programación, representación de conocimientos, investigación y optimización; 3) los sistemas ciberfísicos, incluyendo tanto el internet de las cosas como la robótica.

La IA también supone nuevos tipos de problemas ya que (1b) tiene la capacidad de hacer cosas que solamente los seres vivos (o seres humanos) podían realizar antes. Esto confiere a la IA un rol nuevo y profundo tanto en las prácticas humanas como en la sociedad. La IA podría cuestionar el sentido particular de la experiencia y conciencia humana. Otro conjunto de cuestiones éticas abarca (1c) las interacciones entre la IA y los seres humanos, así como su implicación en nuestra comprensión de los seres humanos y de las tecnologías.

La presente Recomendación (2) otorga una atención particular a las múltiples implicaciones éticas en los sectores centrales de la IA de la UNESCO: (2a) educación, (2b) ciencia, (2c) cultura, (2d) comunicación e información, ya que cada sector está relacionado de manera directa o indirecta con la IA en varios ámbitos. La presente Recomendación está (3) dirigida a los Estados públicos y privados, así como a las organizaciones e individuos que participan en la **investigación, concepción, desarrollo, implementación o uso (RCDDU, por sus siglas en inglés)** de la IA.

## II. PROPÓSITOS Y OBJETIVOS

La recomendación tiene como objetivo (4) la formulación de valores éticos, principios y recomendaciones políticas para la RCDDU de la IA, para que la IA opere por el bien de la humanidad. La (5) complejidad de los problemas éticos derivados de la IA requiere de la cooperación de varias partes interesadas. Este documento (6) está dirigido principalmente a las personas que toman decisiones, pero también puede servir de marco para los actores no estatales, fomentando a la vez un enfoque multipartito basado en un sistema de ética reconocido a nivel mundial

## III. VALORES Y PRINCIPIOS

(7) Los valores y principios no constituyen normas jurídicas, pero sí desempeñan un papel influyente en la elaboración tanto de medidas políticas como de normas jurídicas. Por consiguiente, los valores (8) inspiran comportamientos morales positivos en conformidad con la visión de la comunidad internacional

### III.1. VALORES

**-Dignidad humana:** la RCDDU de la IA debería (9) respetar y preservar la dignidad humana, y este valor (10) debería ser respetado por parte de todos los actores involucrados

**-Derechos humanos y libertades fundamentales:** la RCDDU de la IA debe ser (11) tanto consecuente como conforme con el derecho, los principios y las normas internacionales relativos a los derechos humanos.

**-No dejar a nadie atrás:** la RCDDU de la IA debe darse (12) de modo que se respeten a todos los grupos humanos y que se fomente la creatividad, (13) además de ser compatible con la autonomía de todos los seres humanos.

**-Vivir en armonía:** la RCDDU de la IA debería (14) reconocer la interdependencia de todos los seres humanos (que pertenezcan a un colectivo), (15) debería evitar tanto los conflictos como la violencia, no debería separar, objetivar o poner en riesgo la seguridad de los seres humanos, ni tampoco dividir u oponer a los grupos.

**-Credibilidad:** la IA debería ser creíble y (16) digna de confianza. La RCDDU debería inspirar, en vez de atentar contra, la confianza entre las personas y en la IA. La confianza debe (17) ganarse para cada tipo de uso

**-Protección del medio ambiente:** la RCDDU de la IA debería (18) reconocer la promoción del bienestar medioambiental. Todos los actores deberían observar las leyes medioambientales pertinentes para así minimizar los factores de riesgo propios del cambio climático, además de impedir tanto la explotación como el agotamiento de recursos naturales, los cuales inciden en el deterioro del medio ambiente. Se deberían aprovechar los sistemas de IA (19) para concebir soluciones que protejan el medio ambiente.

## III.2. PRINCIPIOS

Se pueden clasificar (20) los principios en dos grupos: a) (21) aquellos en relación con la interacción humanos-IA b) (22) aquellos en relación con las propiedades de la IA que garanticen que la RCDDU cumpla las expectativas de comportamiento ético:

### GRUPO 1: Principios relativos a la interfaz humanos-tecnología

**-Al servicio de la humanidad y de su desarrollo:** la RCDDU de la IA debe (23) tener como objetivo el desarrollo tanto del ser humano como de su entorno. Se debe mejorar la calidad de vida de todos los seres humanos. Se puede utilizar la RCDDU de la IA (24) para ayudar en las interacciones con personas vulnerables; pero sin objetivar jamás a los seres humanos, ni atentar contra la dignidad humana, ni tampoco vulnerar ni abusar de los derechos humanos.

**-Proporcionalidad:** la RCDDU de la IA (25) no debe extenderse por encima de lo necesario para alcanzar objetivos legítimos, y (26) el método elegido (a) debe ser preferible y proporcional para alcanzar un objetivo dado (b) no debería atentar demasiado contra los valores fundamentales recogidos en la Recomendación (c) adaptada al contexto.

**-Vigilancia y decisiones humanas:** en todo momento es necesario poder (27) atribuir la responsabilidad ética y jurídica de la RCDDU de la IA a una persona física o a una entidad jurídica, (28) el control compartido con la IA debe darse por motivos de eficacia; esta decisión corresponde a los seres humanos en ciertos contextos limitados . La RCDDU de la IA tiene como objetivo orientar a los seres humanos en la toma de decisiones y acciones, sin sustituir nunca la responsabilidad humana.

**-Sostenibilidad:** la RCDDU de la IA debería (29) fomentar el cumplimiento de la sostenibilidad en virtud de marcos aceptados a nivel mundial, tales como los objetivos de desarrollo sostenible.

**-Diversidad e inclusión:** la RCDDU de la IA debería (30) respetar y fomentar tanto la diversidad como la inclusión, tomando como marco básico el derecho internacional de los derechos humanos.

**-Privacidad:** la IA debería (31) respetar, proteger y promover la privacidad. Se deben implementar mecanismos adecuados de gobernanza de datos durante todo el ciclo de vida de la IA.

**-Sensibilización y educación:** (32) Se debería fomentar la sensibilización y comprensión colectiva sobre la IA y el valor de los datos para que la ciudadanía pueda tomar decisiones acertadas respecto al uso de la IA.

**-Gobernanza multipartita y adaptativa:** la gobernanza de la IA debería acomodarse (33) tanto a los cambios tecnológicos como a los modelos comerciales asociados. La gobernanza (34) debería plantearse contar con un abanico de respuestas, desde la gobernanza flexible mediante procesos autoreguladores y de certificación, hasta una gobernanza más estricta en conformidad con las leyes nacionales y, de ser posible y necesario, con los instrumentos internacionales.

#### GRUPO 2: Principios relativos a las propiedades de la IA como tales

**-Equidad:** los actores de la IA deberían (35) respetar tanto la justicia como la equidad para minimizar y evitar reforzar o perpetuar el sesgo sociotécnico, en particular los prejuicios raciales, étnicos, de género, de edad y culturales.

**-Transparencia y explicabilidad:** (36) Tanto el nivel de transparencia como el de explicabilidad deben siempre adecuarse al contexto de uso, (37) la transparencia supone la comprensión de la RCDDU de la IA por parte de las personas correspondiente al contexto de uso y a la sensibilidad del sistema de IA, incluyendo asimismo un resumen de los factores que influyen en una predicción o decisión en particular, (38) la explicabilidad consiste en proporcionar no solo inteligibilidad sino también una vista general de los resultados de la IA (en estrecha relación con la transparencia).

**-Seguridad:** la RCDDU de la IA debería (39) evitar tanto los prejuicios involuntarios (riesgos de seguridad) como las vulnerabilidades ante ataques (funciones de seguridad) con el fin de garantizar la seguridad. Los gobiernos deben (40) desempeñar un papel de primer orden para garantizar la seguridad de la IA, asegurando también una investigación continua de los riesgos potenciales de seguridad para prevenir daños catastróficos.

**-Responsabilidad y rendición de cuentas:** los actores de la IA deberían (41) asumir la responsabilidad moral y jurídica, además de (42) desarrollar mecanismos apropiados que puedan garantizar la responsabilidad de la IA. Se deberían tomar en cuenta las concepciones técnicas e institucionales para garantizar tanto la auditabilidad como la rastreabilidad de la IA.

## **IV. CAMPOS DE ACCIÓN ESTRATÉGICA**

**1.º objetivo de las acciones (43): Alineamiento de la RCDDU de la IA con los valores éticos**

Acción estratégica 1: Promover la diversidad e inclusión: (44) Los Estados miembros deberían trabajar de manera conjunta con las organizaciones internacionales para garantizar

la participación activa de todos los Estados miembros en las discusiones sobre la IA, y también (45) deberían exigir a los actores de la IA que luchen contra los estereotipos tanto culturales como sociales en la IA, deberían (46) cerciorarse de que los actores de la IA demuestren consideración y respeto por la diversidad cultural y social actual, deberían (47) colmar las lagunas en términos de diversidad presenciadas actualmente en el desarrollo de los sistemas de IA, (48) colocar la ética de la IA como una preocupación central en las discusiones sobre la ética relativa a la IA.

## **2.º objetivo de las acciones: evaluación del impacto: (49) capacidades de observación para reaccionar ante consecuencias negativas o involuntarias derivadas de los sistemas de IA**

Acción estratégica 2: Hacer frente a las evoluciones del mercado laboral: los Estados miembros deberían (50) evaluar y procesar tanto el impacto de la IA en los mercados laborales como sus repercusiones en los requisitos del ámbito educativo; "Aprender a aprender", es algo que debería enseñarse al mismo tiempo que las habilidades técnicas, (51) garantizar una transición equitativa para los empleados de riesgo, (52) analizar el impacto de la IA en el mercado laboral local, (53) elaborar políticas laborales que favorezcan a las mujeres y poblaciones subrepresentadas.

Acción estratégica 3: Luchar contra el impacto social y económico de la IA: (54) Impedir tanto el monopolio de la IA como las desigualdades que resulten de ella, (55) brindar al público un conocimiento adecuado de la IA, (56) establecer mecanismos de seguimiento y de evaluación de las iniciativas y políticas en relación con la ética de la IA, (57) considerar un mecanismo de certificación para los sistemas de IA similar a aquellos utilizados por los dispositivos médicos, (58) fomentar en el sector privado la participación de distintas partes interesadas en la gobernanza de la IA y considerar agregar un cargo de responsable de ética de la IA para supervisar la evaluación del impacto, (59) desarrollar estrategias de gobernanza de los datos donde se equilibren tanto los macrodatos como la confidencialidad de los usuarios.

Acción estratégica 4: Repercusiones en la cultura y el medio ambiente: (60) Incorporar la IA, según convenga, en el patrimonio cultural, (61) examinar el impacto de la IA en el lenguaje humano, (62) estudiar los efectos de las interacciones humanas con la IA, (63) fomentar la educación y la sensibilización de la IA en los profesionales creativos, (64) las industrias culturales y las empresas emergentes, (65) evaluar y reducir el impacto medioambiental de la IA.

## **3.º objetivo de las acciones: (66) desarrollar capacidades que consientan tanto la evaluación del impacto como una gobernanza ética.**

Acción estratégica 5: Promover la educación y la sensibilización respecto a la ética de la IA: estimular (67) la integración de la ética de la IA y promover la colaboración entre las habilidades técnicas y las ciencias humanas, (68) promover la sensibilización a la IA y el acceso al conocimiento relativo a las oportunidades y retos derivados de la IA, (69) incorporar el concepto de flexibilidad en la educación superior al considerar las innovaciones que brinda la IA, (70) fomentar programas de sensibilización de la IA, (71) impulsar iniciativas de investigación sobre el uso de la IA en la docencia. Respalda la colaboración entre los sectores universitario e industrial, (72) apoyar la colaboración en la investigación (73) promover la participación de las mujeres y todas las minorías.

Acción estratégica 6: Promover la investigación sobre la ética de la IA: (74) Promover la investigación en la IA, cerciorándose de que (75) los investigadores de IA se capaciten en asuntos sobre la ética, (76) facilitar el acceso a datos investigativos a la comunidad científica y no en detrimento de la privacidad ciudadana, (77) promover la diversidad de

género en la investigación sobre la IA en los sectores tanto universitario como industrial, luchar contra los estereotipos y el acoso, (78) promover una investigación interdisciplinaria y crítica sobre la IA que abarque disciplinas ajenas a las STEM (ciencia, tecnología, ingeniería, matemáticas).

#### **4.º objetivo de las acciones: (79) garantizar un enfoque cooperativo y ético en cuanto al uso de la IA**

Acción estratégica 7: Promover el uso ético de la IA en el desarrollo: (80) fomentar el uso ético de la IA en los sectores del desarrollo, (81) esmerarse en proporcionar plataformas de cooperación internacional relativas a la IA en aras del desarrollo, (82) esforzarse en fomentar las colaboraciones internacionales en cuanto a la investigación sobre la IA.

Acción estratégica 8: Promover la cooperación internacional en el ámbito ético de la IA: (83) Obrar mediante organizaciones internacionales e instituciones de investigación para así llevar a cabo investigaciones sobre la ética de la IA, garantizando que tanto los algoritmos como los datos usados en los sectores de la IA se apliquen de manera equitativa, además de (84) fomentar la cooperación internacional en el desarrollo de la IA para eliminar las divisiones geotécnicas.

#### **5.º objetivo de las acciones: (85) consideraciones éticas respecto a la gobernanza de la IA**

Acción estratégica 9: Establecer mecanismos de gobernanza sobre el ámbito ético de la IA: (86) garantizar que todo mecanismo de gobernanza sobre la IA sea (86a) inclusivo, (86b) transparente, (86c) multidisciplinario, (86d) multilateral, y (87) favorecer tanto el desarrollo como el acceso a un ecosistema digital para conseguir una IA ética, (88) incentivar la elaboración y el uso de directrices sobre la IA, (89) considerando la elaboración de un marco jurídico internacional para la cooperación internacional

Acción estratégica 10: garantizar la fiabilidad de la IA: (90) implementar medidas apropiadas para monitorear todas las fases del ciclo de vida de la IA, y (91) definir los requisitos de transparencia y de explicabilidad de la IA según (91a) el campo de aplicación, (91b) el público meta y (91c) la viabilidad; los estados deberían (92) fomentar la investigación sobre la transparencia y la explicabilidad, (93) elaborar normas internacionales que describan los niveles medibles de transparencia con el fin de realizar una evaluación objetivas de los niveles de cumplimiento

Acción estratégica 11: responsabilidad, obligación de rendición de cuentas y privacidad: (94) adaptar los marcos reguladores y jurídicos para garantizar tanto la responsabilización como la responsabilidad del contenido y de los resultados de la IA, incorporar marcos de responsabilidad para atribuir la responsabilidad de las decisiones y del comportamiento de la IA a una persona física o jurídica; ninguna responsabilidad ni personalidad jurídica debería asociarse con la IA, (95) incorporar evaluaciones de impacto con el fin de identificar y evaluar las ventajas y riesgos de la IA, así como mecanismos de vigilancia adecuados, entre ellos, la auditabilidad, rastreabilidad y explicabilidad; de igual manera, contar con exámenes externos y multidisciplinarios de la IA, (96) hacer participar a todos los actores del ecosistema de la IA en el establecimiento de normas en lugares donde no existan, (97) garantizar que los daños a los usuarios por parte de la IA se puedan examinar, sancionar y resarcir, (98) aplicar las garantías respectivas del derecho fundamental de privacidad de las personas, (99) garantizar que los individuos puedan supervisar el uso de sus informaciones o datos privados, (100) garantizar mejoras en la seguridad relativa a los datos o datos personales identificables, (101) adoptar un enfoque común de datos para fomentar la interoperabilidad de los conjuntos de datos, además de demostrar vigilancia en la supervisión de su respectiva recopilación y uso.

## **V. SEGUIMIENTO Y EVALUACIÓN**

(102) Los miembros deberían seguir y evaluar las políticas, programas y mecanismos en relación con la ética de la IA utilizando(103) los mecanismos posibles como un observatorio de la IA que cumpla con los requisitos éticos en los sectores de competencia de la UNESCO, (104) se deberían desarrollar tanto herramientas como indicadores para medir la eficacia y eficiencia de las políticas relativas a la ética de la IA respecto a las normas convenidas (105) Los procesos de seguimiento y evaluación deberían garantizar una amplia participación de las partes interesadas correspondientes.

## **VI. USO Y APROVECHAMIENTO DE LA RECOMENDACIÓN**

(106) Los miembros deberían cooperar con todas las organizaciones gubernamentales y no gubernamentales tanto nacionales como internacionales, y (107) las partes interesadas descritas deberían tomar todas las medidas posibles para aplicar las disposiciones enumeradas.

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

(108) La UNESCO constituye la agencia principal de las Naciones Unidas que se encarga tanto de promover como de divulgar esta recomendación, y debería también trabajar de manera conjunta con otras entidades no solo de las Naciones Unidas, sino también (109) de otras organizaciones internacionales

## **VIII. DISPOSICIONES FINALES**

(110) Se debe comprender la Recomendación en todo su conjunto, y sus valores y principios deben considerarse como complementarios e interdependientes, (111) y no se puede interpretar elemento alguno de la Recomendación como una aprobación para cualquier actor de incurrir en una actividad contraria a los derechos humanos, las libertades fundamentales, la dignidad humana y la preocupación por la vida en la Tierra y más allá de ella.