

Septiembre 2024

GOBERNANZA DE LA IA EN BENEFICIO DE LA HUMANIDAD



United
Nations



AI
Advisory
Body



**Gobernanza de la Inteligencia Artificial en beneficio de la Humanidad:
Informe final**

Copyright © 2024 Naciones Unidas
Reservados todos los derechos.

No está permitido reproducir ni transmitir parte alguna de esta publicación en ninguna forma y por ningún medio, ni electrónico ni mecánico, como fotocopias, grabaciones o cualquier sistema de almacenamiento y recuperación de información conocido o por inventar, sin la autorización escrita de las Naciones Unidas.

Las solicitudes de reproducción de pasajes o fotocopias de esta publicación deben dirigirse a Copyright Clearance Center (copyright.com).

Todas las consultas sobre derechos y licencias, incluidos los derechos subsidiarios, deben dirigirse a: United Nations Publications, 405 East 42nd Street, S-011FW001, New York, NY 10017, Estados Unidos de América.
Correo electrónico: permissions@un.org; sitio web: shop.un.org.

Las denominaciones empleadas en esta publicación y la forma en que aparecen presentados los datos que contiene no implican, de parte de la Secretaría de las Naciones Unidas, juicio alguno sobre la condición jurídica de países o territorios o zonas, o de sus autoridades, ni respecto a la delimitación de sus fronteras.

eISBN: 9789211068009

Gobernanza de la Inteligencia Artificial en beneficio de la Humanidad

Informe final



**United
Nations**



AI
Advisory
Body

Acerca del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial

El Órgano Asesor de Alto Nivel sobre Inteligencia Artificial, formado por múltiples partes interesadas y propuesto inicialmente en 2020 como parte de la Hoja de ruta para la cooperación digital del Secretario General (A/74/821), se constituyó en octubre de 2023 para realizar análisis y formular recomendaciones sobre la gobernanza internacional de la inteligencia artificial.

Los miembros del Órgano Asesor han participado a título personal, no como representantes de sus respectivas organizaciones. El presente informe representa un consenso mayoritario; no se espera que ningún miembro apruebe todos y cada uno de los puntos contenidos en este documento. Los miembros confirman el acuerdo amplio, pero no unilateral, con sus conclusiones y recomendaciones. Las expresiones que figuran en este informe no implican la aprobación institucional de las respectivas organizaciones de los miembros.

Tabla de Contenidos

Acerca del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial	4
Resumen ejecutivo	7
1. La necesidad de una gobernanza global	7
2. Lagunas en la gobernanza global de la IA	8
3. Mejora de la cooperación global	9
A. Un entendimiento común	10
B. Un terreno común	11
C. Beneficios comunes	15
D. Esfuerzos coherentes	22
E. Reflexiones sobre los modelos institucionales	24
4. Un llamamiento a la acción	24
1. Introducción	26
A. Oportunidades y elementos facilitadores	32
B. Facilitadores clave para aprovechar la IA en beneficio de la humanidad	32
C. La gobernanza como facilitador clave	32
D. Riesgos y desafíos	32
E. Riesgos de la IA	32
F. Retos que hay que afrontar	38
2. La necesidad de una gobernanza global	42
A. Principios rectores y funciones de la gobernanza internacional de la IA	43
B. El nuevo panorama internacional de gobernanza de la IA	46

3. Lagunas en la gobernanza global de la IA	48
A. Lagunas de representación	48
B. Lagunas de coordinación	50
C. Lagunas de implementación	52
4. Mejora de la cooperación global	54
A. Un entendimiento común	55
Un grupo internacional de científicos	55
B. Un terreno común	60
Diálogo sobre políticas de gobernanza de la IA	60
Intercambio de normas sobre la IA	64
C. Beneficios comunes	67
Red de desarrollo de la capacidad	74
Fondo mundial para la IA	76
Marco mundial de datos sobre IA	78
D. Esfuerzos coherentes	82
Una oficina de IA en la Secretaría General de las Naciones Unidas	82
E. Reflexiones sobre los modelos institucionales	86
¿Un organismo internacional de inteligencia artificial?	86
5. Conclusión: Un llamamiento a la acción	91
Anexos	93
Anexo A: Miembros del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial	93
Anexo B: Mandato del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial	94
Anexo C: Lista de participación en las consultas, 2024	95
Anexo D: Lista de “inmersiones en profundidad”	96
Anexo E: Respuestas a la encuesta global AI Risk Pulse Check	97
Anexo F: Respuestas al estudio prospectivo de oportunidades	107
Anexo G: Lista de abreviaciones	113

Resumen ejecutivo

- i** La inteligencia artificial (IA) está transformando el mundo. Este conjunto de tecnologías ofrece un enorme potencial de beneficio, que abarca desde la exploración de nuevas áreas de investigación científica y la optimización de las redes energéticas hasta la mejora de la salud pública y la agricultura, pasando por la contribución al logro de los Objetivos de Desarrollo Sostenible (ODS).
- ii** Sin embargo, si la IA no se regula, las oportunidades que ofrece podrían no manifestarse o distribuirse de forma equitativa. El aumento de las brechas digitales podría limitar los beneficios de la IA a unos pocos Estados, empresas y particulares. Los usos *desaprovechados* (no aprovechar y compartir los beneficios relacionados con la IA debido a falta de confianza o a ausencia de elementos facilitadores, como falta de capacidad o gobernanza ineficaz) podrían limitar el abanico de oportunidades.
- iii** La IA también conlleva otros riesgos. Al sesgo y la vigilancia de la IA se unen nuevas preocupaciones como las confabulaciones (o “alucinaciones”) de los grandes modelos lingüísticos, el aumento de la creación y difusión de desinformación, los riesgos para la paz y la seguridad y el consumo de energía de los sistemas de IA en tiempos de crisis climática.
- iv** Los sistemas de IA rápidos, opacos y autónomos desafían los sistemas reguladores tradicionales, al tiempo que los sistemas cada vez más potentes pueden revolucionar el mundo laboral. Las armas autónomas y los usos de la IA para el orden público plantean serias cuestiones jurídicas, humanitarias y de seguridad.

v En la actualidad hay un gran déficit de gobernanza global de la IA. A pesar de los numerosos debates sobre ética y principios, el mosaico de normas e instituciones es aún incipiente y está lleno de lagunas. La rendición de cuentas suele brillar por su ausencia, en particular cuando se despliegan sistemas de IA que son difícilmente explicables y que afectan a los demás. El cumplimiento suele basarse en la voluntariedad; la práctica desmiente el discurso.

vi Como ya se señaló en nuestro informe provisional¹, la gobernanza de la IA es de vital importancia no solo para afrontar los desafíos y los riesgos, sino también para garantizar que aprovechemos el potencial de la IA sin dejar a nadie atrás.

1. La necesidad de una gobernanza global

vii El imperativo de la gobernanza global, en particular, es irrefutable. Las materias primas de la IA, desde los minerales críticos hasta los datos de entrenamiento, se obtienen en todo el mundo. La IA de uso general, desplegada más allá de las fronteras, genera múltiples aplicaciones a nivel mundial. El desarrollo acelerado de la IA concentra el poder y la riqueza a escala mundial, con implicaciones geopolíticas y geoeconómicas.

viii Además, en la actualidad nadie comprende el funcionamiento interno de la IA lo suficiente como para controlar totalmente sus resultados o predecir su evolución, ni los responsables de las decisiones de desarrollar, desplegar o utilizar sistemas que no comprenden rinden cuentas. Mientras tanto, es probable que las repercusiones negativas y los impactos derivados de tales decisiones sean también globales.

1 Véase <https://un.org/ai-advisory-body>.

- ix** El desarrollo, despliegue y uso de una tecnología de este tipo no puede dejarse únicamente al capricho de los mercados. Los gobiernos nacionales y las organizaciones regionales serán cruciales en este sentido, pero la propia naturaleza de la tecnología - transfronteriza en estructura y aplicación - exige un enfoque global. La gobernanza también puede ser un facilitador clave de la innovación de la IA para los ODS a nivel mundial.
- x** Por tanto, la IA plantea desafíos y oportunidades que requieren un enfoque integral y global que abarque transversalmente los ámbitos político, económico, social, ético, de derechos humanos, técnico, ambiental y otros. Un enfoque de este tipo puede convertir un mosaico de iniciativas en evolución en un todo coherente e interoperable, basado en el derecho internacional y en los ODS, adaptable en distintos contextos y en el transcurso del tiempo.
- xi** En nuestro informe provisional ya esbozamos los principios² que debían regir la creación de nuevas instituciones internacionales de gobernanza de la IA. Estos principios parten del reconocimiento de que la gobernanza de la IA no puede desarrollarse en el vacío, sino que debe estar enmarcada en el derecho internacional, especialmente en el derecho internacional de los derechos humanos.
- xiii** Sin embargo, ninguno de ellos puede ser realmente global en su alcance y exhaustivo en su cobertura. Esto plantea problemas de representación, coordinación e implementación.
- xiv** En términos de representación, partes enteras del mundo han quedado fuera de las conversaciones sobre la gobernanza internacional de la IA³. La Figura a) muestra siete destacadas iniciativas relativas a la IA ajenas a las Naciones Unidas. Hay 7 países que participan en todas las iniciativas de gobernanza de la IA de la muestra, mientras 118 países no participan en ninguna de ellas (principalmente los del Sur Global).
- xv** La equidad exige que más voces desempeñen un papel significativo en las decisiones sobre cómo gobernar la tecnología que nos afecta. La concentración de la adopción de decisiones en el sector tecnológico de la IA no tiene justificación; debemos reconocer que, históricamente, muchas comunidades han sido totalmente excluidas de las conversaciones sobre la gobernanza de la IA.
- xvi** Para ser eficaces, los regímenes de gobernanza de la IA deben también abarcar todo el mundo – y deben serlo para evitar las “carreras armamentistas de la IA” o una “competencia a la baja” en materia de seguridad y derechos, para detectar y responder a los incidentes derivados de decisiones adoptadas a lo largo del ciclo de vida de la IA que abarcan múltiples jurisdicciones, para estimular el aprendizaje, para favorecer la interoperabilidad y para compartir los beneficios de la IA. La tecnología no tiene fronteras y, a medida que se extienda, la idea ilusoria de que un Estado o grupo de Estados puede (o debe) controlarla irá desapareciendo.

2. Lagunas en la gobernanza global de la IA

- xii** No faltan documentos y diálogos centrados en la gobernanza de la IA. Los gobiernos, las empresas, los consorcios y las organizaciones internacionales y regionales han adoptado cientos de directrices, marcos y principios.

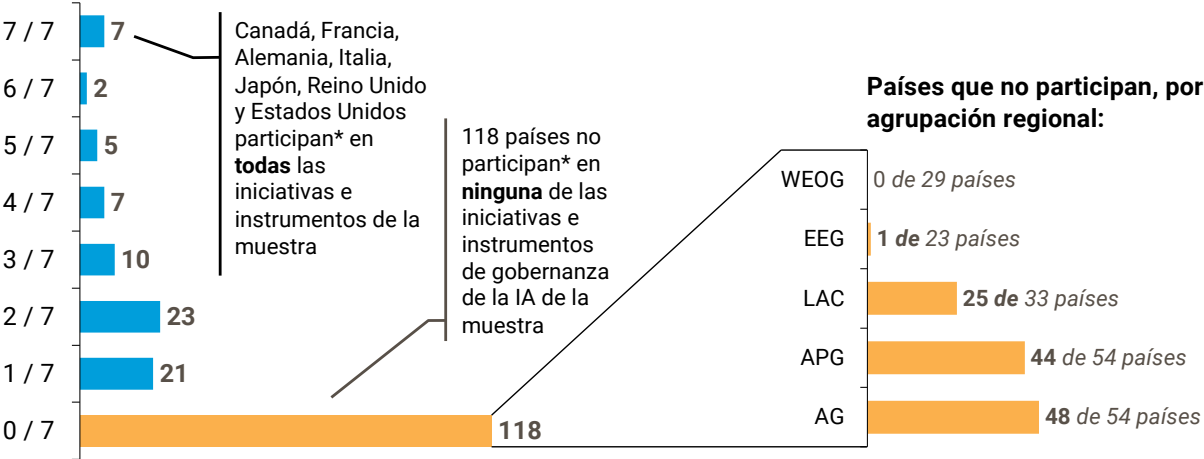
2 Principio rector 1: La gobernanza de la IA debe ser inclusiva y utilizarse para y en beneficio de todos; principio rector 2: La gobernanza de la IA debe estar al servicio del interés público; principio rector 3: La gobernanza de la IA debe construirse en paralelo a la gobernanza de los datos y la promoción de conjuntos de datos comunes; principio rector 4: La gobernanza de la IA debe ser universal, estar interconectada y arraigada en la colaboración flexible de múltiples partes interesadas; y principio rector 5: La gobernanza de la IA debe estar anclada en la Carta de las Naciones Unidas, el derecho internacional de los derechos humanos y otros compromisos internacionales acordados, como los ODS.

3 Excluyendo la Recomendación de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) sobre la ética de la inteligencia artificial (2021) y las dos resoluciones de la Asamblea General sobre IA de 2024: *Aprovechar las oportunidades de sistemas seguros y fiables de inteligencia artificial para el desarrollo sostenible (78/265)* y *Aumentar la cooperación internacional para la creación de capacidad en materia de inteligencia artificial (78/311)*.

Figura (a): Representación en las siete iniciativas internacionales de gobernanza de la IA ajenas a las Naciones Unidas

Muestra: Principios de la OCDE sobre la IA (2019), Principios del G20 sobre la IA (2019), Grupo de redacción del convenio internacional sobre IA del Consejo de Europa (2022-2024), Declaración Ministerial de la Alianza Mundial sobre la Inteligencia Artificial (2022), Declaración de los Ministros del G7 (2023), Declaración de Bletchley (2023) y Declaración Ministerial de Seúl (2024).

Solamente interregional, excluye regional



*Por refrendo de las publicaciones intergubernamentales pertinentes. No se considera que los países participan en una iniciativa plurilateral por el mero hecho de pertenecer a la Unión Europea o a la Unión Africana. Abreviaciones: AG, Grupo de los Estados de África; APG, Grupo de Asia y el Pacífico; EEG, Grupo de los Estados de Europa Oriental; G20, Grupo de los 20; G7, Grupo de los Siete; GPAI, Alianza Mundial sobre la Inteligencia Artificial; ALC, América Latina y el Caribe; OCDE, Organización de Cooperación y Desarrollo Económico; WEOG, Grupo de los Estados de Europa Occidental y Otros Estados.

xvii Las lagunas de coordinación entre las iniciativas y las instituciones amenazan con dividir el mundo en regímenes desconectados e incompatibles de gobernanza de la IA. También falta coordinación dentro del sistema de las Naciones Unidas. Aunque muchas entidades de las Naciones Unidas se ocupan de la gobernanza de la IA, sus mandatos concretos hacen que ninguna lo haga de forma exhaustiva.

xviii Sin embargo, la representación y la coordinación no son suficientes. La rendición de cuentas requiere implementación para que los compromisos con la gobernanza global de la IA se traduzcan en resultados tangibles en la práctica, por ejemplo, en el desarrollo de la capacidad y en el apoyo a las pequeñas y medianas empresas para que compartan las oportunidades. Gran parte de esta labor se llevará a cabo a escalas nacional y regional, pero también es necesario hacer más a escala mundial para hacer frente a los riesgos y para aprovechar los beneficios.

3. Mejora de la cooperación global

xix Nuestras recomendaciones proponen una visión integral de un enfoque global en red, ágil y flexible para gobernar la IA en beneficio de la humanidad, que abarque un entendimiento común, un terreno común y beneficios comunes. Solo un enfoque inclusivo y global de la gobernanza de la IA puede abordar los desafíos y las oportunidades múltiples y cambiantes que presenta la IA a escala global, promoviendo la estabilidad internacional y el desarrollo equitativo.

xx Orientadas por los principios establecidos en nuestro informe provisional, nuestras propuestas pretenden colmar lagunas y aportar coherencia al ecosistema emergente de respuestas e iniciativas internacionales de gobernanza de la IA, ayudando a evitar la fragmentación y la pérdida de oportunidades. Para apoyar estas

medidas de manera eficiente y para colaborar eficazmente con otras instituciones, proponemos una estructura ligera y ágil como expresión de los esfuerzos coherentes: una oficina de IA en la Secretaría General de las Naciones Unidas, cercana al Secretario General, que trabaje como el “aglutinante” eficiente y sostenible de las iniciativas aquí propuestas.

A. Un entendimiento común

xxi Un enfoque global para gobernar la IA comienza con un entendimiento común de sus capacidades y de las oportunidades, los riesgos y las incertidumbres que plantea. Los Estados Miembros deben disponer de información y conocimientos científicos oportunos, imparciales y fiables sobre la IA para poder construir un entendimiento fundacional compartido en todo el mundo y para equilibrar las incoherencias informativas entre las empresas que albergan costosos laboratorios de IA y el resto del mundo (entre otras cosas, mediante el intercambio de información entre las empresas de IA y la comunidad de la IA en general).

xxii La puesta en común de conocimientos científicos es más eficiente a nivel mundial, ya que permite la inversión conjunta en un bien público global y la colaboración de interés público a través de esfuerzos que, de otro modo, estarían fragmentados y duplicados.

Grupo internacional de científicos expertos en IA

xxiii Inspirándose en precedentes como el Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC) y el Comité Científico de las Naciones Unidas para el Estudio de los Efectos de las Radiaciones Atómicas, un grupo internacional y multidisciplinar de científicos expertos en IA podría recopilar e impulsar investigaciones punteras para informar a los científicos, los encargados de formular políticas, los Estados Miembros y otras

partes interesadas que busquen perspectivas científicas sobre la tecnología de la IA o sus aplicaciones en una fuente imparcial y creíble.

xxiv Un panel de científicos bajo los auspicios de las Naciones Unidas podría aportar conocimientos especializados sobre las oportunidades relacionadas con la IA. Esto podría incluir promover una “inmersión en profundidad” en ámbitos de los ODS como la atención de la salud, la energía, la educación, las finanzas, la agricultura, el clima, el comercio y el empleo.

xxv Las evaluaciones de riesgos también podrían basarse en el trabajo de otras iniciativas de investigación sobre IA, ofreciendo las Naciones Unidas un “refugio seguro” de confianza para que los investigadores intercambien ideas sobre “los últimos avances”. Al poner en común los conocimientos de países o empresas que de otro modo no participarían o no estarían incluidos, un panel organizado por las Naciones Unidas puede ayudar a rectificar las percepciones erróneas y reforzar la confianza a nivel mundial.

xxvi Este grupo debería funcionar de forma independiente, con el apoyo de un equipo del sistema de las Naciones Unidas formado por miembros de la oficina de IA que se propone a continuación y de organismos competentes de las Naciones Unidas como la Unión Internacional de Telecomunicaciones (UIT) y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO). Asimismo, debería trabajar en colaboración con otras instituciones internacionales como la Organización de Cooperación y Desarrollo Económicos (OCDE) y la Alianza Mundial sobre la Inteligencia Artificial.

Recomendación 1

Un grupo internacional de científicos expertos en IA

Recomendamos la creación de un grupo internacional e independiente de científicos expertos en IA, formado por diversos expertos multidisciplinares en la materia que presten sus servicios a título personal y de forma voluntaria. Con el apoyo de la oficina de IA propuesta y otros organismos pertinentes de las Naciones Unidas, en colaboración con otras organizaciones internacionales pertinentes, su mandato incluiría:

- a) Publicar un informe anual en el que se analicen las capacidades, las oportunidades, los riesgos y las incertidumbres relacionadas con la IA, y se identifiquen las esferas de consenso científico sobre las tendencias tecnológicas y las áreas en las que se necesita más investigación;
- b) Producir compendios trimestrales de investigación temática sobre las esferas en las que la IA podría contribuir a la consecución de los ODS, centrándose en las áreas de interés público que puedan estar insuficientemente atendidas; y
- c) Publicar informes ad hoc sobre cuestiones emergentes, en particular la aparición de nuevos riesgos o las importantes lagunas existentes en el panorama de la gobernanza.

B. Un terreno común

xxvii Junto a un entendimiento común de la IA, es necesario un terreno común para establecer enfoques de gobernanza interoperables anclados en normas y principios globales en interés de todos los países. Esto es necesario a nivel mundial para evitar una competencia a la baja en la regulación y para reducir la fricción reguladora transfronteriza; para maximizar el aprendizaje y la interoperabilidad técnica; y para responder eficazmente a los retos derivados del carácter transfronterizo de la IA.

Diálogo sobre políticas de gobernanza de la IA

xxviii Es preciso establecer un foro político inclusivo para que todos los Estados Miembros (aprovechando los conocimientos especializados de las partes interesadas) puedan compartir las mejores prácticas que están basadas en los derechos humanos y promueven el desarrollo, que adoptan enfoques de gobernanza interoperables y que tienen en cuenta los desafíos transfronterizos que justifican una mayor consideración política. Esto no significa una gobernanza global de todos los aspectos de la IA, pero puede establecer el marco para la cooperación internacional y alinear mejor los esfuerzos nacionales y sectoriales con las normas y los principios mundiales.

xxix La institucionalización de este tipo de intercambio entre múltiples partes interesadas bajo los auspicios de las Naciones Unidas puede proporcionar un marco fiable e inclusivo para debatir las nuevas prácticas de gobernanza y las respuestas políticas adecuadas. Al ir más allá de las zonas de confort, el diálogo entre países no afines, y entre Estados y partes interesadas, puede acelerar el aprendizaje y sentar las bases de una mayor cooperación, por ejemplo en materia de normas de seguridad y derechos, y en tiempos de crisis mundial. El marco de las Naciones Unidas es esencial para anclar este esfuerzo en el conjunto más amplio posible de normas compartidas.

xxx Combinado con el desarrollo de la capacidad (véanse las recomendaciones 4 y 5), este diálogo inclusivo sobre los enfoques de gobernanza puede ayudar a los Estados y a las empresas a actualizar sus enfoques y metodologías de regulación para responder a la aceleración de la IA. La vinculación con el grupo internacional de científicos potenciaría esa dinámica, de forma comparable a la relación entre el IPCC y la Conferencia de las Naciones Unidas sobre el Cambio Climático.

xxxi En paralelo a las reuniones existentes en Nueva York (como la Asamblea General)⁴ y en Ginebra, podría iniciarse un diálogo sobre políticas. Las reuniones bianuales podrían centrarse en las oportunidades entre los distintos sectores en

una reunión, y en los riesgos en la otra⁵. De cara al futuro, una reunión de este tipo sería un foro adecuado para compartir información sobre incidentes relacionados con la IA, como los que exceden o superan la capacidad de los organismos existentes.

xxxii Una parte de cada sesión de diálogo podría centrarse en los enfoques nacionales de los Estados Miembros, y una segunda parte podría hacerlo en la obtención de especialización y aportaciones de las principales partes interesadas, en particular las empresas tecnológicas y los representantes de la sociedad civil. Además de las sesiones oficiales de diálogo, la participación de las múltiples partes interesadas en las políticas de IA podría aprovechar otros mecanismos existentes más especializados, como la reunión *AI for Good* de la UIT, la reunión anual del Foro para la Gobernanza de Internet, el Foro Mundial de la UNESCO sobre la Ética de la Inteligencia Artificial y la eWeek de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD).

4 De forma análoga al foro político de alto nivel sobre el desarrollo sostenible que tuvo lugar en el contexto de los ODS bajo los auspicios del Consejo Económico y Social).

5 Las partes pertinentes del sistema de las Naciones Unidas podrían participar para poner de relieve las oportunidades y los riesgos existentes; por ejemplo, la UIT podría hacerlo en torno a las normas sobre la IA; la UIT, la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD), el Programa de las Naciones Unidas para el Desarrollo (PNUD) y la Oficina de Coordinación del Desarrollo lo harían en torno a las aplicaciones de la IA para los ODS; la UNESCO podría hacerlo en torno a la ética y la capacidad de gobernanza; la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH) lo haría sobre la rendición de cuentas en materia de derechos humanos basada en las normas y los mecanismos existentes; la Oficina de Asuntos de Desarme podría hacerlo en torno a la regulación de la IA en los sistemas militares; el PNUD lo haría sobre el apoyo a la capacidad nacional de desarrollo; el Foro para la Gobernanza de Internet lo haría en torno a la participación y el diálogo de las múltiples partes interesadas; y la Organización Mundial de la Propiedad Intelectual (OMPI), la Organización Internacional del Trabajo (OIT), la Organización Mundial de la Salud (OMS), la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), el Programa Mundial de Alimentos, el Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR), la UNESCO, el Fondo de las Naciones Unidas para la Infancia (UNICEF), la Organización Meteorológica Mundial y otros podrían hacerlo en torno a las aplicaciones sectoriales y la gobernanza.

Recomendación 2

Diálogo sobre políticas de gobernanza de la IA

Recomendamos el inicio de un diálogo sobre políticas intergubernamental y multilateral sobre la gobernanza de la IA dos veces al año, en paralelo a las reuniones que se celebren en las Naciones Unidas. Su objetivo sería:

- a) Compartir las mejores prácticas sobre una gobernanza de la IA que respete, proteja y promueva los derechos humanos y fomente el desarrollo, y que incluya tanto la búsqueda de oportunidades como la gestión de riesgos;
- b) Promover un entendimiento común sobre la implementación de medidas de gobernanza de la IA por parte de los desarrolladores y los usuarios de los sectores público y privado para mejorar la interoperabilidad internacional de la gobernanza de la IA;
- c) Compartir voluntariamente incidentes significativos de IA que hayan sobrepasado la capacidad de respuesta de los organismos estatales; y
- d) Debatir sobre los informes del grupo internacional de científicos expertos en IA, según proceda.

Intercambio de normas sobre la IA

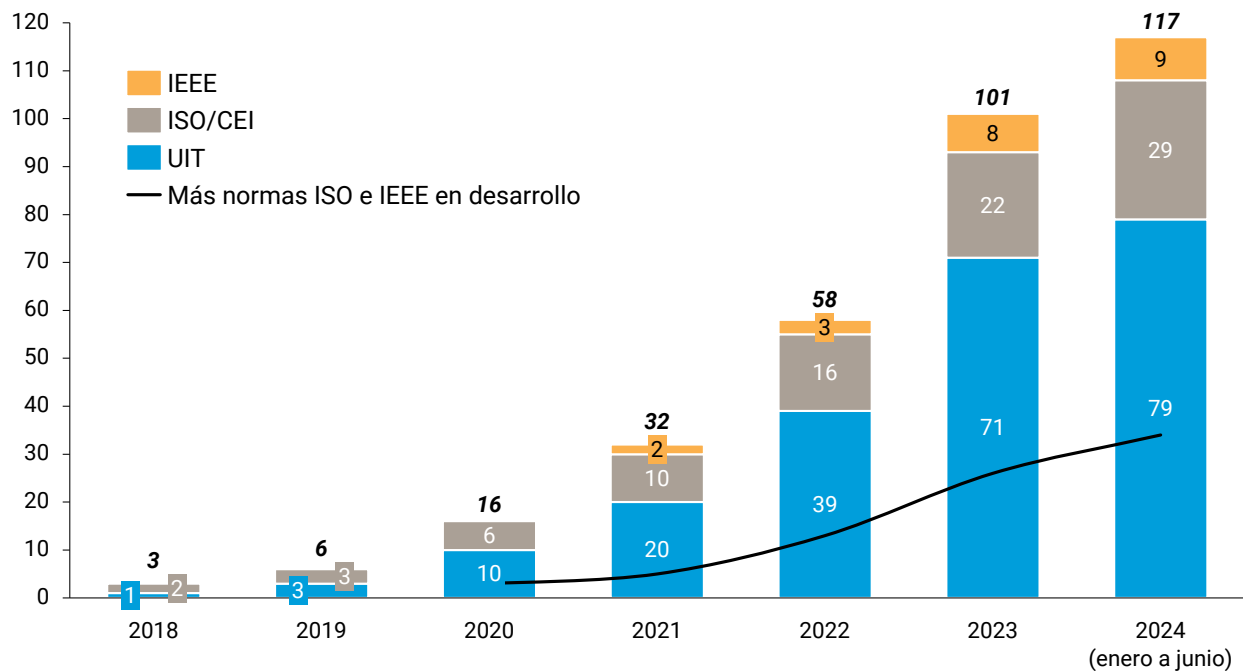
xxxiii Cuando se empezaron a desarrollar los sistemas de IA, existían pocas normas para ayudar a navegar o medir esta nueva frontera. Más recientemente, se ha producido una proliferación de normas. La Figura b) ilustra el creciente número de normas adoptadas por la UIT, la Organización Internacional de Normalización (ISO), la Comisión Electrotécnica Internacional (CEI) y el Instituto de Ingenieros Electricistas y Electrónicos (IEEE).

xxxiv No hay formulaciones comunes a estas organizaciones de normalización, y muchos de los términos utilizados habitualmente en relación con la IA - equidad, seguridad, transparencia - no tienen una definición consensuada. También hay una desconexión entre las normas que se

adoptaron con fines estrictamente técnicos o de validación interna y las que pretenden incorporar principios éticos más amplios. Ahora tenemos un nuevo conjunto de normas que no se basan en un entendimiento común de su significado o que están desvinculadas de los valores que pretendían defender.

xxxv Aprovechando la especialización del grupo internacional de científicos e incorporando a miembros de las diversas entidades nacionales e internacionales que han contribuido al establecimiento de normas, así como a representantes de las empresas tecnológicas y de la sociedad civil, el sistema de las Naciones Unidas podría funcionar como centro de intercambio de normas de aplicación mundial sobre la IA.

Figura (b): Número de normas relacionadas con la IA



Fuentes: IEEE, ISO/CEI, UIT, World Standards Cooperation (en base a la información de junio de 2023, ampliada con la inclusión de normas relacionadas con la IA).

Recomendación 3

Intercambio de normas sobre la IA

Recomendamos el establecimiento de una entidad de intercambio de normas sobre la IA que reúna a representantes de organizaciones nacionales e internacionales de normalización, empresas tecnológicas, la sociedad civil y representantes del grupo internacional de científicos. Este se encargaría de:

- Elaborar y mantener un registro de definiciones y normas aplicables para medir y evaluar los sistemas de IA;
- Debatir y evaluar las normas y los procesos para su elaboración; e
- Identificar las lagunas que requieran nuevas normas.

C. Beneficios comunes

xxxvi La Agenda 2030 para el Desarrollo Sostenible, con sus 17 ODS, puede aportar claridad en cuanto al propósito del desarrollo, el despliegue y los usos de la IA, orientando el arco de las inversiones hacia los desafíos del desarrollo mundial. Sin un enfoque integral e inclusivo de la gobernanza de la IA, podría perderse el potencial de esta para contribuir positivamente a los ODS, y su despliegue podría involuntariamente acentuar o exacerbar las desigualdades y los sesgos.

xxxvii La IA no es una panacea para los desafíos del desarrollo sostenible; es un componente dentro de un conjunto más amplio de soluciones. A fin de liberar realmente el potencial de la IA para hacer frente a los desafíos sociales, es esencial que exista colaboración entre los gobiernos, el mundo académico, la industria y la sociedad civil, de modo que las soluciones basadas en la IA sean inclusivas y equitativas.

xxxviii Gran parte de este objetivo depende del acceso al talento, la capacidad de computación y los datos, de forma que se favorezca el desarrollo de la diversidad cultural y lingüística. La infraestructura básica y los recursos para mantenerla también son condiciones indispensables.

xxxix En cuanto al talento, no todas las sociedades necesitan equipos de expertos informáticos para construir sus propios modelos. Sin embargo, tanto si la tecnología se adquiere como si se toma prestada o se fabrica, es necesaria una capacidad sociotécnica de referencia para comprender las capacidades y limitaciones de la IA y aprovechar adecuadamente los casos de uso de la IA abordando al mismo tiempo los riesgos específicos del contexto.

xi La computación es una de las mayores barreras de entrada en el ámbito de la IA. De los 100 clusters de computación de alto rendimiento más importantes del mundo, capaces de entrenar grandes modelos de IA, ninguno se encuentra en un país en desarrollo⁶. No es realista prometer un acceso a la computación que incluso los países y empresas más ricos tienen dificultades para conseguir. Más bien pretendemos fijar un nivel mínimo en la brecha de la IA para aquellos que no pueden conseguir los facilitadores necesarios por otros medios, por ejemplo, apoyando iniciativas de modelos de desarrollo de IA distribuidos y federados.

xli En cuanto a los datos, en el contexto de la IA es habitual hablar de uso indebido de los datos (violaciones de la privacidad) o de usos desaprovechados de los datos (no explotar los conjuntos de datos existentes). Sin embargo, un problema conexo es la falta de datos, que afecta a grandes partes del planeta, en las que son escasos. El hecho de no reflejar la diversidad lingüística y cultural del mundo se ha relacionado con el sesgo de los sistemas de IA, pero también puede suponer una oportunidad perdida para que esas comunidades accedan a los beneficios de la IA.

xlii Para apoyar la participación inclusiva y efectiva de todos los Estados Miembros en el ecosistema de la IA es necesario un conjunto de recursos compartidos—incluidos modelos abiertos—y en este sentido los enfoques globales tienen claras ventajas.

⁶ Indicador proxy, ya que la mayoría de los clusters de computación de alto rendimiento no disponen de unidades de procesamiento gráfico (GPU) y su uso para la IA avanzada es limitado.

Red de desarrollo de la capacidad

xliii La creciente demanda pública y privada de capacidad humana y de otro tipo en materia de IA coincide con la aparición de centros de excelencia nacionales, regionales y público-privados en IA que desempeñan funciones internacionales de desarrollo de la capacidad. Una red mundial puede servir como plataforma de encuentro que amplíe el abanico de posibles colaboraciones y mejore la interoperabilidad de los enfoques de creación de capacidad.

xliv Desde los Objetivos de Desarrollo del Milenio (ODM) hasta los ODS, las Naciones Unidas llevan mucho tiempo promoviendo el desarrollo de la capacidad de las personas y las instituciones⁷. Una red de instituciones afiliadas a las Naciones Unidas podría aumentar las opciones de los países que buscan alianzas para el desarrollo de la capacidad. También podría incentivar la creación de nuevos centros nacionales de excelencia que estimularan el desarrollo de ecosistemas locales de innovación en materia de inteligencia artificial siguiendo enfoques interoperables acordes con los compromisos normativos de las Naciones Unidas.

xlv Una red de este tipo promovería un paradigma alternativo de desarrollo de tecnologías de IA: ascendente, transversal, abierto y colaborativo. Los esfuerzos a nivel nacional podrían seguir empleando herramientas de diagnóstico como la metodología de la UNESCO de evaluación del estadio de preparación para la IA que ayuden a identificar las lagunas a nivel nacional, y la red internacional ayudaría a colmarlas.

⁷ Gracias a la labor de la UNESCO, la OMPI y otras organizaciones, las Naciones Unidas han contribuido a mantener la rica diversidad de culturas y tradiciones de creación de conocimientos en todo el mundo. La Universidad de las Naciones Unidas lleva mucho tiempo comprometida con la creación de capacidad a través de la educación superior y la investigación, y el Instituto de las Naciones Unidas para Formación Profesional e Investigaciones ha ayudado a capacitar a funcionarios en ámbitos fundamentales para el desarrollo sostenible. La metodología de la UNESCO de evaluación del estadio de preparación es una herramienta clave para apoyar a los Estados Miembros en su aplicación de la Recomendación sobre la Ética de la Inteligencia Artificial de esta Organización. Otros ejemplos son la Academia de la OMS en Lyon (Francia), el Instituto Virtual de Comercio y Desarrollo de la UNCTAD, el Programa de las Naciones Unidas de Becas sobre Desarme dirigido por la Oficina de Asuntos de Desarme y los programas de desarrollo de la capacidad dirigidos por la UIT y el PNUD.

Recomendación 4

Red de desarrollo de la capacidad

Recomendamos la creación de una red de desarrollo de la capacidad en materia de inteligencia artificial que reúna a una serie de centros de desarrollo de la capacidad afiliados a las Naciones Unidas poniendo conocimientos especializados, computación y datos de entrenamiento de IA a disposición de los agentes clave. El objetivo de la red sería:

- a) Acelerar y alinear los esfuerzos regionales y mundiales de capacitación en materia de inteligencia artificial apoyando la creación de redes entre ellos;
- b) Crear capacidad de gobernanza de la IA entre los funcionarios públicos para garantizar el respeto, la protección y el cumplimiento de los derechos humanos y promover el desarrollo;
- c) Poner instructores, computación y datos de entrenamiento de IA de los múltiples centros a disposición de los investigadores y emprendedores sociales que busquen aplicar la IA a casos de interés público local, entre otras cosas, a través de:
 - i) Protocolos que permitan a los equipos de investigación interdisciplinarios y a los empresarios de entornos con escasez de ordenadores acceder a la informática disponible para entrenar/afinar y aplicar sus modelos de forma adecuada a los contextos locales;
 - ii) Entornos controlados para poner a prueba posibles soluciones de IA y aprender sobre la marcha;
 - iii) Un conjunto de oportunidades educativas en línea sobre IA dirigidas a estudiantes universitarios, jóvenes investigadores, emprendedores sociales y funcionarios del sector público; y
 - iv) Un programa de becas para que personas prometedoras pasen una temporada en instituciones académicas o empresas tecnológicas.

Fondo mundial para la IA

xlvi Muchos países se enfrentan a restricciones fiscales y de recursos que limitan su capacidad para utilizar la IA de forma adecuada y eficaz. A pesar de los esfuerzos de desarrollo de la capacidad (recomendación 4), es posible que algunos sigan sin poder acceder a la capacitación, la computación, los modelos y los datos de entrenamiento sin apoyo internacional. También es posible que, sin este apoyo, los esfuerzos de financiación no puedan ampliarse a escala.

xlvii Nuestra intención al proponer un fondo no es garantizar el acceso a recursos y capacidades avanzadas de computación. Puede que la respuesta no siempre sea más computación. También necesitamos mejores formas de conectar talento, computación y datos. El propósito del fondo sería subsanar las carencias subyacentes de capacidad y colaboración de quienes no pueden acceder a los facilitadores necesarios, de modo que:

- a. Los países necesitados puedan acceder a los facilitadores de la IA reduciendo con ello la brecha de la IA;
- b. La colaboración en el desarrollo de la capacidad en materia de IA genere hábitos de cooperación y reduzca la competencia geopolítica;
- c. Los países con enfoques reguladores divergentes tengan incentivos para desarrollar plantillas comunes de gestión de los datos, los modelos y las aplicaciones en favor de los retos sociales relacionados con los ODS y los avances científicos.

xlvi Este enfoque de interés público hace que el fondo complemente la propuesta de una red de desarrollo de la capacidad en materia de inteligencia artificial, a la que el fondo también canalizaría recursos. El fondo proporcionaría una capacidad independiente de seguimiento del impacto y podría obtener y agrupar contribuciones en especie, incluso de entidades del sector privado, para poner a disposición de los interesados programas de capacitación relacionados con la IA, tiempo, computación, modelos y conjuntos de datos seleccionados, a un coste inferior al del mercado. De este modo nos aseguraríamos de que amplias zonas del mundo no se queden atrás y puedan aprovechar la IA para los ODS en diferentes contextos.

xlvi Redunda en interés de todos garantizar la cooperación en el mundo digital, al igual que en el mundo físico. Se pueden establecer analogías con los esfuerzos para combatir el cambio climático, donde los costos de transición, mitigación o adaptación no se distribuyen uniformemente y la asistencia internacional es esencial para ayudar a los países con recursos limitados a que puedan unirse al esfuerzo mundial de hacer frente a un problema planetario.

Recomendación 5

Fondo mundial para la IA

Recomendamos la creación de un fondo mundial para la IA con el fin de reducir la brecha existente en este ámbito. Gestionado por una estructura de gobierno independiente, el fondo recibiría contribuciones financieras y en especie de fuentes públicas y privadas y las desembolsaría, también a través de la red de desarrollo de la capacidad, para mejorar el acceso a facilitadores de la IA orientados a impulsar el empoderamiento local en pro de los ODS como:

- a) Recursos informáticos compartidos para la elaboración y el perfeccionamiento de modelos por parte de los desarrolladores de IA de países que carecen de la capacidad local adecuada o de los medios para adquirirla;
- b) Entornos y herramientas de análisis comparativo y de pruebas para incorporar las mejores prácticas en la elaboración de modelos seguros y fiables y en la gobernanza de los datos;
- c) Soluciones de gobernanza, seguridad e interoperabilidad de aplicación mundial;
- d) Conjuntos de datos e investigaciones sobre cómo combinar datos y modelos para proyectos relacionados con los ODS; y
- e) Un repositorio de modelos de IA y conjuntos de datos seleccionados para los ODS.

Marco mundial de datos sobre IA

- I** El acceso a los datos de entrenamiento de IA, a través del mercado u otros mecanismos, es un facilitador esencial para el desarrollo de ecosistemas locales de IA, especialmente en países, comunidades, regiones y grupos demográficos con “falta” de datos (véase la sección anterior sobre “beneficios comunes”).
- li** Solamente la acción colectiva mundial puede incentivar la interoperabilidad, la administración, la preservación de la privacidad, el empoderamiento y la mejora de los derechos de manera que se promueva una competencia al alza entre jurisdicciones en favor de la protección de los derechos humanos y otros compromisos acordados, la disponibilidad de datos y la compensación justa a los titulares de los datos en la gestión de la recopilación, la generación, el uso y la monetización de los datos de entrenamiento de IA. Este objetivo justifica nuestra propuesta de un marco mundial de datos sobre IA.
- lii** Este marco no crearía nuevos derechos relacionados con los datos. Más bien abordaría cuestiones de disponibilidad, interoperabilidad y uso de los datos de entrenamiento de IA. Ayudaría a establecer un entendimiento común sobre cómo armonizar los distintos marcos nacionales y regionales de protección de datos. También podría promover el desarrollo de ecosistemas locales de IA que apoyen la diversidad cultural y lingüística y que contengan la concentración económica.
- liii** Estas medidas podrían complementarse con la promoción de conjuntos de datos comunes y disposiciones para albergar fideicomisos de datos en esferas relevantes para los ODS, a partir de plantillas de acuerdos para mantener y compartir datos de manera justa, segura y equitativa. La preparación de estas plantillas

y el almacenamiento y análisis de los datos conservados en conjuntos de datos comunes o en fideicomisos podrían contar con el apoyo de la red de desarrollo de la capacidad y el fondo mundial para la IA propuestos (véanse las recomendaciones 4 y 5).

- liv** Las Naciones Unidas están en una posición única para apoyar el establecimiento de principios globales y acuerdos prácticos para la gobernanza y el uso de los datos de entrenamiento de IA en consonancia con los compromisos internacionales acordados en materia de derechos humanos, propiedad intelectual y desarrollo sostenible, basándose en años de trabajo de la comunidad de datos e integrándolos con los recientes avances en materia de ética y gobernanza de la IA. Se trata de una función análoga a la que desempeña la Comisión de las Naciones Unidas para el Derecho Mercantil Internacional en la promoción del comercio internacional mediante el desarrollo de marcos jurídicos y no jurídicos transfronterizos.
- lv** Del mismo modo, la Comisión de Ciencia y Tecnología para el Desarrollo y la Comisión de Estadística tienen en su agenda datos para el desarrollo y datos sobre los ODS. La Organización Mundial de la Propiedad Intelectual (OMPI) también está estudiando importantes cuestiones relacionadas con los contenidos, los derechos de autor y la protección de los conocimientos y la expresión cultural autóctonos.

Recomendación 6

Marco mundial de datos sobre IA

Recomendamos la creación de un marco mundial de datos sobre IA, desarrollado mediante un proceso iniciado por un organismo competente como la Comisión de las Naciones Unidas para el Derecho Mercantil Internacional y fundamentado en el trabajo de otras organizaciones internacionales, para:

- a) Establecer definiciones y principios relacionados con los datos para la gobernanza global de los datos de entrenamiento de IA, incluidos los extraídos de las mejores prácticas existentes, y promover la diversidad cultural y lingüística;
- b) Establecer normas comunes sobre la procedencia y el uso de los datos de entrenamiento de IA para una rendición de cuentas transparente y basada en los derechos en todas las jurisdicciones; e
- c) Implantar mecanismos de administración e intercambio de datos de configuración del mercado que permitan el desarrollo de ecosistemas locales de IA a nivel mundial, como por ejemplo:
 - i) Fideicomisos de datos;
 - ii) Mercados mundiales bien regulados de intercambio de datos anónimos para el entrenamiento de modelos de IA; y
 - iii) Acuerdos modelo para facilitar el acceso internacional a los datos y la interoperabilidad mundial, posiblemente como protocolos técnico-jurídicos del marco.

D. Esfuerzos coherentes

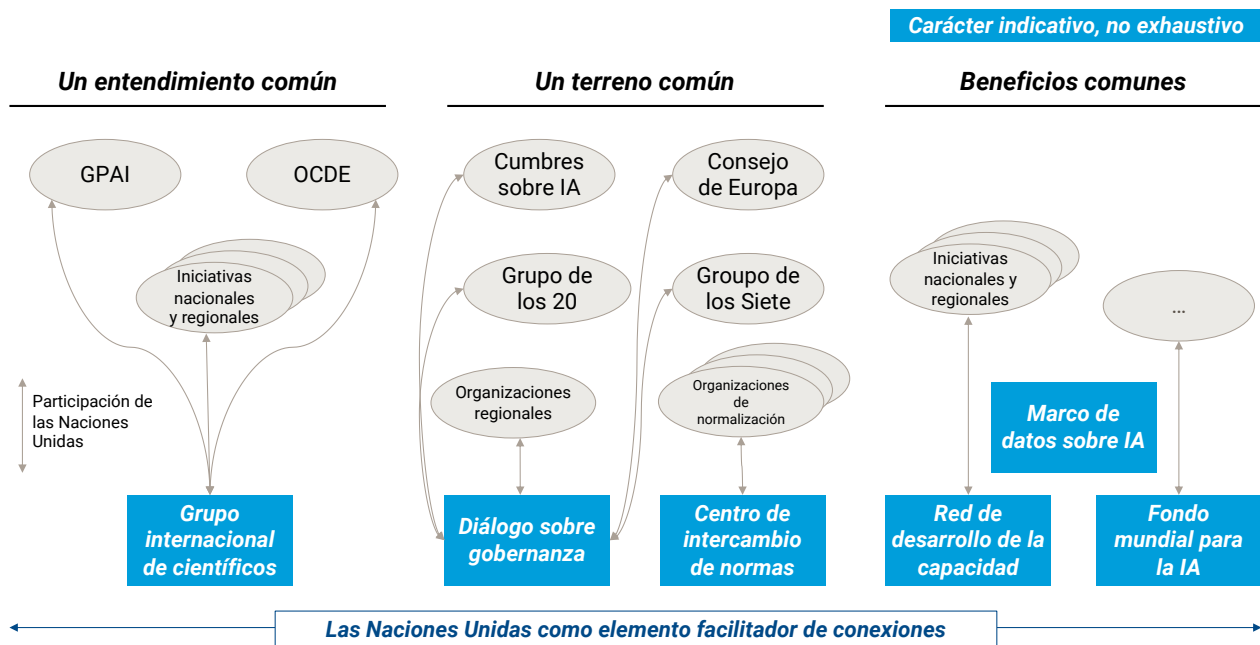
- lvi** Las propuestas mencionadas pretenden colmar las lagunas de representación, coordinación e implementación detectadas en el incipiente régimen internacional de gobernanza de la IA. Estas lagunas pueden tratar de subsanarse mediante alianzas y colaboración con las instituciones y los mecanismos existentes dedicados a promover un entendimiento común, un terreno común y beneficios comunes.
- lvii** Sin embargo, sin un punto focal específico en las Naciones Unidas que apoye y permita la coordinación entre estos y otros esfuerzos, el mundo carecerá del enfoque inclusivo en red, ágil y coherente necesario para una gobernanza eficaz y equitativa de la IA como tecnología transfronteriza, en rápida evolución y de uso general.
- lviii** El mosaico de normas e instituciones descrito en la sección anterior “Lagunas en la gobernanza global de la IA” refleja el reconocimiento generalizado de que la gobernanza de la IA es

una necesidad mundial. La heterogeneidad de esa respuesta exige cierto grado de esfuerzo coherente.

Una oficina de IA en la Secretaría General de las Naciones Unidas

- lix** Proponemos por tanto un mecanismo sencillo que actúe como elemento “aglutinante” para apoyar e impulsar las propuestas de este informe, por ejemplo mediante el establecimiento de alianzas, y que permita a su vez al sistema de las Naciones Unidas hablar con una sola voz en el cambiante ecosistema de gobernanza de la IA.
- lx** Esta sencilla y ágil capacidad, en forma de oficina de IA dentro de la Secretaría General de las Naciones Unidas, dependería del Secretario General y tendría la ventaja de estar conectada con todo el sistema de las Naciones Unidas sin estar vinculada a una parte del mismo. Esto es importante debido al futuro incierto de la IA y a la gran probabilidad de que impregne todos los aspectos del quehacer humano.

Figura (c): Función propuesta para las Naciones Unidas en el ecosistema internacional de gobernanza de la IA



Abreviaciones: GPAI: Alianza Mundial sobre la Inteligencia Artificial; OCDE: Organización de Cooperación y Desarrollo Económicos; SDO: organizaciones de normalización.

lxi Este órgano debería ser ágil, defender la inclusión y entablar activamente alianzas para acelerar la coordinación y la implementación, aprovechando con carácter prioritario los recursos y las funciones existentes en el sistema de las Naciones Unidas. El foco de atención debería estar en las aplicaciones civiles de la IA.

lxii El órgano debería estar formado, en parte, por personal de las Naciones Unidas cedido por organismos especializados y otras entidades del sistema de las Naciones Unidas como la UIT, la UNESCO, la Oficina del Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUDH), la UNCTAD, la Universidad de las Naciones Unidas y el Programa de las Naciones Unidas para el Desarrollo (PNUD). Asimismo, debería contar con la participación de las múltiples partes interesadas, incluidas las empresas, la

sociedad civil y el mundo académico, y trabajar en colaboración con las principales organizaciones ajenas a las Naciones Unidas (véase la Figura c)). De esa forma, las Naciones Unidas estarían en condiciones de habilitar las conexiones para lograr un entendimiento común, un terreno común y beneficios comunes en el ecosistema internacional de gobernanza de la IA.

lxiii La recomendación 7 se basa en una cuidadosa evaluación de los ámbitos en los que las Naciones Unidas pueden aportar valor añadido, incluidos aquellos en los que pueden liderar y ayudar a coordinar y aquellos en los que deben hacerse a un lado. La ONU también aporta las ventajas de los acuerdos institucionales existentes, incluidos los prenegociados sobre la financiación y los procedimientos administrativos bien establecidos y aceptados.

Recomendación 7

Una oficina de IA en la Secretaría General de las Naciones Unidas

Recomendamos la creación de una oficina de IA dentro de la Secretaría General, dependiente del Secretario General. Su organización debería ser sencilla y **ágil, aprovechando en la medida de lo posible las entidades competentes de las Naciones Unidas ya existentes**. Actuando como elemento “aglutinante” para apoyar e impulsar las propuestas de este informe, estableciendo alianzas e interactuando con otros procesos e instituciones, el mandato de la oficina incluiría:

- a)** Brindar apoyo al grupo internacional de científicos expertos en IA, al diálogo sobre políticas, al centro de intercambio de normas, a la red de desarrollo de la capacidad y, en la medida en que sea necesario, al fondo mundial y al marco mundial de datos sobre IA;
- b)** Participar en actividades de divulgación dirigidas a las múltiples partes interesadas, incluidas las empresas tecnológicas, la sociedad civil y el mundo académico, sobre cuestiones emergentes relacionadas con la IA; y
- c)** Asesorar al Secretario General sobre cuestiones relacionadas con la IA en coordinación con otras partes pertinentes del sistema de las Naciones Unidas para ofrecer una respuesta conjunta “de todo el sistema de las Naciones Unidas”.

E. Reflexiones sobre los modelos institucionales

lxiv Los debates sobre la IA a menudo se dividen en extremos. En nuestras consultas por todo el mundo, hemos hablado con quienes ven un futuro de bienes ilimitados proporcionados por sistemas de IA cada vez más baratos y útiles. También hablamos con quienes temen un futuro más oscuro, la división y el desempleo, e incluso la extinción⁸.

lxv No sabemos cuál es el futuro más probable, si el utópico o el distópico. Pero también somos conscientes de que la tecnología puede ir en una dirección que acabe con esta dualidad. Este informe se centra en las oportunidades y los riesgos a corto plazo, y se basa en datos científicos y hechos reales.

lxvi Las siete recomendaciones expuestas representan nuestra mayor esperanza de beneficiarnos de la IA al tiempo que minimizamos y mitigamos sus riesgos conforme va evolucionando. También somos conscientes de los desafíos prácticos que plantea la creación de instituciones internacionales a gran escala. Por eso proponemos un enfoque institucional en red, con un apoyo sencillo y ágil. Si los riesgos se agravan y las oportunidades aumentan, estos cálculos pueden cambiar.

lxvii Las guerras mundiales condujeron al moderno sistema internacional; el desarrollo de armas químicas, biológicas y nucleares cada vez más potentes desembocó en regímenes que limitan su propagación y promueven los usos pacíficos de las tecnologías subyacentes. La evolución en la comprensión de nuestra humanidad común condujo al sistema moderno de derechos humanos y a nuestros compromisos actuales con los ODS para todos. El cambio climático pasó de ser un problema aislado a convertirse en un desafío mundial.

lxviii Del mismo modo, la IA puede llegar a un nivel que requiera más recursos y más autoridad de lo que se propone en estas recomendaciones, funciones más estrictas de elaboración de normas, aplicación, supervisión, verificación y validación, ejecución forzosa, rendición de cuentas, reparación de daños y respuestas de emergencia. Así pues, consideramos prudente reflexionar sobre estos modelos institucionales. La sección final de este informe pretende contribuir a ese esfuerzo.

4. Un llamado a la acción

lxix Seguimos siendo optimistas sobre el futuro de la IA y su potencial beneficioso. Ese optimismo depende, sin embargo, del realismo sobre los riesgos y la insuficiencia de estructuras e incentivos actualmente existentes. La tecnología es demasiado importante y hay demasiado en juego como para confiar únicamente en las fuerzas del mercado y en un mosaico fragmentado de acciones nacionales y multilaterales.

lxx Las Naciones Unidas pueden ser el vehículo de un nuevo contrato social para la IA que garantice la aceptación mundial de un régimen de gobernanza que nos proteja y nos empodere a todos. Este contrato social garantizará que las oportunidades se distribuyan equitativamente y que los riesgos no recaigan sobre los más vulnerables ni se transmitan a las generaciones futuras, como ha pasado con el cambio climático.

lxxi Como grupo y como personas procedentes de diferentes campos de especialización, organizaciones y partes del mundo, esperamos continuar este esencial debate. Junto con las muchas otras personas con las que nos hemos relacionado en este viaje y la comunidad global a la que representan, esperamos que este informe contribuya a nuestros esfuerzos combinados por gobernar la IA en beneficio de la humanidad.

⁸ Véase <https://safe.ai/work/statement-on-ai-risk>.

Figura (d): Reunión del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial, celebrada en Singapur el 29 de mayo de 2024



1. Introducción

- 1 El Órgano Asesor de Alto Nivel sobre Inteligencia Artificial del Secretario General se creó para realizar análisis y formular recomendaciones sobre la gobernanza internacional de la inteligencia artificial. Nuestros miembros son diversos en cuanto a geografía y género, disciplina y edad; contamos con expertos procedentes de los gobiernos, la sociedad civil, el sector privado y el mundo académico. Los intensos y amplios debates han dado lugar a un amplio consenso (como refleja nuestro [informe provisional](#))¹, sobre la existencia de un déficit de gobernanza global de la IA. En ese informe articulamos los principios rectores de la gobernanza de la IA y las funciones que podrían ser necesarias a escala internacional.
- 2 En los meses siguientes recibimos numerosos comentarios y consultas. Entre ellos cabe citar 18 “inmersiones en profundidad” en esferas temáticas concretas con más de 500 participantes expertos, más de 250 comunicaciones escritas de más de 150 organizaciones y 100 particulares de todas las regiones, una encuesta global denominada AI risk pulse check con unos 350 expertos encuestados de todas las regiones, y consultas y reuniones informativas periódicas con los Estados Miembros, las entidades de las Naciones Unidas y otros grupos de partes interesadas producidas en más de 40 encuentros en todas las regiones.² Los miembros del Órgano Asesor también han participado ampliamente en foros de todo el mundo, han mantenido más de 100 debates virtuales y han celebrado tres reuniones plenarias presenciales en Nueva York, Ginebra y Singapur.
- 3 El presente informe final tiene, por tanto, muchos autores. Aunque no puede reflejar toda la riqueza y diversidad de las opiniones expresadas, demuestra nuestro compromiso común de garantizar que la IA se desarrolle, despliegue y utilice de forma que beneficie a toda la humanidad, y de garantizar que la IA sea gobernada de forma eficaz e inclusiva a nivel internacional.
- 4 Este informe reafirma las conclusiones del informe provisional del Órgano Asesor sobre las oportunidades y los elementos facilitadores, los riesgos y los desafíos; también reitera la necesidad de una gobernanza global de la IA y formula siete recomendaciones.
- 5 Entre ellas figura la creación de un grupo de científicos expertos en IA que promueva un entendimiento común de la capacidad, las oportunidades, los riesgos y las incertidumbres de la IA. A partir de este entendimiento común, necesitamos mecanismos para encontrar un terreno común sobre cómo debe gobernarse la IA a nivel internacional. Conseguirlo depende de un diálogo regular y del desarrollo de normas aceptables y aplicables para todos.
- 6 El informe también formula recomendaciones sobre los beneficios comunes que están orientadas a garantizar el reparto equitativo de los beneficios de la IA, lo cual puede depender del acceso a modelos o capacidades como el talento, la capacidad de computación y los datos. Entre ellas se incluye una red para el desarrollo de la capacidad, un fondo mundial para la IA y un marco mundial de datos sobre IA.

1 Véase <https://un.org/ai-advisory-body>.

2 Véase, en el anexo C, una sinopsis de las consultas mantenidas.

- 7** Para hacer posible estos esfuerzos, colaborar con otras iniciativas e instituciones a fin de abordar las preocupaciones y las oportunidades de la IA y garantizar que el sistema de las Naciones Unidas hable con una sola voz sobre la IA, proponemos la creación de una oficina de IA dentro de la Secretaría de las Naciones Unidas.
- 8** Aunque hemos considerado la posibilidad de recomendar la creación de un organismo internacional para la IA, por el momento no recomendamos esta medida; sin embargo, reconocemos la necesidad de que la gobernanza siga el ritmo de los avances tecnológicos.
- 9** Más allá de los debates y procesos multilaterales inmediatos en los que participan los gobiernos, nuestro informe también va dirigido a la sociedad civil y al sector privado, a los investigadores y a las personas interesadas de todo el mundo. Somos plenamente conscientes de que los ambiciosos objetivos que hemos reseñado solo podrán alcanzarse con una participación mundial multisectorial.
- 10** En general, creemos que el futuro de esta tecnología sigue abierto. Así lo ha corroborado nuestra inmersión en profundidad en la dirección que está tomando la tecnología y en el debate entre los enfoques abiertos y cerrados de su desarrollo (véase el recuadro 9). Una alternativa de futuro sería el desarrollo de modelos más grandes y potentes en un número cada vez menor de empresas. Otra podría ser un panorama de innovación global más diverso dominado por modelos de IA interoperables de dimensiones pequeñas y medianas que ofrezcan multitud de aplicaciones sociales y económicas. Nuestras recomendaciones pretenden hacer más probable esta última opción, reconociendo al mismo tiempo los riesgos.
- 11** Desde su fundación, las Naciones Unidas se han comprometido a promover el progreso económico y social de todos los pueblos³. En primer lugar, los Objetivos de Desarrollo del Milenio trataron de establecer metas ambiciosas para que todos los pueblos del mundo tuvieran oportunidades económicas; después, los Objetivos de Desarrollo Sostenible (ODS) trataron de conciliar la necesidad de desarrollo con las limitaciones ambientales de nuestro planeta. El creciente desarrollo, despliegue y uso de herramientas y sistemas de IA plantea el próximo gran desafío para garantizar que abracemos juntos nuestro futuro digital en lugar de ampliar nuestra brecha digital.
- 12** La gobernanza inclusiva de la IA es posiblemente uno de los retos de gobernanza más difíciles a los que se enfrentarán las Naciones Unidas. Se observa un desajuste entre el papel dominante del sector privado en la IA y el orden westfaliano de la política internacional. Los Estados se sienten tentados por el potencial de poder y prosperidad de la IA en un momento de intensa competencia geopolítica. Muchas sociedades se encuentran todavía al margen del desarrollo, despliegue y uso de la IA, mientras que unas pocas están presas de un entusiasmo mezclado con preocupación por el impacto transversal de la IA.
- 13** A pesar de los desafíos, no hay otra opción. Lo que está en juego es demasiado importante para las Naciones Unidas, sus Estados Miembros y la comunidad en general, cuyas aspiraciones representa la Organización. Esperamos que este informe proporcione algunas pautas que contribuyan a nuestros esfuerzos concertados para gobernar la IA en beneficio de la humanidad.

3 Este compromiso incluye el comercio, la inversión extranjera directa y la transferencia de tecnología como facilitadores del desarrollo a largo plazo.

A. Oportunidades y elementos facilitadores

14 La IA está transformando el mundo. Este conjunto de tecnologías⁴ ofrece un enorme potencial de beneficio, que abarca desde la exploración de nuevas áreas de investigación científica (véase el recuadro 1) y la optimización de las redes energéticas, hasta la mejora de la salud pública o la agricultura.⁵ Si se materializan, las grandes

posibilidades que ofrece el uso de herramientas de IA para las personas, los sectores económicos, la investigación científica y otros ámbitos de interés público podrían contribuir significativamente a impulsar nuestras economías (véase el recuadro 2) así como a mejorar nuestras sociedades. La IA al servicio del interés público - por ejemplo, para pronosticar y hacer frente a pandemias, inundaciones, incendios forestales e inseguridad alimentaria - podría incluso contribuir a avanzar hacia la consecución de los ODS.

Recuadro 1: Potencial de la IA para los avances científicos

La IA bien podría ser el siguiente gran paso en los avances científicos basándose en el legado transformador de Internet. La World Wide Web facilitó el intercambio de grandes cantidades de datos experimentales, artículos científicos y documentación entre científicos. La IA se construye sobre esta base, posibilitando el análisis de amplios conjuntos de datos, descubriendo patrones ocultos, construyendo nuevas hipótesis y asociaciones, y acelerando el ritmo de los hallazgos, incluso mediante experimentos a escala con robótica automatizada.

El impacto de la IA en la ciencia abarca grandes disciplinas. De la biología a la física, de las ciencias ambientales a las ciencias sociales, la IA se está integrando en los flujos de trabajo de investigación y está acelerando la producción de conocimientos científicos. Algunas de las afirmaciones que hoy se vierten podrían resultar exageradas, mientras que otras ya se han demostrado y su potencial a largo plazo parece prometedor^a.

Por ejemplo, en biología, el problema de los últimos 50 años relativo al plegamiento de las proteínas y la predicción de su estructura se ha resuelto en gran medida gracias a la IA. Este logro incluye la predicción de la estructura de más de 200 millones de proteínas, y la base de datos de acceso libre resultante es utilizada por más de 2 millones de científicos en más de 190 países en el momento de redactar este informe, muchos de los cuales trabajan en las enfermedades olvidadas. Desde entonces se ha ampliado a otras biomoléculas de la vida, el ADN, el ARN y los ligandos y sus interacciones.

a Véase John Jumper y otros, "Highly accurate protein structure prediction with AlphaFold", *Nature*, vol. 596 (julio de 2021), págs. 583 a 589; véase también Josh Abramson y otros, "Accurate structure prediction of biomolecular interactions with AlphaFold 3", *Nature*, vol. 630, págs. 493 a 500 (mayo de 2024).

4 Según la Organización de Cooperación y Desarrollo Económicos (OCDE): "Un sistema de IA es un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de la información que recibe, cómo generar información de salida como contenidos, predicciones, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los diferentes sistemas de IA varían en sus niveles de autonomía y adaptabilidad después de su implementación" (Véase <https://oecd.ai/en/wonk/ai-system-definition-update>).

5 Sin embargo, para valorar las afirmaciones sobre los beneficios de la IA, creemos que es necesaria una evaluación rigurosa por parte de los expertos en la materia. El desarrollo de la IA en beneficio de la humanidad debe basarse en pruebas científicas y en una evaluación exhaustiva de sus ventajas e inconvenientes y de las alternativas. Además de la investigación científica, las ciencias sociales también se están transformando.

Recuadro 1: Potencial de la IA para los avances científicos (Continuación)

En el caso del alzhéimer, el párkinson y la esclerosis lateral amiotrófica, los expertos que utilizan la IA están identificando biomarcadores de la enfermedad y prediciendo las respuestas al tratamiento, lo que mejora notablemente la precisión y la rapidez del diagnóstico y el desarrollo de tratamientos^b. En general, la IA está ayudando a avanzar en la medicina de precisión (por ejemplo en las enfermedades neurodegenerativas), adaptando los tratamientos a los perfiles genéticos y clínicos. La tecnología de la IA también está ayudando a acelerar el descubrimiento y desarrollo de nuevos compuestos químicos^c.

En radioastronomía, la velocidad y la escala de los datos recogidos por instrumentos modernos como la Matriz de un Kilómetro Cuadrado pueden desbordar los métodos tradicionales. La IA puede marcar la diferencia, entre otras cosas, ayudando a seleccionar en qué parte de los datos hay que centrarse para obtener nuevos conocimientos. A través de la técnica de la “agrupación no supervisada”, la IA puede detectar patrones en los datos sin que se le indique específicamente lo que debe buscar^d. La aplicación de la IA a la investigación en ciencias sociales podría ofrecer una visión profunda de la compleja dinámica humana, lo que mejoraría la comprensión de las tendencias socioeconómicas.

Con el tiempo, al posibilitar niveles sin precedentes de interdisciplinariedad, la IA podría diseñarse y desplegarse para generar nuevos ámbitos científicos, del mismo modo que la bioinformática y la neuroinformática surgieron de la integración de técnicas computacionales con la investigación biológica y neurológica. La capacidad de la IA para integrar y analizar diversos conjuntos de datos de ámbitos como el cambio climático, la seguridad alimentaria y la salud pública podría abrir vías de investigación que tiendan puentes entre estos campos tradicionalmente separados, si se hace de forma responsable.

La IA también puede incrementar el impacto de la investigación científica en las políticas públicas al permitir la validación de hipótesis complejas, por ejemplo, combinando modelos climáticos con datos agrícolas para predecir los riesgos para la seguridad alimentaria y relacionando estos conocimientos con los resultados en materia de salud pública. Otra perspectiva es el impulso de la ciencia ciudadana y el aprovechamiento de los conocimientos y datos locales para los desafíos mundiales.

b Isaias Ghebrehwet y otros, “Revolutionizing personalized medicine with generative AI: a systematic review”, *Artificial Intelligence Review*, vol. 57, núm. 127 (abril de 2024).

c Amil Merchant y otros, “Scaling deep learning for materials discovery”, *Nature*, vol. 624, págs. 80 a 85 (noviembre de 2023).

d Zack Savitsky, “Astronomers are enlisting AI to prepare for a data downpour”, *MIT Technology Review*, 20 de mayo de 2024.

Recuadro 2: Oportunidades económicas de la IA

Desde la Revolución Industrial, unas pocas innovaciones han acelerado espectacularmente el progreso económico. Estas primeras “tecnologías de uso general” han reconfigurado múltiples sectores e industrias. El último gran cambio llegó con los ordenadores y la era digital. Estas tecnologías transformaron las economías y aumentaron la productividad en todo el mundo, pero su impacto tardó decenios en dejarse sentir plenamente.

La inteligencia artificial generativa está rompiendo la tendencia de adopción lenta. Los expertos creen que sus efectos transformadores se verán en esta década. Esta rápida integración significa que los nuevos avances en IA podrían remodelar rápidamente las industrias, cambiar los procesos de trabajo y aumentar la productividad. Así pues, la rápida adopción de la IA podría transformar nuestras economías y sociedades de un modo sin precedentes.

Los beneficios económicos de la IA pueden ser considerables. Aunque es difícil predecir todas las ramificaciones de la IA en nuestras complejas economías, las previsiones indican que la IA podría aumentar significativamente el producto interno bruto mundial, lo que tendría un impacto relevante en casi todos los sectores. Para las empresas, especialmente las microempresas y las pequeñas y medianas empresas, la IA puede ofrecer acceso a herramientas avanzadas de análisis y automatización que antes solo estaban al alcance de las grandes corporaciones. La amplitud de su aplicabilidad sugiere que la IA podría ser una tecnología de uso general. Como tal, la IA podría facilitar la productividad de los particulares, las pequeñas y grandes empresas y otras organizaciones de sectores tan diversos como el comercio minorista, la fabricación y las operaciones, la atención de la salud y el sector público, en economías desarrolladas y en desarrollo^a. Para ello será necesaria una amplia adopción de la IA dentro de los sectores y entre ellos, su aplicación en usos que mejoren la productividad, y una IA que haga más productivos a los trabajadores y que introduzca nuevas actividades económicas a escala. También se requerirán inversión y aumento de la intensidad de capital, innovaciones colaborativas, cambios en los procesos y la organización, disposición operativa de la fuerza de trabajo y políticas favorables.

Sin embargo, aunque la IA puede mejorar la productividad, impulsar el comercio internacional y aumentar los ingresos, también se espera que afecte al trabajo. Las investigaciones sugieren que la IA podría ayudar a los trabajadores en algunos casos y eliminar puestos de trabajo en otros^b. Los estudios realizados, entre ellos los de la Organización Internacional del Trabajo (OIT), sugieren que, en un futuro próximo, la IA, probablemente, ayudará más que sustituirá a los trabajadores^c.

Los estudios han demostrado que se espera que la probable pérdida de puestos de trabajo se produzca de forma diferente en economías que se encuentran en distintas fases de desarrollo^d. Mientras que las economías avanzadas están más expuestas, también están mejor preparadas para aprovechar la IA y complementar su fuerza de trabajo. Los países de ingreso bajo y mediano pueden tener menos capacidad para aprovechar esta tecnología. Además, la integración de la IA en la fuerza de trabajo puede afectar de forma desproporcionada a determinados grupos demográficos, ya que las mujeres podrían correr mayor riesgo de perder su puesto de trabajo en algunos sectores.

a James Manyika y Michael Spence, “The coming AI economic revolution: can artificial intelligence reverse the productivity slowdown?”, *Foreign Affairs*, 24 de octubre de 2023.

b Erik Brynjolfsson y otros, “Generative AI at work”, National Bureau of Economic Research, documento de trabajo 31161, 2023; véase también Shakked Noy y Whitney Zhang, “Experimental evidence on the productivity effects of generative artificial intelligence”, *Science*, vol. 381, núm. 6654, págs. 187 a 192 (julio de 2023).

c Pawel Gmyrek y otros, *Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality* (Ginebra, OIT, 2023).

d Mauro Cazzaniga y otros, “Gen-AI: artificial intelligence and the future of work”, nota de debate del personal SDN2024/001 (Washington, D. C.: Fondo Monetario Internacional, 2024).

Recuadro 2: Oportunidades económicas de la IA (Continuación)

Si no se realizan esfuerzos específicos y coordinados para cerrar la brecha digital, no se aprovechará la capacidad potencial de la IA para apoyar el desarrollo sostenible y la mitigación de la pobreza, lo que provocará que amplios segmentos de la población mundial sigan estando en desventaja en un entorno tecnológico en rápida evolución y agudizará las desigualdades actuales.

Para integrar con éxito la IA en la economía mundial, necesitamos una gobernanza eficaz que gestione los riesgos y garantice resultados justos. Esto implica, entre otras opciones, la creación de entornos controlados de regulación para probar los sistemas de IA, el fomento de la cooperación internacional en materia de normas y el establecimiento de mecanismos para evaluar continuamente el impacto de la IA en los mercados de trabajo y en la sociedad. Aparte de estrategias nacionales sólidas de IA y apoyo internacional, ello exige específicamente:

- **Desarrollo de aptitudes:** Implementar programas de educación y capacitación que permitan desarrollar competencias de IA en toda la fuerza de trabajo, desde la alfabetización digital básica hasta los conocimientos técnicos avanzados, para preparar a los trabajadores para un futuro enriquecido por la IA.
- **Infraestructura digital:** Inversión significativa en infraestructura digital, especialmente en los países en desarrollo, para reducir la brecha de la IA y facilitar su adopción generalizada.
- **Integración en el lugar de trabajo:** Aprovechar el diálogo social y las alianzas público-privadas para gestionar la integración de la IA en el lugar de trabajo, garantizar la participación de los trabajadores en el proceso y proteger los derechos laborales.
- **Consideraciones sobre la cadena de valor:** Garantizar condiciones de trabajo decentes a lo largo de toda la cadena de valor de la IA, incluidas áreas que a menudo se pasan por alto, como la anotación de datos y la moderación de contenidos, para un desarrollo equitativo de la IA.

Figura 1: Algunas oportunidades de desarrollo y riesgos de la IA en los mercados emergentes



Oportunidades

- Nuevos productos y modelos de negocio que incluyen soluciones de aceleración, soluciones para la base de la pirámide y un acceso más fácil al crédito
- Automatización de los principales procesos empresariales, reduciendo los costos de producción
- Desarrollo del capital humano
- Innovación en los servicios públicos



Riesgos

- Obsolescencia de la vía tradicional de crecimiento económico basada en las exportaciones
- Aumento de la brecha digital y tecnológica
- Transformación de los requisitos laborales y alteración de las funciones laborales tradicionales
- Privacidad, seguridad y confianza pública

Fuente: Corporación Financiera Internacional.

B. Facilitadores clave para aprovechar la IA en beneficio de la humanidad

- 15** Las posibles oportunidades que surjan del desarrollo y el uso de la IA no necesariamente se materializarán o aprovecharán de forma equitativa. En mayo de 2024, un estudio sobre la financiación de proyectos de inteligencia artificial para avanzar hacia la consecución de los ODS reveló que solo el 10 % de las subvenciones asignadas se habían destinado a organizaciones con sede en países de ingreso bajo o mediano; en el caso del capital privado, la cifra fue del 25 % (más del 90 % de este en China).⁶

C. La gobernanza como facilitador clave

- 16** Para que los beneficios de la IA se materialicen plenamente y vayan más allá de unas pocas personas en unos pocos países, es necesario que existan facilitadores a nivel mundial. Garantizar que la IA se utiliza para el bien común y que sus oportunidades se distribuyen equitativamente requerirá una acción gubernamental e intergubernamental que incentive la participación del sector privado, el mundo académico y la sociedad civil. Cualquier marco de gobernanza debería establecer incentivos a nivel mundial para promover objetivos más amplios e inclusivos y ayudar a identificar y afrontar las desventajas.

D. Riesgos y desafíos

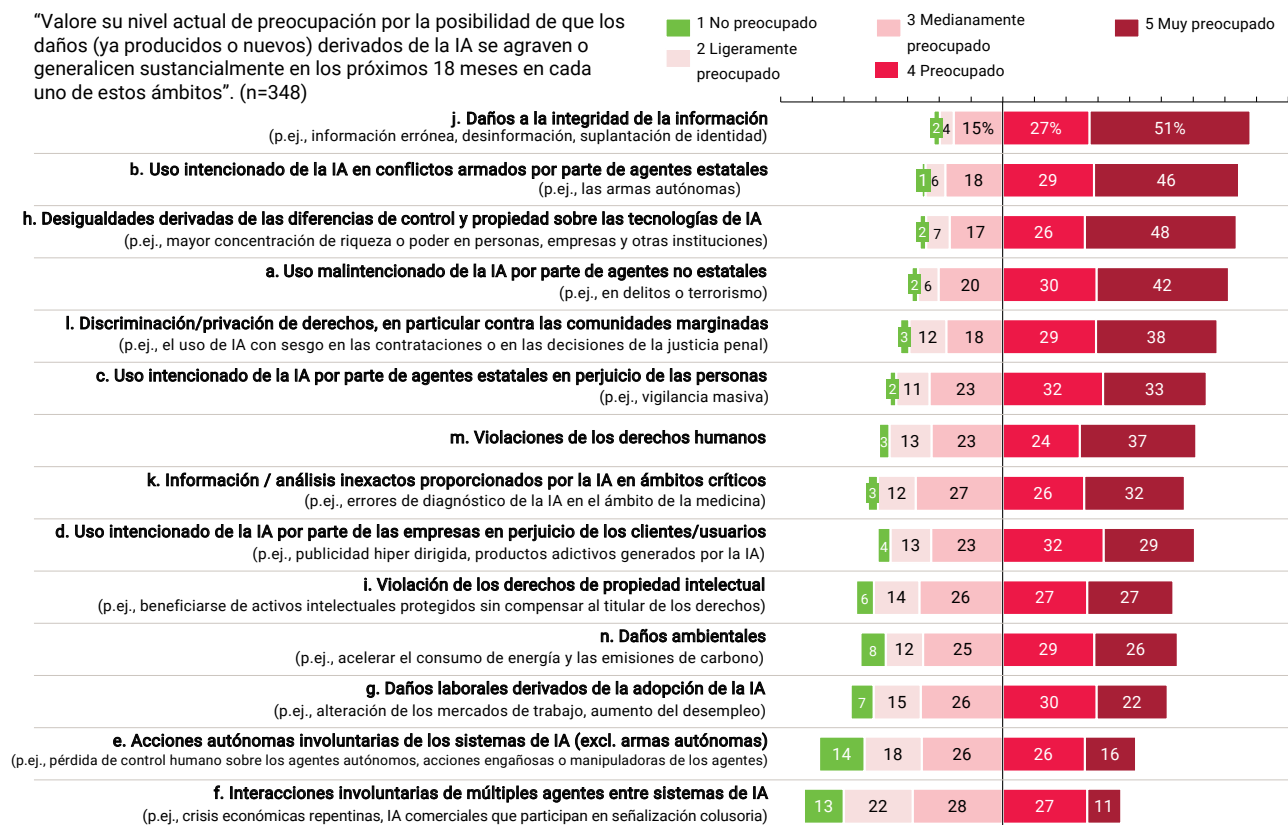
- 17** El desarrollo, despliegue y uso de la IA conlleva riesgos, que pueden afectar a muchos ámbitos a la vez. Conceptualizamos los riesgos relacionados con la IA en relación con las vulnerabilidades; esto ofrece una forma de definir las agendas políticas basada en la vulnerabilidad.
- 18** La velocidad, la opacidad y la autonomía de la IA plantean desafíos a los sistemas reguladores tradicionales. La aceleración del desarrollo técnico y el despliegue de la IA también pone en juego la gobernanza internacional, ya que su naturaleza de uso general tiene implicaciones transfronterizas en múltiples ámbitos al mismo tiempo.

E. Riesgos de la IA

- 19** Problemas como el sesgo en los sistemas de IA y la vigilancia indebida que posibilita la IA están cada vez más documentados. El uso de IA avanzada conlleva otros riesgos, como la confabulación de los grandes modelos lingüísticos, el elevado consumo de recursos y los riesgos para la paz y la seguridad. La desinformación generada por la IA amenaza a las instituciones democráticas.
- 20** Elaborar una lista exhaustiva de los riesgos de la IA de forma permanente es una tarea inútil; dada la naturaleza ubicua y en rápida evolución de la IA y de sus usos, creemos que es más útil examinar los riesgos desde la perspectiva de las comunidades vulnerables y los bienes comunes (véanse los párrs. 26 a 28 a continuación).
- 21** Una imagen de la percepción actual de los riesgos por parte de los expertos queda ilustrada en los resultados de un ejercicio de análisis prospectivo encargado para nuestro trabajo (la encuesta global AI risk pulse check; véase el anexo E), una encuesta en la que se recogieron las percepciones sobre

⁶ “Un estudio de la ubicación de las sedes de los beneficiarios de subvenciones a partir de una base de datos de fundaciones de mayoría estadounidense revela que, de 2018 a 2023, solo el 10 % de las subvenciones asignadas a iniciativas de IA que abordan uno o más de los ODS se destinaron a organizaciones con sede en países de ingreso bajo o mediano. El estudio sobre el capital privado muestra que el 36 % de las 9.000 empresas que abordan los ODS tiene su sede en Estados Unidos, pero estas empresas recibieron el 54 % de la financiación total. También observamos que, aunque el 20 % de las 9.000 empresas que se ocupan de los ODS tiene su sede en países de ingreso bajo o mediano, estas recibieron un porcentaje mayor (25 %) de la financiación total. Una de las razones es que las empresas chinas reciben un elevado porcentaje de la inversión... El resto de países en desarrollo de la muestra recibieron solo el 3 % de la financiación aunque representaban el 7 % de la muestra”. (Medha Bankhwal y otros, “AI for social good: improving lives and protecting the planet”, McKinsey & Company, mayo de 2024).

Figura 2: Niveles de preocupación de los expertos sobre los riesgos de la IA en múltiples ámbitos



Nota: Excluye «No sabe» / «Sin opinión» y las respuestas en blanco.

Fuente: AI Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024.

las tendencias y los riesgos relacionados con la IA de 348 expertos en IA de distintas disciplinas y 68 países⁷ de todas las regiones. En general, 7 de cada 10 expertos encuestados se mostraron preocupados o muy preocupados por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agravaran o generalizaran sustancialmente en los siguientes 18 meses (véase el anexo E).

22 Entre una lista de ejemplos de áreas de riesgo relacionadas con la IA, muchos de los expertos se mostraron preocupados o muy preocupados por los daños relacionados con lo siguiente (véase también la figura 2)⁸:

- Las implicaciones sociales de la IA: el **78 %** se refirió al daño a la integridad de la información [pregunta j], el **74 %** lo hizo a desigualdades como la concentración de riqueza y poder en pocas manos [pregunta h], y el **67 %** lo hizo a la discriminación y la privación de derechos, especialmente entre las comunidades marginadas [pregunta i]);
- Un uso intencionado de la IA que perjudica a otros: el **75 %** se refirió al uso por agentes estatales en los conflictos armados [pregunta b], el **72 %** lo hizo al uso malintencionado por parte de agentes no estatales [pregunta a] y el **65 %** lo hizo al uso por agentes estatales que perjudica a particulares [pregunta c].

⁷ La lista de participantes en la encuesta se elaboró a partir de las redes de la Oficina del Enviado del Secretario General para la Tecnología y del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial, e incluyó a expertos en “inmersión en profundidad”. Para mejorar la representación, durante todo el período de la encuesta se siguió invitando a más expertos. El total de 348 encuestados constituye una muestra global, sólida y equilibrada de encuestados que cuentan con la especialización necesaria para ofrecer una opinión fundada sobre los riesgos de la IA (véase la metodología completa en el anexo E).

⁸ A partir de la clasificación de riesgos basada en la vulnerabilidad del recuadro 4, una versión anterior de la cual aparecía en nuestro informe provisional.

23 En todas las áreas de riesgo, salvo en dos, la mayoría de los expertos en IA encuestados se mostraron preocupados o muy preocupados por la materialización de daños. Aunque menos de la mitad de los expertos expresaron tal preocupación por los daños no intencionados de la IA [preguntas e y f], 1 de cada 6 de los que estaban muy preocupados por los daños no intencionados de la IA mencionaron que esperaban que los sistemas agenciales tuvieran algunos de los impactos más sorprendentes o significativos en los riesgos relacionados con la IA en 2025⁹.

24 Las percepciones de los expertos variaron también dependiendo de la región y el género (véanse los resultados más detallados en el anexo E). Esto puso

de relieve la importancia de una representación inclusiva en los ejercicios relativos a la definición de los riesgos compartidos. A pesar de las diferencias, los resultados revelaron preocupaciones compartidas sobre los daños que la IA puede causar en el siguiente año, lo que indica la urgencia de los expertos por abordar los riesgos en múltiples ámbitos y vulnerabilidades en un futuro próximo.

25 Además, las armas autónomas en los conflictos armados o la delincuencia o el terrorismo y el uso de la IA para el orden público plantean en particular graves cuestiones jurídicas, humanitarias y de seguridad (véase el recuadro 3).¹⁰

Recuadro 3: La IA y la seguridad nacional e internacional

Muchas tecnologías de IA no son simplemente de doble uso, sino intrínsecamente “reutilizables”. Las aplicaciones de IA para el cumplimiento de la ley y los controles de las fronteras van en aumento, y suscitan preocupación las debidas garantías procesales, la vigilancia y la falta de rendición de cuentas respecto a los compromisos de los Estados con las normas de derechos humanos consagradas en la Declaración Universal de Derechos Humanos y otros instrumentos.

Entre los retos que plantea el uso de la IA en el ámbito militar se encuentran las nuevas carreras armamentísticas, la reducción del umbral de conflicto, la difuminación de las líneas que separan la guerra de la paz, la proliferación entre agentes no estatales y la derogación de principios del derecho internacional humanitario establecidos desde hace mucho tiempo, como la necesidad militar, la distinción, la proporcionalidad y la limitación del sufrimiento innecesario. Por razones jurídicas y morales, las decisiones de matar no deberían automatizarse mediante la IA. Los Estados deben comprometerse a abstenerse de desplegar y utilizar aplicaciones militares de IA en conflictos armados de forma que no se ajusten plenamente al derecho internacional, incluido el derecho internacional humanitario y de derechos humanos.

En la actualidad, 120 Estados Miembros apoyan un nuevo tratado sobre las armas autónomas y tanto el Secretario General como el Presidente del Comité Internacional de la Cruz Roja han hecho un llamamiento para que las negociaciones de dicho tratado concluyan en 2026. El Órgano Asesor insta a los Estados Miembros a atender este llamamiento.

⁹ Pregunta: “¿Qué tendencias emergentes actuales cree que podrían tener el impacto más sorprendente o significativo sobre los riesgos relacionados con la IA en los próximos 18 meses?”

¹⁰ Esta lista pretende ser meramente ilustrativa y solo menciona algunos de los riesgos a los que se enfrentan las personas y las sociedades.

Recuadro 3: La IA y la seguridad nacional e internacional (Continuación)

El Órgano Asesor considera fundamental establecer límites claros que delimiten los casos de uso ilícito, incluido el uso de IA para seleccionar y atacar objetivos de forma autónoma. Basándose en los compromisos existentes en materia de examen de la legalidad de las armas en el derecho internacional humanitario, los Estados deberían exigir a los fabricantes de armas, mediante obligaciones contractuales y otros medios, que lleven a cabo exámenes jurídicos y técnicos para evitar el diseño y desarrollo poco éticos de aplicaciones militares de la IA. Los Estados deberían también realizar exámenes jurídicos y técnicos sobre el uso de la IA, así como de las armas y los medios de guerra, y compartir las mejores prácticas al respecto.

Además, los Estados deberían llegar a un entendimiento común acerca de los mecanismos de prueba, evaluación, verificación y validación de la IA en el ámbito militar y de la seguridad. Deberían cooperar para crear capacidad y compartir conocimientos mediante el intercambio de buenas prácticas y la promoción de una gestión responsable del ciclo de vida de las aplicaciones de la IA en el ámbito militar y de la seguridad. Para evitar la adquisición de sistemas de IA potentes y potencialmente autónomos por parte de agentes no estatales peligrosos como grupos criminales o terroristas, los Estados deberían establecer controles y procesos adecuados en todo el ciclo de vida de los sistemas de IA, incluidos procesos sobre el final del ciclo de vida (es decir, el desmantelamiento) de las aplicaciones militares de IA.

En aras de la transparencia, podrían crearse “consejos asesores” que ofrecieran asesoramiento y control de expertos independientes a lo largo de todo el ciclo de vida de las aplicaciones militares y de seguridad de la IA. La industria y otros actores deberían estudiar mecanismos para evitar el uso indebido de la tecnología de IA con fines militares malintencionados o no previstos.

26 Sin embargo, la gestión de riesgos requiere ir más allá de enumerar o priorizar los riesgos. Enmarcar los riesgos en función de las vulnerabilidades puede desplazar el centro de atención de las agendas políticas del “qué” de cada riesgo (por ejemplo, “riesgo para la seguridad”) al “quién” está en peligro y “dónde”, así como quién debe rendir cuentas en cada caso.

27 Esto es importante, ya que los riesgos cambiantes se manifiestan de forma diferente según las personas y las sociedades. Un enfoque basado en la vulnerabilidad, también propuesto en nuestro informe provisional, ofrece un marco abierto para centrarse en aquellas personas que podrían verse perjudicadas por la IA, lo que puede servir de base para una gestión dinámica de los riesgos (véase el recuadro 4).

Recuadro 4: Clasificación de los riesgos relacionados con la IA en función de la vulnerabilidad existente o potencial (Continuación)

Personas

- Dignidad humana, valor o capacidad de acción (por ejemplo, manipulación, engaño, nudging, condena, explotación, discriminación, igualdad de trato, persecución penal, vigilancia, pérdida de autonomía humana y selección de objetivos asistida por inteligencia artificial).
- Integridad física y mental, salud y seguridad (por ejemplo, nudging, soledad y aislamiento, neurotecnología, armas autónomas letales, vehículos autónomos, diagnósticos médicos, acceso a la atención de la salud e interacción con sistemas químicos, biológicos, radiológicos y nucleares).
- Oportunidades de vida (por ejemplo, educación, empleo y vivienda).
- (Otros) derechos humanos y libertades civiles, como el derecho a la presunción de inocencia (p. ej., actuación policial predictiva), el derecho a un juicio imparcial (p. ej., predicción de la reincidencia, culpabilidad, reincidencia, predicción y juicios autónomos), la libertad de expresión e información (p. ej., *nudging*, información personalizada, burbujas de información), la privacidad (p. ej., tecnología de reconocimiento facial) y la libertad de reunión y circulación (p. ej., tecnología de rastreo en espacios públicos).

Política y sociedad

- Discriminación y trato injusto de grupos, en particular por motivos de rasgos individuales o grupales como el género, el aislamiento del grupo y la marginación.
- Impacto diferencial en niños, personas de edad avanzada, personas con discapacidad y grupos vulnerables.
- Seguridad nacional e internacional (por ejemplo, armas autónomas, control policial y fronterizo de migrantes y refugiados, delincuencia organizada, terrorismo y proliferación y escalada de conflictos).
- Democracia (por ejemplo, elecciones y confianza).
- Integridad de la información (por ejemplo, información errónea o desinformación, ultrafalsificaciones e información personalizada).
- Estado de derecho (por ejemplo, funcionamiento de las instituciones y confianza en ellas, cumplimiento de la ley y poder judicial).
- Diversidad cultural y cambios en las relaciones humanas (por ejemplo, homogeneidad y falsos amigos).
- Cohesión social (p. ej., burbujas de filtros, disminución de la confianza en las instituciones y las fuentes de información).
- Valores y normas (p. ej., éticas, morales, culturales y jurídicas).

Economía

- Concentración de poder.
- Dependencia tecnológica.
- Desigualdad de oportunidades económicas, acceso al mercado, distribución y asignación de recursos.
- Infratilización de la IA.
- Uso excesivo de la IA o “tecnosolucionismo”.
- Estabilidad de los sistemas financieros, las infraestructuras críticas y las instituciones.
- Protección de la propiedad intelectual.

Medio ambiente

- Consumo excesivo de energía, agua y recursos materiales (incluidos minerales raros y otros recursos naturales).

28 La pertinencia política de adoptar un enfoque de los riesgos relacionados con la IA basado en la vulnerabilidad se ilustra examinando las

consideraciones relativas a la gobernanza de la IA desde la perspectiva de un grupo vulnerable concreto, como los niños (véase el recuadro 5).

Recuadro 5: La infancia, en el foco de la gobernanza de la IA

Garantizar que las empresas y las escuelas atienden las necesidades y los derechos de los niños requiere un enfoque de gobernanza integral centrado en sus circunstancias singulares. Los niños generan una tercera parte de los datos y crecerán en una economía y un mundo acostumbrados al uso de la IA. Este recuadro resume algunas de las medidas relativas a este tema debatidas durante nuestras inmersiones en profundidad.

Dar prioridad a los derechos y la voz de los niños:

La gobernanza de la IA debe reconocer a los niños como partes interesadas prioritarias, haciendo hincapié en su derecho a desarrollarse libres de los efectos adictivos de la tecnología y en su derecho a desvincularse de ella. A diferencia de los enfoques generales centrados en el ser humano, la gobernanza centrada en el niño debe tener en cuenta las repercusiones a largo plazo en las perspectivas, la propia imagen y las opciones y oportunidades vitales de los niños. Incluir a los niños en los procesos de diseño y gobernanza es fundamental para garantizar que los sistemas de IA sean seguros y adecuados para su uso.

Investigación y formulación de políticas:

Es preciso realizar una investigación exhaustiva para comprender cómo afecta la IA al desarrollo social, cognitivo y emocional de los niños a lo largo del tiempo. Esta investigación debe servir de base a los debates sobre políticas y orientar las medidas de protección en todos los países.

Protección y privacidad:

Los niños no deben ser utilizados como sujetos de experimentación con IA. Proteger la privacidad de los niños es primordial. Las tecnologías de IA deben incorporar protocolos estrictos de protección de datos y ofrecer contenidos adecuados a cada edad.

Evaluación del impacto sobre los niños y diseño apropiado para ellos:

La obligatoriedad de realizar evaluaciones de impacto infantil en los sistemas de IA es esencial para garantizar su idoneidad y seguridad. Los sistemas de IA deben diseñarse teniendo en cuenta las necesidades de los niños e incorporar desde el principio características de seguridad y restricciones. Las decisiones de diseño deben contar con la participación de los propios niños.

Inclusión digital y equidad:

El acceso a la IA debe dotar a los niños de autonomía, opciones y voz, haciendo hincapié en un enfoque integral de la inclusión digital. Esto incluye proporcionar contenidos de IA en varios idiomas y garantizar que sean culturalmente apropiados para los niños que no hablan inglés.

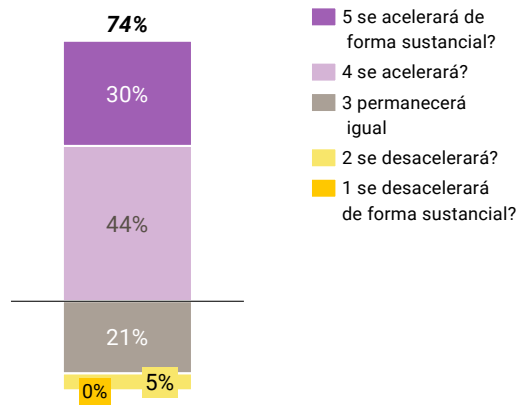
Cooperación y normas internacionales:

Para proteger a los niños en los distintos entornos educativos y de desarrollo es necesaria la interoperabilidad mundial de las normas que regulan la participación de los niños en las tecnologías de IA. Las normas mundiales resultarán esenciales para abordar los flujos transfronterizos de datos y el uso de la inteligencia artificial ética para los niños.

Figura 4: Expectativas de los expertos en cuanto al desarrollo tecnológico de la IA

El 74% espera que el ritmo del cambio tecnológico se acelerará (el 30% espera que lo hará de forma sustancial)

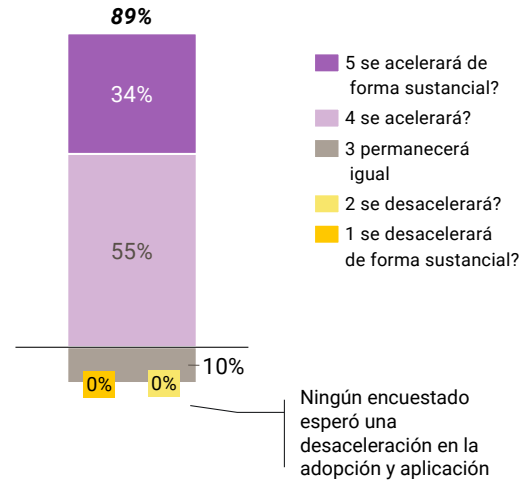
"En los próximos 18 meses, en comparación con los últimos 3 meses, ¿espera que el ritmo de los cambios tecnológicos en la IA (por ejemplo, el desarrollo y lanzamiento de nuevos modelos):" (n=348)



Nota: Las cifras pueden no sumar el 100% debido al redondeo. Excluye "No sabe" / "No contesta" y respuestas en blanco.
Fuente: Al Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024.

El 89% espera que el ritmo de adopción y aplicación se acelerará (el 34% espera que lo hará de forma sustancial)

"En los próximos 18 meses, en comparación con los últimos 3 meses, ¿espera que el ritmo de adopción y aplicación de la IA (por ejemplo, nuevos usos de la IA en empresas y administraciones públicas):" (n=348)



33 Como se ha señalado en el párrafo 23, algunos expertos prevén el despliegue de sistemas agenciales en 2025. Además, los principales expertos técnicos reconocen que muchos modelos de IA siguen siendo opacos y que sus resultados no son totalmente predecibles ni controlables, incluso cuando los efectos secundarios negativos pueden repercutir en otros a escala mundial.

34 La creciente dependencia de la adopción de decisiones automatizada y de la creación de contenidos mediante algoritmos opacos puede socavar el trato equitativo y la seguridad. Aunque los seres humanos suelen ser jurídicamente responsables de las decisiones de automatizar procesos que afectan a otros, es posible que los mecanismos de rendición de cuentas no evolucionen con la suficiente rapidez para que dicha rendición de cuentas tenga un efecto rápido y significativo.

35 Así pues, surge el riesgo social de que cada vez menos personas acaben siendo consideradas responsables de los daños derivados de sus decisiones de automatizar procesos utilizando

la IA, mientras se van incorporando al mundo sistemas cada vez más potentes. Esto requiere una gobernanza ágil para garantizar que los mecanismos de rendición de cuentas sigan el ritmo de la aceleración de la IA.

36 Si el ritmo de desarrollo y despliegue de la IA desafía a las instituciones existentes, también lo hace su amplitud. La IA avanzada, que es una tecnología de uso general y alcance mundial, puede desplegarse en distintos ámbitos y repercutir en las sociedades de múltiples maneras, con amplias implicaciones políticas.

37 Las implicaciones y el impacto potencial de la intersección de la IA con múltiples esferas como las finanzas, el mercado de trabajo, la educación y los regímenes políticos presagian amplias consecuencias que exigen un planteamiento que abarque a toda la sociedad (véanse ejemplos en el recuadro 6). Las instituciones existentes deben preparar respuestas integrales e intersectoriales que aborden las amplias repercusiones sociales de la IA.

Recuadro 6: Impactos sociales de la IA

Como parte de su compromiso más amplio, los miembros del Órgano Asesor consultaron a una serie de interesados para debatir las implicaciones de la inteligencia artificial en la sociedad. Este recuadro resume las principales preocupaciones e iniciativas posibles planteadas como parte de las inmersiones en profundidad en este tema.

Impacto social, psicológico y comunitario:

A medida que la IA aumente su potencia y generalización, su desarrollo, despliegue y uso serán más personalizados, lo que puede fomentar la alienación y la adicción. Para algunos miembros del Órgano Asesor, la IA entrenada en los datos personales y su consiguiente papel como interlocutor principal e intermediario puede suponer un punto de inflexión para los seres humanos, con el potencial de crear nuevos problemas sociales urgentes y agravar los ya existentes.

Por ejemplo, los futuros sistemas de inteligencia artificial podrían ser capaces de generar una fuente inagotable de contenidos de video de alta calidad adaptados a las preferencias personales de las personas. El aumento del aislamiento social, la alienación, los problemas de salud mental, la pérdida de capacidad de acción humana y el impacto en la inteligencia emocional y el desarrollo social son solo algunas de las posibles consecuencias.

Estas cuestiones están hoy insuficientemente estudiadas por los encargados de formular políticas en el contexto de tecnologías como los dispositivos inteligentes e Internet, y prácticamente inexploradas en el contexto de la IA, ya que los actuales marcos de gobernanza dan prioridad a los riesgos para los particulares y no para la sociedad en su conjunto.

Cuando los encargados de formular políticas se planteen futuras respuestas a la IA, deberán sopesar también estos factores y elaborar políticas que promuevan el bienestar de la sociedad, en particular de los jóvenes. Las intervenciones gubernamentales podrían propiciar entornos que den prioridad a las interacciones presenciales entre seres humanos, facilitando el acceso a servicios de salud mental e invirtiendo más en instalaciones deportivas, bibliotecas públicas y arte.

Sin embargo, más vale prevenir que curar: los desarrolladores de la industria deberían diseñar sus productos sin características personalizadas adictivas, asegurarse de que los productos no dañan la salud mental y promover (en lugar de debilitar) un sentimiento de pertenencia compartida en la sociedad. Las empresas tecnológicas deberían establecer políticas para gestionar los riesgos sociales en condiciones de igualdad con otros riesgos como parte de los esfuerzos por identificar y mitigar los riesgos en todo el ciclo de vida de los productos de IA.

Desinformación y confianza:

Las ultrafalsificaciones, los clones de voz y las campañas automatizadas de desinformación plantean una amenaza específica y grave para las instituciones y para procesos democráticos como las elecciones, así como para las sociedades democráticas y la confianza social en general, entre otras cosas mediante la manipulación e interferencia de información extranjera. El desarrollo de ecosistemas de información de circuito cerrado, reforzados por la IA y que utilizan los datos personales, puede tener profundos efectos en las sociedades, haciéndolas potencialmente más proclives a aceptar la intolerancia y la violencia hacia los demás.

Recuadro 6: Impactos sociales de la IA (Continuación)

Proteger la integridad de las instituciones y los procesos gubernamentales representativos requiere sistemas sólidos de verificación y detección de ultrafalsificaciones, así como procedimientos rápidos de detección y retirada de contenidos que puedan engañar y causar daño o división social, o que inciten a la propaganda de guerra, los conflictos y el discurso de odio. Las personas que no son figuras públicas deben estar protegidas contra la creación de ultrafalsificaciones de su imagen con fines fraudulentos, difamatorios o abusivos. Las ultrafalsificaciones sexualizadas son especialmente preocupantes para las mujeres y las niñas y pueden constituir una forma de violencia de género.

El compromiso voluntario de los agentes del sector privado con el etiquetado de las ultrafalsificaciones o con la posibilidad de que los usuarios las señalen y retiren las que se hayan creado o distribuido con ánimo malicioso es un primer paso importante. Sin embargo, no basta para mitigar los riesgos sociales. En vez de ello, es necesario un enfoque global de las múltiples partes interesadas, junto con compromisos vinculantes. Unas normas comunes para la autenticación de contenidos y la procedencia digital permitirían un enfoque mundialmente reconocido para identificar imágenes, videos y audios sintéticos y modificados por IA.

Además, el intercambio de conocimientos en tiempo real entre agentes públicos y privados, basado en normas internacionales, permitiría una capacidad de respuesta rápida para eliminar inmediatamente los contenidos engañosos o la manipulación e interferencia de información extranjera, antes de que tengan la oportunidad de hacerse virales. No obstante, estos procesos deben incorporar salvaguardias que garanticen que no se manipulan ni se abusa de ellos para favorecer la censura.

Estas acciones deben ir acompañadas de medidas preventivas para aumentar la resiliencia de la sociedad ante la desinformación y la propaganda impulsadas por la IA, como campañas de concienciación pública sobre el potencial de la IA para debilitar la integridad de la información. Además, los Estados Miembros deberían promover campañas de alfabetización mediática y digital, apoyar iniciativas de verificación de datos e invertir en la creación de capacidad contra la manipulación e interferencia de información extranjera.

38 El ritmo, la amplitud y la incertidumbre del desarrollo, despliegue y uso de la IA ponen de relieve la importancia de adoptar un enfoque integral, transversal y ágil de la IA. A escala internacional, una perspectiva integral debe reflejarse en un enfoque institucional en red de la gobernanza de la IA entre sectores y fronteras, que implique a las partes interesadas sin dejarse capturar por ellas.

39 En cuanto al cambio climático, el mundo se ha dado cuenta tardíamente de que es necesario un enfoque integral de la acción colectiva mundial. Ahora podemos hacerlo con el propio diseño de la IA.

40 Los retos anteriores se ven agravados por la concentración de la riqueza y la adopción de decisiones en un puñado de desarrolladores y aplicadores privados de IA, en particular en empresas multinacionales. Esto plantea otra cuestión: cómo pueden participar los interesados en la gobernanza de la IA sin menoscabar el interés público.

2. La necesidad de una gobernanza global

- 41** En la actualidad hay un gran déficit de gobernanza global de la IA. A pesar de los numerosos debates sobre ética y principios, el mosaico de normas, instituciones e iniciativas es aún incipiente y está lleno de lagunas. La rendición de cuentas y la reparación de los daños suelen brillar por su ausencia. El cumplimiento se basa en la voluntariedad. Existe una desconexión de fondo entre la retórica de alto nivel, los sistemas que se están desarrollando, desplegando y utilizando y las condiciones necesarias para la seguridad y la inclusividad. Como ya señalamos en nuestro informe provisional, la gobernanza de la IA es de vital importancia - no solo para afrontar los desafíos y los riesgos, sino también para garantizar que aprovechemos su potencial sin dejar a nadie atrás¹¹.
- 42** El imperativo de la gobernanza global, en particular, es irrefutable. Las materias primas de la IA, desde los minerales críticos hasta los datos de entrenamiento, se obtienen en todo el mundo. La IA de uso general, desplegada más allá de las fronteras, genera múltiples aplicaciones a nivel mundial. El desarrollo acelerado de la IA concentra el poder y la riqueza a escala mundial, con implicaciones geopolíticas y geoeconómicas. Además, en la actualidad nadie comprende el funcionamiento interno de la IA lo suficiente como para controlar totalmente sus resultados o predecir su evolución. Los responsables de las decisiones de desarrollar, desplegar o utilizar sistemas que no comprenden tampoco rinden cuentas. Mientras tanto, es probable que las repercusiones negativas y los impactos derivados de tales decisiones sean también globales.
- 43** A pesar del alcance mundial de la IA, las estructuras y normativas institucionales nacionales y regionales terminan en las fronteras físicas. Esto reduce la capacidad de un solo país para gobernar las aplicaciones derivadas de la IA que provocan daños transfronterizos, o para abordar los problemas a lo largo de las complejas cadenas de suministro transfronterizo de infraestructuras informáticas, flujos de datos de entrenamiento y fuentes de energía que hay detrás del desarrollo y uso de la IA. Las principales empresas de IA suelen tener una influencia más directa en las aplicaciones derivadas (a través de la mitigación previa de los riesgos) que la mayoría de los países que actúan por sí solos.
- 44** El desarrollo, despliegue y uso de una tecnología de este tipo no puede dejarse únicamente al albur de los mercados. Los gobiernos nacionales y las organizaciones regionales serán cruciales. Sin embargo, además de las consideraciones de equidad, acceso y prevención y reparación de daños, la propia naturaleza de la tecnología - transfronteriza en estructura y aplicación - requiere un enfoque multisectorial global. Sin un marco global inclusivo que implique a las partes interesadas, y dada la dinámica competitiva en juego, tanto los gobiernos como las empresas podrían verse tentados a recortar gastos o a dar prioridad al interés propio.
- 45** Por tanto, la IA plantea desafíos y oportunidades globales que requieren un enfoque integral y global que abarque transversalmente los ámbitos político, económico, social, ético, de derechos humanos, técnico, ambiental y otros. Un enfoque de este tipo puede convertir un mosaico de iniciativas en evolución en un todo coherente e interoperable, basado en el derecho internacional y adaptable en distintos contextos y momentos.

11 Véase <https://un.org/ai-advisory-body>.

46 La necesidad de una gobernanza global de la IA surge en un momento de competencia geopolítica y geoeconómica por la influencia y los mercados. Sin embargo, para hacer frente a los riesgos de la IA y al mismo tiempo hacer posible que las oportunidades se aprovechen de forma equitativa, se requiere una acción mundial concertada. El aumento de la brecha digital podría limitar los beneficios de la IA a unos pocos Estados y personas, mientras que los riesgos y los daños afectarían a muchos, especialmente a los grupos vulnerables.

A. Principios rectores y funciones de la gobernanza internacional de la IA

47 En nuestro informe provisional ya esbozamos los cinco principios que deben regir la creación de nuevas instituciones internacionales de gobernanza de la IA:

- **Principio rector 1:** La gobernanza de la IA debe ser inclusiva y utilizarse para y en beneficio de todos
- **Principio rector 2:** La gobernanza de la IA debe estar al servicio del interés público
- **Principio rector 3:** La gobernanza de la IA debe construirse en paralelo a la gobernanza de los datos y la promoción de conjuntos de datos comunes
- **Principio rector 4:** La gobernanza de la IA debe ser universal, estar interconectada y arraigada en la colaboración flexible de múltiples partes interesadas
- **Principio rector 5:** La gobernanza de la IA debe estar anclada en la Carta de las Naciones Unidas, el derecho internacional de los derechos humanos y otros compromisos internacionales acordados, como los ODS

48 El recuadro 7 resume los comentarios sobre estos principios, que hicieron hincapié en la importancia de los derechos humanos y en la necesidad de una mayor claridad en la aplicación efectiva de los principios rectores, en particular en lo relativo a la

gobernanza de los datos. En estos comentarios se nos pidió que abordáramos el problema de garantizar que el apoyo a la inclusividad estuviera respaldado por la acción y que los grupos marginados estuvieran debidamente representados.

49 En nuestro informe provisional propusimos también varias funciones institucionales que podrían ser desempeñadas a nivel internacional (véase la Figura 5). La mayoría de los comentarios recibidos confirmaron la necesidad de contar con estas funciones a nivel mundial, al tiempo que pedían funciones complementarias relacionadas con la gobernanza de los datos y la IA para hacer efectivo el principio rector 3 (La gobernanza de la IA debe construirse en paralelo a la gobernanza de los datos y la promoción de conjuntos de datos comunes).

50 En cuanto a las funciones institucionalmente “más difíciles” de la gobernanza de la IA, como la supervisión, la verificación, la presentación de informes, el cumplimiento, la consolidación de la rendición de cuentas, la respuesta y la exigencia de cumplimiento, los participantes señalaron que, en primer lugar y antes de institucionalizar dichas funciones, sería necesario cumplir las obligaciones de los tratados internacionales, y que aún no se había demostrado la conveniencia de institucionalizar dichas funciones en la gobernanza de la IA como tecnología.

51 No todas las funciones deben ser desempeñadas exclusivamente por las Naciones Unidas. Sin embargo, para que el mosaico de normas e instituciones se transforme en una red de seguridad que promueva y apoye la innovación sostenible en beneficio de toda la humanidad, es necesario que detrás de las reglas y normas con las que evaluamos si la gobernanza está logrando o no sus objetivos haya un entendimiento científico compartido y un terreno común en torno a la IA.

52 Durante nuestras consultas recibimos peticiones de un análisis más detallado del panorama de los esfuerzos existentes y emergentes para gobernar la IA a nivel internacional, y de las lagunas que deben colmarse para lograr una gobernanza internacional de la IA equitativa, eficaz y eficiente.

Recuadro 7: Comentarios sobre los principios rectores

Énfasis en la gobernanza de la IA basada en los derechos humanos:

Basándose en las amplias consultas realizadas por el Órgano Asesor de Alto Nivel sobre Inteligencia Artificial tras la publicación de su informe provisional, el principio rector 5 (La gobernanza de la IA debe basarse en la Carta de las Naciones Unidas, el derecho internacional de los derechos humanos, el derecho internacional humanitario y otras obligaciones internacionales) obtuvo el mayor apoyo en todos los sectores de las partes interesadas, incluidos los gobiernos, la sociedad civil, la comunidad técnica, el mundo académico y el sector privado. Este principio implica respetar, promover y cumplir los derechos humanos y enjuiciar las violaciones de los mismos, así como cumplir la resolución 78/265 de la Asamblea General relativa a aprovechar las oportunidades de sistemas seguros y fiables de inteligencia artificial para el desarrollo sostenible, adoptada por unanimidad en marzo de 2024.

En sus deliberaciones, el Órgano Asesor se mostró convencido de que para mitigar los riesgos y daños de la IA, abordar los nuevos casos de uso y garantizar que la IA pueda beneficiar realmente a toda la humanidad sin dejar a nadie atrás, la gobernanza de la IA debe estar centrada en los derechos humanos y garantizar una rendición de cuentas basada en los derechos en todas las jurisdicciones. Este compromiso fundamental con los derechos humanos es transversal y se aplica a todas las recomendaciones formuladas en este informe final.

Mecanismos específicos de aplicación y claridad en las directrices:

Muchas partes interesadas hicieron hincapié en la necesidad de contar con planes de acción detallados y directrices claras para garantizar la aplicación efectiva de los principios rectores del Órgano Asesor para la gobernanza internacional de la IA. Las entidades gubernamentales propusieron elaborar recomendaciones claras para definir y garantizar el interés público, junto con mecanismos de participación y supervisión pública. Las entidades del sector privado subrayaron con frecuencia la necesidad de contar con políticas claras y de aprovechar los marcos reguladores existentes para mantener unos mercados de IA competitivos e innovadores. Muchas organizaciones internacionales y de la sociedad civil también pidieron sistemas ágiles de gobernanza diseñados para responder oportunamente a la evolución de las tecnologías. Algunos solicitaron específicamente una nueva entidad que tuviera “fuerza y garra” y que no se limitara a la coordinación.

Mecanismos para la rendición de cuentas de los principales agentes:

Una preocupación común fue la rendición de cuentas por IA discriminatoria, sesgada o perjudicial por otros motivos, para lo que se propusieron mecanismos que garantizaran la rendición de cuentas y la reparación de los daños, y que abordaran la concentración de capacidad tecnológica y poder de mercado. Muchas organizaciones subrayaron la necesidad de abordar el poder sin control y garantizar los derechos de los consumidores y la competencia leal. Las instituciones académicas reconocieron los puntos fuertes de los principios rectores en su universalidad e inclusividad, pero sugirieron mejoras en la participación de las partes interesadas. Los agentes del sector privado hicieron hincapié en el uso responsable de la IA y en la eliminación de las barreras de acceso.

Recuadro 7: Comentarios sobre los principios rectores (Continuación)

Funciones más específicas en la gobernanza de datos de IA:

La ausencia de sistemas de gobernanza de datos se mencionó en múltiples consultas, y las partes interesadas señalaron a las Naciones Unidas como espacio natural para el diálogo sobre la gobernanza de datos. Los gobiernos subrayaron la necesidad de contar con marcos sólidos de gobernanza de datos que den prioridad a la privacidad, la protección de datos y el uso equitativo de los mismos, y abogaron por directrices internacionales para gestionar la complejidad de los datos en el desarrollo de la IA. Se pidió que los marcos se elaboraran a través de un proceso transparente e inclusivo que incluyera consideraciones éticas como el consentimiento y la privacidad.

El mundo académico destacó que la gobernanza de los datos debería tratarse como una prioridad a corto plazo. Las entidades del sector privado señalaron que las medidas de gobernanza de los datos deberían complementar la gobernanza de la IA, haciendo hincapié en leyes integrales sobre privacidad y en el uso responsable de la IA. Las organizaciones internacionales y las organizaciones de la sociedad civil subrayaron que la gobernanza de los datos de entrenamiento de IA debería proteger los derechos de los consumidores y apoyar la competencia leal entre los desarrolladores de IA a través de un acceso no exclusivo a los datos de entrenamiento de IA, lo que subraya la necesidad de adoptar medidas concretas y viables para la gobernanza de los datos. Las Naciones Unidas fueron identificadas como espacio clave para abordar estos desafíos de gobernanza y reducir la desigualdad de recursos.

Figura 5: Funciones de gobernanza de la IA propuestas a nivel internacional

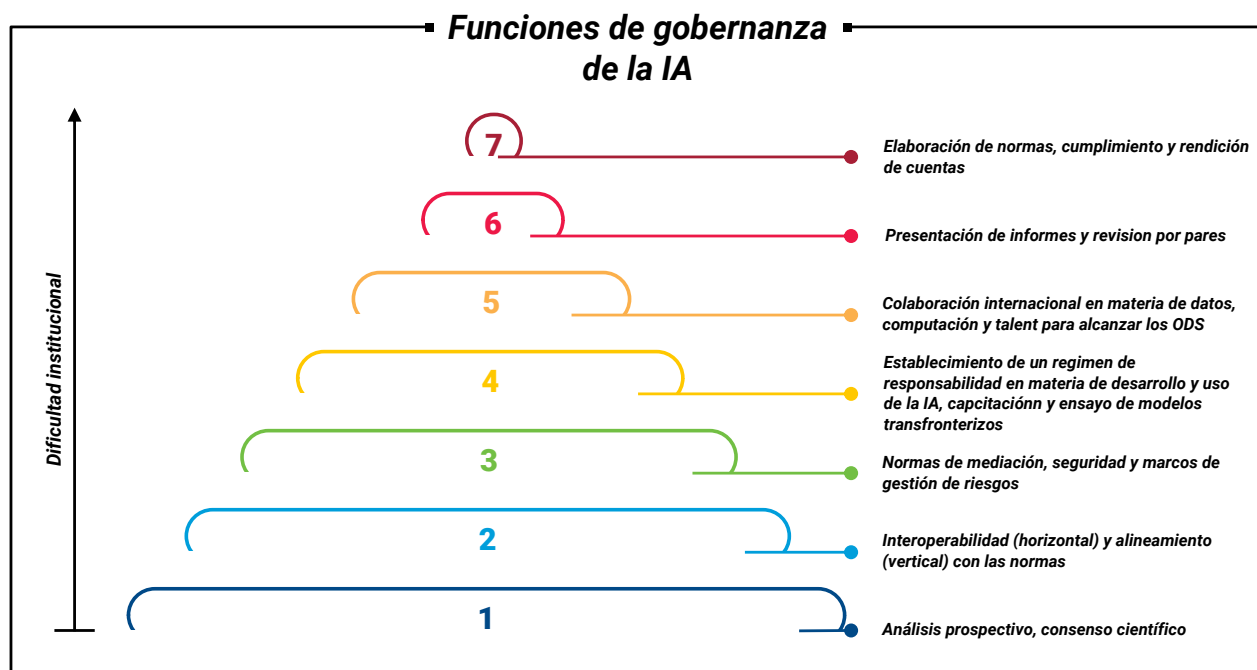
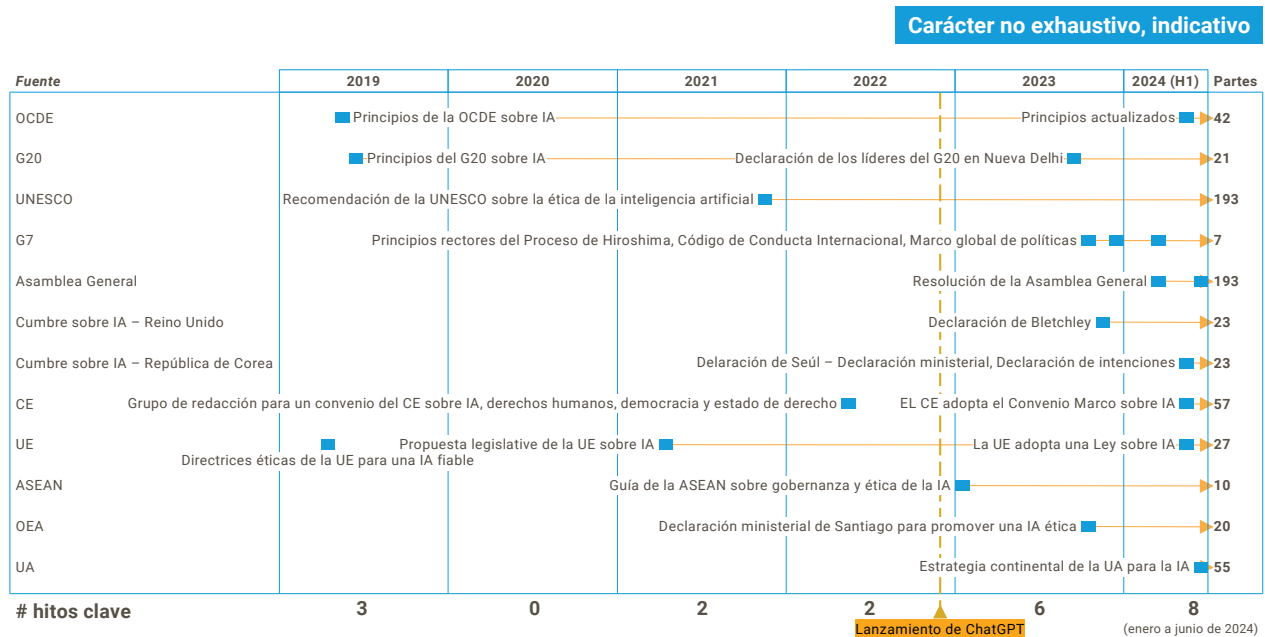


Figura 6: Iniciativas regionales e interregionales de gobernanza de la IA, hitos clave 2019–2024 (H1)



Abreviaciones: ASEAN, Asociación de Naciones de Asia Sudoriental; UA, Unión Africana; CE, Consejo de Europa; UE, Unión Europea; G20, Grupo de los 20; G7, Grupo de los Siete; GPAI, Alianza Mundial sobre la Inteligencia Artificial; OEA, Organización de los Estados Americanos; OCDE, Organización de Cooperación y Desarrollo Económicos; UNESCO, Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

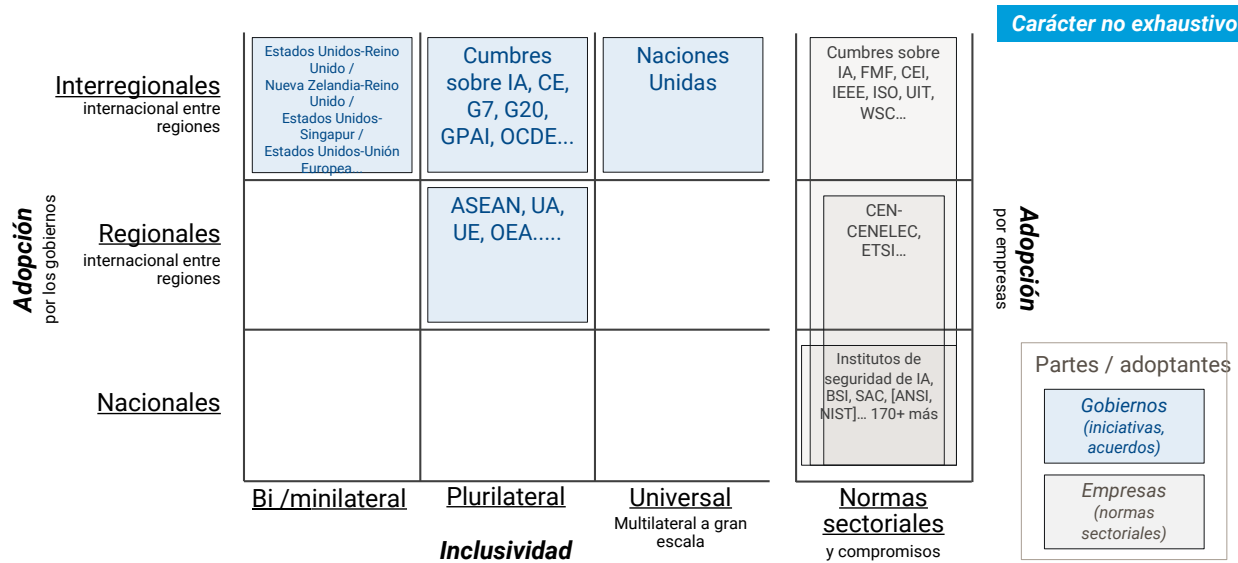
B. El nuevo panorama internacional de gobernanza de la IA

53 No cabe duda de que no faltan documentos y diálogos centrados actualmente en la gobernanza de la IA. Los gobiernos, las empresas y los consorcios, así como las organizaciones internacionales y regionales, han adoptado cientos de directrices, marcos y principios. Decenas de foros reúnen a diversos agentes, desde procesos intergubernamentales establecidos y organismos de expertos hasta iniciativas ad hoc de múltiples partes interesadas. Todo ello va acompañado de la regulación existente y en fase de elaboración a escalas nacional y regional.

54 Proliferan las iniciativas internacionales de los gobiernos (véase la Figura 6). Estas nuevas iniciativas siguen cada vez más un enfoque transversal de la gobernanza de la IA a nivel internacional, constituido por principios, declaraciones, manifestaciones y otras publicaciones que abordan la IA de forma integral, más que en ámbitos específicos. Estas iniciativas se han acelerado considerablemente desde 2023, impulsadas por el lanzamiento de múltiples grandes modelos lingüísticos de IA de uso general a raíz de la aparición de ChatGPT en noviembre de 2022.

55 Paralelamente, se han elaborado y publicado normas sectoriales sobre IA para su adopción a escala internacional. Otras iniciativas multisectoriales también han tratado de salvar la brecha entre los sectores público y privado, por ejemplo en foros de debate como el Foro para la Gobernanza de Internet.

Figura 7: Fuentes de las iniciativas de gobernanza centradas específicamente en la IA



Abreviaciones: ANSI, Instituto Nacional Americano de Normalización; ASEAN, Asociación de Naciones de Asia Sudoriental; UA, Unión Africana; BSI, British Standards Institution; CEN, Comité Europeo de Normalización; CENELEC, Comité Europeo de Normalización Electrotécnica; CE, Consejo de Europa; ETSI, Instituto Europeo de Normas de Telecomunicación; UE, Unión Europea; FMF, Foro de modelos fronterizos; G20, Grupo de los 20; G7, Grupo de los Siete; GPAI, Alianza Mundial sobre la Inteligencia Artificial; CEI, Comisión Electrotécnica Internacional; IEEE, Instituto de Ingenieros Electricistas y Electrónicos; ISO, Organización Internacional de Normalización; UIT, Unión Internacional de Telecomunicaciones; NIST, Instituto Nacional de Normas y Tecnología; OEA, Organización de los Estados Americanos; OCDE, Organización de Cooperación y Desarrollo Económicos; SAC, Administración China de Normalización; WSC, World Standards Cooperation.

56 En la Figura 7 se resumen algunas de las fuentes de las iniciativas de gobernanza de la IA y las normas sectoriales, clasificadas por su alcance geográfico y su carácter inclusivo (en esta reciente labor reconocemos muchos años de esfuerzos por parte del mundo académico, la sociedad civil y los órganos profesionales).

57 Entre los ejemplos de iniciativas plurilaterales regionales e interregionales pertinentes cabe citar las lideradas por la Unión Africana, diversos anfitriones de las cumbres sobre inteligencia artificial, la Asociación de Naciones de Asia Sudoriental, el Consejo de Europa, la Unión Europea, el Grupo de los Siete (G7), el Grupo de los Veinte (G20), la Alianza Mundial sobre la Inteligencia Artificial, la Organización de los Estados Americanos y la Organización de Cooperación y Desarrollo Económicos (OCDE), entre otros.

58 Nuestro análisis de los actuales mecanismos de gobernanza puede quedar obsoleto en cuestión de meses. Sin embargo, sí puede ayudar a ilustrar cómo se relacionan las iniciativas internacionales actuales y emergentes de gobernanza de la IA con nuestros principios rectores para la formación de nuevas instituciones de gobernanza global de la IA, entre otros, el principio rector 1 (La gobernanza de la IA debe ser inclusiva y utilizarse para y en beneficio de todos).

3. Lagunas en la gobernanza global de la IA

- 59** Las múltiples iniciativas nacionales, regionales, de múltiples partes interesadas y de otro tipo ya mencionadas han producido logros significativos y han orientado nuestro trabajo; muchos de sus representantes han contribuido a nuestras deliberaciones por escrito o han participado en nuestras consultas.
- 60** Sin embargo, más allá de un par de iniciativas procedentes de las Naciones Unidas, ninguna de las iniciativas puede ser realmente global en su alcance¹². Estas lagunas de representación en la gobernanza de la IA a nivel internacional son un problema porque la tecnología es global y su impacto será global.
- 61** Las lagunas de coordinación entre las iniciativas y las instituciones amenazan con dividir el mundo en regímenes desconectados e incompatibles de gobernanza de la IA.
- 62** Además, las lagunas en la implementación y la rendición de cuentas reducen la capacidad de los Estados, el sector privado, la sociedad civil, el mundo académico y la comunidad técnica para traducir los compromisos, por muy representativos que sean, en resultados tangibles.
- 64** Muchas excluyen a partes enteras del mundo. Como muestra la Figura 8 si se analizan siete iniciativas interregionales plurilaterales de gobernanza de la IA ajenas a las Naciones Unidas con miembros que se solapan, 7 países participan en todas ellas, mientras que 118 países no participan en ninguna (la mayoría del Sur Global, donde incluso los países líderes en IA están representados de forma desigual; véase la Figura 12).
- 65** La selectividad es comprensible en una fase temprana de la gobernanza, cuando existe cierto grado de experimentación, competencia en torno a las normas y diversos niveles de familiaridad con las nuevas tecnologías. Sin embargo, a medida que la gobernanza internacional de la IA va madurando, la representación global adquiere mayor importancia en términos de equidad y eficacia.
- 66** Además de la falta de inclusividad de los esfuerzos actuales, también existen lagunas de representación en las iniciativas nacionales y regionales centradas en lograr un entendimiento científico común de la IA. Estas lagunas de representación pueden manifestarse en los procesos de adopción de decisiones sobre la forma de delimitar el alcance de las evaluaciones, dotarlas de recursos y llevarlas a cabo.

A. Lagunas de representación

- 63** Nuestro análisis de las diversas iniciativas de gobernanza de la IA ajenas a las Naciones Unidas que abarcan regiones muestra que la mayoría de las iniciativas no son plenamente representativas en sus dimensiones intergubernamentales.

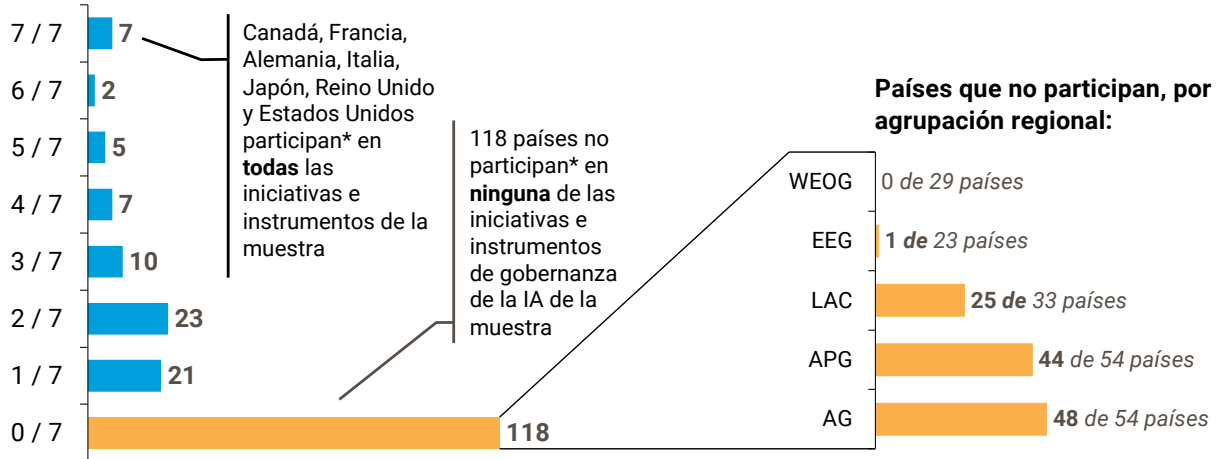
- 67** La equidad exige que más voces desempeñen un papel significativo en las decisiones sobre cómo gobernar la tecnología que nos afecta a todos, así como reconocer que muchas comunidades han sido totalmente excluidas hasta ahora de esas

¹² Recomendación de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) sobre la Ética de la Inteligencia Artificial (2021) y dos resoluciones de la Asamblea General sobre IA.

Figura 8: Representación en siete iniciativas internacionales de gobernanza de la IA ajenas a las Naciones Unidas

Muestra: Principios de la OCDE sobre la IA (2019), Principios del G20 sobre la IA (2019), Grupo de redacción del convenio internacional sobre IA del Consejo de Europa (2022-2024), Declaración Ministerial de la Alianza Mundial sobre la Inteligencia Artificial (2022), Declaración de los Ministros del G7 (2023), Declaración de Bletchley (2023) y Declaración Ministerial de Seúl (2024).

Solamente interregional, excluye regional



*Por refrendo de las publicaciones intergubernamentales pertinentes. No se considera que los países participan en una iniciativa plurilateral por el mero hecho de pertenecer a la Unión Europea o a la Unión Africana. Abreviaciones: AG, Grupo de los Estados de África; APG, Grupo de Asia y el Pacífico; EEG, Grupo de los Estados de Europa Oriental; G20, Grupo de los 20; G7, Grupo de los Siete; GPAI, Alianza Mundial sobre la Inteligencia Artificial; ALC, América Latina y el Caribe; OCDE, Organización de Cooperación y Desarrollo Económico; WEOG: Grupo de los Estados de Europa Occidental y Otros Estados.

conversaciones. La relativa escasez de temas en las agendas de las principales iniciativas que son prioridad en ciertas regiones señala un desequilibrio derivado de la infrarrepresentación¹³.

68 Los regímenes de gobernanza de la IA deben abarcar todo el mundo para ser eficaces a la hora de generar confianza, evitar las “carreras armamentísticas de la IA” o la “competencia a la baja” en materia de seguridad y derechos, y para responder eficazmente a los desafíos derivados del carácter transfronterizo de la IA, estimular el aprendizaje, fomentar la interoperabilidad y compartir los beneficios de la IA¹⁴. Además, la inclusión de opiniones diversas - incluidas las opiniones no afines - es beneficiosa para anticiparse a las amenazas y considerar respuestas que sean creativas y adaptables.

69 Limitando el número de países incluidos en los procesos clave de elaboración de las agendas, establecimiento de relaciones e intercambio de información, el pluralismo selectivo puede limitar la consecución de sus propios objetivos. Entre ellos se incluyen la compatibilidad de los nuevos enfoques de gobernanza de la IA, la seguridad global de la IA y el entendimiento científico compartido de la IA a nivel global (véanse las recomendaciones 1, 2 y 3 sobre lo que hace que un enfoque global sea especialmente eficaz en este caso).

70 Las dos resoluciones de la Asamblea General sobre la IA adoptadas en 2024 hasta la fecha¹⁵ reflejan el reconocimiento por parte de las naciones líderes en IA de que es necesario abordar las lagunas de representación en relación con la gobernanza

¹³ Por ejemplo, la gobernanza de los conjuntos de datos de entrenamiento de IA, el acceso a la capacidad de computación, el desarrollo de la capacidad en materia de IA, los riesgos relacionados con la IA en lo que respecta a la discriminación de grupos marginados y el uso de la IA en conflictos armados (véase el anexo E sobre los resultados de la encuesta global AI risk pulse check, que muestra las diferentes percepciones de los riesgos por parte de los países que pertenecen al Grupo de los Estados de Europa Occidental y otros Estados frente a los que no pertenecen a él). Muchos Estados y comunidades marginadas también han sido excluidos de los beneficios de la IA o pueden sufrir sus daños de forma desproporcionada. La equidad exige un enfoque diverso e inclusivo que tenga en cuenta los puntos de vista de todas las regiones y que distribuya las oportunidades uniformemente al tiempo que mitiga los riesgos.

¹⁴ En caso de que se establezcan líneas rojas - análogas, quizá, a la prohibición de la clonación humana -, estas solo serán exigibles si existe una aceptación global de la norma, así como un control de su cumplimiento. Esto sigue siendo así a pesar de que, paradójicamente, en el paradigma actual, mientras los costos de algún sistema de IA bajan, los costos de los sistemas avanzados de IA (posiblemente los más importantes de controlar) suben.

¹⁵ Resoluciones 78/265 (aprovechar las oportunidades de sistemas seguros y fiables de inteligencia artificial para el desarrollo sostenible) y 78/311 (aumentar la cooperación internacional para la creación de capacidad en materia de inteligencia artificial).

internacional de la IA, y de que las Naciones Unidas podrían ser el foro adecuado para reunir al mundo en este sentido.

- 71** El pacto digital global en septiembre de 2024, y el Foro de la Cumbre Mundial sobre la Sociedad de la Información en 2025, ofrecen dos posibilidades adicionales de aplicar políticas en las que podría institucionalizarse un conjunto representativo a escala mundial de procesos de gobernanza de la IA para abordar las lagunas de representación¹⁶.

B. Lagunas de coordinación

- 72** La aparición y la evolución de las iniciativas de gobernanza de la IA no garantizan un trabajo conjunto eficaz en beneficio de la humanidad. Al contrario, han aparecido lagunas de coordinación. La conexión efectiva entre las iniciativas plurilaterales selectivas (véase la Figura 8) y otras iniciativas regionales tampoco está garantizada, por lo que existe riesgo de incompatibilidad entre regiones.
- 73** Tampoco existen mecanismos mundiales para que todas las organizaciones internacionales de normalización (véase la Figura 7), las iniciativas internacionales de investigación científica o las iniciativas de desarrollo de la capacidad en materia

de IA puedan coordinarse entre sí, lo que debilita la interoperabilidad de los enfoques y provoca fragmentación. Las lagunas de coordinación entre las diversas iniciativas submundiales se abordan mejor, en algunos casos, a nivel mundial.

- 74** Dentro del sistema de las Naciones Unidas también hay lagunas de coordinación, que se reflejan en la gran variedad de documentos e iniciativas existentes en relación con la IA. La Figura 9 muestra 26 instrumentos de las Naciones Unidas en ámbitos específicos que pueden aplicarse a la IA, 25 de los cuales son vinculantes y que requerirán una interpretación en lo que respecta a la IA. Otros 32 documentos del ámbito de las Naciones Unidas y organizaciones afines se centran específicamente en la IA, pero ninguno de ellos es vinculante¹⁷. En algunos casos abordan los riesgos de la IA y cómo aprovechar sus ventajas en ámbitos específicos.
- 75** El nivel de actividad demuestra la importancia de la IA para los programas de las Naciones Unidas. A medida que la IA se generalice y afecte a aspectos cada vez más amplios de la sociedad, aumentarán los llamamientos a la acción de diversas partes del sistema de las Naciones Unidas, en particular a través de normas vinculantes. También demuestra la naturaleza *ad hoc* de las respuestas, que en gran medida se han elaborado orgánicamente en ámbitos específicos y sin una estrategia general. Las lagunas de coordinación resultantes provocan solapamientos y merman la interoperabilidad y el impacto.

¹⁶ Distintas iniciativas plurilaterales, como los Principios de la OCDE sobre la IA, el Proceso de Hiroshima sobre IA del G7 y el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, están abiertas a partidarios o adeptos más allá de los países que las emprendieron. Sin embargo, es posible que esta apertura no proporcione la representación y la legitimidad al ritmo y con la amplitud necesarios para responder a la aceleración de la proliferación de la IA en todo el mundo. Mientras tanto, persisten las lagunas de representación en los procesos internacionales de gobernanza de la IA, con la adopción de decisiones concentrada en manos de unos pocos países y empresas.

¹⁷ Una encuesta realizada por la Junta de los Jefes Ejecutivos en febrero de 2024 a 57 entidades de las Naciones Unidas registró 50 documentos relativos a la gobernanza de la IA; respondieron a ella 44 de las 57 entidades, entre ellas: la Comisión Económica para América Latina y el Caribe, la Comisión Económica y Social para Asia y el Pacífico, la Comisión Económica y Social para Asia Occidental, la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), el Organismo Internacional de Energía Atómica (OIEA), la Organización de Aviación Civil Internacional (OACI), el Fondo Internacional de Desarrollo Agrícola, la Organización Internacional del Trabajo (OIT), el Fondo Monetario Internacional, la Organización Internacional para las Migraciones, el Centro de Comercio Internacional (ITC), la Unión Internacional de Telecomunicaciones (UIT), la Entidad de las Naciones Unidas para la Igualdad de Género y el Empoderamiento de las Mujeres (ONU-Mujeres), el Programa Conjunto de las Naciones Unidas sobre el VIH/Sida (ONUSIDA), la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD), el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas, el Departamento de Comunicación Global de las Naciones Unidas, la Oficina Ejecutiva del Secretario General, la Oficina de Coordinación de Asuntos Humanitarios, la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, la Oficina de las Naciones Unidas de Lucha contra el Terrorismo, la Oficina de Asuntos de Desarme, la Oficina de Tecnología de la Información y las Comunicaciones, la Oficina del Enviado del Secretario General para la Tecnología (OSET), el Programa de las Naciones Unidas para el Desarrollo (PNUD), la Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres, el Programa de las Naciones Unidas para el Medio Ambiente, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), la Convención Marco de las Naciones Unidas sobre el Cambio Climático, el Fondo de Población de las Naciones Unidas, la Oficina del Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR), el Fondo de las Naciones Unidas para la Infancia, el Instituto Interregional de las Naciones Unidas para Investigaciones sobre la Delincuencia y la Justicia, la Organización de las Naciones Unidas para el Desarrollo Industrial, la Oficina de las Naciones Unidas contra la Droga y el Delito / Oficina de las Naciones Unidas en Viena, la Oficina de las Naciones Unidas de Servicios para Proyectos, el Organismo de Obras Públicas y Socorro de las Naciones Unidas para los Refugiados de Palestina en el Cercano Oriente, la Universidad de las Naciones Unidas, el programa de los Voluntarios de las Naciones Unidas, la Organización Mundial del Comercio, la Unión Postal Universal, el Grupo Banco Mundial, el Programa Mundial de Alimentos, la Organización Mundial de la Salud (OMS) y la Organización Mundial de la Propiedad Intelectual (OMPI). Véase "United Nations system white paper on AI governance: an analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance" (disponible en <https://unsceb.org/united-nations-system-white-paper-ai-governance>).

Figura 9: Selección de documentos de las Naciones Unidas y organizaciones afiliadas relacionados con la gobernanza de la IA

Carácter no exhaustivo					
<p>Ética y política</p> <p>UNESCO</p> <ul style="list-style-type: none"> Recomendación sobre la ética de la inteligencia artificial <p>OMS</p> <ul style="list-style-type: none"> Guidance on Ethics & Governance of Artificial Intelligence for Health <p>Fondo de las Naciones Unidas para la Infancia (UNICEF)</p> <ul style="list-style-type: none"> Orientación de políticas sobre el uso de la inteligencia artificial en favor de la infancia The Case for Better Governance of Children's Data: A Manifesto Responsible Data for Children (rd4c.org) <p>Programa de las Naciones Unidas para los Asentamientos Humanos (ONU-HÁBITAT)</p> <ul style="list-style-type: none"> Mainstreaming human rights in the digital transformation of cities - A guide for local governments Policy framework for centering people, inclusion, and human rights in smart city development <p>ONU-Mujeres</p> <ul style="list-style-type: none"> Conclusiones acordadas en el 67º período de sesiones de la Comisión de la Condición Jurídica y Social de la Mujer <p>Fondo de Población de las Naciones Unidas</p> <ul style="list-style-type: none"> Programa de Acción de la Conferencia Internacional sobre la Población y el Desarrollo - Un marco basado en los derechos humanos y centrado en las personas <p>KEY: De aplicación a la IA Puede aplicarse a la IA * Vinculante</p>	<p>Derechos humanos</p> <p>ACNUDH</p> <ul style="list-style-type: none"> Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial* Pacto Internacional de Derechos Civiles y Políticos* Pacto Internacional de Derechos Económicos, Sociales y Culturales* Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer* Convención contra la Tortura y Otros Tratos o Penas Crueles, Inhumanos o Degradantes* <p>OIT</p> <ul style="list-style-type: none"> Convenio sobre las Peores Formas de Trabajo Infantil, 1999 (núm. 182)* Convenio sobre Seguridad y Salud de los Trabajadores, 1981 (núm. 155)* Convenio sobre el Marco Promocional para la Seguridad y Salud en el Trabajo, 2006 (núm. 187)* Convenio sobre la Discriminación (Empleo y Ocupación), 1958 (núm. 111)* Convenio sobre los Representantes de los Trabajadores, 1971 (núm. 135)* Convenio sobre la Política del Empleo, 1964 (núm. 122)* Repertorio de recomendaciones prácticas de la OIT sobre la protección de los datos personales de los trabajadores <p>UNICEF</p> <ul style="list-style-type: none"> Convención sobre los Derechos del Niño* 	<p>Normas técnicas</p> <p>UIT</p> <ul style="list-style-type: none"> AI in Telecom Operations and Management AI in Smart Systems and Cities AI in Network Management and Services - AI in Specific Technologies or Applications <p>PNUD</p> <ul style="list-style-type: none"> The Digital Public Goods standard for AI systems (elaborada conjuntamente con la Alianza de Bienes Públicos Digitales) <p>OACI</p> <ul style="list-style-type: none"> Anexos al Convenio de Chicago* <p>Paz y seguridad</p> <p>UNODA</p> <ul style="list-style-type: none"> Artículo 36 del Protocolo Adicional I a los Convenios de Ginebra* Convención sobre las Armas Biológicas* Convención sobre las Armas Químicas <p>Salud</p> <p>OMS</p> <ul style="list-style-type: none"> Regulatory considerations on artificial intelligence for health Generating Evidence for Artificial Intelligence Based Medical Devices: A Framework for Training Validation and Evaluation Guidance on Ethics & Governance of Artificial Intelligence for Health 	<p>Comunicaciones</p> <p>Departamento de Comunicación Global</p> <ul style="list-style-type: none"> Developing work on principles on information integrity <p>UNESCO</p> <ul style="list-style-type: none"> Directrices para la gobernanza de las plataformas digitales <p>Comercio</p> <p>OMC</p> <ul style="list-style-type: none"> Acuerdo General sobre el Comercio de Servicios* Acuerdo sobre Obstáculos Técnicos al Comercio* Acuerdos sobre Tecnología de la Información* Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual Relacionados con el Comercio* Acuerdo sobre Facilitación del Comercio <p>CNUDMI</p> <ul style="list-style-type: none"> Proyecto de disposiciones sobre la contratación automatizada <p>Propiedad intelectual</p> <p>OMPI</p> <ul style="list-style-type: none"> Convención Internacional sobre la Protección de los Artistas Intérpretes o Ejecutantes, los Productores de Fonogramas y los Organismos de Radiodifusión* Convenio de Berna para la Protección de las Obras Literarias y Artísticas* Tratado de Beijing sobre Interpretaciones y Ejecuciones Audiovisuales* Tratado de Cooperación en materia de Patentes* 	<p>Drogas y delincuencia</p> <p>Oficina de las Naciones Unidas contra la Droga y el Delito</p> <ul style="list-style-type: none"> Declaración de Kioto <p>UNICRI</p> <ul style="list-style-type: none"> Policy Framework for Responsible Limits on Facial Recognition. Use Case: Law Enforcement Investigations Toolkit for Responsible Artificial Intelligence Innovation in Law Enforcement <p>OLCT</p> <ul style="list-style-type: none"> Estrategia Global de las Naciones Unidas contra el Terrorismo: octavo examen (A/RES/77/298) <p>Educación</p> <p>UNESCO</p> <ul style="list-style-type: none"> Guía para el uso de IA generativa en educación e investigación Proyecto de marcos de competencias de IA para estudiantes y profesores Marco de transformación digital y competencias en Inteligencia Artificial para funcionarios públicos <p>Otros</p> <p>ONU-HÁBITAT</p> <ul style="list-style-type: none"> AI Risk Assessment Framework Directrices internacionales sobre ciudades inteligentes centradas en las personas <p>Organización de las Naciones Unidas para el Desarrollo Industrial</p> <ul style="list-style-type: none"> Declaración de Abu Dabi, ONUDI GC.18 <p>Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres</p> <ul style="list-style-type: none"> Marco de Sendai para la Reducción del Riesgo de Desastres* 	

Fuente: "United Nations system white paper on AI governance: an analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance", 28 de febrero de 2024.

76 El número y la diversidad de enfoques son una señal de que el sistema de las Naciones Unidas está respondiendo a un problema emergente. Con una orquestación adecuada y en combinación con procesos que adopten un enfoque integral, estos esfuerzos pueden brindar una vía eficiente y sostenible hacia una gobernanza internacional inclusiva de la IA en ámbitos específicos. Esto podría tener un impacto significativo, armonizado y coordinado en ámbitos como la salud, la educación, las normas técnicas y la ética, más allá de la proliferación de iniciativas e instituciones en este campo en expansión. El derecho internacional, incluido el derecho internacional de los derechos humanos, proporciona una base normativa compartida para todos los esfuerzos relacionados con la IA, facilitando así la coordinación y la coherencia.

77 Aunque la labor de muchas entidades de las Naciones Unidas afecta a la gobernanza de la IA, sus mandatos específicos implican que ninguna de ellas lo hace de forma exhaustiva; y sus puntos focales gubernamentales designados están igualmente especializados¹⁸. Esto limita la capacidad de las actuales entidades de las Naciones Unidas para abordar por sí solas las múltiples implicaciones de la IA a escala mundial. A niveles nacional y regional, estas lagunas han sido abordadas por nuevas instituciones como los institutos de seguridad de la IA o las oficinas de IA con un enfoque adecuadamente transversal¹⁹.

18 Por ejemplo, los ministerios de educación, ciencia y cultura (UNESCO); telecomunicaciones o tecnología de la información y las comunicaciones (UIT); industria (Organización de las Naciones Unidas para el Desarrollo Industrial); y trabajo (OIT).

19 Así lo hacen en el Canadá, el Japón, Singapur, la República de Corea, el Reino Unido, los Estados Unidos y la Unión Europea.

C. Lagunas de implementación

- 78** La representación y la coordinación, sin embargo, no son suficientes. Se necesitan procesos de acción y seguimiento para garantizar que los compromisos de buena gobernanza se traducen en resultados tangibles en la práctica. Debemos hacer más para garantizar la rendición de cuentas. La presión grupal y el aprendizaje entre pares son dos elementos que pueden favorecer la rendición de cuentas.
- 79** La colaboración con el sector privado será igualmente importante para la rendición de cuentas y la reparación de los daños. Las Naciones Unidas tienen experiencia al respecto con sus Principios Rectores sobre las Empresas y los Derechos Humanos. En este sentido, necesitaremos una sólida colaboración de la sociedad civil y de los expertos científicos para que los gobiernos y las empresas privadas rindan cuentas sobre sus compromisos y afirmaciones.
- 80** La falta de facilitadores para aprovechar los beneficios de la IA en pro del bien público dentro de los países y entre ellos constituye una de las principales lagunas en la implementación. Muchos países han puesto en marcha estrategias nacionales para promover las infraestructuras y el talento relacionados con la IA y están surgiendo algunas iniciativas de ayuda internacional²⁰. Sin embargo, estas iniciativas no disponen de suficientes redes ni recursos.
- 81** A escala mundial, la conexión de las iniciativas nacionales y regionales de desarrollo de la capacidad y la puesta en común de recursos para apoyar a los países que quedan al margen de tales esfuerzos pueden contribuir a garantizar que ningún país se quede atrás en el aprovechamiento compartido de las oportunidades asociadas a la IA. Otra laguna clave en la implementación es la ausencia de un fondo específico para el desarrollo de la capacidad en materia de IA, a pesar de la existencia de algunos mecanismos de financiación para la capacidad digital (recuadro 8).

²⁰ Los esfuerzos a nivel nacional podrían seguir empleando herramientas de diagnóstico como la metodología de la UNESCO de evaluación del estado de preparación para la IA que ayudan a identificar las lagunas a nivel nacional, y la red internacional ayudaría a colmarlas.

Recuadro 8: Lagunas en la financiación mundial de la capacidad en materia de IA

El Órgano Asesor cree que no existen fondos mundiales para el desarrollo de la capacidad en materia de IA que tengan la envergadura y el mandato necesarios para financiar la importante inversión que se requiere para reducir la brecha de la IA.

Las estimaciones indicativas sitúan el importe necesario entre los 350 millones y los 1.000 millones de dólares de los Estados Unidos al año^a, incluidas las contribuciones en especie del sector privado, con el mandato de centrarse en el desarrollo de la capacidad en materia de IA a través de todos sus facilitadores, como el talento, la computación, los datos de entrenamiento, el desarrollo de modelos y la colaboración interdisciplinar para las aplicaciones. Entre los actuales mecanismos multilaterales de financiación se incluyen los siguientes:

a) El Fondo Conjunto para los ODS: Este fondo es amplio y abarca todos los ODS, así como la respuesta de emergencia. Financia iniciativas nacionales de integración de las políticas de las Naciones Unidas y el apoyo financiero estratégico a los países para avanzar en la consecución de los ODS. El fondo ayuda a las Naciones Unidas a proporcionar financiación y a impulsar la programación relativa a los ODS. Desde 2017, 30 entidades participantes de las Naciones Unidas han recibido un total de 223 millones de dólares. El fondo no financia directamente a gobiernos, comunidades o entidades nacionales, y tampoco financia iniciativas transfronterizas.

En 2023, el fondo tenía unos 16 donantes que aportaron un total de 57,7 millones de dólares, y se calcula que en 2024 esta cifra ascenderá a 58,8 millones. El sector privado ha aportado 83.155 dólares desde 2017, y no ha habido aportaciones en 2023 ni en 2024 hasta la fecha.

La mayoría de los fondos, el 60 %, se destinan a actividades en cinco ODS: el ODS 2 (hambre cero), el ODS 5 (igualdad de género), el ODS 7 (energía asequible y limpia), el ODS 9 (industria, innovación e infraestructuras) y el ODS 17 (alianzas).

La ventana de transformación digital del fondo (iniciada en 2023) ha financiado un proyecto de 250.000 dólares, desembolsados a partes iguales por la Unión Internacional de Telecomunicaciones y el Programa de las Naciones Unidas para el Desarrollo. A finales del ejercicio económico de 2023, su tasa de ejecución era del 2,27 %. Las actividades de transformación digital constituyen una pequeña parte de las actividades del fondo, y normalmente están relacionadas con otros ODS (por ejemplo, la conectividad y la infraestructura digital para ayudar a la prestación de servicios, por ejemplo, en los pequeños Estados insulares en desarrollo).

b) La Alianza para el Desarrollo Digital del Banco Mundial: Este fondo ayuda a los países a desarrollar e implementar la transformación digital con un enfoque en las infraestructuras y el acceso y uso de la banda ancha, la implantación de infraestructuras públicas digitales y la producción, accesibilidad y uso de datos. A finales de 2022 el fondo había invertido 10.700 millones de dólares en más de 80 países.

La alianza incluye un fondo fiduciario de múltiples donantes asociado a la ciberseguridad (Estonia, Alemania, Japón y el Reino de los Países Bajos, destinado a apoyar el desarrollo de la capacidad nacional en materia de ciberseguridad).

a Menos del 1 % de la inversión anual estimada en IA del sector privado en 2023.

4. Mejora de la cooperación global

- 82** Una vez presentado el déficit de gobernanza global, pasamos ahora a las recomendaciones para abordar las lagunas prioritarias a corto plazo.
- 83** Nuestras recomendaciones proponen una visión integral de un enfoque global en red, ágil y flexible para gobernar la IA en beneficio de la humanidad, que abarque un entendimiento común, un terreno común y beneficios comunes para mejorar la representación, permitir la coordinación y reforzar la implementación (véase la Figura 1)). Solo un enfoque inclusivo y global de la gobernanza de la IA como este puede abordar los desafíos y las oportunidades múltiples y cambiantes que presenta la IA a escala global, promoviendo la estabilidad internacional y el desarrollo equitativo.
- 84** Orientadas por los principios enumerados en nuestro informe provisional (véase el párr. 47), nuestras propuestas pretenden colmar lagunas y aportar coherencia al ecosistema emergente de respuestas e iniciativas internacionales de gobernanza de la IA, ayudando a evitar la fragmentación y la pérdida de oportunidades. Para apoyar estas medidas de manera eficiente y colaborar eficazmente con otras instituciones, proponemos una estructura ligera y ágil como expresión de los esfuerzos coherentes: una oficina de IA en la Secretaría de las Naciones Unidas, cercana al Secretario General, que trabaje como “aglutinante” para mantener unidas estas otras piezas.

Figura 10: Resumen de las recomendaciones y de cómo estas abordan las lagunas en la gobernanza global de la IA

Propósito	Mejorar la representación	Permitir la coordinación	Reforzar la aplicación
Un entendimiento común Grupo internacional de científicos expertos en IA	✓	✓	
Un terreno común Diálogo sobre políticas de gobernanza de la IA Intercambio de normas sobre la IA	✓	✓	(✓)
Beneficios comunes Red de desarrollo de la capacidad Fondo mundial para la IA Marco mundial de datos sobre IA	✓	✓	✓
Esfuerzo coherente Una oficina de IA en la Secretaría General de las Naciones Unidas	Asesoramiento al Secretario General, una voz coherente dentro del sistema de las Naciones Unidas comprometerá a los Estados y a las partes interesadas y apoyará otras propuestas		

85 La ONU dista mucho de ser perfecta. Sin embargo, la legitimidad que se deriva de su carácter inclusivo único, unida a sus fundamentos normativos vinculantes en el derecho internacional, incluido el derecho internacional de los derechos humanos, permite albergar esperanzas de gobernar la IA en beneficio y protección de la humanidad de una manera equitativa, efectiva y justa²¹.

A. Un entendimiento común

86 Un enfoque global para gobernar la IA comienza con un entendimiento común de sus capacidades y de las oportunidades, los riesgos y las incertidumbres que esta plantea.

87 El ámbito de la IA ha evolucionado rápidamente, produciendo una cantidad abrumadora de información y complicando descifrar el bombo publicitario de la realidad. Esto puede alimentar la confusión, impedir el entendimiento común y beneficiar a las grandes empresas de IA a expensas de los responsables de formular políticas, la sociedad civil y los ciudadanos.

88 Además, la poca colaboración científica internacional y de intercambio de información puede generar percepciones globales erróneas y socavar la confianza internacional.

89 Los Estados Miembros deben disponer de información y conocimientos científicos oportunos, imparciales y fiables sobre la IA para construir un entendimiento fundacional compartido en todo el mundo y equilibrar las incoherencias informativas entre las empresas que albergan costosos laboratorios de IA y el resto del mundo, entre otras cosas, mediante el intercambio de información entre las empresas de IA y la comunidad de la IA en general.

90 Todo ello resulta más eficiente a nivel mundial, ya que permite la inversión conjunta en un bien público global y la colaboración de interés público a través de esfuerzos que, de otro modo, estarían fragmentados y duplicados.

Un grupo internacional de científicos

Recomendación 1: Un grupo internacional de científicos expertos en IA

Recomendamos la creación de un grupo internacional e independiente de científicos expertos en IA, formado por diversos expertos multidisciplinares en la materia que presten sus servicios a título personal y de forma voluntaria. Con el apoyo de la oficina de IA propuesta y otros organismos pertinentes de las Naciones Unidas, en colaboración con otras organizaciones internacionales pertinentes, su mandato incluiría:

- a. Publicar un informe anual en el que se analicen las capacidades, las oportunidades, los riesgos y las incertidumbres relacionadas con la IA, y se identifiquen las esferas de consenso científico sobre las tendencias tecnológicas y las áreas en las que se necesita más investigación;
- b. Producir compendios trimestrales de investigación temática sobre las esferas en las que la IA podría contribuir a la consecución de los ODS, centrándose en las áreas de interés público que puedan estar insuficientemente atendidas; y
- c. Publicar informes ad hoc sobre cuestiones emergentes, en particular la aparición de nuevos riesgos o las importantes lagunas existentes en el panorama de la gobernanza.

91 Existen precedentes de este tipo de instituciones. Algunos ejemplos son el Comité Científico de las Naciones Unidas para el Estudio de los Efectos de las Radiaciones Atómicas, la Plataforma Intergubernamental Científico-Normativa sobre Diversidad Biológica y Servicios de los Ecosistemas (IPBES), el Comité Científico de Investigaciones Antárticas y el Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC).

- 92** Estos modelos son conocidos por sus enfoques sistemáticos de cuestiones complejas y generalizadas que afectan a diversos sectores y poblaciones mundiales. Sin embargo, aunque pueden servir de inspiración, ninguno es totalmente idóneo para evaluar la tecnología de la IA, por lo que no deberían reproducirse directamente. En su lugar se requiere un enfoque a medida.
- 93** Aprendiendo de estos precedentes, un grupo independiente, internacional y multidisciplinar de científicos expertos en IA podría recopilar y dinamizar investigaciones punteras que sirvan de base a quienes busquen perspectivas científicas sobre la tecnología de la IA o sus aplicaciones en una fuente imparcial y creíble. Un ejemplo del tipo de cuestiones a las que podría contribuir el grupo de expertos es el actual debate sobre los sistemas de IA abiertos frente a los cerrados, tratado en el recuadro 9.
- 94** Un grupo de expertos científicos bajo los auspicios de las Naciones Unidas adoptaría un enfoque amplio para abarcar de forma integral una amplia gama de prioridades. Este enfoque podría incluir la obtención de conocimientos especializados sobre oportunidades relacionadas con la IA y la promoción de una “inmersión en profundidad” en ámbitos de los ODS como la atención de la salud, la energía, la educación, las finanzas, la agricultura, el clima, el comercio y el empleo.
- 95** Las evaluaciones de riesgos también podrían basarse en el trabajo de otras iniciativas de investigación sobre IA, ofreciendo las Naciones Unidas un “refugio seguro” de confianza para que los investigadores intercambien ideas sobre los últimos avances. El derecho internacional, incluido el derecho internacional de los derechos humanos, proporcionaría una brújula para definir los riesgos pertinentes. Al poner en común los conocimientos de países o empresas que de otro modo no participarían o no estarían incluidos, un grupo de expertos albergado en las Naciones Unidas puede ayudar a rectificar las percepciones erróneas y reforzar la confianza a nivel mundial.
- 96** Un grupo de expertos de este tipo no llevaría a cabo necesariamente su propia investigación, sino que actuaría como catalizador de la acción en red²². Podría recopilar, extraer y trasladar los avances en el ámbito de la IA a sus destinatarios, resaltando los posibles casos de uso. Reduciría la asimetría de la información, ayudaría a impedir inversiones mal orientadas y mantendría el flujo de información a través de una red mundial de expertos.
- 97** El grupo de expertos tendría tres principales grupos destinatarios:
- El primero es la comunidad científica mundial²³. El traslado de la investigación fundamental sobre la IA a las empresas privadas, impulsado en parte por el costo de la capacidad de computación, ha suscitado la preocupación de que dicha investigación pueda estar indebidamente guiada por intereses financieros. Un grupo de expertos científicos podría promover una mayor investigación en instituciones públicas de todo el mundo centrada en el bien común.
 - En segundo lugar, las evaluaciones periódicas independientes servirían de fundamento a los Estados Miembros, los responsables de formular políticas y los demás procesos recomendados en este informe. Una encuesta anual de riesgos realizada por expertos de todo el mundo ayudaría a conformar la agenda del diálogo que sobre gobernanza de la IA se propone en la recomendación 2. El informe sobre los últimos avances serviría de base para la elaboración de normas propuesta en la recomendación 3, así como para la red de desarrollo de la capacidad propuesta en la recomendación 4.
 - En tercer lugar, a través de sus informes públicos, podría ser una fuente imparcial de información de alta calidad para la ciudadanía.

²² Podría basarse, en particular, en los grupos sectoriales o regionales de expertos ya existentes y en funcionamiento.
²³ También podría llegar a un público más amplio, incluida la sociedad civil y el público en general.

- 98** El alcance global de las redes, accesible únicamente a través de las Naciones Unidas, permitiría el entendimiento común de la base más amplia, poniendo a disposición los hallazgos de forma pertinente a los diversos contextos socioeconómicos y geográficos. De este modo, el grupo de expertos puede convertir a las Naciones Unidas en una plataforma fiable para un entendimiento inclusivo en red y multidisciplinar de las partes interesadas.
- 99** El grupo de expertos podría establecerse por un período inicial de 3 a 5 años (con una prórroga sujeta a revisión por parte del Secretario General), y podría funcionar con arreglo a lo siguiente:
- El grupo podría comenzar con un número de entre 30 y 50 miembros nombrados mediante una combinación de designación por los Estados Miembros y autodesignación, de forma similar a como se creó el Órgano Asesor. El grupo de expertos debería centrarse en los conocimientos científicos de las distintas disciplinas, garantizar una representación diversa por regiones y género, y reflejar la naturaleza interdisciplinar de la IA. Los miembros podrían rotar periódicamente dentro del mandato general de 3 a 5 años.
 - El grupo de expertos podría reunirse virtualmente (y presencialmente en sesión plenaria, quizá dos veces al año). Las reuniones podrían celebrarse de forma rotatoria en las ciudades que albergan las entidades pertinentes de las Naciones Unidas, entre ellas las del Sur Global. Se debería alentar al grupo a formar grupos de trabajo temáticos, añadiendo miembros adicionales según sea necesario e implicando a redes de asociados académicos. En estos grupos de trabajo se podría invitar a participar a las entidades pertinentes de las Naciones Unidas.²⁴
- El grupo actuaría de forma independiente, en particular en lo relativo a sus constataciones y conclusiones, con el apoyo de un equipo del sistema de las Naciones Unidas procedente de la oficina de IA y de organismos pertinentes de las Naciones Unidas como la UIT y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO).
 - El grupo debería colaborar con otras instituciones internacionales como la OCDE y la Alianza Mundial sobre la Inteligencia Artificial, con otros procesos pertinentes como el reciente informe científico sobre los riesgos de la IA avanzada encargado por el Reino Unido²⁵, y con las organizaciones regionales pertinentes, y trabajar aprovechando sus esfuerzos de investigación.
 - Un comité directivo elaboraría un programa de investigación que garantizara la inclusión de puntos de vista y la incorporación de consideraciones éticas, supervisaría la asignación de recursos, fomentaría la colaboración con una red de instituciones académicas y otras partes interesadas, y revisaría las actividades y los resultados del grupo.
- 100** Aprovechando el poder de convocatoria único de las Naciones Unidas y su alcance mundial inclusivo entre los grupos de partes interesadas, un grupo internacional de científicos expertos podría ofrecer procesos y resultados de colaboración científica fiables y corregir las asimetrías de información de forma que se aborden las lagunas de representación y coordinación identificadas en los párrafos 66 y 73, promoviendo con ello una gobernanza internacional de la IA equitativa y eficaz.

²⁴ Para consultar una lista de entidades de las Naciones Unidas activas en este ámbito, véase la Figura 9.

²⁵ International Scientific Report on the Safety of Advanced AI: Interim Report. Disponible en: <https://gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai>.

Recuadro 9: Sistemas de IA abiertos frente a cerrados

Uno de los temas que tratamos en nuestras consultas fue el actual debate sobre los sistemas de IA abiertos frente a cerrados. Los sistemas de IA que son abiertos en diversos grados suelen denominarse “IA de código abierto”, pero este es un término algo equívoco si se compara con el software de código abierto. Es importante reconocer que la apertura de los sistemas de IA es más un conjunto de características que un único atributo.

En un artículo se explicó que “un sistema de IA totalmente cerrado solamente es accesible para un grupo concreto. Puede tratarse de una empresa desarrolladora de IA o de un grupo específico dentro de ella, principalmente con fines de investigación y desarrollo internos. Por otro lado, los sistemas más abiertos pueden autorizar el acceso público o poner a su disposición algunas partes, como los datos, el código o las características del modelo, para facilitar el desarrollo externo de la IA”^a.

Los sistemas de IA de código abierto en el ámbito de la IA generativa presentan tanto riesgos como oportunidades. Las empresas aducen a menudo la “seguridad de la IA” como razón para no revelar las especificaciones del sistema, lo que refleja la tensión existente entre los enfoques abiertos y cerrados en el sector. Los debates suelen girar en torno a dos extremos: la apertura total, que implica compartir todos los componentes y conjuntos de datos del modelo, y la apertura parcial, que implica revelar únicamente las ponderaciones del modelo.

Los sistemas de IA de código abierto fomentan la innovación y suelen ser un requisito para obtener financiación pública. En el extremo abierto del espectro, cuando el código subyacente se pone a libre disposición, desarrolladores de todo el mundo pueden experimentar, mejorar y crear nuevas aplicaciones. Esto fomenta un entorno de colaboración en el que las ideas y los conocimientos especializados se comparten fácilmente. Algunos líderes del sector sostienen que esta apertura es vital para la innovación y el crecimiento económico.

Sin embargo, en la mayoría de los casos, los modelos de IA de código abierto están disponibles como interfaz de programación de aplicaciones. En este caso, el código original no se comparte, las ponderaciones originales nunca se cambian y las actualizaciones de los modelos se convierten en nuevos modelos.

Además, los modelos de código abierto suelen ser más pequeños y transparentes. Esta transparencia puede generar confianza, permitir que las consideraciones éticas se aborden de forma proactiva y apoyar la validación y reproducción de los modelos, ya que los usuarios pueden examinar el funcionamiento interno del sistema de IA, comprender su proceso de toma de decisiones e identificar posibles sesgos.

Los sistemas de IA cerrados ofrecen un mayor control a sus desarrolladores. Además, los sistemas de código cerrado pueden ser más ágiles y eficientes, ya que el código base no está en constante evolución por las contribuciones públicas. Muchas empresas consideran poco práctica la apertura total y promueven la apertura parcial como única opción viable. Sin embargo, este punto de vista pasa por alto el potencial que una apertura significativa ofrece para lograr un “enfoque equilibrado”^b.

a Angela Luna, “The open or closed AI dilemma”, 2 de mayo de 2024. Disponible en <https://bipartisanpolicy.org/blog/the-open-or-closed-ai-dilemma>.
b Inspirado por Andreas Liesenfeld y Mark Dingemans, “Rethinking open source generative AI: open-washing and the EU AI Act”, Conferencia ACM de 2024 sobre equidad, responsabilidad y transparencia (FAccT '24) (junio de 2024).

Recuadro 9: Sistemas de IA abiertos frente a cerrados (continuado)

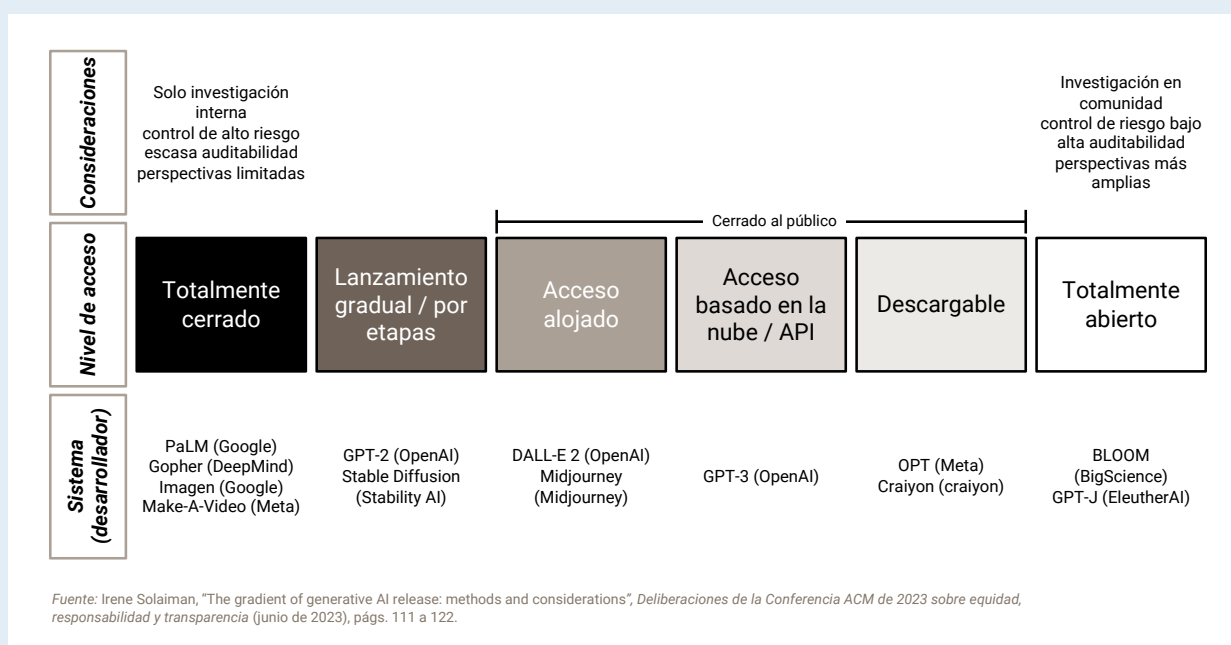
Una apertura significativa se sitúa entre los dos extremos del espectro y puede adaptarse a diferentes casos de uso. Este método equilibrado fomenta el desarrollo seguro, innovador e inclusivo de la IA al permitir el escrutinio público y la auditoría independiente de los datos de entrenamiento y ajuste revelados. La apertura, que va más allá del mero hecho de compartir las ponderaciones de los modelos, puede impulsar la innovación y la inclusión, ayudando a las aplicaciones en los ámbitos de la investigación y la educación.

La definición de “IA de código abierto” está evolucionando⁶ y a menudo se ve influida por intereses corporativos, como se ilustra en la Figura 11. Para hacer frente a esta situación, recomendamos iniciar un proceso, coordinado por el grupo internacional de científicos propuesto, para desarrollar un enfoque completo y gradual de la apertura. Esto permitiría adoptar un enfoque sólido y basado en pruebas de la apertura, que ayudaría a los usuarios y a los responsables de formular políticas a tomar decisiones informadas sobre los modelos y arquitecturas de la IA.

La divulgación de datos - aunque se limite a elementos clave - es esencial para comprender el desempeño de los modelos, garantizar la capacidad de reproducción y evaluar los riesgos jurídicos. La claridad en torno a los grados de apertura puede ayudar a contrarrestar el “compromiso superficial” de las empresas y fomentar un ecosistema tecnológico transparente.

También es importante que, a medida que la tecnología vaya madurando, examinemos los regímenes de gobernanza para la aplicación de sistemas de IA tanto abiertos como cerrados. Debemos elaborar directrices de IA responsables, normas vinculantes y estándares susceptibles de medición para los desarrolladores y diseñadores de productos y servicios que incorporen tecnologías de IA, así como para sus usuarios y todos los agentes implicados a lo largo de su ciclo de vida.

Figura 11: Intereses corporativos y apertura



c The Open Source AI Definition – draft v. 0.0.3. Disponible en <https://opensource.org/deepdive/drafts/the-open-source-ai-definition-draft-v-0-0-3>.

B. Un terreno común

101 Junto a un entendimiento común de la IA, es necesario un terreno común para establecer enfoques de gobernanza que sean interoperables en todas las jurisdicciones y se fundamenten en normas internacionales, como la Declaración Universal de Derechos Humanos (véase el principio rector 5).

102 Esto es necesario a nivel mundial no solo para lograr una representación equitativa, sino también para evitar una “competencia a la baja” en la regulación y reducir la fricción reguladora transfronteriza, maximizar la interoperabilidad técnica y ontológica, y detectar y responder a los incidentes derivados de las decisiones que a lo largo del ciclo de vida de la IA afectan a múltiples jurisdicciones.

Diálogo sobre políticas de gobernanza de la IA

Recomendación 2: Diálogo sobre políticas de gobernanza de la IA

Recomendamos el inicio de un diálogo sobre políticas intergubernamental y multilateral sobre la gobernanza de la IA dos veces al año, en paralelo a las reuniones que se celebren en las Naciones Unidas. Su objetivo sería:

- a. Compartir las mejores prácticas sobre gobernanza de la IA que respeten, protejan y promuevan los derechos humanos y fomenten el desarrollo, y que incluyan tanto la búsqueda de oportunidades como la gestión de riesgos;
- b. Promover un entendimiento común sobre la implementación de medidas de gobernanza de la IA por parte de los desarrolladores y los usuarios de los sectores público y privado para mejorar la interoperabilidad internacional de la gobernanza de la IA;

- c. Compartir voluntariamente incidentes significativos de IA que hayan sobrepasado la capacidad de respuesta de los organismos estatales; y
- d. Debatir sobre los informes del grupo internacional de científicos expertos en IA, según proceda.

103 La gobernanza internacional de la IA está actualmente fragmentada, y es un mosaico en el mejor de los casos. Hay 118 países que no participan en ninguna de las siete iniciativas destacadas recientes de gobernanza de la IA ajenas a las Naciones Unidas y de carácter intergubernamental²⁶ (véase la Figura 8). Las lagunas de representación se dan incluso entre los 60 países con mayor capacidad en IA, lo que pone de manifiesto la selectividad de la gobernanza internacional de la IA en la actualidad (véase la Figura 12).

104 Es preciso establecer un foro político inclusivo para que todos los Estados Miembros, aprovechando los conocimientos especializados de las partes interesadas, puedan compartir las mejores prácticas que protegen los derechos humanos y promueven el desarrollo, adoptar enfoques de gobernanza interoperables y hacer un seguimiento de los riesgos comunes que justifican nuevas intervenciones políticas.

105 Esto no implica una gobernanza global de todos los aspectos de la IA (por otro lado imposible y no deseable, dados los diferentes intereses y prioridades de los Estados). Sin embargo, el intercambio de puntos de vista sobre los avances de la IA y las respuestas políticas puede establecer el marco para la cooperación internacional.

26

Estas iniciativas no siempre son directamente comparables. Algunas reflejan el trabajo de las organizaciones internacionales o regionales existentes, mientras que otras se basan en invitaciones ad hoc de países con ideas afines.

Figura 12: Los 60 países con mayor capacidad de IA (Índice Tortoise 2023) en la muestra de las principales iniciativas plurilaterales de gobernanza de la IA de carácter intergubernamental

parte de

País / parte*	Índice Tortoise Global AI (2023)	Número de iniciativas en las que participa	Orden Cronológico →						
			Principios de la OCDE sobre la IA (2019)	Principios del G20 sobre la IA (2019)	Redactores del Consejo de Europa (2022)	Declaración Ministerial de la GPA (2022)	Declaración de Bletchley (2023)	Declaración Ministerial del G7 relativa al proceso de Hiroshima sobre IA (2023)	Declaración Ministerial de Seúl (2024)
Estados Unidos de América	1	7							
China	2	2							
Singapur	3	4							
Reino Unido de Gran Bretaña e Irlanda del Norte	4	7							
Canadá	5	7							
República de Corea	6	5							
Israel	7	5							
Alemania	8	7							
Suiza	9	4							
Finlandia	10	2							
Países Bajos	11	5							
Japón	12	7							
Francia	13	7							
India	14	4							
Australia	15	6							
Dinamarca	16	3							
Suecia	17	3							
Luxemburgo	18	2							
Irlanda	19	4							
Austria	20	2							
España	21	5							
Belgium	22	3							
Italia	23	7							
Noruega	24	2							
Estonia	25	2							
Emiratos Árabes Unidos	27	2							
Portugal	28	2							
Federación de Rusia	29	1							
Arabia Saudita	30	3							
Malta	31	2							
Brasil	33	4							
Nueva Zelanda	34	3							
Eslovenia	35	3							
Hungría	36	2							
Turquía	37	6							
Islandia	38	2							
Chile	39	3							
Qatar	40	0							
Lituania	41	2							
Malasia	42	0							
Grecia	43	2							
Indonesia	44	3							
Viet Nam	45	0							
Colombia	46	1							
Argentina	47	4							
Eslovaquia	48	2							
México	49	5							
Egipto	50	1							
Uruguay	51	1							
Armenia	52	1							
Sudáfrica	53	1							
Túnez	54	0							
Marruecos	55	0							
Bahrein	56	0							
Pakistán	57	0							
Sri Lanka	58	0							
Nigeria	59	2							
Kenya	60	2							
Unión Europea	n/a	5							
Total (incluidos los no mostrados)			47	20	58	29	29	7	28

*Including jurisdictions such as the Holy See and the European Union.

Sources:

- OECD Recommendation of the Council on Artificial Intelligence (adopted 21 May 2019), available at <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
- G20, AI Principles (June 2019), available at https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf.
- GPAI, 2022 ministerial declaration (22 November 2022), available at [https://one.oecd.org/document/GPAI/C\(2022\)17/FINAL/en.pdf](https://one.oecd.org/document/GPAI/C(2022)17/FINAL/en.pdf).
- Bletchley Declaration (1 Nov 2023), available at <https://gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>.
- G7, Hiroshima AI Process G7 Digital & Tech Ministers' Statement (1 Dec 2023), available at https://www.soumu.go.jp/hiroshimaiiprocess/pdf/document02_en.pdf.
- Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (adopted 17 May 2024), available at <https://coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>.
- Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity, AI Seoul Summit (22 May 2024).
- Tortoise Media, Global AI Index (2023), available at <https://tortoisemedia.com/intelligence/global-ai/#rankings>.

106 Las Naciones Unidas se encuentran en una posición única para facilitar estos diálogos de forma inclusiva y ayudar a los Estados Miembros a trabajar juntos de forma eficaz. El conjunto de normas existentes y nuevas del sistema de las Naciones Unidas puede ofrecer sólidos fundamentos normativos para la acción concertada, basados en la Carta de las Naciones Unidas, los derechos humanos y otras normas internacionales como el derecho ambiental y el derecho internacional humanitario, así como en los ODS y otros compromisos internacionales²⁷.

107 En combinación con los conocimientos especializados del grupo internacional de científicos expertos y el desarrollo de la capacidad (véanse las recomendaciones 1, 4 y 5), el diálogo inclusivo en las Naciones Unidas puede ayudar a los Estados y a las empresas a actualizar sus enfoques y metodologías de regulación para seguir el ritmo de la aceleración de la IA de una manera interoperable que promueva un terreno común. Algunas de las características distintivas de las Naciones Unidas pueden ser útiles a este respecto:

- a. Anclar el diálogo inclusivo en el conjunto de normas de las Naciones Unidas, como la Carta de las Naciones Unidas y el derecho internacional de los derechos humanos, puede promover una “competencia al alza” en los enfoques de gobernanza. Y a la inversa, sin la universalidad de las Naciones Unidas, la acción colectiva internacional se enfrenta a una mayor presión para sucumbir a una “competencia reguladora a la baja” entre las jurisdicciones en materia de seguridad y ámbito de uso de la IA.
- b. La universalidad de las Naciones Unidas también puede facilitar la coordinación entre las iniciativas submundiales existentes para lograr una mayor compatibilidad entre ellas. Muchos de los participantes en nuestras

consultas pidieron que las Naciones Unidas se convirtieran en un espacio clave para facilitar la coordinación entre las iniciativas regionales y plurilaterales existentes, teniendo en cuenta los diversos valores de las distintas culturas, lenguas y regiones.

- c. Los procedimientos predecibles, transparentes, basados en normas y justificables de la Organización posibilitan un compromiso político continuo para tender puentes entre países no afines y moderar la peligrosa conflictividad. Además de fomentar la confianza, las relaciones y las líneas de comunicación en tiempos de crisis, los diálogos inclusivos y fiables pueden promover nuevas normas, derecho consuetudinario y acuerdos que mejoren la cooperación entre los Estados.

108 Operacionalmente:

- a. Debería iniciarse un diálogo sobre políticas en paralelo a las reuniones que se celebran en Nueva York, como la Asamblea General²⁸, Ginebra y otros lugares del Sur Global.
- b. Una parte de cada sesión de diálogo podría centrarse en los enfoques nacionales de los Estados Miembros, y una segunda parte podría hacerlo en la obtención de especialización y aportaciones de las principales partes interesadas, en particular las empresas tecnológicas y los representantes de la sociedad civil.
- c. La participación gubernamental podría estar abierta a todos los Estados Miembros, a una agrupación equilibrada por regiones (para un debate más centrado entre un subconjunto interesado representativo y rotatorio) o a una combinación de ambas, ajustada según convenga a los distintos puntos o segmentos de las agendas, a medida que evolucione la

²⁷ Véase, por ejemplo, la Carta de las Naciones Unidas (párrafo, propósitos y principios y artículos 13, 55, 58 y 59); véanse también los principales instrumentos internacionales en materia de derechos humanos (Declaración Universal de Derechos Humanos, Pacto Internacional de Derechos Civiles y Políticos, Pacto Internacional de Derechos Económicos, Sociales y Culturales, Convención Internacional sobre la Eliminación de Todas las Formas de Discriminación Racial, Convención sobre los Derechos del Niño, Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer, Convención contra la Tortura, Convención sobre los Derechos de las Personas con Discapacidad, Convención Internacional sobre la Protección de los Derechos de Todos los Trabajadores Migratorios y de Sus Familiares, Convención Internacional para la Protección de Todas las Personas contra las Desapariciones Forzadas); los instrumentos del derecho internacional de los derechos humanos (Convenciones de Ginebra, Convención sobre Ciertas Armas Convencionales, Convención sobre el Genocidio, Convención de La Haya, etc.); los instrumentos relativos a los principios conexos como la distinción, la proporcionalidad y la precaución, y los 11 principios sobre los sistemas de armas autónomos letales adoptados en la Convención sobre Ciertas Armas Convencionales; los instrumentos de desarme y control de armamentos en lo que respecta a la prohibición de las armas de destrucción masiva (Tratado sobre la No Proliferación de las Armas Nucleares, Convención sobre las Armas Químicas, Convención sobre las Armas Biológicas y Toxínicas); los instrumentos del derecho ambiental (Convención Marco de las Naciones Unidas sobre el Cambio Climático, la Convención sobre la Prohibición de Utilizar Técnicas de Modificación Ambiental con Fines Militares u Otros Fines Hostiles, el Acuerdo de París y los principios conexos como el principio de precaución, el principio de integración y la participación pública; y compromisos no vinculantes sobre la Agenda 2030 para el Desarrollo Sostenible, género y ética, como la Recomendación de la UNESCO sobre la ética de la inteligencia artificial.

²⁸ Análogo al foro político de alto nivel sobre el desarrollo sostenible que tiene lugar bajo los auspicios del Consejo Económico y Social en el contexto de los ODS.

tecnología y surjan o adquieran relevancia las preocupaciones globales. Dado el carácter dinámico de la tecnología y del contexto político, una estructura rígida podría no resultar útil.

- d. Además de las sesiones oficiales de diálogo, la participación de las múltiples partes interesadas en las políticas de IA también podría aprovechar otros mecanismos existentes como la reunión AI for Good de la UIT, la reunión anual del Foro para la Gobernanza de Internet, el Foro sobre la Ética de la Inteligencia Artificial de la UNESCO y la eWeek de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD), abierta a la participación voluntaria de representantes de todos los Estados Miembros.
- e. En consonancia con la naturaleza inclusiva del diálogo, las agendas de los debates podrían ser amplias para abarcar diversas perspectivas y preocupaciones. Por ejemplo, las reuniones bianuales podrían centrarse en las oportunidades entre los distintos sectores en una reunión, y en las tendencias de riesgos en la otra.²⁹ Las agendas podrían incluir los usos de la IA para lograr los ODS, cómo proteger a los niños o minimizar el impacto del cambio climático, y un intercambio de enfoques para gestionar los riesgos. Las reuniones también podrían incluir un debate sobre la definición de los términos utilizados en la gobernanza y las normas técnicas de la IA, así como en los informes del grupo internacional de científicos expertos en IA, según proceda.
- f. Además, se podría invitar a diversas partes interesadas, en particular a las empresas tecnológicas y a representantes de la sociedad civil, a participar a través de las instituciones que se detallan a continuación, así como en talleres de política sobre aspectos concretos

de la gobernanza de la IA, como los límites (si los hubiera) de los enfoques de código abierto para las formas más avanzadas de IA, los umbrales para el seguimiento y la notificación de incidentes de IA, la aplicación de la legislación sobre derechos humanos a los nuevos casos de uso, o el uso de la legislación sobre competencia y antimonopolio para hacer frente a las concentraciones de poder entre las empresas tecnológicas³⁰.

- g. La oficina de IA propuesta podría crear también un repositorio de ejemplos de gobernanza de la IA que incluya legislación, políticas e instituciones de todo el mundo para su consideración en el diálogo sobre políticas, en colaboración con los actuales foros como la OCDE.

109 A pesar de las dos resoluciones de la Asamblea General sobre la IA en 2024, actualmente no existe un diálogo institucionalizado con mandato sobre la gobernanza de la IA en las Naciones Unidas que se corresponda con la visión inclusiva y fiable de esta recomendación. Existen procesos similares a nivel internacional, pero principalmente en ámbitos regionales o plurilaterales (párr. 57) que no son inclusivos ni globales de forma fiable.

110 Como complemento a un proceso fluido de cumbres plurilaterales y regionales sobre IA, las Naciones Unidas pueden ofrecer una sede estable para los diálogos sobre la gobernanza de la IA³¹. La inclusión por naturaleza - un requisito crucial para desempeñar un papel estabilizador en tiempos geopolíticamente delicados - también puede abordar las lagunas de representación y coordinación identificadas en los párrafos 64 y 72, promoviendo una acción colectiva más eficaz sobre la gobernanza de la IA en el interés común de todos los países.

29 Las partes pertinentes del sistema de las Naciones Unidas podrían participar para poner de relieve las oportunidades y los riesgos existentes; por ejemplo, la UIT podría hacerlo en torno a las normas sobre la IA; la UIT, la UNCTAD, el PNUD y la Oficina de Coordinación del Desarrollo lo harían en torno a las aplicaciones de la IA para los ODS; la UNESCO podría hacerlo en torno a la ética y la capacidad de gobernanza; la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH) lo haría sobre la rendición de cuentas en materia de derechos humanos basada en las normas y los mecanismos existentes; la Oficina de Asuntos de Desarme podría hacerlo en torno a la regulación de la IA en los sistemas militares; el PNUD lo haría sobre el apoyo a la capacidad nacional de desarrollo; el Foro para la Gobernanza de Internet lo haría en torno a la participación y el diálogo de las múltiples partes interesadas; y la OMPI, la OIT, la OMS, la FAO, el Programa Mundial de Alimentos, el ACNUR, la UNESCO, el Fondo de las Naciones Unidas para la Infancia, la Organización Meteorológica Mundial y otros lo harían en torno a las aplicaciones sectoriales y la gobernanza.

30 Una reunión de este tipo podría brindar también la oportunidad de debatir con las múltiples partes interesadas cualquier endurecimiento de la gobernanza global de la IA. Ello podría incluir, por ejemplo, prohibiciones sobre el desarrollo de sistemas de IA incontenibles o incontrolables, o la obligación de que todos los sistemas de IA sean lo suficientemente transparentes como para que sus consecuencias puedan ser rastreadas hasta un agente jurídico que asuma la responsabilidad de las mismas.

31 Aunque las múltiples cumbres sobre IA han ayudado a un subconjunto de 20 a 30 países a alinearse en cuestiones de seguridad de la IA, la participación ha sido desigual: el Brasil, China e Irlanda suscribieron la Declaración de Bletchley en noviembre de 2023, pero no la Declaración Ministerial de Seúl seis meses después (véase la Figura 12). Por el contrario, México y Nueva Zelanda suscribieron la Declaración Ministerial de Seúl, pero no se adhirieron a la Declaración de Bletchley.

Intercambio de normas sobre la IA

Recomendación 3: Intercambio de normas sobre la IA

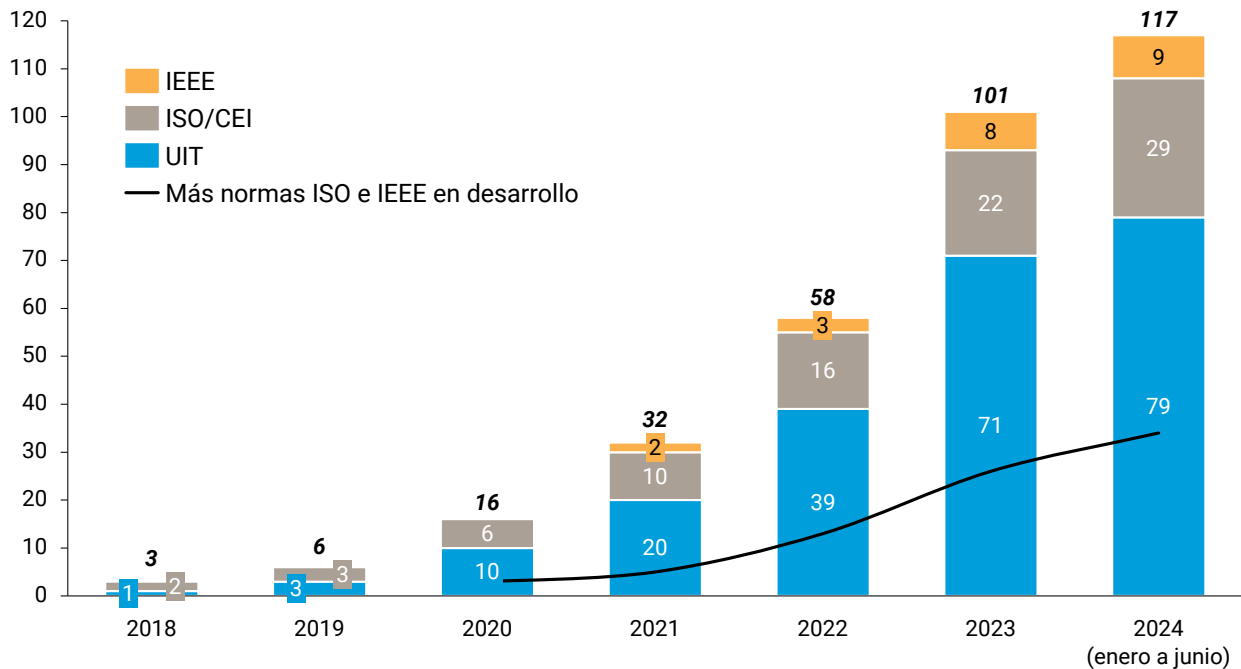
Recomendamos el establecimiento de una entidad de intercambio de normas sobre la IA que reúna a representantes de organizaciones nacionales e internacionales de normalización, empresas tecnológicas, la sociedad civil y el grupo internacional de científicos expertos. La entidad se encargaría de:

- Elaborar y mantener un registro de definiciones y normas aplicables para medir y evaluar los sistemas de IA;
- Debatir y evaluar las normas y los procesos para su elaboración; e
- Identificar las lagunas que requieran nuevas normas.

111 Cuando se empezaron a desarrollar los sistemas de IA, existían pocas normas para ayudar a navegar o medir esta nueva frontera. La prueba de Turing, que consiste en determinar si una máquina puede mostrar un comportamiento equivalente (o indistinguible) del de un ser humano, cautivó la imaginación popular, pero tiene más importancia cultural que científica. De hecho, es revelador que algunos de los mayores avances computacionales se hayan medido por su éxito en los juegos, como cuando un ordenador podía ganar a los humanos al ajedrez, al Go, al póquer o a Jeopardy. Tales medidas eran fácilmente comprensibles para los no especialistas, pero no eran ni rigurosas ni especialmente científicas.

112 Más recientemente se ha producido una proliferación de normas. La Figura 13 ilustra el creciente número de normas pertinentes adoptadas por la UIT, la Organización Internacional de Normalización (ISO), la Comisión Electrotécnica Internacional (CEI) y el Instituto de Ingenieros Electricistas y Electrónicos (IEEE)³².

Figura 13: Número de normas relacionadas con la IA



Fuentes: IEEE, ISO/CEI, UIT, World Standards Cooperation (en base a la información de junio de 2023, ampliada con la inclusión de normas relacionadas con la IA).

³² También están proliferando nuevas normas a escalas nacional y multinacional, como los compromisos voluntarios de la Casa Blanca de los Estados Unidos en materia de IA y los códigos de buenas prácticas de la Unión Europea previstos en la AI Act.

- 113** Se observan dos tendencias. En primer lugar, estas normas se elaboraron en su mayoría para abordar cuestiones específicas. No hay una formulación común y muchos términos que se utilizan habitualmente en relación con la IA—equidad, seguridad, transparencia—no tienen una definición consensuada ni son susceptibles de medición (a pesar de la reciente labor de la OCDE y del Instituto Nacional de Normas y Tecnología, que han adoptado un nuevo enfoque para sistemas dinámicos como la IA).
- 114** En segundo lugar, existe una disociación entre las normas que se adoptaron con fines estrictamente técnicos o de validación interna y las que pretenden incorporar principios éticos más amplios. Los informáticos y los especialistas en ciencias sociales suelen avanzar diferentes interpretaciones de un mismo concepto, por lo que un paradigma común de normas sociotécnicas es prometedor, pero sigue siendo una aspiración (véase el recuadro 10).
- 115** La consecuencia es que tenemos un nuevo conjunto de normas que no se basan en un entendimiento común de su significado o que están desvinculadas de los valores que pretendían defender. Sobre todo, existen pocas normas consensuadas en materia de consumo de energía y de IA. La falta de integración de las consideraciones de derechos humanos en los procesos de elaboración de normas es otra laguna que debe colmarse³³.
- 116** Esto tiene un costo real. Además de las preocupaciones de los Estados Miembros y de los particulares, las empresas (incluidas las pequeñas y medianas empresas de los países en desarrollo) expresaron en muchas de nuestras consultas su preocupación por el hecho de que la fragmentación de la gobernanza y la incoherencia de las normas puedan elevar los costos de la actividad empresarial en un mundo cada vez más globalizado.
- 117** Este informe no está proponiendo que las Naciones Unidas se sumen a la proliferación de normas. Al contrario, aprovechando los conocimientos especializados del grupo internacional de científicos (propuesto en la recomendación 1) e incorporando a miembros de las diversas entidades nacionales e internacionales que han contribuido al establecimiento de normas, así como a representantes de las empresas tecnológicas y de la sociedad civil, el sistema de las Naciones Unidas podría funcionar como centro de intercambio de normas de aplicación mundial sobre la IA³⁴.
- 118** El valor añadido de la Organización consistiría en fomentar el intercambio entre el conjunto más amplio de organizaciones de normalización con el fin de maximizar la interoperabilidad mundial de las normas técnicas, al tiempo que se incorporan, a los debates sobre las normas de la IA, los nuevos conocimientos sobre la elaboración de normas sociotécnicas.
- 119** La recopilación y distribución de información sobre normas de IA, aprovechando y trabajando con iniciativas existentes como el *AI Standards Hub*³⁵, permitiría a los participantes de todas las organizaciones de normalización converger en un lenguaje común en las esferas clave.
- 120** Con el apoyo de la oficina de IA propuesta, la entidad de intercambio de normas podría también aprovechar los fuertes vínculos con el grupo internacional de expertos científicos en las cuestiones técnicas y con el diálogo sobre políticas en las cuestiones morales, éticas, normativas, jurídicas y políticas.

33 Véase A/HRC/53/42 (Derechos humanos y procesos de establecimiento de normas técnicas para las tecnologías digitales nuevas y emergentes: Informe de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos) y Resolución 53/29 del Consejo de Derechos Humanos (Tecnologías digitales nuevas y emergentes y derechos humanos).

34 Esto puede parecer una tarea difícil, pero el avance hacia un acuerdo sobre el impuesto mínimo global demuestra la posibilidad de la acción colectiva incluso en ámbitos económica y políticamente complejos.

35 Véase <https://aistandardshub.org>.

121 Si así se acordara, la UIT, la ISO, la CEI y el IEEE podrían liderar conjuntamente una cumbre inicial sobre normas de IA, con un seguimiento anual para mantener la relevancia y el impulso. Para sentar las bases de un enfoque sociotécnico que incorpore consideraciones económicas, éticas y de derechos humanos, también deberían participar la OCDE, la Organización Mundial de la Propiedad Intelectual (OMPI), la Organización Mundial del Comercio (OMC), la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (ACNUDH), la OIT, la UNESCO y otras entidades pertinentes de las Naciones Unidas.³⁶

122 La entidad de intercambio de normas también debería sustentar la labor de desarrollo de la capacidad prevista en la recomendación 4, para garantizar que las normas respalden la práctica sobre el terreno. Y se podría compartir información sobre las herramientas desarrolladas a escala nacional o regional que permitan autoevaluar el cumplimiento de las normas.

123 El informe no propone por el momento que las Naciones Unidas hagan nada más que servir de foro para debatir y acordar normas. En la medida en que las normas de seguridad se vayan formalizando con el tiempo, podrían servir de base para su seguimiento y verificación por parte de un eventual organismo.

Recuadro 10: Normas aplicables a la seguridad de la IA

Un enfoque integral de la seguridad de la IA implica comprender la capacidad de los modelos avanzados de IA, adoptar normas para el diseño y despliegue seguros, y evaluar tanto los sistemas como sus repercusiones más amplias.

En el pasado, las normas sobre la IA se centraban principalmente en las especificaciones técnicas, y detallaban cómo debían construirse y funcionar los sistemas. Sin embargo, dado que las tecnologías de IA tienen un impacto cada vez mayor en la sociedad, es necesario cambiar a un paradigma sociotécnico. Este cambio reconoce que los sistemas de IA no existen en el vacío; interactúan con los usuarios humanos y afectan a las estructuras sociales. Las normas modernas de IA pueden integrar consideraciones éticas, culturales y sociales junto a los requisitos técnicos. En el contexto de la seguridad, esto incluye garantizar la fiabilidad y la interpretabilidad, así como evaluar y mitigar los riesgos para los derechos individuales^a y colectivos, la seguridad nacional e internacional y la seguridad pública en diferentes contextos.

Uno de los principales objetivos de los recientemente creados institutos nacionales de seguridad de la IA es garantizar enfoques coherentes y eficaces de la seguridad de la IA. La armonización de estos enfoques permitiría que los sistemas de IA cumplieran unos elevados criterios de seguridad a escala internacional, lo que posibilitaría la innovación y el comercio transfronterizo manteniendo al mismo tiempo unos protocolos de seguridad rigurosos.

En la medida en que la “seguridad” es contextual, la participación de las diversas partes interesadas y culturas en la elaboración de tales normas mejora su pertinencia y eficacia y contribuye al entendimiento compartido de definiciones y conceptos. Al incorporar diversas perspectivas, los protocolos pueden abordar más a fondo la amplia gama de riesgos y beneficios potenciales asociados a las tecnologías de la IA.

^a Véase A/HRC/53/42 (Derechos humanos y procesos de establecimiento de normas técnicas para las tecnologías digitales nuevas y emergentes: Informe de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos) y Resolución 53/29 del Consejo de Derechos Humanos (Tecnologías digitales nuevas y emergentes y derechos humanos).

36 En la cumbre podrían participar las organizaciones de normalización sectoriales, nacionales y regionales pertinentes.

C. Beneficios comunes

124 La Agenda 2030, con sus 17 ODS, puede aportar un propósito único a la IA, distanciándola de un uso derrochador y perjudicial y orientando el arco de las inversiones hacia los desafíos del

desarrollo mundial. De lo contrario, las inversiones perseguirán beneficios incluso a costa de imponer externalidades negativas a los demás. Otra contribución distintiva que pueden hacer las Naciones Unidas es vincular la aplicación positiva de la IA a la garantía de una distribución equitativa de sus oportunidades (recuadro 11).

Recuadro 11: La IA y los ODS

El potencial de la IA para los avances científicos (recuadro 1) y la creación de oportunidades económicas (recuadro 2) son la base de la esperanza de que la IA pueda acelerar el progreso en la consecución de los ODS. Una revisión de los datos pertinentes realizada en 2023 sostenía que la IA puede actuar como facilitadora en 134 metas (el 79 %) relativas a todos los ODS, generalmente a través de una mejora tecnológica que permite superar ciertas limitaciones existentes.^a

Una visión general de las percepciones de los expertos queda ilustrada por los resultados de un ejercicio de estudio de las oportunidades encomendado para nuestra labor, en el que se encuestó a más de 120 expertos de 38 países acerca de sus expectativas sobre el impacto positivo de la IA en términos de avances científicos, actividades económicas y ODS. En la encuesta solo se preguntaba por las posibles implicaciones positivas de la IA.

En general, los expertos tenían expectativas dispares sobre la rapidez con la que la IA podría tener un impacto positivo importante (véase también la Figura 14):

- Se mostraron más optimistas sobre la **aceleración de los descubrimientos científicos**: 7 de cada 10 afirmaron que es probable que la IA tenga un impacto positivo importante en los próximos tres años o antes en los países de ingreso alto o mediano alto, y el 28 % pronosticó lo mismo para los países de ingreso mediano bajo o bajo.
- Alrededor de 5 de cada 10 encuestados estimaron probable un impacto positivo importante en el **aumento de la actividad económica** en los próximos tres años o antes en los países de ingreso alto o mediano alto, y el 32 % pronosticó lo mismo en los países de ingreso mediano bajo o bajo.
- Un total del 46 % esperó probable un impacto positivo importante en el **progreso de los ODS** en los próximos tres años o antes en los países de ingreso alto o mediano alto. Sin embargo, solo el 21 % vio probable que esto ocurriera en los países de ingreso mediano bajo o bajo, y 4 de cada 10 expertos estimaron que en estos lugares el impacto positivo importante sobre los ODS tardaría al menos diez años en producirse.

^a Véase Ricardo Vinuesa y otros, "The role of artificial intelligence in achieving the Sustainable Development Goals". *Nature Communication*, vol. 11, núm. 233 (enero de 2020). Este estudio también sostiene que 59 metas (35%, también en todos los ODS) pueden experimentar un impacto negativo del desarrollo de la IA.

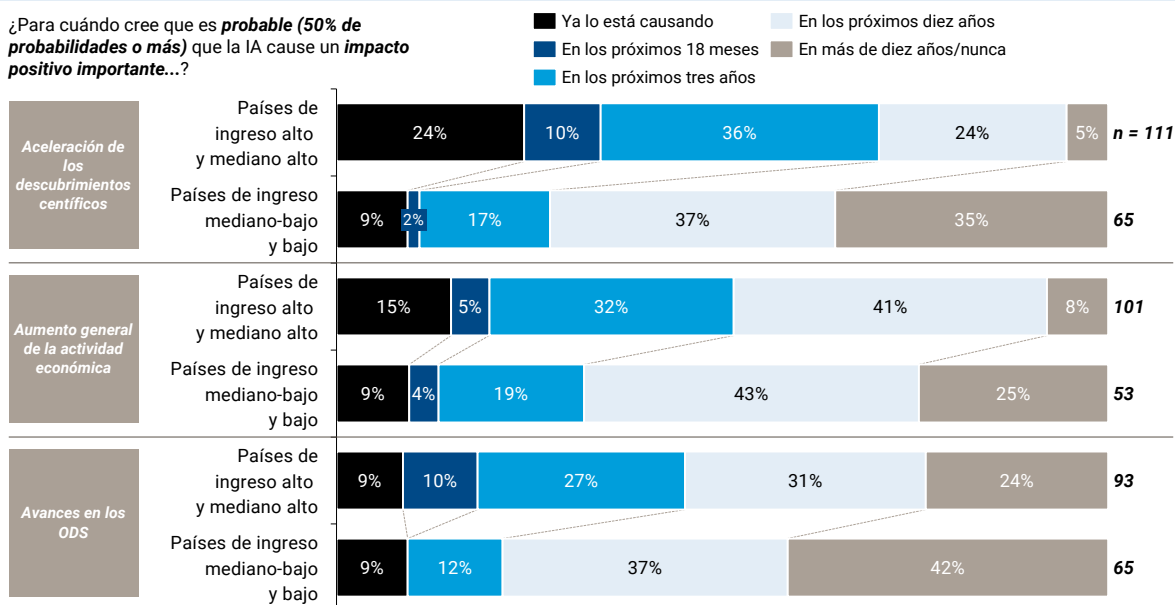
Recuadro 11: La IA y los ODS (Continuado)

Los expertos pronosticaron un mayor impacto positivo de la IA en los próximos tres años en los países de más alto ingreso en todos los ámbitos de la encuesta, incluida la aceleración de los descubrimientos científicos, el aumento de la actividad económica^b y en las 14 esferas de los ODS sobre las que se preguntó (véase la Figura 15). Los expertos se mostraron más optimistas sobre el impacto positivo de la IA en la salud y la educación (ODS 3 y 4), donde entre el 20 y el 25 % de los expertos esperaron un impacto positivo importante o transformador de la IA en los próximos tres años en los países de ingreso alto y mediano alto. Se mostraron menos optimistas en cuanto al impacto positivo de la IA en la igualdad de género y las desigualdades (ODS 5 y 10), y 2 de cada 3 consideraron que la IA no tendría ningún impacto positivo en la reducción de las desigualdades dentro de los países o entre ellos, tanto en los países de ingreso alto como en los de ingreso bajo.

Cabe esperar que la IA tenga un impacto mayor y más temprano en los países de ingreso alto, en parte debido a los obstáculos que impiden su avance en los países de ingreso mediano bajo y bajo (véase la Figura 16). Más de la mitad de los encuestados citaron la falta de facilitadores -desde infraestructuras más deficientes hasta la falta de políticas nacionales y de gobernanza internacional- como un factor importante que dificulta aún más a los países de ingreso bajo aprovechar la IA para la actividad económica y el progreso de los ODS.

Estos resultados subrayan el carácter incierto de la posible contribución de la IA a los ODS y el hecho de que esta siga dependiendo en gran medida de los factores facilitadores que faltan. Esto es especialmente así en los países menos adelantados, que ya carecen de mucho de lo que tienen los países más desarrollados, desde infraestructuras hasta políticas. Sin colaboración para crear capacidad y facilitar el acceso a los principales factores facilitadores, las divisiones existentes en materia de IA podrían ampliarse aún más y quedar enquistadas, lo que limitaría la capacidad de la IA para contribuir de forma significativa al progreso de la ciencia, los beneficios económicos y los avances en los ODS antes de 2030.

Figura 14: Expectativas de los expertos sobre el tiempo necesario para un impacto positivo importante de la IA en los distintos ámbitos



Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. El número específico de encuestados se indica en cada pregunta.

Fuente: Encuesta AI Risk Pulse Check de 21 de agosto de 2024 (n=121, de los cuales 78 respondieron a preguntas sobre países de ingreso mediano bajo y bajo).

^b El porcentaje de expertos que esperan un "impacto positivo importante" en el aumento de la actividad económica y la aceleración de los descubrimientos científicos en un plazo de tres años es mayor en el primer gráfico que en el segundo. Esto puede deberse al énfasis añadido en la pregunta "¿Para cuándo cree que es probable (50% de probabilidades o más) que la IA cause un impacto positivo importante?" del primer gráfico, que no aparece en el segundo.

Recuadro 11: La IA y los ODS (Continuado)

Figura 15: Expectativas de los expertos sobre los principales efectos positivos de la IA en los próximos tres años, por ámbito y ODS

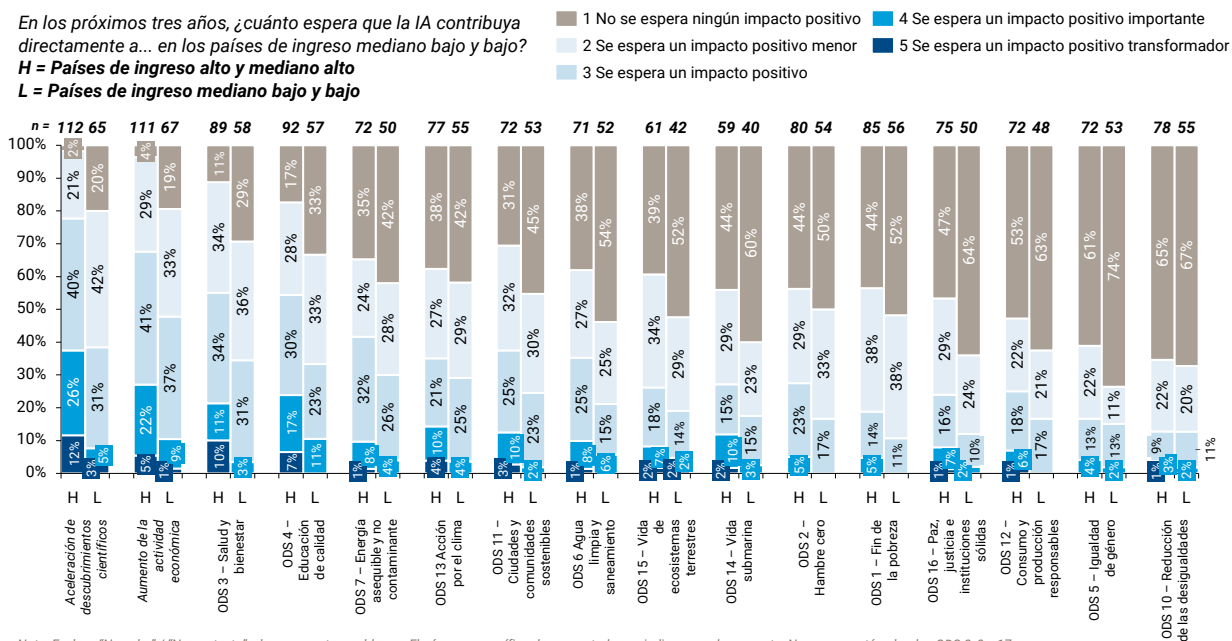
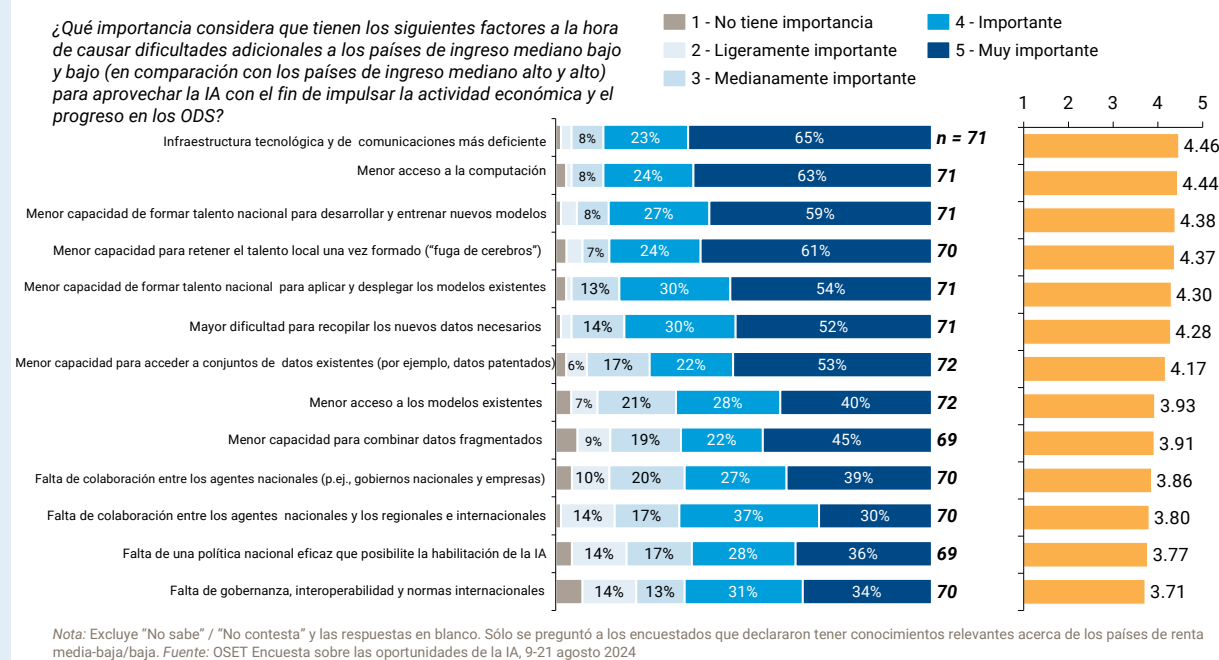


Figura 16: Clasificación de los expertos de los obstáculos que impiden aprovechar la IA para impulsar la actividad económica y el progreso de los ODS en los países de ingreso mediano bajo y bajo



125 Como afirmábamos en nuestro informe provisional, este objetivo depende en gran medida del acceso al talento, la capacidad de computación y los datos de forma que se favorezca el desarrollo de la diversidad cultural y lingüística. La propia gobernanza puede ser un facilitador clave para alinear incentivos, generar confianza y adoptar prácticas sostenibles, al tiempo que fomenta la colaboración a través de fronteras y ámbitos temáticos. Sin un enfoque integral e inclusivo de la gobernanza de la IA, podría perderse el potencial de esta para contribuir positivamente a los ODS, y su despliegue podría involuntariamente acentuar las desigualdades y los sesgos existentes.

126 Durante las amplias consultas llevadas a cabo por el Órgano Asesor sobre temas como la educación, la salud, los datos, el género, los niños, la paz y la seguridad, las industrias creativas y el trabajo, se hizo evidente que la IA tiene un potencial sustancial para acelerar significativamente el progreso de los ODS debido a su capacidad para impulsar la innovación y los resultados en diversas esferas críticas.

127 Sin embargo, la IA no es una panacea para los desafíos del desarrollo sostenible; es un componente dentro de un conjunto más amplio de soluciones, y puede incluso agravar algunos de estos desafíos, como el cambio climático. A fin de liberar realmente el potencial de la IA para hacer frente a los desafíos sociales, es esencial que exista colaboración entre los gobiernos, el mundo académico, la industria y la sociedad civil.

128 La eficacia de las soluciones de IA depende de la calidad y disponibilidad de los datos, y existe una gran preocupación por la calidad y representatividad de los conjuntos de datos pertinentes a los ODS, que pueden no reflejar la realidad de determinadas poblaciones. Además, las soluciones de IA diseñadas por expertos en IA sin pleno conocimiento de los ámbitos de

aplicación interseccionales suelen probarse o desarrollarse en un entorno digital o virtual y no son suficientemente sólidas o impactantes en entornos reales de desarrollo. Por eso, las soluciones de IA deben diseñarse en colaboración y aplicarse con un profundo conocimiento de sus contextos sociales, económicos y culturales. Deben encajar en estrategias locales y nacionales más amplias de transformación digital y de lucha contra las brechas digitales.

129 Por ejemplo, los países de ingreso bajo y mediano bajo no pueden lograr capacidad en materia de IA sin tener garantizadas una electricidad y una conectividad a Internet fiables para el funcionamiento de los centros de datos, el mantenimiento de operaciones informáticas uniformes, el acceso a conjuntos de datos mundiales, la participación en colaboraciones internacionales de investigación y el uso de herramientas de IA basadas en la nube. Es por ello que nos sumamos a los llamamientos para invertir en infraestructuras digitales básicas, que son un requisito previo para que los países en desarrollo puedan participar en los avances de la IA y beneficiarse de ellos.

130 La creación de capacidad en materia de IA es crucial para garantizar que las personas de todo el mundo, independientemente de la fase de desarrollo en que se encuentre su región, puedan beneficiarse de los avances de la IA. La creación de capacidad estratégica, respaldada con financiación suficiente, es también esencial para que las tecnologías de IA sean eficaces, sostenibles y de interés público, lo que es clave para las iniciativas mundiales de desarrollo. A continuación examinamos tres facilitadores esenciales de la capacidad nacional en materia de IA: la disponibilidad de conocimientos técnicos, el acceso a la computación y la disponibilidad de datos de calidad.

Talento

- 131** La capacidad de las sociedades de todo el mundo para participar en los beneficios de la IA depende, ante todo, de las personas. Es importante reconocer que no todas las sociedades necesitan equipos de expertos informáticos para construir sus propios modelos. Sin embargo, independientemente de que la tecnología se pueda adquirir, tomar prestada o fabricar, es necesario contar con recursos humanos para comprender las capacidades y limitaciones de la IA y aprovechar adecuadamente los casos de uso que esta posibilita.
- 132** Esta capacidad—principalmente en el sector público, pero también en el mundo académico, las empresas y la sociedad civil—mejorará la eficacia de las estrategias de IA y su implementación en los distintos sectores. Cultivar la capacidad humana relacionada con la IA también será fundamental para preservar la diversidad cultural y lingüística del mundo y crear conjuntos de datos de alta calidad para el futuro desarrollo de la IA. En esencia, se trata del desarrollo de la capacidad de la IA en pro del interés público.
- 133** Aumentar los recursos humanos en los diversos entornos con una demografía joven, como en África (una tercera parte de la fuerza de trabajo mundial será africana en la primera mitad de este siglo), también será crucial para la futura reserva mundial de talentos. La mejora de la capacidad de las mujeres en tecnología debe centrarse, por un lado, en cerrar la brecha de género existente y, por otro, en evitar la brecha de género en la IA. El sector de la IA también necesita más mujeres en puestos de liderazgo para integrar la perspectiva de género en la gobernanza de la IA. Para ello, hay que empezar por ofrecer a las niñas más oportunidades de talento en el ámbito de la IA.

Computación

- 134** A pesar de los esfuerzos que se están realizando para encontrar enfoques de IA que requieran menos recursos de computación, la necesidad de poder acceder a una computación asequible sigue siendo acuciante si queremos entrenar modelos de IA competentes³⁷. Esta es una de las mayores barreras de entrada al ámbito de la IA para las empresas del Sur Global, pero también para muchas nuevas empresas y pequeñas y medianas empresas del Norte Global. De los 100 principales clusters de computación de alto rendimiento del mundo capaces de entrenar grandes modelos de IA, ninguno se encuentra en un país en desarrollo³⁸. Solo hay un país africano entre los 300 primeros. Dos países concentran la mitad de los centros de datos a hiperescala del mundo³⁹.
- 135** La mayoría de los desarrolladores acceden a la infraestructura informática a través de servicios en la nube; muchos han optado por aliarse con las grandes empresas de la nube para asegurarse un acceso fiable a la computación. Es posible que los problemas de la cadena de suministro se resuelvan con el tiempo y que la competencia dé lugar a fuentes de hardware más diversas, como chips de alto rendimiento para modelos de entrenamiento y chips aceleradores de IA para su despliegue en dispositivos móviles. Sin embargo, en un futuro próximo esta limitación seguirá siendo una gran barrera a la hora de lograr un ecosistema de innovación de IA más inclusivo a nivel mundial.
- 136** Irónicamente, la capacidad de computación puede mantenerse inactiva o quedarse obsoleta rápidamente. La plena utilización de esta capacidad a lo largo de los ciclos de amortización tiene un valor potencial. Sin embargo, hay obstáculos que superar en cuanto a la interoperabilidad de las distintas configuraciones de hardware y la programación de tareas exigentes, preservando al mismo tiempo la prioridad de uso en tiempo crítico (como en el caso de las predicciones meteorológicas).

37 El Órgano Asesor tiene conocimiento de un caso reciente en el que una empresa con sede en el Sur Global gastó 70 millones de dólares en un curso de tres meses de capacitación en grandes modelos lingüísticos. Poseer las unidades de procesamiento gráfico (GPU) en lugar de alquilarlas a proveedores de servicios en la nube habría costado mucho menos.

38 Véase <https://top500.org/statistics/sublist>; indicador proxy, ya que la mayoría de los clusters de computación de alto rendimiento no disponen de GPU y su uso para la IA avanzada es limitado.

39 UNCTAD, *Informe sobre la Economía Digital 2021* (Ginebra, 2021).

137 Además, sin talento ni datos, la computación por sí sola carece de valor. En el propuesto fondo mundial para la IA, nos planteamos cómo abordar los tres aspectos mediante una combinación de apoyo financiero y en especie.

Datos

138 Aunque muchos debates sobre la economía de la IA se centran en la “guerra por el talento” y la competencia por equipos como las unidades de procesamiento gráfico (GPU), los datos no son menos importantes. Facilitar el acceso a datos de entrenamiento de calidad a escala para el entrenamiento de modelos de IA por parte de las nuevas empresas y las pequeñas y medianas empresas, así como contar con mecanismos que respeten los derechos para compensar a los titulares de los datos y a los creadores de datos de entrenamiento, podría ser el facilitador más importante de una floreciente economía de la IA. La puesta en común de datos en pro del interés público en la consecución de ODS específicos es un aspecto clave (descrito en el recuadro 12), pero no suficiente.

139 En el contexto de la IA, es habitual hablar de “uso indebido” de los datos (violaciones de la privacidad) o de “usos desaprovechados” de los datos (no explotar los conjuntos de datos existentes), pero un problema conexo es la “falta” de datos, que afecta a grandes partes del planeta que carecen de ellos. Un ejemplo es el ámbito de la atención de la salud, en el que aproximadamente la mitad de los principales conjuntos de datos corresponden a una decena de organizaciones, de las cuales una está en Europa, otra en Asia y el resto en América del Norte⁴⁰.

140 Otro ejemplo lo encontramos en la agricultura, donde, para que los modelos de IA resulten útiles, se necesitan datos de una compleja interacción de factores (como el clima, el suelo y las prácticas de gestión de los cultivos). La agricultura también suele adolecer de escasez de datos y de infraestructuras de recopilación de datos. Es preciso dedicar esfuerzos específicos a la conservación de conjuntos de datos agrícolas, especialmente en el contexto de la resiliencia al cambio climático de los sistemas alimentarios.

141 De forma análoga al problema del capital informal, aquellos cuyos datos no se recopilan - desde los registros de nacimiento hasta las transacciones financieras - pueden verse excluidos de la posibilidad de participar en los beneficios de la economía de la IA, obtener prestaciones del gobierno o acceder al crédito. El uso de datos sintéticos puede compensar solo parcialmente la necesidad de nuevos conjuntos de datos.

142 Los comentarios a nuestro informe provisional señalaban que no se había articulado suficientemente la forma en que las actuales prácticas interjurisdiccionales en torno a la obtención, el uso y la no divulgación de los datos de entrenamiento de IA amenazan los derechos y provocan concentración económica. Se nos recomendó estudiar cómo la gobernanza internacional de la IA podría hacer posible e impulsar una participación más diversa en el aprovechamiento de los datos para la IA.

40 Véase <https://2022.internethealthreport.org/facts>.

Recuadro 12: Puesta en común de datos para el interés público en las áreas de los ODS

Los datos colaborativos y los bienes comunes de la IA -en los que los modelos compartidos se entrenan de forma cruzada con datos agrupados- pueden desempeñar un papel clave en la promoción del interés público cuando, de otro modo, faltarían datos o serían demasiado escasos para obtener beneficios de la IA. Los fondos comunes de datos multifuncionales y multidominio podrían permitir el desarrollo de conjuntos de datos transdisciplinarios que abarquen varios dominios de los ODS, derivados de una variedad de fuentes.

A modo de ejemplo, podemos considerar la compleja cuestión de evaluar las repercusiones del cambio climático en la salud. Para abordar eficazmente este reto, es esencial un enfoque transdisciplinar que integre datos epidemiológicos sobre la prevalencia de enfermedades con datos meteorológicos que rastreen las variaciones climáticas. Al agrupar estos distintos tipos de datos de países de todo el mundo, preservando la privacidad, los investigadores pueden utilizar la IA para identificar patrones y correlaciones que no son evidentes a partir de conjuntos de datos aislados.

La inclusión de datos de todos los países garantiza una cobertura completa, que refleja la naturaleza global del cambio climático y capta diversos impactos ambientales y resultados sanitarios en las distintas regiones. El origen transdisciplinar de los datos aumenta la precisión predictiva de los modelos destinados a prever futuras crisis de salud pública o catástrofes naturales provocadas por el cambio climático.

Desarrollo de una capacidad pública internacional básica en materia de IA en beneficio de todos

143 En relación con los tres facilitadores señalados, las economías avanzadas tienen tanto el potencial como el deber de promover el desarrollo de la capacidad en materia de IA mediante la colaboración internacional. Por su parte, se beneficiarán de una economía digital de base más amplia, así como de flujos de talento y datos de calidad. Y lo que es más importante, todo el mundo se beneficiará de la generalización de la buena gobernanza de la IA a través de dicha colaboración.

144 La cooperación debe centrarse en fomentar el talento para la IA, impulsar la alfabetización de la sociedad en IA, mejorar la capacidad de gobernanza de la IA, aumentar el acceso a las

infraestructuras de IA, promover plataformas de datos y conocimientos adaptadas a las diversas necesidades culturales y regionales, y mejorar la adopción de aplicaciones y la capacidad de servicio de la IA. Solo un enfoque integral de este tipo puede garantizar un acceso equitativo a los beneficios de la IA, de forma que ninguna nación se quede atrás.

145 Muchas de las partes interesadas a las que consultamos hicieron hincapié en que deberían diseñarse estrategias concretas para aunar recursos mundiales con el fin de crear capacidad, impulsar la acción colectiva para una distribución equitativa de las oportunidades y cerrar la brecha digital.

Red de desarrollo de la capacidad

Recomendación 4: Red de desarrollo de la capacidad

Recomendamos la creación de una red de desarrollo de la capacidad en materia de inteligencia artificial que vincule a una serie de centros de desarrollo de la capacidad afiliados a las Naciones Unidas poniendo conocimientos especializados, computación y datos de entrenamiento de IA a disposición de los agentes clave. El objetivo de la red sería:

- a. Acelerar y alinear los esfuerzos regionales y mundiales de capacitación en materia de inteligencia artificial apoyando la creación de redes entre ellos;
- b. Crear capacidad de gobernanza de la IA entre los funcionarios públicos para garantizar el respeto, la protección y el cumplimiento de los derechos humanos y promover el desarrollo;
- c. Poner instructores, computación y datos de entrenamiento de IA de los múltiples centros a disposición de los investigadores y emprendedores sociales que busquen aplicar la IA a casos de interés público local, entre otras cosas, a través de:
 - i. Protocolos que permitan a los equipos de investigación interdisciplinarios y a los empresarios de entornos con escasez de recursos informáticos acceder a la computación disponible para entrenar, perfeccionar y aplicar sus modelos de forma adecuada al contexto local;
 - ii. Entornos controlados para poner a prueba posibles soluciones de IA y aprender sobre la marcha;

- iii. Un conjunto de oportunidades educativas sobre la IA en línea dirigidas a estudiantes universitarios, jóvenes investigadores, emprendedores sociales y funcionarios del sector público; y
- iv. Un programa de becas para que personas con talento pasen un tiempo en instituciones académicas o empresas tecnológicas.

146 Desde los Objetivos de Desarrollo del Milenio hasta los ODS, las Naciones Unidas llevan mucho tiempo contribuyendo al desarrollo de la capacidad de las personas y las instituciones⁴¹. Gracias a la labor de la UNESCO, la OMPI y otras organizaciones, las Naciones Unidas han contribuido a mantener la rica diversidad de culturas y tradiciones de creación de conocimientos en todo el mundo.

147 Al mismo tiempo, el desarrollo de la capacidad en materia de IA requeriría un nuevo enfoque, en particular una capacitación interdisciplinaria para crear una nueva generación de expertos multidisciplinares en ámbitos como la salud pública y la IA, o los sistemas alimentarios y energéticos y la IA.

148 La capacidad también tendría que vincularse a los resultados a través de la capacitación práctica⁴² en entornos de pruebas y proyectos de colaboración que pongan en común datos y computación para resolver problemas compartidos. En el diseño de esta infraestructura de capacitación colaborativa deberían estar incorporadas las evaluaciones de riesgos, las pruebas de seguridad y otras metodologías de gobernanza.

149 Dadas la urgencia y la escala del desafío, proponemos seguir un enfoque estratégico que centralice y articule el acceso a la computación a

⁴¹ La Universidad de las Naciones Unidas lleva mucho tiempo comprometida con la creación de capacidad a través de la educación superior y la investigación, y el Instituto de las Naciones Unidas para Formación Profesional e Investigaciones ha contribuido a capacitar a funcionarios en ámbitos fundamentales para el desarrollo sostenible. La metodología de la UNESCO de evaluación del estado de preparación para la IA es una herramienta clave para apoyar a los Estados Miembros en su aplicación de la Recomendación sobre la Ética de la Inteligencia Artificial. Otros ejemplos son la Academia de la OMS en Lyon, el Instituto Virtual de Comercio y Desarrollo de la UNCTAD, el Programa de las Naciones Unidas de Becas sobre Desarme dirigido por la Oficina de Asuntos de Desarme y los programas de desarrollo de la capacidad dirigidos por la UIT y el PNUD.

⁴² Diversas instituciones nacionales, entre ellas autoridades financieras y médicas, como la Infocomm Media Development Authority de Singapur, han desarrollado entornos de pruebas.

través de una red de nodos de computación de alto rendimiento, incentive el desarrollo de conjuntos de datos críticos en los ámbitos relevantes para los ODS, promueva el intercambio de modelos de IA, incorpore las mejores prácticas en materia de gobernanza de la IA y genere talento transversal para la IA de interés público, garantizando así la integración transversal de los conocimientos especializados en materia de derechos humanos.

150 En otras palabras, en lugar de ir persiguiendo uno por uno los facilitadores críticos mediante través de proyectos inconexos, proponemos una **estrategia integral aplicada de una sola vez a través de una cadena de centros colaboradores**. Las iniciativas emergentes sobre desarrollo de la capacidad y la IA para los ODS, como la Red Internacional de Computación e Inteligencia Artificial (ICAIN) creada por Suiza, pueden ayudar a establecer la masa crítica inicial para esta estrategia.

151 Lo ideal sería que hubiera al menos uno o dos nodos en cada región del mundo. Los dos centros de competencias que participan en la Alianza Mundial sobre la Inteligencia Artificial podrían unirse a las Naciones Unidas en el apoyo a la red. Las instituciones académicas y los contribuyentes del sector privado al desarrollo de la capacidad podrían solicitar su afiliación a través del nodo regional más cercano o de una organización internacional que apoye la red.

152 Nos anima especialmente la perspectiva de la cooperación entre países, por ejemplo, a través del acceso federado a infraestructuras de computación y afines. Como señalábamos en nuestro informe provisional, la Organización Europea de Investigación Nuclear (CERN) ofrece lecciones de utilidad. Una “CERN distribuida” reimaginada para la IA, conectada en red a través de diversos Estados y regiones, podría ampliar las oportunidades para un mayor acceso a las herramientas y los conocimientos especializados de la IA.

153 Concebimos la red de desarrollo de la capacidad como catalizador de las capacidades nacionales y regionales y no como concentrador de hardware,

talento y datos. La aceleración del aprendizaje podría impulsar a los centros nacionales de excelencia a estimular el desarrollo de ecosistemas locales de innovación en IA, abordando las lagunas subyacentes de coordinación e implementación mencionadas en los párrafos 73, 80 y 81. Los esfuerzos a nivel nacional podrían seguir empleando herramientas de diagnóstico como la metodología de la UNESCO de evaluación del estadio de preparación para la IA que ayuden a evaluar la madurez inicial de los países, identificar las lagunas y orientar cómo pueden adaptarse las hojas de ruta para el desarrollo de la capacidad por país y región, con la red internacional ayudando a colmar estas lagunas.

154 La oficina de IA propuesta puede ser la mejor situada para centrarse en la estrategia, las alianzas y la afiliación orientadas a vincular los nodos con la red, sirviendo para conectar más que para reinventar. También podría ayudar a negociar el acceso a la computación en toda la red. Uno o varios nodos de la red podrían liderar aspectos concretos de la capacitación, y albergar entornos de pruebas o clusters de computación de alto rendimiento para el desarrollo de modelos de inteligencia artificial. Los nodos podrían colaborar en programas de investigación sobre temas como el uso de datos preservando la privacidad, nuevos métodos para vincular diferentes tipos de hardware o conjuntos de datos para el entrenamiento de modelos, así como formas de utilizar modelos de IA combinados entre sí.

155 Tenemos la esperanza de que la red también promueva un *paradigma alternativo de desarrollo de tecnología de IA*: ascendente, interdisciplinar, interregional, abierto y colaborativo. Teniendo en cuenta el aumento de los costos energéticos y de otro tipo del entrenamiento y despliegue de los modelos de IA, y la perspectiva de que la computación pueda no utilizarse, tiene sentido vincular los recursos computacionales para el acceso en base al tiempo compartido y al tiempo que se aprovecha dicho acceso para promover el talento, los datos y los modelos de IA en pro de la consecución de los ODS.

Fondo mundial para la IA

Recomendación 5: Fondo mundial para la IA

Recomendamos la creación de un fondo mundial para la IA con el fin de reducir la brecha existente en este ámbito. Gestionado por una estructura de gobierno independiente, el fondo recibiría contribuciones financieras y en especie de fuentes públicas y privadas y las desembolsaría, también a través de la red de desarrollo de la capacidad, para mejorar el acceso a facilitadores de la IA orientados a impulsar el empoderamiento local en pro de los ODS como:

- a. Recursos informáticos compartidos para la elaboración y el perfeccionamiento de modelos por parte de desarrolladores de IA de países que carecen de la capacidad local adecuada o de los medios para adquirirla;
- b. Entornos y herramientas de análisis comparativo y de pruebas para incorporar las mejores prácticas en la elaboración de modelos seguros y fiables y en la gobernanza de los datos;
- c. Soluciones de gobernanza, seguridad e interoperabilidad de aplicación mundial, basadas en el derecho internacional, incluido el derecho internacional de los derechos humanos;
- d. Conjuntos de datos e investigaciones sobre cómo combinar datos y modelos para proyectos relacionados con los ODS; y
- e. Un repositorio de modelos de IA y conjuntos de datos seleccionados para los ODS.

156 El modelo de desarrollo y uso de la IA que aquí se propone es análogo a la visión original de Internet: una infraestructura distribuida pero conectada, interoperable y empoderadora. El interés público estaría mejor servido por un mercado en el que los modelos de IA y la infraestructura y los datos de los que dependen sean interoperables, estén bien gobernados y sean fiables. Pero esto no se conseguiría automáticamente. Sería indispensable dedicar esfuerzos específicos respaldados por recursos suficientes.

157 Abordamos esta recomendación con humildad, conscientes de las poderosas fuerzas del mercado que determinan el acceso al talento y a la computación, y de la competencia geopolítica que se opone a la colaboración en el ámbito de la ciencia y la tecnología. Lamentablemente, muchos países no podrán acceder a la capacitación, la computación, los modelos y los datos de entrenamiento sin apoyo internacional. Los actuales esfuerzos de financiación tampoco pueden ampliarse sin este apoyo.

158 La igualdad de condiciones es, en parte, una cuestión de equidad. También redundaría en nuestro interés colectivo crear un mundo en el que todos contribuyan a un ecosistema compartido y se beneficien de él. Y no solo entre Estados. Garantizar la diversidad en el acceso al desarrollo de modelos de IA y a la infraestructura de pruebas también ayudaría a abordar las preocupaciones sobre la concentración desproporcionada de poder en manos de unas pocas empresas tecnológicas.

Propósito y objetivo del fondo

159 Nuestra intención al proponer un fondo no es garantizar el acceso a recursos y capacidades informáticas que incluso los países y empresas más ricos luchan por adquirir. Puede que la respuesta no sea siempre más computación. Puede que también necesitemos formas diferentes de aprovechar las infraestructuras de computación de alto rendimiento existentes, que están construidas para picos de uso y no necesariamente diseñadas para la IA. Tal vez podría haber mejores formas de conectar talento, computación y datos.

160 El propósito del fondo es por tanto subsanar las deficiencias subyacentes de coordinación e implementación señaladas en los párrafos 73, 80 y 81 en favor de aquellos que no pueden acceder a los facilitadores necesarios por otras vías, para garantizar que:

- a. Los países necesitados puedan acceder a los facilitadores de la IA reduciendo con ello la brecha de la IA;
- b. La colaboración en el desarrollo de la capacidad en materia de IA genere hábitos de cooperación y reduzca la competencia geopolítica;

- c. Los países con enfoques reguladores divergentes tengan incentivos para desarrollar plantillas comunes de gestión de los datos, los modelos y las aplicaciones en favor de los retos sociales relacionados con los ODS y los avances científicos.

- 161** La capacidad desarrollada con los recursos del fondo mundial se orientaría hacia los ODS y la gobernanza global compartida de la IA (recuadro 13). Por ejemplo, esta podría incorporar un “acervo de gobernanza” para las pruebas de seguridad. Esto ayudaría a generalizar las mejores prácticas en toda la base de usuarios, al tiempo que reduciría la carga de validación para los pequeños usuarios.
- 162** Este enfoque de interés público hace que el fondo mundial complemente la propuesta de una red de desarrollo de la capacidad en materia de inteligencia

artificial, a la que el fondo canalizaría recursos. El fondo proporcionaría además una capacidad independiente de seguimiento del impacto. De este modo nos aseguraríamos de que amplias zonas del mundo no se queden atrás, sino que puedan aprovechar la IA para los ODS en diferentes contextos.

- 163** Redunda en interés de todos garantizar la cooperación en el mundo digital, al igual que en el mundo físico. Se pueden establecer analogías con los esfuerzos para combatir el cambio climático, donde los costos de transición, mitigación o adaptación no se distribuyen uniformemente y la asistencia internacional es esencial para ayudar a los países con recursos limitados a que puedan unirse al esfuerzo mundial de hacer frente a un problema planetario.

Recuadro 13: Fondo mundial para la IA – Ejemplos de posibles inversiones

Un fondo relativamente modesto podría ayudar a crear una infraestructura informática mínima compartida para el entrenamiento de modelos pequeños y medianos. Estos modelos tienen un gran potencial para los ODS, por ejemplo, para la capacitación de los agricultores en su idioma local.

Esta inversión también crearía un entorno de pruebas para que los desarrolladores pudieran perfeccionar los modelos de código abierto existentes con sus propios datos contextuales y de alta calidad. El acceso a la infraestructura informática y al entorno de pruebas podría hacerse en régimen de tiempo compartido, con cuotas de uso razonables que contribuyeran a sufragar los costos de mantenimiento y funcionamiento.

Un tercer uso de la financiación sería ayudar a conservar conjuntos de datos de referencia para determinados ODS en los que no exista un incentivo comercial. El desarrollo de modelos, las pruebas y los esfuerzos de conservación de datos podrían unirse estratégicamente en un potente enfoque práctico de empoderamiento de la IA vinculado a resultados concretos.

Por último, el fondo podría estimular la investigación y el desarrollo no solo de aplicaciones de la IA relacionadas con el desarrollo y los ODS, sino también de la interconexión de la informática y los modelos y de nuevas evaluaciones de gobernanza.

164 En este caso, la atención se centra en el uso de la financiación para garantizar que, en países de diferentes regiones, se pueda crear una capacidad mínima para comprender el potencial que la IA tiene para el desarrollo sostenible, adaptar y construir modelos para las necesidades locales y unirse a los esfuerzos de colaboración internacional en materia de IA.

La gobernanza del fondo

165 El fondo obtendría y agruparía las contribuciones en especie, incluidas las procedentes de entidades del sector privado. La coordinación de las contribuciones financieras y en especie requiere niveles adecuados de supervisión independiente y rendición de cuentas. Los mecanismos de gobernanza deberían ser inclusivos y contar con miembros de los gobiernos, el sector privado, filántropos, la sociedad civil y los organismos de las Naciones Unidas. Deberían recabar aportaciones científicas y de expertos, canalizadas (por ejemplo) a través del grupo internacional de expertos científicos ya propuesto, y generar neutralidad y confianza para la colaboración en torno al desarrollo de datos y modelos.

Operaciones del fondo

166 El modelo operativo del fondo debería basarse en las lecciones aprendidas de las colaboraciones internacionales para la investigación y el desarrollo, como la CERN y Gavi, la Alianza para las Vacunas, así como en las enseñanzas extraídas de las plataformas comerciales para infraestructuras de tiempo compartido. También debería aprender de organismos como el Fondo Mundial (creado en 2002 para recabar recursos con el fin de acabar con el VIH, la tuberculosis y la malaria)⁴³ y el Fondo para el Análisis de Riesgos Complejos (que agrupa datos para apoyar a todas las partes interesadas en sus actividades de anticipación, prevención y respuesta a las crisis).

Marco mundial de datos sobre IA

Recomendación 6: Marco mundial de datos sobre IA

Recomendamos la creación de un marco mundial de datos sobre IA, desarrollado mediante un proceso iniciado por un organismo competente como la Comisión de las Naciones Unidas para el Derecho Mercantil Internacional y fundamentado en el trabajo de otras organizaciones internacionales, para:

- a. Establecer definiciones y principios relacionados con los datos para la gobernanza global de los datos de entrenamiento de IA, incluidos los extraídos de las mejores prácticas existentes, y promover la diversidad cultural y lingüística;
- b. Establecer normas comunes sobre la procedencia y el uso de los datos de entrenamiento de IA para una rendición de cuentas transparente y basada en los derechos en todas las jurisdicciones; y
- c. Implantar mecanismos de administración e intercambio de datos de configuración del mercado que permitan el desarrollo de ecosistemas locales de IA a nivel mundial, como por ejemplo:
 - i. Fideicomisos de datos;
 - ii. Mercados mundiales bien regulados de intercambio de datos anónimos para el entrenamiento de modelos de IA; y
 - iii. Acuerdos modelo para facilitar el acceso internacional a los datos y la interoperabilidad mundial, posiblemente como protocolos técnico-jurídicos del marco.

167 En nuestras consultas se nos dijo que, aunque ha habido muchas propuestas para promover un acceso más amplio a los datos y a los acuerdos de intercambio de datos para crear ecosistemas de IA más diversos, hasta ahora pocas se han materializado. Se trata de una laguna crítica en el desarrollo de ecosistemas de IA inclusivos y dinámicos.

43 Véase <https://www.theglobalfund.org/en/about-the-global-fund>.

- 168** Parte de la respuesta se encuentra en la transparencia sobre los rasgos culturales, lingüísticos y de otro tipo de los datos de entrenamiento de IA. También resulta útil identificar los datos infrarrepresentados o “ausentes”. En relación con este aspecto está la promoción de “conjuntos de datos comunes” que incentiven la conservación de los datos de entrenamiento para múltiples agentes. Tales iniciativas podrían establecer las mejores prácticas demostrando cómo el diseño puede incorporar marcos técnico-jurídicos para la privacidad, la protección de datos, la interoperabilidad y el uso equitativo de los datos, así como los derechos humanos.
- 169** Los mercados de datos para la IA son hoy una especie de “salvaje oeste”. La idea de “coge lo que puedas y escóndelo en algoritmos opacos” parece ser un principio operativo; otro son los arreglos contractuales de exclusividad para el acceso a datos patentados, aplicables en determinadas jurisdicciones. Tales relaciones de exclusividad explican la preocupación de la Autoridad de los Mercados y la Competencia del Reino Unido por el hecho de que “el sector [de los modelos fronterizos] se esté desarrollando de un modo que entraña el riesgo de producir resultados negativos en el mercado”⁴⁴.
- 170** Por tanto, consideramos esencial iniciar un proceso mundial en el que participen una gran variedad de agentes, entre ellos países con distintos niveles de desarrollo, que cuente con el respaldo de las organizaciones internacionales pertinentes del sistema de las Naciones Unidas y de otros organismos (OCDE, OMPI y la Organización Mundial del Comercio), con el fin de crear mecanismos de protección y vías comunes para el desarrollo de ecosistemas de datos de entrenamiento de IA. Los resultados de este proceso no tienen por qué ser leyes vinculantes, sino modelos de contratos y acuerdos técnico-jurídicos. Estos acuerdos facilitadores pueden elaborarse uno a uno, como protocolos para un marco de principios y definiciones.
- 171** Aunque la información pormenorizada está fuera de nuestro alcance, los principios clave para un marco mundial de datos sobre IA incluirían: interoperabilidad, administración, preservación de la privacidad, empoderamiento, mejora de los derechos y habilitación de ecosistemas de IA.
- 172** Somos conscientes de que las políticas antimonopolio y de la competencia siguen estando en manos de las autoridades nacionales y regionales. Sin embargo, la acción colectiva internacional puede facilitar el acceso transfronterizo a los datos de entrenamiento, no disponibles a nivel nacional, para las empresas emergentes locales de IA.
- 173** Las Naciones Unidas están en una posición única para apoyar el establecimiento de principios globales y acuerdos prácticos para la gobernanza y el uso de los datos de entrenamiento de IA, basándose en años de trabajo de la comunidad de datos e integrándolos con los recientes avances en materia de ética y gobernanza de la IA. Esto es análogo a los esfuerzos de la Comisión de las Naciones Unidas para el Derecho Mercantil Internacional en materia de comercio internacional, que incluyen la elaboración de marcos transfronterizos jurídicos y no jurídicos y la facilitación del comercio y la inversión digitales a través de leyes modelo sobre comercio electrónico, computación en la nube y gestión de la identidad.
- 174** Del mismo modo, la Comisión de Ciencia y Tecnología para el Desarrollo y la Comisión de Estadística de las Naciones Unidas tienen en su agenda datos para el desarrollo y datos sobre los ODS. La OMPI también está estudiando importantes cuestiones relacionadas con los contenidos, los derechos de autor y la protección de los conocimientos y la expresión cultural autóctonos.

- 175** El marco aquí propuesto se entendería sin perjuicio de los marcos nacionales o regionales de protección de datos existentes y no crearía nuevos derechos relacionados con los datos ni prescribiría cómo se aplican internacionalmente los actuales derechos, sino que tendría que diseñarse de forma que se impidiera su apropiación por intereses comerciales o de otro tipo que pudieran mermar o excluir la protección de los derechos. El marco mundial de datos sobre IA propuesto abordaría cuestiones transversales de disponibilidad, interoperabilidad y uso de los datos de entrenamiento de IA. Ayudaría a establecer un entendimiento común sobre cómo armonizar los distintos marcos nacionales y regionales de protección de datos.
- 176** Las medidas para abordar estos problemas a escala nacional y regional son prometedoras, y los sectores público y privado prestan ahora más atención a las mejores prácticas. Sin embargo, sin un marco mundial que regule los conjuntos de datos de entrenamiento de IA, la competencia comercial invita a una convergencia a la baja entre jurisdicciones en cuanto a los requisitos de acceso y uso, lo que dificulta la gobernanza internacional de la cadena de valor de la IA. Solamente la acción colectiva mundial puede promover una “competencia al alza” en la gobernanza de la recopilación, creación, uso y monetización de los datos de entrenamiento de IA de manera que se favorezca la interoperabilidad, la administración, la preservación de la privacidad, el empoderamiento y la mejora de los derechos.
- 177** Del mismo modo, estas medidas son necesarias para promover el desarrollo de ecosistemas locales de inteligencia artificial y contener la concentración económica. Estas medidas podrían complementarse con la promoción de conjuntos de datos comunes y disposiciones para albergar fideicomisos de datos en ámbitos relevantes para los ODS (véase el recuadro 14). La red de desarrollo de la capacidad y el fondo mundial para la IA podrían prestar apoyo para la elaboración de estas plantillas y para el almacenamiento y análisis de los datos almacenados en conjuntos de datos comunes o en fideicomisos de datos.

Recuadro 14: Protección de los datos para el entrenamiento de modelos de IA: empoderamiento de datos, fideicomisos de datos y acuerdos de flujo de datos transfronterizos

Hay muchas circunstancias en las que es necesario proteger los datos (por ejemplo, por motivos de privacidad, confidencialidad comercial, propiedad intelectual y seguridad), pero en las que también sería beneficioso, para las personas y la sociedad, ponerlos a disposición para el entrenamiento de modelos de IA.

Los derechos legales sobre los datos son, por lo general, derechos encaminados a impedir acciones en relación con los datos. Los derechos de privacidad de los datos también tienen un carácter personal. El establecimiento de los derechos sobre los datos puede dificultar el ejercicio de estos derechos de una forma flexible que permita utilizar los datos para algunos fines sin perder los derechos, y hacerlo de forma colectiva como grupo. Incluso cuando es posible controlar los permisos de forma flexible y positiva, esto suele requerir más tiempo, conocimientos técnicos y confianza de los que disponen la mayoría de las personas y organizaciones.

Los mecanismos que permiten a los propietarios y titulares de los datos un uso seguro y limitado de los mismos manteniendo al mismo tiempo sus derechos pueden describirse como medios de empoderamiento de datos. El empoderamiento de datos puede convertir a muchas más personas y grupos de la sociedad en asociados activos y partes interesadas en la IA, y no solo en titulares de los datos. Ya se están desarrollando herramientas para gestionar el acceso de forma segura, como los fideicomisos de datos y las aplicaciones de protección de la privacidad para dirigir los flujos de datos transfronterizos.

Los fideicomisos de datos son mecanismos que permiten a particulares y organizaciones facilitar el acceso a sus datos de forma generalizada bajo control de fideicomisarios. Los propietarios de los datos pueden establecer las condiciones de acceso, uso y finalidad que los fideicomisarios obligan a cumplir. Los propietarios y los titulares de los datos conservan sus derechos legales al tiempo que contribuyen a objetivos compartidos. Un modelo de inteligencia artificial entrenado con estos datos podría tener un rendimiento más preciso que otro que careciera de esta información específica, pudiendo por tanto servir mejor al bienestar de ese grupo concreto o de la sociedad en general.

Los mecanismos para gestionar el acceso y uso, y el acceso transfronterizo en particular, se basan en marcos jurídicos específicos. El uso de estos mecanismos en la práctica también requiere una adaptación a las necesidades y los contextos de los sectores y las comunidades. Es preciso identificar y subsanar las deficiencias en la gestión de los datos. El éxito y la generalización de estos mecanismos dependerán de la seguridad técnica y del mantenimiento de la confianza de quienes aportan los datos.

Proponemos por tanto que se preste más apoyo al desarrollo de estas herramientas y a la identificación de los ámbitos en los que su uso para el entrenamiento de la IA podría aportar el mayor valor para los ciudadanos.

D. Esfuerzos coherentes

- 178** Al promover un entendimiento común, un terreno común y unos beneficios comunes, las propuestas anteriores pretenden colmar las lagunas identificadas en el incipiente régimen internacional de gobernanza de la IA. Las lagunas de representación, coordinación e implementación pueden afrontarse mediante alianzas y colaboración con las instituciones y los mecanismos existentes.
- 179** Sin embargo, sin un punto focal específico en las Naciones Unidas que apoye y permita la coordinación entre estos y otros esfuerzos y que garantice que el sistema de las Naciones Unidas habla con una sola voz sobre la IA, el mundo carecerá del enfoque inclusivo en red, ágil y coherente necesario para una gobernanza eficaz y equitativa de la IA.
- 180** Por estas razones, proponemos la creación de una capacidad que sea pequeña y ágil, en forma de oficina de IA dentro de la Secretaría de las Naciones Unidas.

Una oficina de IA en la Secretaría General de las Naciones Unidas

Recomendación 7: Una oficina de IA en la Secretaría General

Recomendamos la creación de una oficina de IA dentro de la Secretaría General, dependiente del Secretario General. Su organización debería ser sencilla y ágil, aprovechando en la medida de lo posible las entidades competentes de las Naciones Unidas ya existentes. Actuando como elemento “aglutinante” para apoyar e impulsar las propuestas de este informe, estableciendo alianzas e interactuando con otros procesos e instituciones, el mandato de la oficina incluiría:

- a. Brindar apoyo al grupo internacional de científicos expertos en IA, al diálogo sobre políticas, al intercambio de normas, a la red de

desarrollo de la capacidad y, en la medida en que sea necesario, al fondo mundial y al marco mundial de datos sobre IA;

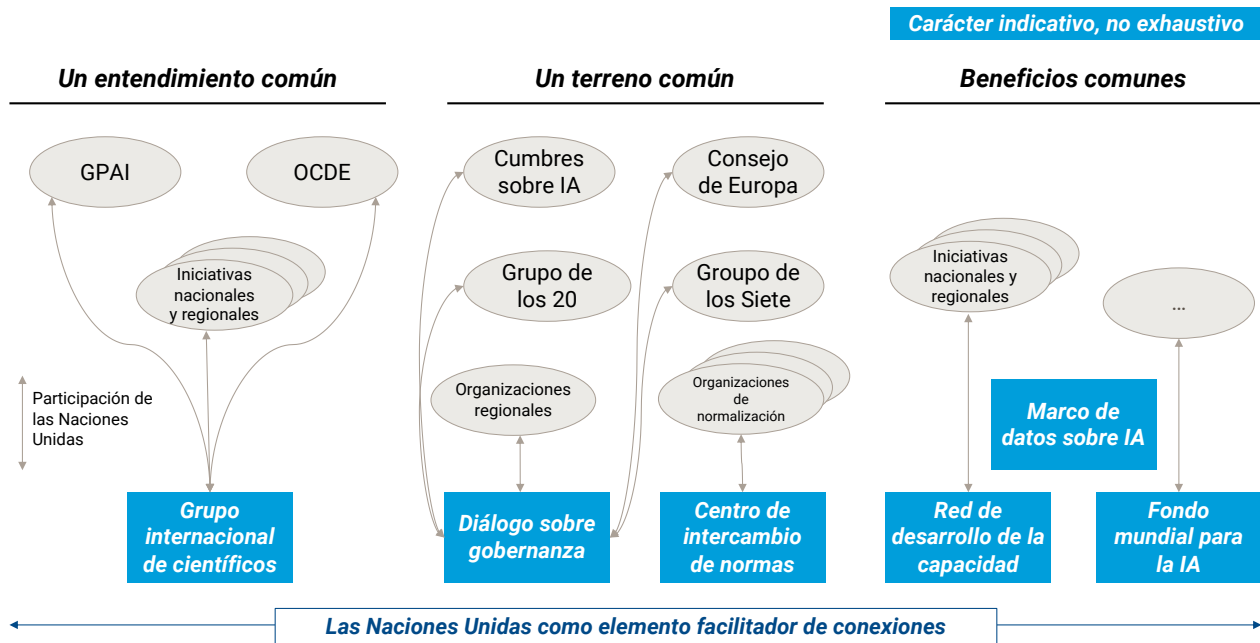
- b. Participar en actividades de divulgación dirigidas a múltiples partes interesadas, incluidas las empresas tecnológicas, la sociedad civil y el mundo académico, sobre cuestiones emergentes relacionadas con la IA;
- c. Asesorar al Secretario General sobre cuestiones relacionadas con la IA, en coordinación con otras partes pertinentes del sistema de las Naciones Unidas para ofrecer una respuesta “de todo el sistema”.

- 181** Durante nuestras consultas quedó claro que aún no se ha demostrado la conveniencia de crear un organismo con competencias de presentación de informes, seguimiento, verificación y exigencia de cumplimiento, y que los Estados Miembros tampoco están dispuestos a crear una nueva y costosa organización.
- 182** Por tanto, nos centramos en el valor que pueden ofrecer las Naciones Unidas, conscientes de las deficiencias del sistema de Naciones Unidas, así como en lo que podría lograrse de forma realista en el plazo de un año. En este sentido, proponemos un mecanismo ligero y ágil que actúe como “aglutinante” para mantener unidos los procesos de entendimiento común, terreno común y beneficios comunes, y que permita al sistema de las Naciones Unidas hablar con una sola voz en el ecosistema internacional de gobernanza de la IA, que está en plena evolución.
- 183** Del mismo modo que los países han creado institutos y oficinas dedicados a la gobernanza nacional, regional e internacional de la IA⁴⁵, vemos la necesidad de crear una capacidad que preste servicios y apoyo al grupo internacional de científicos expertos en IA y al diálogo sobre políticas de IA, y que impulse el intercambio de normas sobre la IA y la red de desarrollo de la capacidad, con menos gastos generales y costos de transacción que si cada uno de ellos recibiera el apoyo de organizaciones diferentes.

45 Así lo hacen en el Canadá, los Estados Unidos, el Japón, la República de Corea, el Reino Unido, Singapur y la Unión Europea.

- 184** Una oficina de IA dentro de la Secretaría General de las Naciones Unidas, dependiente del Secretario General, tendría la ventaja de estar conectada con todo el sistema de las Naciones Unidas sin estar vinculada a una parte del mismo. Esto es importante debido al futuro incierto de la IA y a la gran probabilidad de que impregne todos los aspectos del quehacer humano.
- 185** Una oficina de inteligencia artificial pequeña y ágil estaría bien posicionada para conectar los distintos ámbitos y organizaciones en cuestiones de gobernanza de la IA y ayudar a subsanar las deficiencias de forma dinámica, trabajando para ampliar los esfuerzos existentes dentro y fuera de las Naciones Unidas. Al tender puentes y conectar otras iniciativas, como las lideradas por organizaciones regionales y otras iniciativas plurilaterales, puede ayudar a reducir los costos de la cooperación entre ellas.
- 186** Este órgano debería defender la inclusión y colaborar activamente para acelerar la coordinación y la implementación, aprovechando con carácter prioritario los recursos y las funciones existentes en el sistema de las Naciones Unidas. El órgano podría estar formado, en parte, por personal de las Naciones Unidas cedido por los organismos especializados y otras partes del sistema de las Naciones Unidas. Debería contar con la participación de las múltiples partes interesadas, incluidas la sociedad civil, la industria y el mundo académico, y entablar alianzas con las principales organizaciones ajenas a las Naciones Unidas, como la OCDE.
- 187** La oficina de IA garantizaría el intercambio de información en todo el sistema de las Naciones Unidas y posibilitaría que este hablara con autoridad y con una sola voz. El recuadro 15 enumera las posibles funciones y los primeros resultados de dicha oficina.
- 188** Esta recomendación se basa en una cuidadosa evaluación de los ámbitos en los que las Naciones Unidas pueden aportar valor añadido, incluidos aquellos en los que pueden liderar, colmar lagunas y ayudar a coordinar y aquellos en los que deben hacerse a un lado, trabajando en estrecha colaboración con los esfuerzos existentes (véase la Figura 17). La ONU también aporta las ventajas de los acuerdos institucionales existentes, incluidos los prenegociados sobre la financiación y los procedimientos administrativos bien aceptados.
- 189** Es preciso tener en cuenta la evolución de las características de la tecnología de la IA. Existe una alta probabilidad de que se produzcan avances técnicos que cambien drásticamente el panorama actual de los modelos de IA. Una oficina de IA de este tipo debería estar preparada para ajustar los marcos de gobernanza a la evolución del panorama y responder a acontecimientos imprevistos relacionados con la tecnología de la inteligencia artificial.

Figura 17: Función propuesta para las Naciones Unidas en el ecosistema internacional de gobernanza de la IA



Abreviaciones: GPAI: Alianza Mundial sobre la Inteligencia Artificial; OCDE: Organización de Cooperación y Desarrollo Económicos; SDO: organizaciones de normalización.

Recuadro 15: Posibles funciones y resultados a un año de la oficina de IA

La oficina de IA debería tener una estructura ligera y aspirar a ser ágil, fiable y estar conectada en red. En caso necesario, debería operar con un sistema radial para conectarse con otras partes del sistema de las Naciones Unidas y otras instancias.

Sus actividades de divulgación podrían incluir la función de nodo central en la denominada arquitectura de coordinación ágil entre los Estados Miembros, las redes plurilaterales, las organizaciones de la sociedad civil, el mundo académico y las empresas tecnológicas en un complejo régimen que se entrelaza para resolver problemas de forma colaborativa a través de la creación de redes, y como lugar seguro y de confianza para tratar temas relevantes. De forma ambiciosa, podría convertirse en el elemento que aglutinara y ayudara a mantener unidas otras redes en desarrollo.

El apoyo a las diversas iniciativas propuestas en este informe incluye la importante función de garantizar la rápida inclusividad en la obtención de productos como informes científicos, diálogos sobre gobernanza e identificación de entidades de seguimiento apropiadas.

Recuadro 15: Posibles funciones y resultados a un año de la oficina de IA (Continuación)

Entendimiento común:

- Facilitación de la contratación del grupo internacional de expertos científicos y prestación de apoyo al grupo.

Terreno común:

- Diálogos sobre políticas de servicios con aportaciones de múltiples partes interesadas en apoyo de la interoperabilidad y el aprendizaje en materia de políticas. Un primer tema prioritario es la articulación de los umbrales de riesgo y los marcos de seguridad entre jurisdicciones.
- Apoyo a la UIT, la ISO/CEI y el IEEE en el establecimiento de una entidad de intercambio de normas sobre la IA.

Beneficios comunes:

- Apoyo a la red de desarrollo de la capacidad en materia de IA, centrándose inicialmente en la creación de capacidad en inteligencia artificial de interés público entre los funcionarios públicos y los empresarios sociales. Definición de la visión inicial de la red, los resultados, la estructura de gobernanza, las alianzas y los mecanismos operacionales.
- Definición de la visión, los resultados, la estructura de gobernanza y los mecanismos operacionales del fondo mundial para la IA, y petición de opiniones sobre la propuesta a los Estados Miembros, la industria y las partes interesadas de la sociedad civil con miras a financiar los proyectos iniciales en un plazo de seis meses a partir de su creación.
- Elaboración y publicación de una lista anual de áreas de inversión prioritarias para orientar tanto el fondo mundial para la IA como las inversiones ajenas a esta estructura.

Esfuerzos coherentes:

- Establecimiento de mecanismos sencillos que ayuden a los Estados Miembros y a otras organizaciones pertinentes a estar más conectados y coordinados y a ser más eficaces a la hora de realizar sus esfuerzos de gobernanza global de la IA.
- Preparación de un marco inicial que sirva para orientar y supervisar el trabajo de la oficina de IA y que incluya una taxonomía de riesgos para la gobernanza global, una revisión del panorama político mundial de la IA y un mapa mundial de las partes interesadas.
- Elaboración de informes trimestrales y realización de presentaciones periódicas presenciales a los Estados Miembros sobre los progresos de la oficina de IA con respecto a su plan de trabajo, y establecimiento de canales de información que permitan realizar los ajustes necesarios.
- Establecimiento de un comité directivo dirigido conjuntamente por la oficina de IA, la UIT, la UNCTAD, la UNESCO y otras entidades y organizaciones pertinentes de las Naciones Unidas para acelerar el trabajo de las Naciones Unidas al servicio de las funciones mencionadas, y examen trimestral de los progresos de estos esfuerzos acelerados.
- Promoción de oportunidades conjuntas de aprendizaje y desarrollo para los representantes de los Estados Miembros con el fin de ayudarles a desempeñar sus responsabilidades en la gobernanza global de la IA, en cooperación con las entidades y organizaciones pertinentes de las Naciones Unidas, como el Instituto de las Naciones Unidas para Formación Profesional e Investigaciones y la Universidad de las Naciones Unidas.

E. Reflexiones sobre los modelos institucionales

190 Los debates sobre la IA a menudo se dividen en extremos. En nuestras consultas por todo el mundo, hemos hablado con quienes ven un futuro de oportunidades ilimitadas proporcionadas por sistemas de IA cada vez más baratos y útiles. También hablamos con quienes temen un futuro más oscuro, la división y el desempleo, e incluso la extinción.

191 No sabemos qué nos deparará el futuro. Somos conscientes de que la tecnología puede ir en una dirección que acabe con esta dualidad. En este informe nos hemos centrado en las oportunidades y los riesgos a corto plazo, basándonos en datos científicos. Las recomendaciones expuestas en este informe representan nuestra mayor esperanza de beneficiarnos de la IA al tiempo que minimizamos y mitigamos sus riesgos. También somos conscientes de los desafíos prácticos que plantea la creación de instituciones internacionales a gran escala. Por eso proponemos un enfoque institucional en red, con un apoyo sencillo y ágil.

192 Sin embargo, si los riesgos se agravan y las oportunidades aumentan, estos cálculos cambiarán. Las guerras mundiales condujeron al moderno sistema internacional; el desarrollo de armas cada vez más potentes desembocó en regímenes que limitan su propagación y promueven el uso pacífico de las tecnologías subyacentes.

193 La evolución en la comprensión de nuestra humanidad común condujo al sistema moderno de derechos humanos y a nuestros compromisos actuales con los ODS para todos. El cambio climático pasó de ser un problema aislado a convertirse en un desafío mundial. La IA puede evolucionar de forma similar hasta un nivel que requiera más recursos y más autoridad de la que se propone en este informe.

194 Nuestro mandato incluía plantearnos las funciones, la forma y los plazos de un nuevo organismo internacional para la IA. Concluimos el presente informe con algunas reflexiones sobre la cuestión, aunque en este momento no recomendamos la creación de dicho organismo.

¿Un organismo internacional de inteligencia artificial?

195 Si los riesgos de la IA se agravan y se concentran más, podría ser necesario que los Estados Miembros consideraran la posibilidad de crear una institución internacional más sólida con competencias de supervisión, presentación de informes, verificación y exigencia de cumplimiento.

196 Existen precedentes de este tipo de evolución. Desde las Convenciones de La Haya de 1899 y 1907 hasta el Protocolo de Ginebra de 1925 y culminando con la Convención sobre las Armas Químicas de 1993, las sustancias químicas de doble uso han estado sujetas durante mucho tiempo a restricciones de acceso, protocolos de almacenamiento y uso, y a la prohibición de su utilización como arma.

197 También se han prohibido las armas biológicas y se han puesto límites a la investigación, como los límites relacionados con el ADN artificial o el empalme de genes en 1975. En estos límites se hacía hincapié en la contención como elemento esencial del diseño de los experimentos, y se vinculaba el nivel de contención al riesgo estimado. Determinados tipos de experimentos de alto riesgo en los que no podía garantizarse la contención quedaron básicamente prohibidos. También se ponen límites a las investigaciones que amenazan con traspasar líneas éticas fundamentales, como las actuales restricciones a la clonación humana, un ejemplo del tipo de “línea roja” que algún día puede ser necesaria en el contexto de la investigación sobre la IA, junto con una cooperación eficaz en lo que respecta a la exigencia de su cumplimiento.

- 198** Las evaluaciones científicas continuas también son una característica de algunos de estos marcos, por ejemplo, del Consejo Consultivo Científico de la Organización para la Prohibición de las Armas Químicas y del artículo XII de la Convención sobre las Armas Biológicas.
- 199** La comparación entre la IA y la energía nuclear es bien conocida. Desde que se dividió el átomo, los científicos tuvieron claro que esta tecnología podía utilizarse para hacer el bien, aunque sus investigaciones se orientaran a la construcción de una nueva y terrible arma. Entonces, como ahora, era revelador que los que pedían con más insistencia la limitación de esta nueva tecnología contaran entre sus filas con destacados científicos.
- 200** El gran acuerdo en el que se basó el Organismo Internacional de Energía Atómica (OIEA) consistió en compartir los fines beneficiosos de la energía nuclear - en la producción de energía, la agricultura y la medicina - a cambio de garantías de no volver a utilizarla jamás como arma. Como demuestra el régimen de no proliferación de las armas nucleares, unas buenas normas son necesarias pero no suficientes para una regulación eficaz.
- 201** Los límites de la analogía están claros. La energía nuclear implica un conjunto bien definido de procesos relacionados con materiales específicos que están desigualmente distribuidos, y gran parte de los materiales y la infraestructura necesarios para crear capacidad nuclear están controlados por los Estados nacionales; la IA es un término difuso; sus aplicaciones son extremadamente amplias; sus capacidades más poderosas abarcan los sectores productivos y los Estados. El gran acuerdo del OIEA se centró en armas costosas de construir y difíciles de ocultar; la utilización de la IA como arma promete no ser ni lo uno ni lo otro.
- 202** Una primera idea - la gestión mancomunada del combustible nuclear con fines pacíficos - no funcionó como se esperaba. En cuanto a esto último, una analogía más apropiada para la IA podría ser el CERN, que agrupa financiación, talento e infraestructuras. Sin embargo, la comparación tiene sus límites, dada la diferencia entre la física fundamental experimental y la IA, que requiere un enfoque más distribuido.
- 203** Otra analogía imperfecta son organizaciones como la Organización de Aviación Civil Internacional (OACI) y la Organización Marítima Internacional (OMI). Las tecnologías subyacentes al transporte están bien establecidas, y sus aplicaciones civiles pueden diferenciarse fácilmente de las militares, lo que no ocurre con la IA de uso general. La red de organismos nacionales de reglamentación que aplican las normas internacionales elaboradas en el marco de la OACI y la OMI también está consolidada. Esta red se centra en la seguridad, la facilitación de la actividad comercial y la interoperabilidad. El cumplimiento no se gestiona de manera descendente.
- 204** Hay otros enfoques del cumplimiento que pueden servir de inspiración. La gestión del riesgo financiero se beneficia de mecanismos como el Consejo de Estabilidad Financiera (CEF) y el Grupo de Acción Financiera (GAFI), sin recurrir a los tratados.
- 205** Con el tiempo, podría resultar esencial contar con algún tipo de mecanismo a escala mundial para formalizar las líneas rojas si queremos que la regulación de la IA sea aplicable. Un mecanismo de este tipo podría incluir compromisos formales similares a los de la CERN para concentrar los recursos destinados a la colaboración en la investigación de la IA y compartir los beneficios como parte del acuerdo.
- 206** Sin embargo, dada la velocidad, la autonomía y la opacidad de los sistemas de IA, esperar a que surja una amenaza puede significar que cualquier respuesta llegue demasiado tarde. La continuación de las evaluaciones científicas y del diálogo sobre políticas garantizaría que el mundo no se viera sorprendido. Cualquier decisión de iniciar un proceso formal correspondería, naturalmente, a los Estados Miembros.

207 Entre los posibles umbrales para tal decisión podría incluirse la perspectiva de que se desarrollen sistemas de IA incontrolables o incontenibles, o el despliegue de sistemas que no puedan ser rastreados hasta agentes humanos, corporativos o estatales. También podrían incluirse los indicios de que los sistemas de IA presentan cualidades que sugieren la aparición de una “superinteligencia”, aunque esto no ocurre en los sistemas de IA actuales.

208 Una primera medida razonable sería establecer un grupo de observación, compuesto por expertos diversos y destacados, para vigilar la situación. El grupo de expertos científicos podría encargarse de realizar investigaciones sobre esta cuestión, como parte de su serie de compendios trimestrales de investigación temática. Con el tiempo, el diálogo sobre políticas podría ser un foro adecuado para compartir información sobre incidentes de IA, como los que amplían o superan las capacidades de los organismos existentes, de forma análoga a las prácticas del OIEA sobre garantías mutuas en materia de seguridad nuclear, o de la Organización Mundial de la Salud (OMS) en materia de control de enfermedades.

209 Las funciones del organismo de IA que se propone podrían basarse en la experiencia de otros organismos pertinentes, como el OIEA, la Organización para la Prohibición de las Armas Químicas, la OACI, la OMI, la CERN y la Convención sobre las Armas Biológicas. Entre sus funciones se podrían incluir:

- El desarrollo y la promulgación de estándares y normas para la seguridad de la IA;
- La supervisión de los sistemas de IA que puedan amenazar la paz y la seguridad internacionales, o violar gravemente los derechos humanos o el derecho internacional humanitario;
- La recepción e investigación de informes sobre incidentes o usos indebidos, y la elaboración de informes sobre infracciones graves;
- La verificación del cumplimiento de las obligaciones internacionales;
- La coordinación de la rendición de cuentas, las respuestas de emergencia y la reparación de daños en relación con incidentes de seguridad de la IA;
- La promoción de la cooperación internacional para los usos pacíficos de la IA.

210 Para diseñar un futuro organismo dedicado a la IA, sería necesario adoptar un enfoque específico basado en las enseñanzas extraídas de otras instituciones según proceda (recuadro 16).

Recuadro 16: Lecciones aprendidas de anteriores instituciones de gobernanza global

La IA es un conjunto único de tecnologías con riesgos e impactos sociales que trascienden las fronteras. Sin embargo, no es el primer conjunto de tecnologías que ha dado lugar a acuerdos globales de gobernanza. La aviación civil, el cambio climático, la energía nuclear y la financiación del terrorismo son también ámbitos complejos y multidimensionales que han justificado una respuesta mundial.

Algunos de estos ámbitos, como la aviación civil, el cambio climático y la energía nuclear, han dado lugar a la creación de nuevas instituciones de las Naciones Unidas. Otros, como la protección de los flujos financieros mundiales, han dado lugar a órganos que no se basan en tratados y que, sin embargo, han proporcionado marcos normativos sólidos, mecanismos de aplicación eficaces basados en el mercado y fuertes alianzas público-privadas.

Al establecer paralelismos entre estas respuestas institucionales y los esfuerzos incipientes por hacer lo mismo con la IA, no deberíamos centrarnos demasiado en qué analogía institucional es la más adecuada para el conjunto de problemas de la IA. Nuestro informe provisional ya anticipó que deberíamos analizar qué funciones de gobernanza son necesarias para una gobernanza global de la IA eficaz e inclusiva, y qué podemos aprender de los esfuerzos de gobernanza global del pasado.

Una de las lecciones aprendidas es que es necesario desarrollar un entendimiento científico y técnico compartido del problema para que la respuesta política esté comúnmente aceptada. En este sentido, el IPCC, que sigue ocupándose de los riesgos del cambio climático, es un modelo útil. Ofrece un ejemplo de cómo un enfoque inclusivo para elaborar informes y desarrollar el consenso científico en un ámbito en constante evolución puede establecer condiciones equitativas para los investigadores y los responsables políticos y crear el entendimiento compartido esencial para una formulación de políticas eficaz. El proceso de elaboración y difusión de los informes del IPCC y de los balances mundiales, aunque no ha estado exento de dificultades, ha sido de vital importancia para construir un entendimiento compartido y una base común de conocimientos, reducir los costos de la cooperación y dirigir la Conferencia de las Partes en la Convención Marco de las Naciones Unidas sobre el Cambio Climático hacia resultados políticos concretos.

En el caso de la IA, a medida que evolucione la tecnología, será igual de importante desarrollar un entendimiento científico compartido. Como la capacidad de los sistemas de IA sigue aumentando y los riesgos potenciales pueden superar los enfoques eficaces conocidos para mitigarlos, el grupo internacional de expertos científicos podría evolucionar para adaptarse a las nuevas necesidades.

Una segunda lección es que la colaboración entre múltiples partes interesadas puede ofrecer normas sólidas y promover respuestas rápidas. En este sentido, la OACI y el GAFI ofrecen ejemplos útiles de cómo gobernar una cuestión altamente técnica a través de las fronteras. En la aviación civil, las normas de seguridad de la OACI, elaboradas por expertos del sector y de los gobiernos y aplicadas mediante restricciones de acceso al mercado, garantizan que un avión que despegue por ejemplo de Nueva York pueda aterrizar en Ginebra sin tener que someterse a nuevas auditorías de seguridad. Una combinación de auditorías de seguridad dirigidas por la OACI y auditorías impulsadas por los Estados Miembros garantiza una aplicación coherente, incluso a medida que evoluciona la tecnología.

El GAFI - creado por el en 1989 para hacer frente al blanqueo de dinero - ofrece otro ejemplo de cómo las instituciones de derecho indicativo pueden promover normas comunes y su aplicación. Su sistema de revisión por pares para la supervisión es flexible; y la aceptación generalizada de sus recomendaciones ha generado un costo de reputación para las empresas y los Estados Miembros que no las cumplen. Incluso a medida que han ido evolucionando los riesgos para los flujos financieros internacionales, sobre todo con el auge del terrorismo y la financiación de la proliferación, la ágil estructura y el marco normativo del GAFI le han permitido responder con rapidez y seguir el ritmo de los complejos desafíos.

Recuadro 16: Lecciones aprendidas de anteriores instituciones de gobernanza global (Continuación)

A su manera, tanto la OACI como el GAFI han creado normas internacionales ampliamente reconocidas, marcos nacionales para medir el cumplimiento y sistemas interoperables para responder a determinadas clases de riesgos y desafíos que se manifiestan en todas las jurisdicciones. La OACI hace cumplir las normas mediante incentivos y restricciones de acceso al mercado, mientras que el GAFI crea un riesgo para la reputación en caso de incumplimiento. Ambos ofrecen plantillas útiles para la IA, ya que demuestran cómo los gobiernos y otras partes interesadas pueden trabajar juntos para crear una red de normas y reglamentos interconectados y establecer penalizaciones en forma de costos en caso de incumplimiento.

Una tercera lección es que la coordinación mundial resulta esencial para supervisar y tomar medidas en respuesta a riesgos graves que tienen un potencial de impacto generalizado. Los modelos del CEF y del OIEA ofrecen excelentes ejemplos. El CEF fue creado en 2009 por el G20 para vigilar y alertar sobre los riesgos sistémicos para el sistema financiero internacional. Su composición única de funcionarios de finanzas del G20 y organizaciones financieras y de desarrollo internacionales le ha permitido ser ágil, hábil e inclusivo a la hora de coordinar los esfuerzos para identificar los riesgos financieros mundiales.

El enfoque del OIEA en materia de salvaguardias nucleares ofrece un modelo diferente. Sus acuerdos de salvaguardias amplias, firmados por 182 Estados, forman parte del régimen más amplio de las Naciones Unidas para garantizar el cumplimiento. Mediante una combinación de inspecciones y vigilancia - y la amenaza de una acción del Consejo de Seguridad de la ONU -, el OIEA ejerce quizás la censura más visible sobre los Estados Miembros que no cumplen las normas.

Tanto el CEF como el OIEA demuestran que la coordinación internacional es fundamental para vigilar los riesgos graves. A medida que los riesgos de la IA se hacen más evidentes y pronunciados, puede surgir una necesidad similar de crear una nueva institución centrada en la IA para maximizar los esfuerzos de coordinación y supervisar los riesgos graves y sistémicos, de modo que los Estados Miembros puedan intervenir siempre que sea posible para adelantarse a esos riesgos.

Una cuarta lección es que es importante crear un acceso inclusivo a los recursos necesarios para la investigación y el desarrollo, y a sus beneficios. Las experiencias de la CERN y del OIEA son aleccionadoras. La CERN reúne a académicos y físicos de talla mundial para llevar a cabo complejas investigaciones sobre aceleradores de partículas y otros proyectos orientados a beneficiar a la humanidad. También imparte capacitación a físicos e ingenieros.

De forma similar, el OIEA facilita el acceso a la tecnología, en este caso a la energía nuclear y las radiaciones ionizantes. El equilibrio es sencillo: Los Estados Miembros cumplen las salvaguardias nucleares y el OIEA ofrece asistencia técnica para el uso de la energía nuclear con fines pacíficos. En este sentido, el OIEA proporciona un enfoque inclusivo para hacer llegar los beneficios de la tecnología a los países en desarrollo. Su facilitación de una red de centros de excelencia sobre seguridad nuclear se asemeja a nuestra recomendación de un enfoque en red para la creación de capacidad.

Como hemos explicado, la IA es un conjunto de tecnologías cuyos beneficios deben compartirse de forma más inclusiva y equitativa, especialmente con los países del Sur Global. Y por ello hemos recomendado tanto una red de desarrollo de la capacidad en materia de IA como un fondo mundial para la IA. A medida que aprendemos más sobre la IA a través del trabajo del grupo internacional de científicos expertos en IA, y conforme el despliegue responsable de la IA en apoyo de los ODS se hace aún más apremiante, puede que los Estados Miembros de las Naciones Unidas quieran institucionalizar esta función de forma más amplia. Si lo hacen, deberían aprovechar las lecciones de la CERN y del OIEA como modelos útiles para apoyar un acceso más amplio a los recursos como parte de una estructura global de gobernanza de la IA.

5. Conclusión: Un llamamiento a la acción

- 211** Como expertos, seguimos siendo optimistas sobre el futuro de la IA y su potencial para hacer el bien. Ese optimismo depende, sin embargo, del realismo sobre los riesgos y la insuficiencia de estructuras e incentivos actualmente existentes. También debemos ser realistas sobre los recelos internacionales que podrían obstaculizar la acción colectiva mundial necesaria para una gobernanza eficaz y equitativa. La tecnología es demasiado importante y hay demasiado en juego como para confiar únicamente en las fuerzas del mercado y en un mosaico fragmentado de acciones nacionales y multilaterales.
- 212** Tenemos que ser proactivos y decididos. Más allá de la dualidad de oportunidades y riesgos está el desafío del cambio rápido y transversal. El impacto descendente de la IA puede dejar a pocas personas intactas. Poner su gobernanza en manos de unos pocos desarrolladores, o de los países que los acogen, creará una situación profundamente injusta en la que los impactos del desarrollo, despliegue y uso de la IA se impondrán a la mayoría de las personas sin que hayan tenido voz ni voto en las decisiones adoptadas para hacerlo.
- 213** El último año de interés y debate mundiales sobre la gobernanza de la IA nos ha dado esperanzas. Hay discrepancias entre países y sectores, pero también un fuerte deseo de diálogo. La participación de numerosos expertos, responsables de formular políticas, empresarios, investigadores y defensores de la IA de distintas regiones, géneros y disciplinas nos ha demostrado que la diversidad no tiene por qué conducir a la discordia, y que el diálogo puede llevar a un terreno común y a la colaboración.
- 214** En cierto modo, dudamos: ¿debemos ser pragmáticos y centrarnos en lo que parece factible? ¿O debemos apuntar alto con grandes ambiciones? Al final, decidimos hacer ambas cosas. Nuestras propuestas reflejan una visión amplia de un régimen de gobernanza global de la IA equitativo y eficaz, con una atenta reflexión sobre cómo puede aplicarse paso a paso.
- 215** Expresamos nuestro agradecimiento a las numerosas personas, organizaciones y Estados Miembros que han contribuido a nuestras deliberaciones, así como a los representantes de los organismos de las Naciones Unidas y al personal de la Secretaría, que evaluaron objetivamente la capacidad y las limitaciones de las Naciones Unidas en este complejo ámbito. La cuestión de la gobernanza de la IA no consiste únicamente en gestionar las implicaciones de esta tecnología. También está en juego el futuro de la cooperación multilateral y entre las múltiples partes interesadas.
- 216** Cuando miremos atrás dentro de cinco años, el panorama tecnológico podría ser radicalmente distinto del actual. Sin embargo, si mantenemos el rumbo y superamos las vacilaciones y las dudas, dentro de cinco años podremos contemplar un panorama de gobernanza de la IA inclusivo y empoderador para las personas, las comunidades y los países de todo el mundo. En última instancia, lo que importa no es el cambio tecnológico en sí, sino cómo responde a él la humanidad.

- 217** Creemos que las funciones y los mecanismos recomendados en este informe, si se implementan de buena fe, pueden ofrecer un régimen ágil y adaptable que siga el ritmo de la IA y ayude a cosechar sus beneficios y a hacer frente a sus riesgos. Estas funciones pueden ayudarnos a detectar a tiempo los problemas y las oportunidades, a utilizar principios y marcos compartidos para alinear la acción internacional, a promover la cooperación internacional y a desarrollar la capacidad de las personas y las instituciones para afrontar el cambio.
- 218** La materialización de las recomendaciones del presente informe también puede incentivar nuevas formas de pensar: una mentalidad de colaboración y aprendizaje, la participación de múltiples partes interesadas y un amplio compromiso público. La ONU puede ser el vehículo de un nuevo contrato social para la IA que garantice la aceptación mundial de un régimen de gobernanza que nos proteja y nos empodere a todos. Este contrato garantizará que se pueda acceder a las oportunidades y que estas se distribuyan equitativamente, y que los riesgos no recaigan sobre los más vulnerables ni se transmitan a las generaciones futuras, como hemos visto trágicamente con el cambio climático.
- 219** Como grupo y como personas procedentes de muchos campos de especialización, organizaciones y partes del mundo, esperamos continuar este esencial debate. Junto con las muchas personas con las que nos hemos relacionado en este viaje y la comunidad global a la que representan, esperamos que este informe contribuya a nuestros esfuerzos combinados por gobernar la IA en beneficio de la humanidad.

Anexos

Anexo A: Miembros del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial

Anna Abramova

Omar Sultan Al Olama

Latifa Al-Abdulkarim

Estela Aranha

Carme Artigas (Copresidenta)

Ran Balicer

Paolo Benanti

Abeba Birhane

Ian Bremmer (CopONENTE)

Anna Christmann

Natasha Crampton

Nighat Dad

Vilas Dhar

Virginia Dignum

Arisa Ema

Mohamed Farahat

Amandeep Singh Gill

Wendy Hall

Rahaf Harfoush

Ruimin He

Hiroaki Kitano

Haksoo Ko

Andreas Krause

James Manyika (Copresidente)

María Vanina Martínez Posse

Seydina Moussa Ndiaye

Mira Murati

Petri Myllymäki

Alondra Nelson

Nazneen Rajani

Craig Ramlal

Emma Ruttkamp-Bloem

Marietje Schaake (CopONENTE)

Sharad Sharma

Jaan Tallin

Philip Thigo

Jimena Sofía Viveros Álvarez

Zeng Yi

Zhang Linghan

Anexo B: Mandato del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial

El Órgano Asesor de Alto Nivel sobre Inteligencia Artificial, convocado por el Secretario General, realizará análisis y formulará recomendaciones sobre la gobernanza internacional de la inteligencia artificial (IA). Los informes iniciales del Órgano Asesor ofrecerán aportaciones independientes y de expertos de alto nivel para los debates nacionales, regionales y multilaterales en curso.

El Órgano Asesor estará compuesto por 38 miembros procedentes de los gobiernos, el sector privado, la sociedad civil y el mundo académico, así como por un Secretario. Su composición será equilibrada en cuanto a género, edad, representación geográfica y ámbito de especialización relacionado con los riesgos y las aplicaciones de la IA. Los miembros del Órgano Asesor ejercerán sus funciones a título personal.

El Órgano Asesor trabajará y consultará ampliamente con los gobiernos, el sector privado, el mundo académico, la sociedad civil y las organizaciones internacionales. Será ágil e innovador a la hora de interactuar con los procesos y plataformas existentes, así como de aprovechar las aportaciones de las diversas partes interesadas. Podría crear grupos de trabajo o grupos sobre temas específicos.

Los miembros serán seleccionados por el Secretario General a partir de las designaciones de los Estados Miembros y de una convocatoria pública de candidaturas. Tendrá dos copresidentes y un comité ejecutivo. Todos los grupos de partes interesadas estarán representados en el Comité Ejecutivo.

El Órgano Asesor se reunirá durante un período inicial de un año, con posibilidad de prórroga por parte del Secretario General. Celebrará reuniones presenciales y virtuales.

El Órgano Asesor presentará un primer informe antes del 31 de diciembre de 2023 para someterlo a la consideración del Secretario General y de los Estados Miembros de las Naciones Unidas. Este informe presentará un análisis de alto nivel de las opciones para la gobernanza internacional de la IA.

En base a las reacciones al primer informe, el Órgano Asesor presentará un segundo informe antes del 31 de agosto de 2024, en el que podrá formular recomendaciones detalladas sobre las funciones, la forma y los plazos de un nuevo organismo internacional de gobernanza de la IA.

El Órgano Asesor debería evitar la duplicación con los foros y procesos que actualmente se ocupan de la IA. En su lugar, debería tratar de aprovechar las plataformas y los asociados existentes, incluidas las entidades de las Naciones Unidas, que trabajan en ámbitos relacionados. Debería respetar plenamente la actual estructura de las Naciones Unidas, así como las prerrogativas nacionales, regionales y sectoriales en la gobernanza de la IA.

Las deliberaciones del Órgano Asesor contarán con el apoyo de una pequeña secretaría con sede en la Oficina del Enviado del Secretario General para la Tecnología y se financiarán con recursos extrapresupuestarios procedentes de donantes.

Anexo C: Lista de participación en las consultas, 2024

Participación	Fecha 2024	Región
UNESCO Eslovenia	5 de enero	Europa
Junta de Asesoramiento Científico del Secretario General	8 de enero	Mundial
Presentación a los Estados Miembros del informe provisional	12 de enero	Mundial
Foro Económico Mundial de Davos	24 de enero	Europa
Reunión virtual de Altos Funcionarios de la Asociación de Naciones de Asia Sudoriental (ASEAN)	30 de enero	Asia
Cumbre Mundial de Gobierno	12 de febrero	Oriente Medio
Instituto de Montreal para Algoritmos de Aprendizaje (Mila - Quebec AI Institute)	14 de febrero	América del Norte
Consultas en Berlín	15 de febrero	Europa
Foro Euroasiático sobre la Tecnología de la Información	20 de febrero	Mundial
Congreso Mundial de Móviles	26 de febrero	Europa
Instituto Estatal de Relaciones Internacionales de Moscú	28 de febrero	Europa
Taller de la Royal Society sobre gobernanza internacional de la IA	28 de febrero	Europa
Red de Asesores Científicos y Tecnológicos en los Ministerios de Relaciones Exteriores	28 de febrero	Mundial
Diálogo OCDE-Unión Africana sobre IA	4 de marzo	Europa
Consultas en Bruselas	5 de marzo	Europa
Banco Mundial - Cumbre Digital Mundial	5 de marzo	América del Norte
Ciencia Abierta e Inteligencia Artificial: Seminario web sobre cuestiones éticas	5 de marzo	Europa Oriental
Diálogo de la UNESCO sobre la Transformación Digital	6 de marzo	Europa
Unión Parlamentaria Internacional	6 de marzo	Mundial
47º período de sesiones del Comité de Alto Nivel sobre Programas de la ONU	11 de marzo	Mundial
Cumbre Mundial de la Juventud sobre Derechos Digitales	13 de marzo	América Latina
Cumbre del Grupo de los Siete (G7) sobre IA en Trento (Italia)	15 de marzo	Europa
Reuniones iniciales de la red de consulta los días 18 y 19 de marzo	18 de marzo	Mundial
68º período de sesiones de la Comisión sobre la Condición Jurídica y Social de la Mujer	21 de marzo	América del Norte
Información actualizada a los Estados Miembros del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial	25 de marzo	Mundial
Observatorio Africano de Inteligencia Artificial Responsable	25 de marzo	África
Conferencia sobre la IA para un futuro sostenible e inclusivo - Agencia Francesa de Desarrollo (AFD)	26 de marzo	Europa
Elaboración de normas mundiales: reacción colectiva	28 de marzo	África
Innovate Switzerland	2 de abril	Europa
Visita de la Oficina del Enviado del Secretario General para la Tecnología a China - 9 a 12 de abril	9 de abril	Asia
Foro Ruso para la Gobernanza de Internet	9 de abril	Europa Oriental
Wharton Cypher Days - Finanzas	12 de abril	América del Norte
Visita a Silicon Valley	15 de abril	América del Norte
Stanford - Simposio sobre políticas de IA: un balance mundial	16 de abril	América del Norte
Comisión de Ciencia y Tecnología para el Desarrollo	16 de abril	Europa
Grupo de Trabajo sobre Economía Digital del G20 - 16 a 18 de abril - Brasil	17 de abril	América Latina
Información actualizada a los Estados Miembros del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial	22 de abril	Mundial
Conferencia sobre Inteligencia Artificial de la Universidad de las Naciones Unidas en Macao, 24 y 25 de abril	24 de abril	Asia
Visita de la Oficina del Enviado del Secretario General para la Tecnología a Bruselas y París, 25 y 26 de abril	26 de abril	Europa
Presentación del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial ante el Comité Asesor Nacional de IA (Estados Unidos)	2 de mayo	América del Norte
Cumbre Mundial de Riad sobre la IA (GAIN) con la Organización del Mundo Islámico para la Educación, la Ciencia y la Cultura (53 países, cuatro regiones)	14 de mayo	Oriente Medio
La IA al servicio del desarrollo sostenible: contribución de Kazajistán a la Agenda 2030	20 de mayo	Asia
Grupo de los Estados de América Latina y el Caribe	21 de mayo	América Latina
Foro Académico del grupo BRICS (Brasil, Rusia, India, China y Sudáfrica)	22 de mayo	Mundial
Sesión de Seúl sobre la gobernanza de la IA	23 de mayo	Asia
Cumbre Asia Tech x Singapore - 29 a 31 de mayo	29 de mayo	Asia
Cumbre mundial AI for Good, 29 a 31 de mayo	29 de mayo	Europa

Anexo D: Lista de “inmersiones en profundidad”

Esfera temática	Fecha (hora de verano del este)
Educación	29 de marzo
Propiedad intelectual y contenidos	2 de abril
Infancia	4 de abril
Paz y seguridad (1)	12 de abril
Paz y seguridad (2)	29 de abril
Agricultura (sesión 1)	30 de abril
Agricultura (sesión 2)	30 de abril
Religión	1 de mayo
Código abierto y dirección técnica	1 de mayo
Impacto en la sociedad	3 de mayo
Género	7 de mayo
Datos	13 de mayo
El futuro del trabajo	13 de mayo
Normas (sesión 1)	14 de mayo
Normas (sesión 2)	14 de mayo
Paz y seguridad (3)	20 de mayo
Medio ambiente	20 de mayo
Salud	22 de mayo
Estado de derecho, derechos humanos, democracia	24 de mayo

Anexo E: Respuestas a la encuesta global AI Risk Pulse Check

A petición del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial, la Oficina del Enviado del Secretario General para la Tecnología (OSET) llevó a cabo una encuesta global sobre los riesgos de la IA denominada AI risk pulse check como parte de un ejercicio de análisis prospectivo de la IA, con el fin de recoger las percepciones de expertos de todo el mundo sobre los riesgos de la inteligencia artificial. Se pidió a los expertos que respondieran con sus opiniones a título personal (no en representación de su institución o empleador). Se pidió a los expertos que calificaran el nivel que esperaban de aceleración o desaceleración del cambio tecnológico de la IA y (por separado) de la adopción y aplicación de la IA.

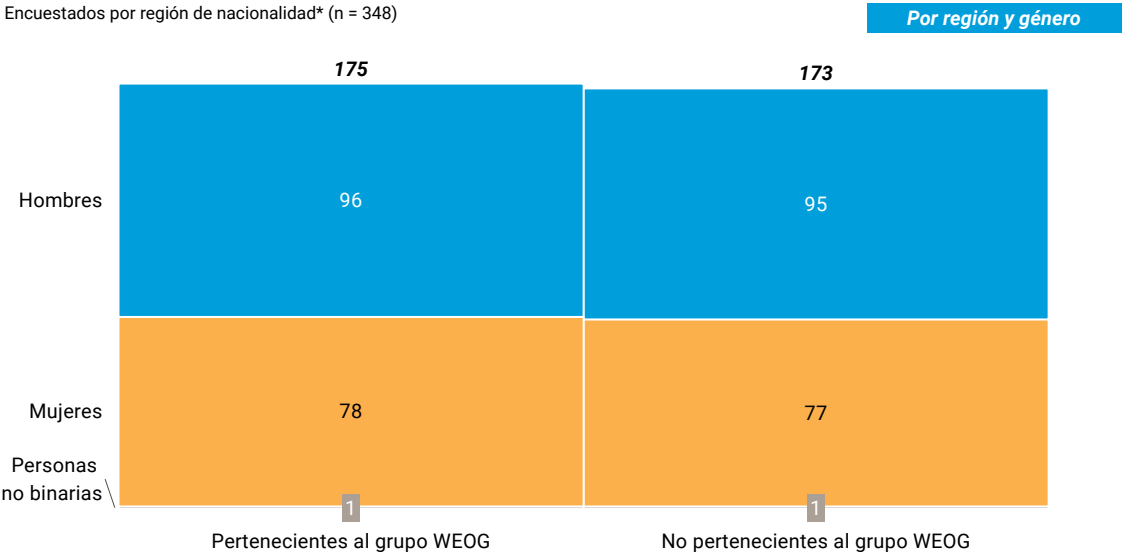
También se les pidió que calificaran su nivel general de preocupación por el agravamiento o generalización sustancial de los daños (ya producidos o nuevos) derivados de la IA, y en qué medida había aumentado o disminuido esa preocupación recientemente. A los encuestados se les proporcionó una lista de 14 ámbitos de daños (como el “uso malintencionado de la IA por parte de agentes no estatales”) para que valoraran su nivel de preocupación. Por último, se incluyeron numerosas preguntas de respuesta libre en las que se invitaba a los expertos a comentar sobre las nuevas tendencias y sobre las personas, los grupos y los (eco)sistemas especialmente amenazados por la IA, así como a profundizar en sus respuestas.

La encuesta se llevó a cabo del 13 al 25 de mayo de 2024, y la lista de participantes se elaboró a partir de las redes de la OSET y del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial e incluyó a expertos del Órgano Asesor en “inmersión en profundidad”. Durante el período de realización de la encuesta, se invitó a más expertos, en particular de regiones que suelen estar menos representadas en los debates sobre la IA, a partir de las referencias de los encuestados iniciales y de las redes regionales. Más de 340 personas respondieron a la encuesta, aportando una perspectiva rica y diversa (incluso entre regiones y géneros) sobre los riesgos que plantea la IA.

Sinopsis de la muestra

Las variables por género y región están equilibradas entre sí

El análisis univariante por género y región no está inmediatamente contaminado por la otra variable



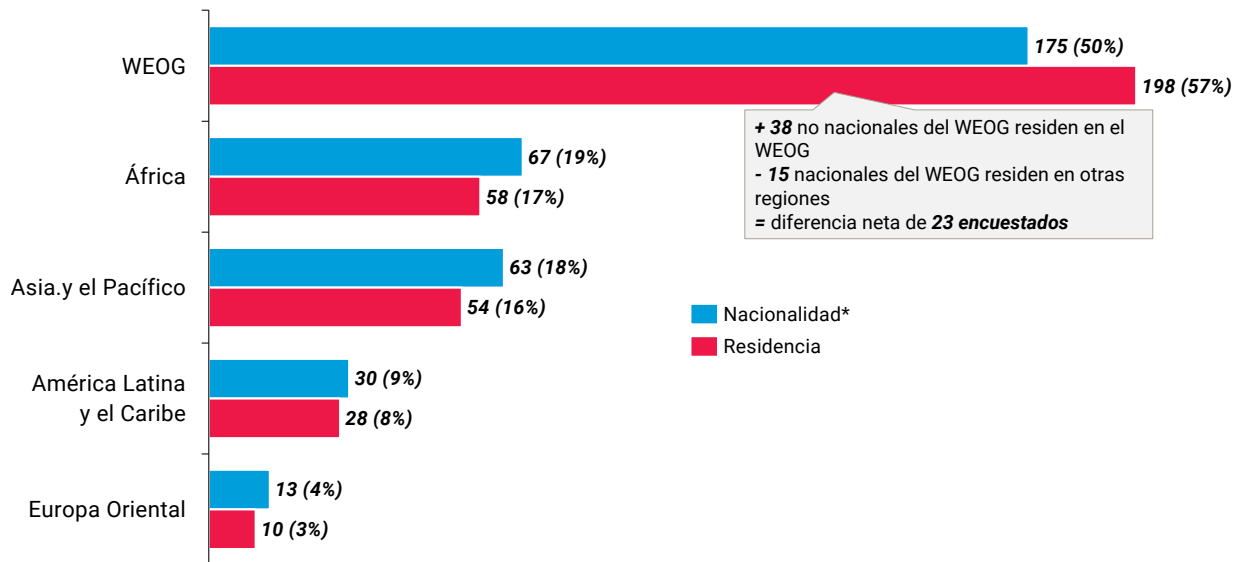
43 encuestados (12%) indicaron varias nacionalidades. Si los encuestados residían en uno de sus países de nacionalidad, se utilizó esa nacionalidad (34 de 43). En caso contrario, se utilizó la nacionalidad menos representada (9 de 43). Fuente: AI Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

La muestra sigue siendo global si se considera según el criterio de residencia

El 84% de los encuestados reside en la región de su nacionalidad

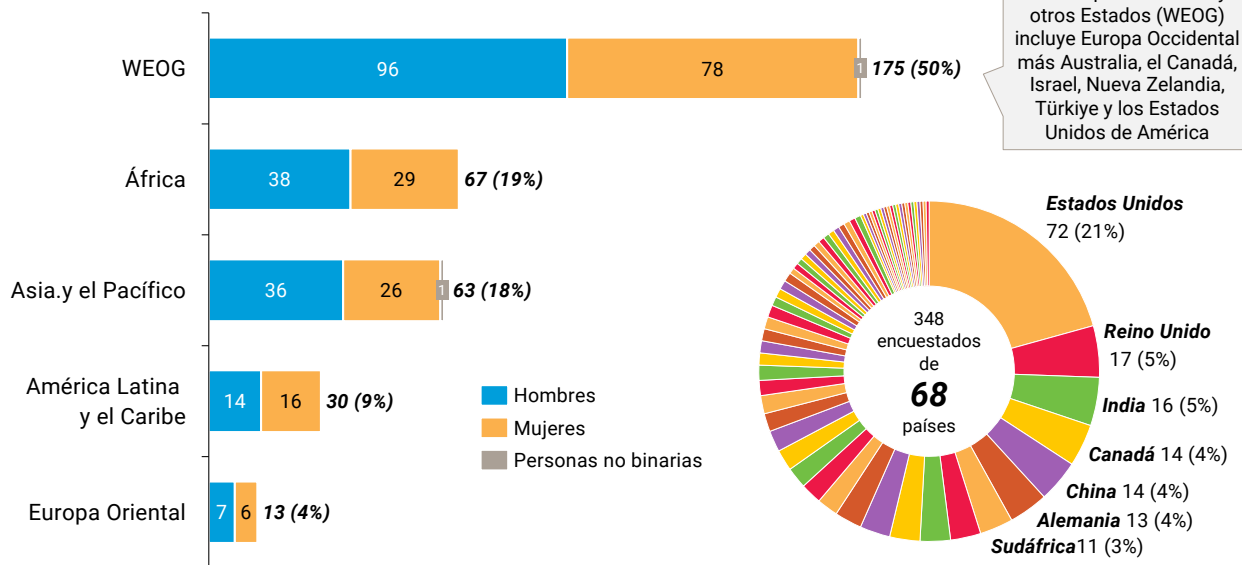
Encuestados por región de nacionalidad* (n = 348)

Por región



43 encuestados (12%) indicaron varias nacionalidades. Si los encuestados residían en uno de sus países de nacionalidad, se utilizó esa nacionalidad (34 de 43). En caso contrario, se utilizó la nacionalidad menos representada (9 de 43). Fuente: AI Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

Encuestados por región de nacionalidad* (n = 348)



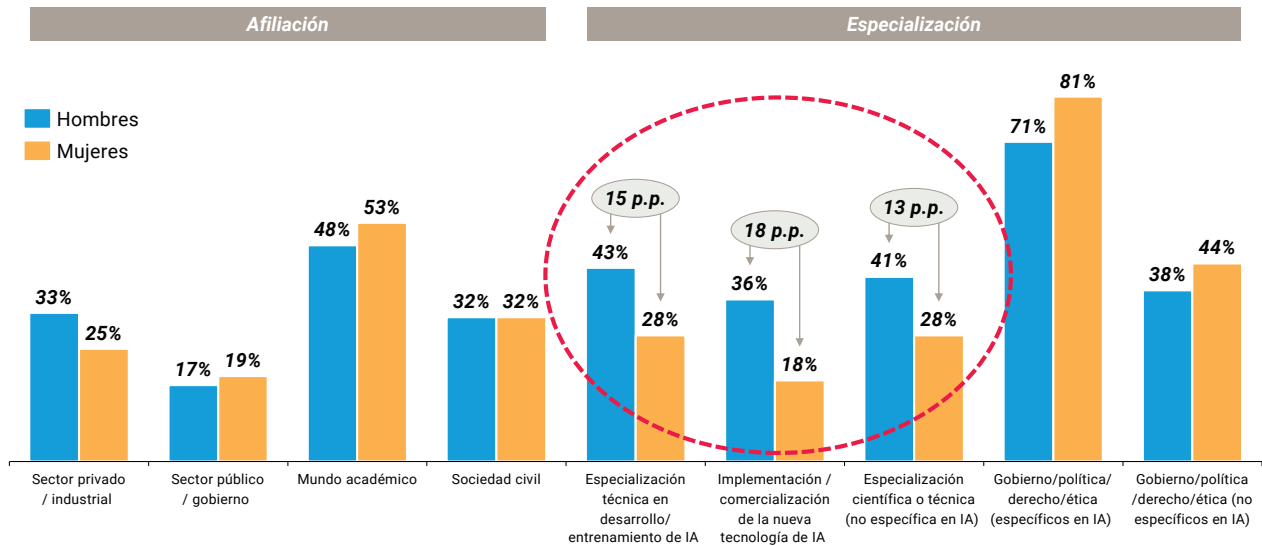
43 encuestados (12%) indicaron varias nacionalidades. Si los encuestados residían en uno de sus países de nacionalidad, se utilizó esa nacionalidad (34 de 43). En caso contrario, se utilizó la nacionalidad menos representada (9 de 43). Fuente: AI Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

Los perfiles de encuestados hombres y mujeres presentan algunas diferencias

Más hombres dicen tener conocimientos técnicos: más mujeres, gobernanza/política/derecho/ética

% Porcentaje de encuestados que comunicaron tener afiliación/ especialización por región de nacionalidad* (n = 348)

Por género



Nota: p.p. = puntos porcentuales.

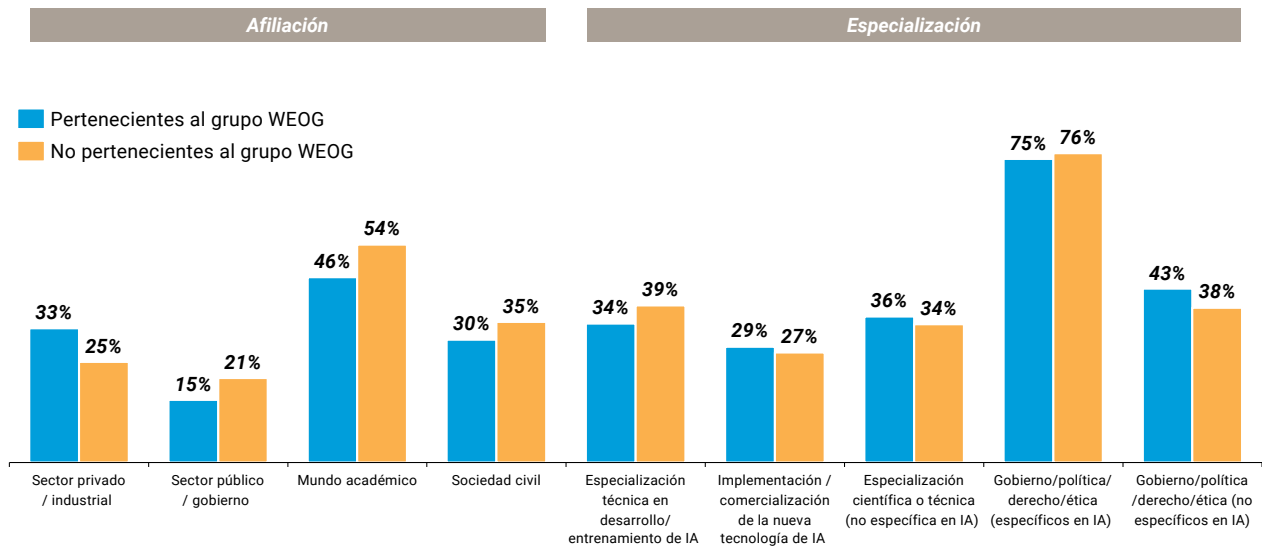
Fuente: AI Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

Perfiles razonablemente similares de los encuestados pertenecientes y no pertenecientes al grupo WEOG

Los encuestados no pertenecientes al grupo WEOG suelen provenir más del sector público o académico que del sector privado o industrial

Porcentaje de encuestados que comunicaron tener afiliación/ especialización por región de nacionalidad * (n = 348)

Por región



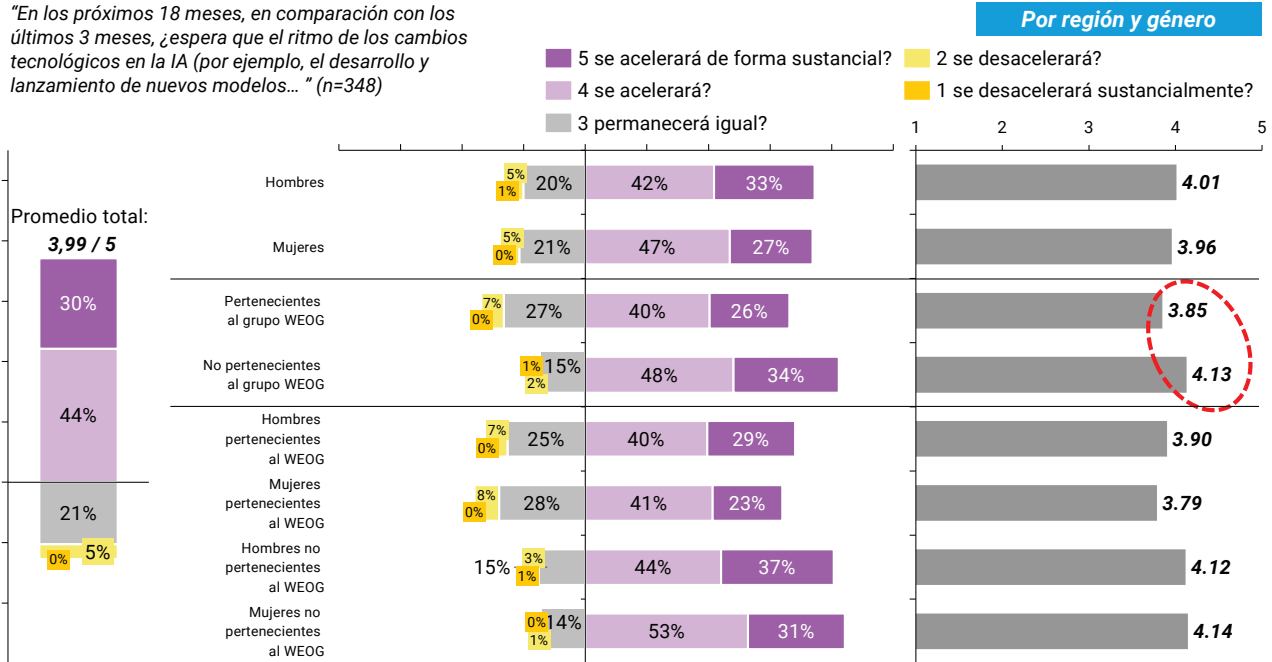
43 encuestados (12%) indicaron varias nacionalidades. Si los encuestados residían en uno de sus países de nacionalidad, se utilizó esa nacionalidad (34 de 43). En caso contrario, se utilizó la nacionalidad menos representada (9 de 43). Fuente: AI Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

Percepciones sobre la aceleración de la IA

El 74% de los encuestados espera una aceleración del cambio tecnológico

Un mayor porcentaje de encuestados que no pertenecen al grupo WEOG esperan una aceleración frente a los que sí pertenecen a él

"En los próximos 18 meses, en comparación con los últimos 3 meses, ¿espera que el ritmo de los cambios tecnológicos en la IA (por ejemplo, el desarrollo y lanzamiento de nuevos modelos..." (n=348)



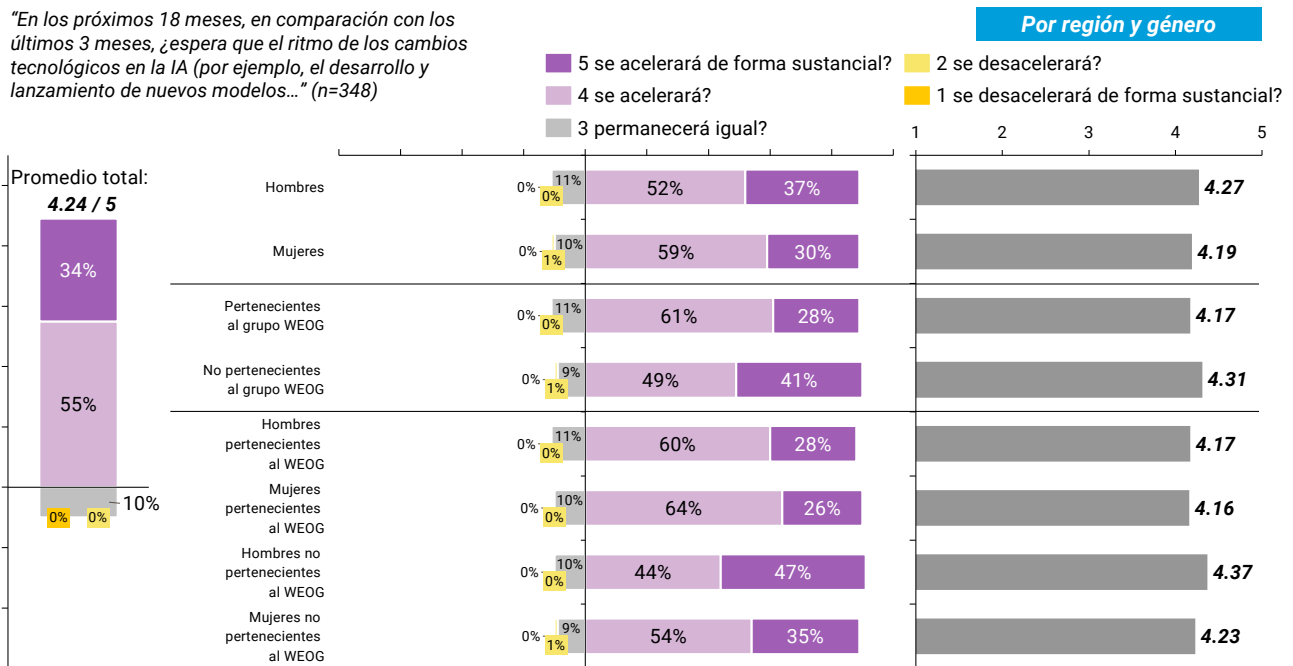
Nota: Excluye "No sabe" / "No contesta" y respuestas en blanco.

Fuente: Al Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

El 89% de los encuestados espera una aceleración de la adopción y aplicación

Un número ligeramente superior de encuestados no pertenecientes al WEOG esperan una aceleración sustancial (especialmente hombres)

"En los próximos 18 meses, en comparación con los últimos 3 meses, ¿espera que el ritmo de los cambios tecnológicos en la IA (por ejemplo, el desarrollo y lanzamiento de nuevos modelos..." (n=348)



Nota: Excluye "No sabe" / "No contesta" y respuestas en blanco.

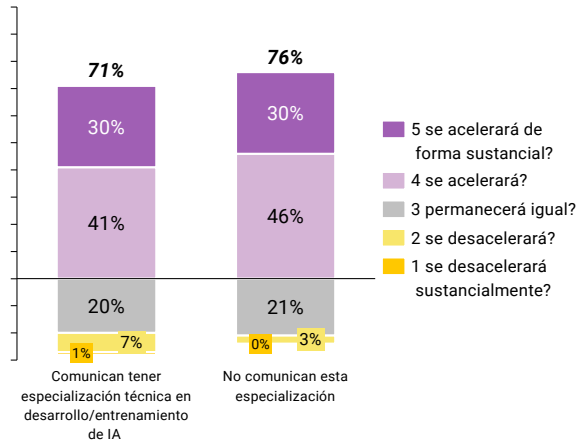
Fuente: Al Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

Impacto limitado de la especialización técnica (desarrollo/entrenamiento de IA)

Los encuestados fueron ligeramente más pesimistas sobre el cambio tecnológico y ligeramente más optimistas sobre la adopción/aplicación

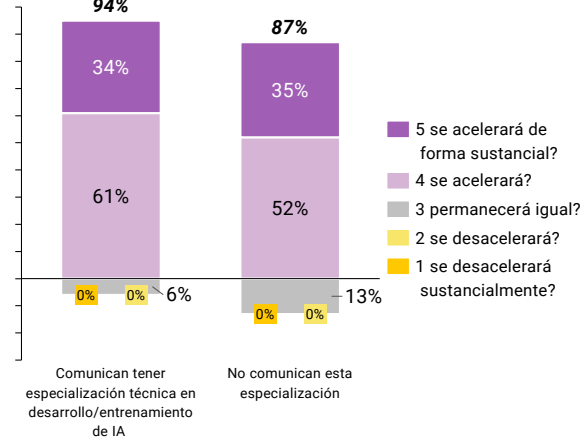
Cambio tecnológico

"En los próximos 18 meses, en comparación con los últimos 3 meses, ¿espera que el ritmo de los cambios tecnológicos en la IA (por ejemplo, el desarrollo y lanzamiento de nuevos modelos..." (n=348)



Adopción y aplicación

"En los próximos 18 meses, en comparación con los últimos 3 meses, ¿espera que el ritmo de adopción y aplicación de la IA (por ejemplo, nuevos usos de la IA en empresas y administraciones públicas):" (n=348)



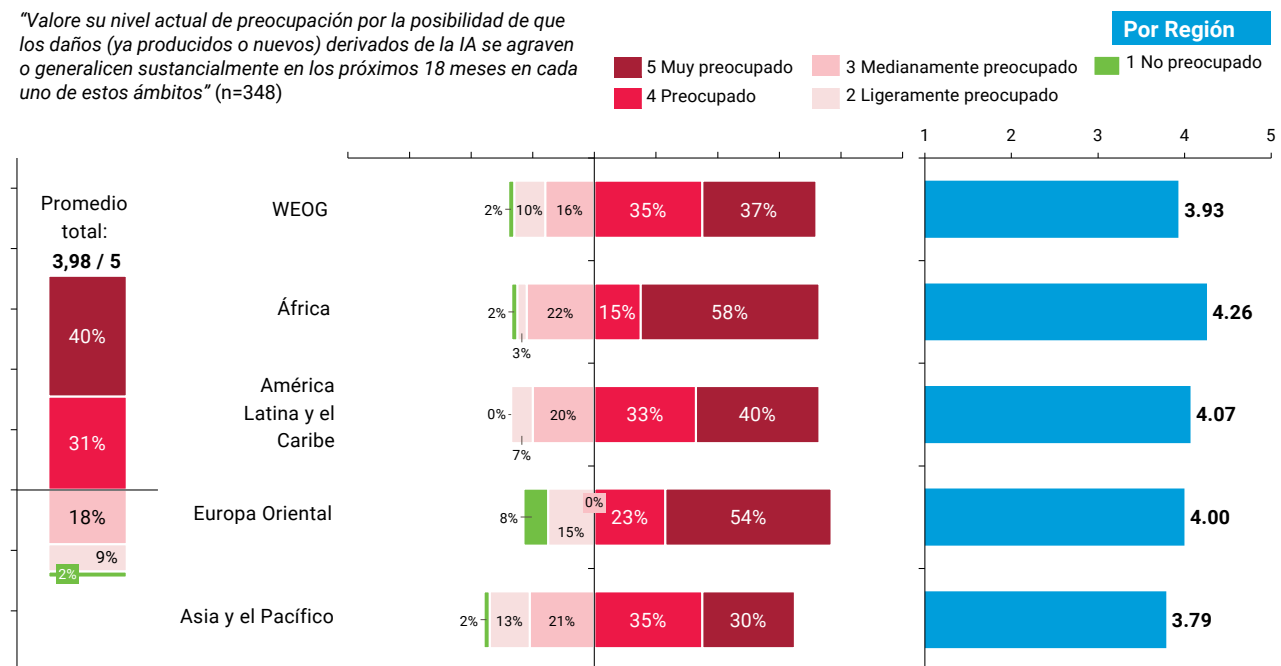
Nota: Las cifras pueden no sumar el 100% debido al redondeo. Excluye "No sabe" / "No contesta" y respuestas en blanco.
Fuente: AI Risk Pulse Check de la Oficina del Enviado del Secretario General para la Tecnología, 13 a 25 de mayo de 2024

Percepciones sobre los riesgos de daños producidos por la IA a 18 meses (desde mayo de 2024)

El 71% de los encuestados está preocupado o muy preocupado por los daños de la IA en los próximos 18 meses

Los encuestados africanos están más preocupados que los demás; los de Asia y el Pacífico lo están menos que los pertenecientes al grupo WEOG

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos" (n=348)

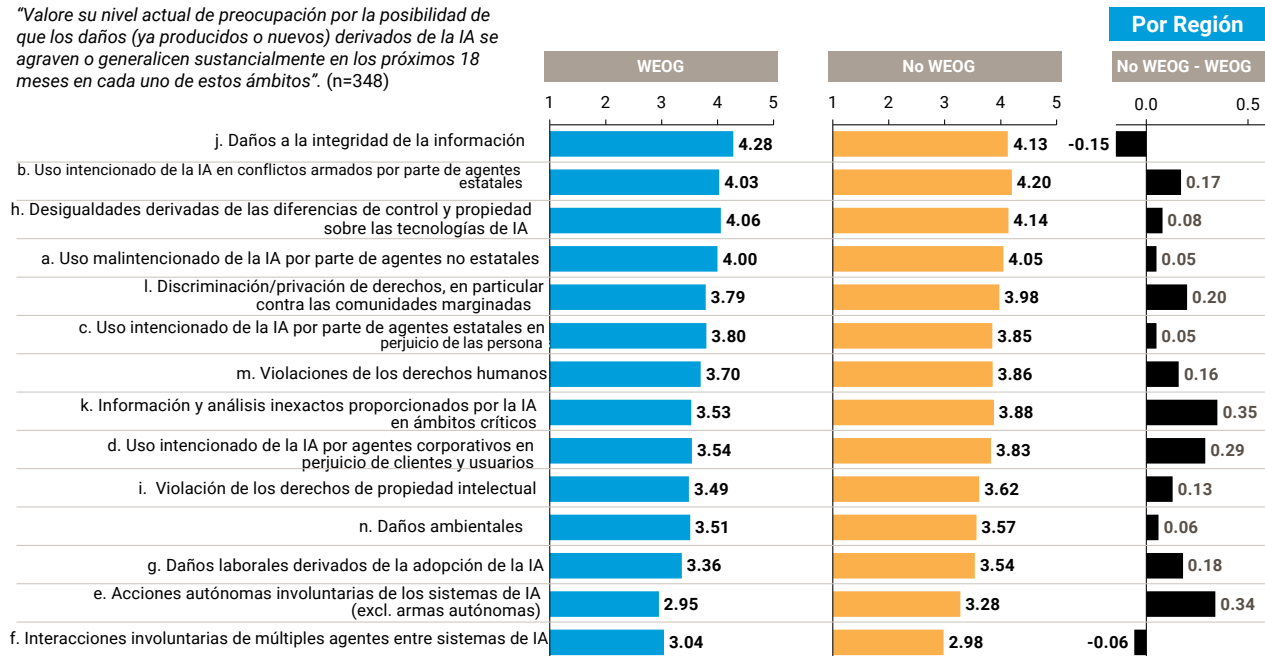


Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco.
Fuente: OSET Pulso de Riesgo IA, 13-25 mayo 2024.

Los no pertenecientes al WEOG están más preocupados que los pertenecientes al WEOG en la mayoría de los ámbitos

Lagunas especialmente graves en cuanto a información inexacta, acciones autónomas no intencionadas y uso corporativo intencionado

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos". (n=348)

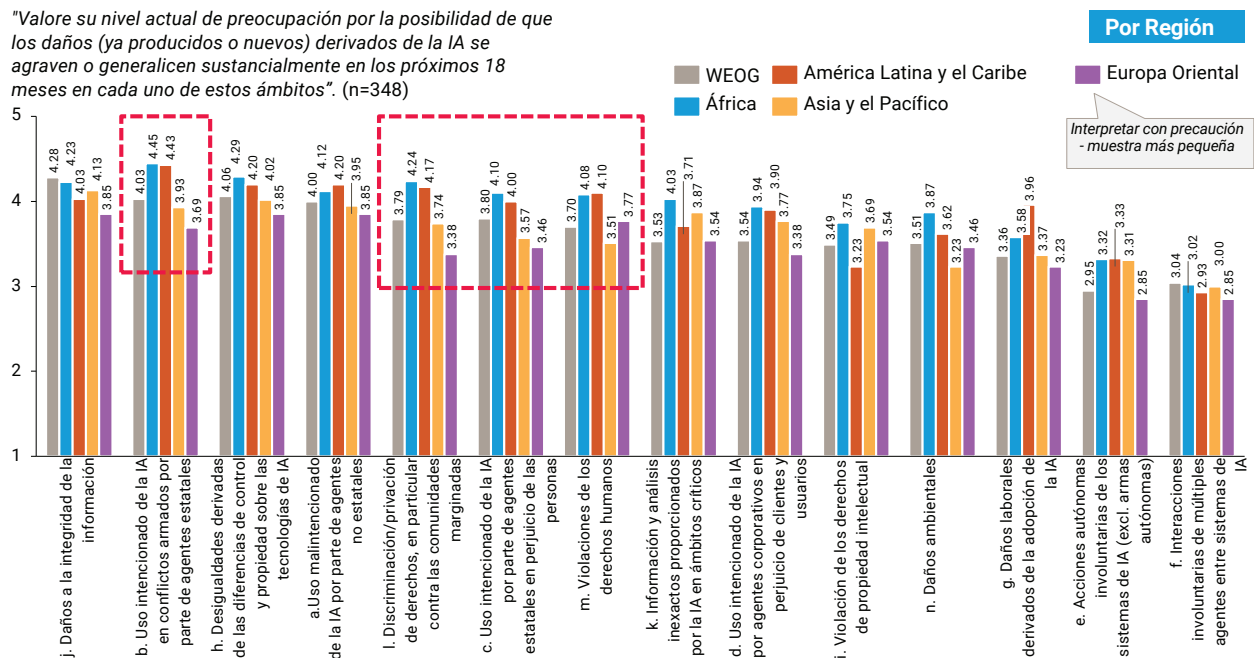


Se muestra: Promedio, donde: 1 = No preocupado; 2 = Ligeramente preocupado; 3 = Medianamente preocupado; 4 = Preocupado; 5 = Muy preocupado. Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Fuente: OSET Pulso de Riesgo IA, 13-25 mayo 2024.

Mucha preocupación, la mayor en África y América Latina y el Caribe

Especialmente en torno al uso estatal en conflictos armados, que favorece la discriminación y las violaciones de los derechos humanos

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos". (n=348)

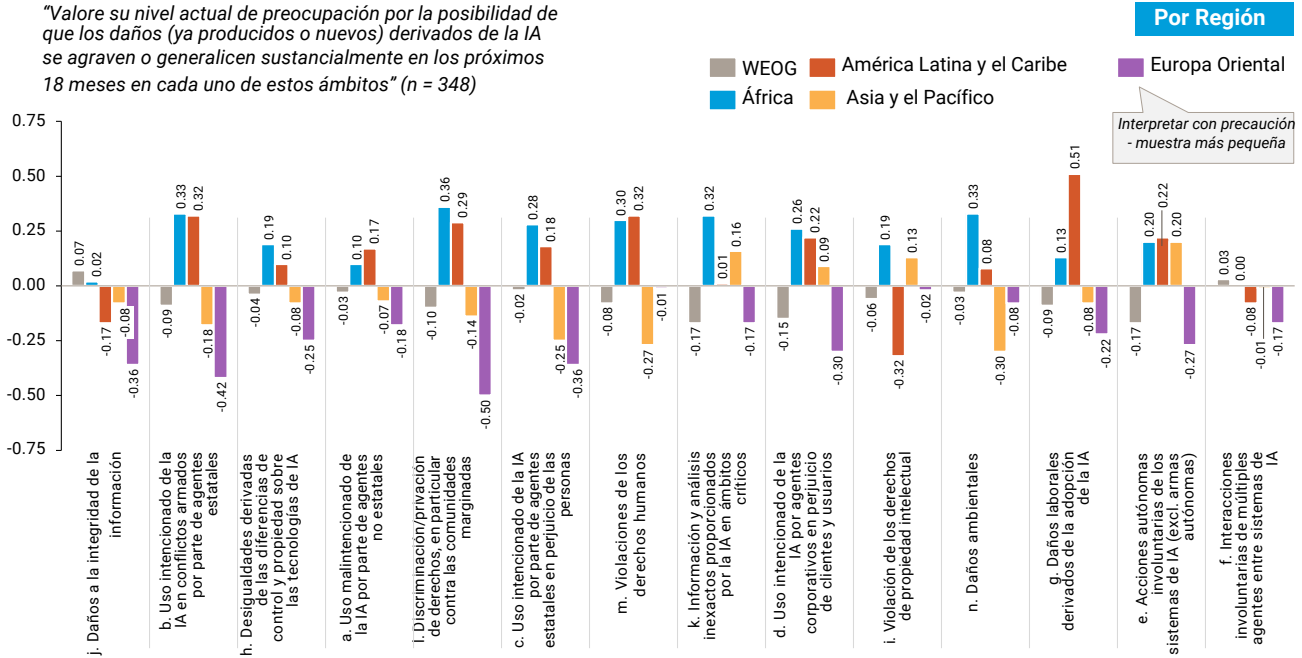


Se muestra: Promedio, donde: 1 = No preocupado; 2 = Ligeramente preocupado; 3 = Medianamente preocupado; 4 = Preocupado; 5 = Muy preocupado. Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Fuente: OSET Pulso de Riesgo IA, 13-25 mayo 2024.

Mucha preocupación, la mayor en África y América Latina y el Caribe

Especialmente en torno al uso estatal en conflictos armados, que favorece la discriminación y las violaciones de los derechos humanos

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos" (n = 348)

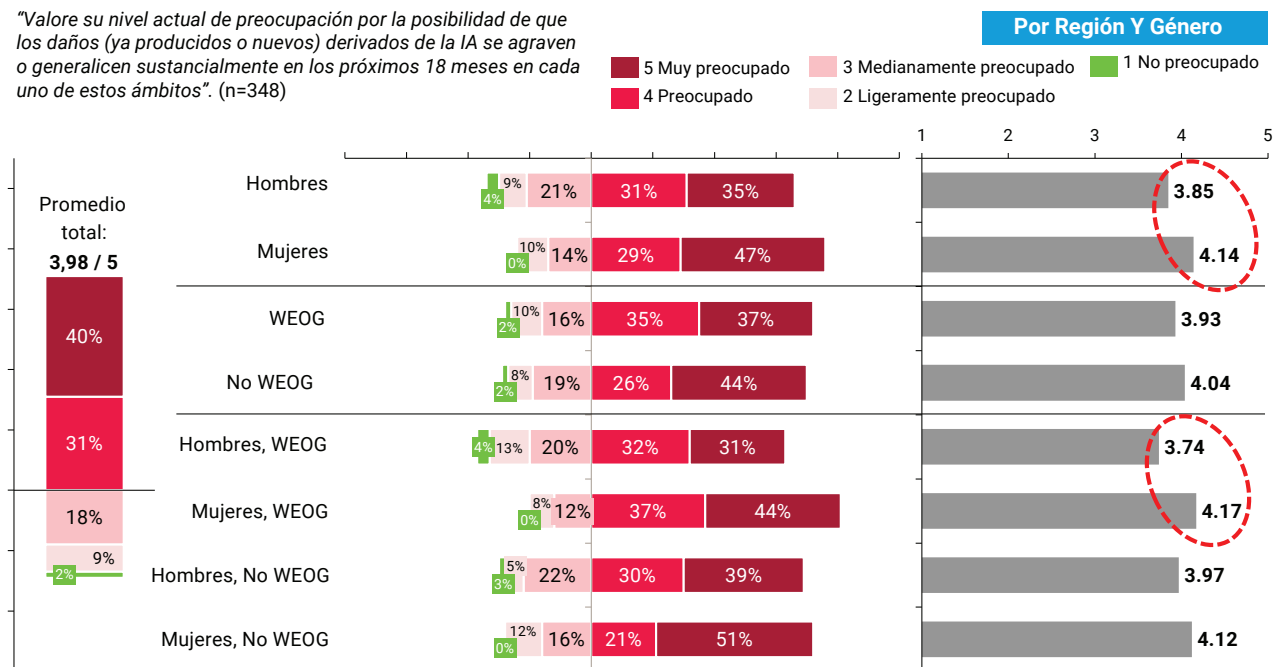


Se muestra: 1 = No preocupado, 2 = Ligeramente preocupado, 3 = Algo preocupado, 4 = Preocupado, 5 = Muy preocupado. Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Fuente: OSET Pulso de Riesgo IA, 13-25 mayo 2024.

El 71% de los encuestados está preocupado o muy preocupado por los daños causados por la IA en los próximos 18 meses

Las mujeres están más preocupadas que los hombres, sobre todo las mujeres pertenecientes al WEOG

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos". (n=348)

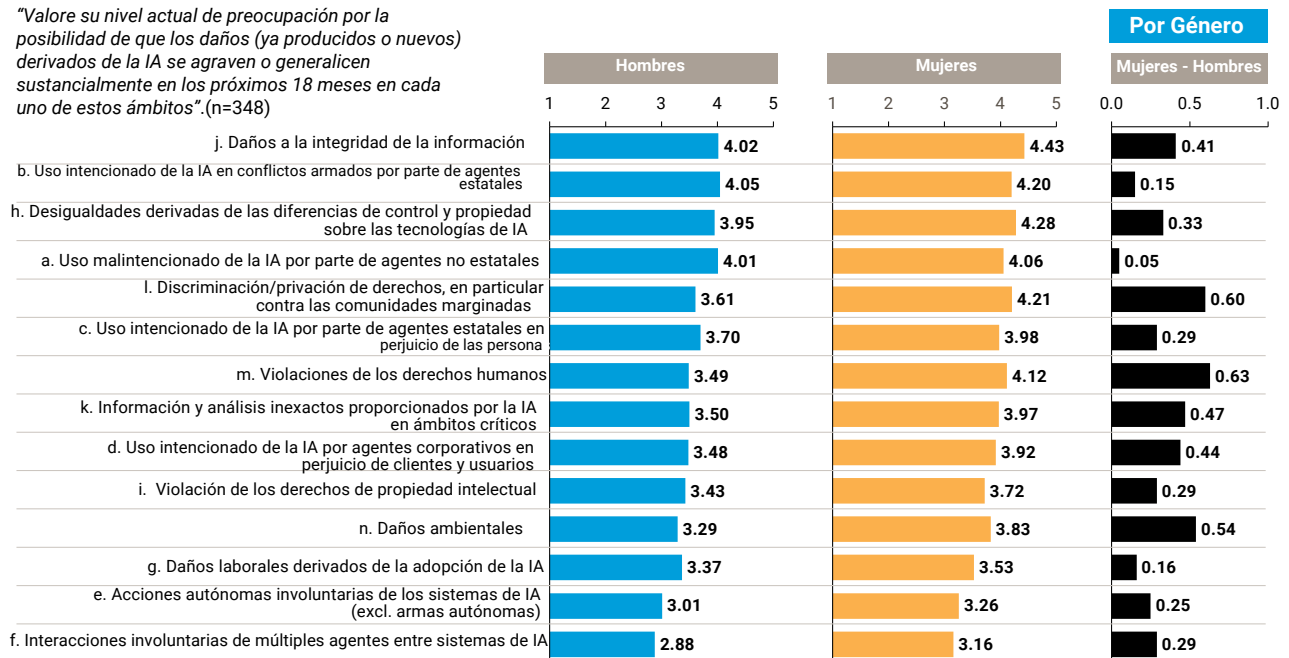


Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Fuente: OSET AI Risk Pulse Check, 13 A 25 DE Mayo de 2024.

Las mujeres están más preocupadas que los hombres en todos los ámbitos

Lagunas especialmente graves en materia de violación de los derechos humanos, discriminación y medio ambiente

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos". (n=348)

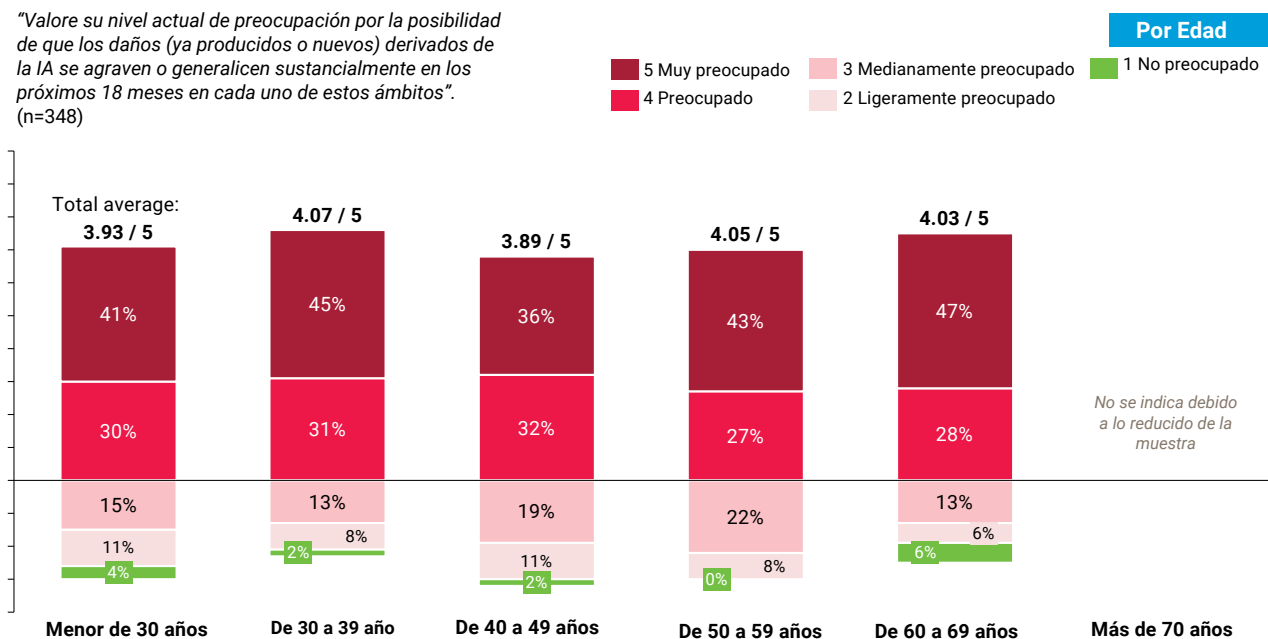


Se muestra: Promedio, donde: 1 = No preocupado; 2 = Ligeramente preocupado; 3 = Medianamente preocupado; 4 = Preocupado; 5 = Muy preocupado.
 Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Fuente: OSET Pulso de Riesgo IA, 13-25 mayo 2024.

El 71% de los encuestados está preocupado o muy preocupado por los daños causados por la IA en los próximos 18 meses

Diferencias relativamente pequeñas de preocupación según la edad de los encuestados

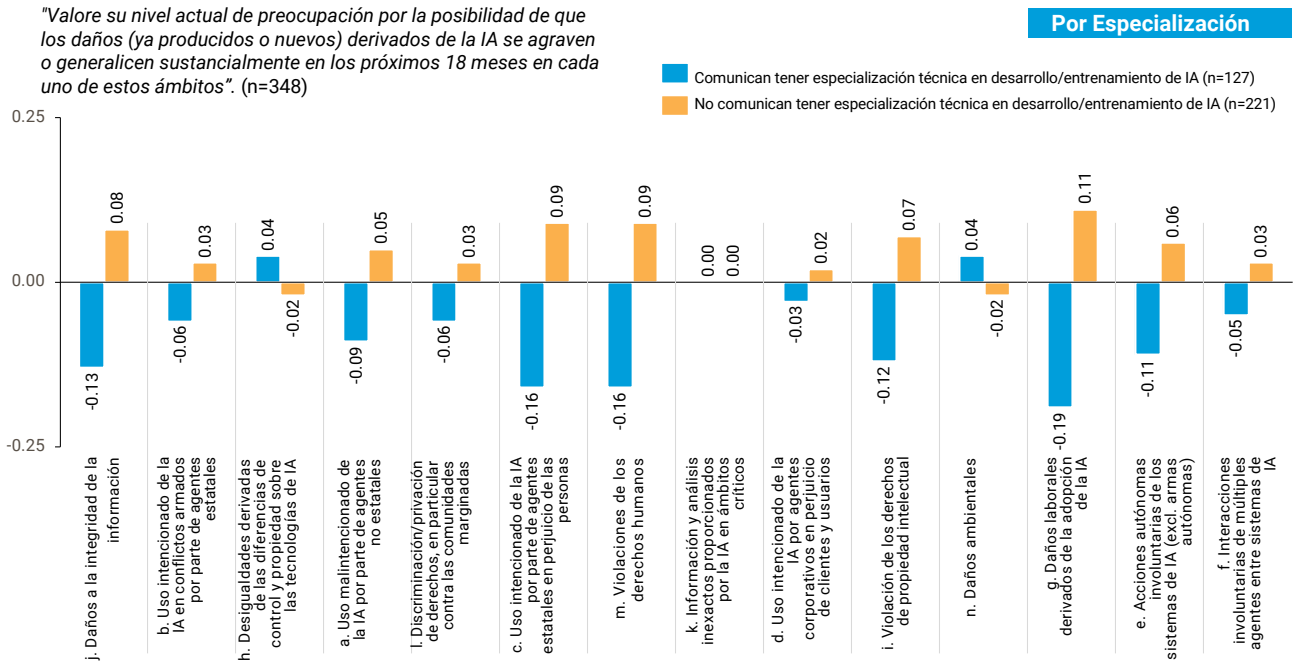
"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos". (n=348)



Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco.
 Fuente: OSET AI Risk Pulse Check, 13 A 25 DE Mayo de 2024.

Los encuestados con especialización técnica (desarrollo/entrenamiento de IA) están menos preocupados en la mayoría de los ámbitos

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos". (n=348)

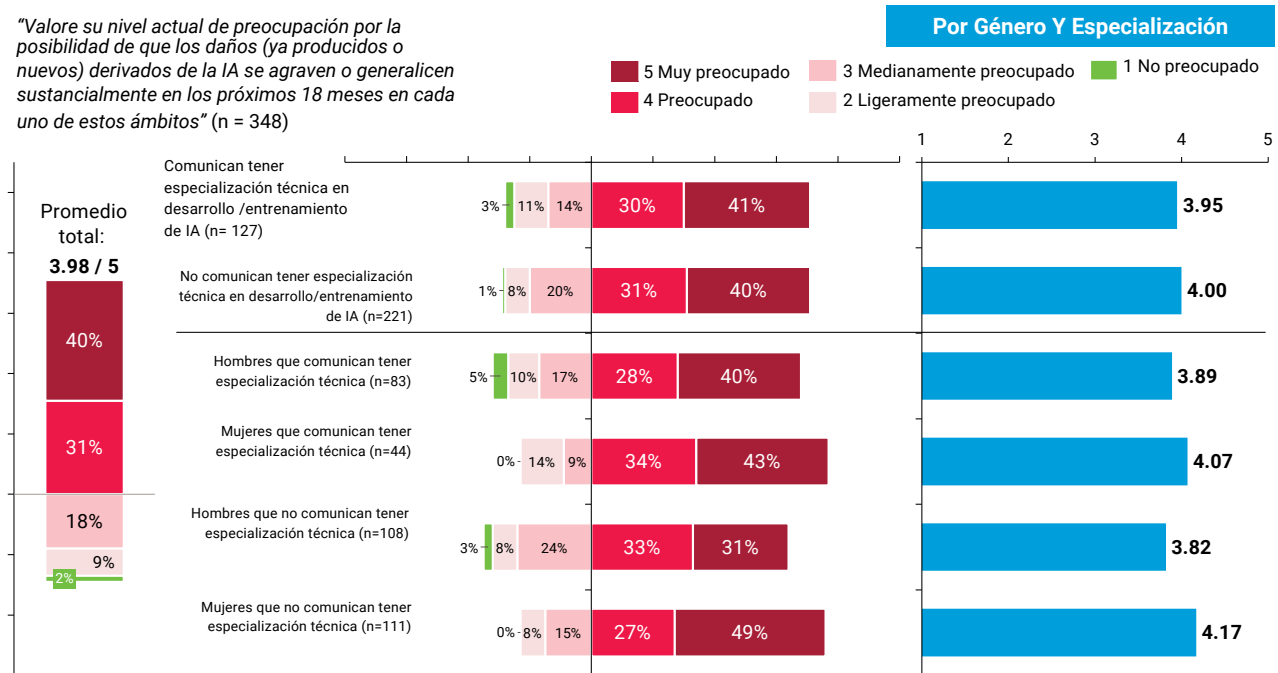


Se muestra: 1 = No preocupado, 2 = Ligeramente preocupado, 3 = Algo preocupado, 4 = Preocupado, 5 = Muy preocupado. Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Fuente: OSET Pulso de Riesgo IA, 13-25 mayo 2024.

Impacto limitado de la especialización técnica (desarrollo/entrenamiento de IA)

Los hombres están menos preocupados que las mujeres, independientemente de lo que comuniquen

"Valore su nivel actual de preocupación por la posibilidad de que los daños (ya producidos o nuevos) derivados de la IA se agraven o generalicen sustancialmente en los próximos 18 meses en cada uno de estos ámbitos" (n = 348)



Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Fuente: OSET AI Risk Pulse Check, 13 A 25 DE Mayo de 2024.

Cambio de percepción sobre el nivel de preocupación, en los últimos tres meses, acerca de los riesgos de daños por IA

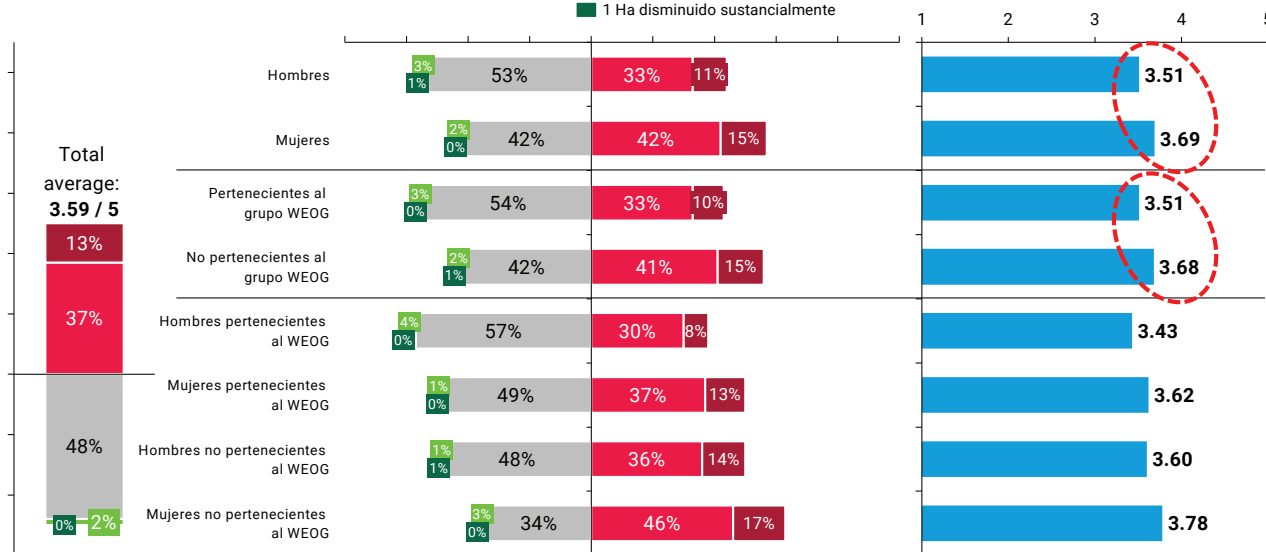
En los últimos tres meses ha aumentado la preocupación del 50% de los encuestados; el 48% permanece igual

La preocupación no ha disminuido en casi ningún caso; y ha aumentado más entre las mujeres encuestadas que no pertenecen al WEOG

"¿Cómo ha cambiado su nivel de preocupación en los últimos tres meses?" (n=348)

- 5 Ha aumentado sustancialmente
- 4 Ha aumentado
- 3 Ha permanecido igual
- 2 Ha disminuido
- 1 Ha disminuido sustancialmente

POR GÉNERO Y ESPECIALIZACIÓN



Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco.
Fuente: Encuesta AI Risk Pulse Check, 13 a 25 de mayo de 2024.

Anexo F: Respuestas al estudio prospectivo de oportunidades

A petición del Órgano Asesor de Alto Nivel sobre Inteligencia Artificial, la Oficina del Enviado del Secretario General para la Tecnología (OSET) llevó a cabo un estudio prospectivo mundial sobre las oportunidades de la IA. Se pidió a los expertos que respondieran con sus opiniones a título personal (no en nombre de su institución o empleador). La encuesta se dividió en secciones que abordaban las oportunidades en los países de ingreso alto y mediano alto y en los países de ingreso mediano bajo y bajo, y solo respondieron a esas preguntas los encuestados que tenían conocimientos específicos sobre los contextos de los países de ingreso mediano bajo y bajo. En la encuesta únicamente se preguntó por las posibles implicaciones positivas de la IA.

Se preguntó a los encuestados en qué medida conocían, hasta la fecha, ejemplos concretos de IA que aumentarían la actividad económica, acelerarían los descubrimientos científicos y contribuirían al progreso de determinados ODS¹. Se les pidió que proporcionaran información detallada, así como estudios de caso, nombres de organizaciones, datos y enlaces a artículos/publicaciones/documentos pertinentes. A continuación se preguntó a los encuestados cuánto progreso esperaban, en las mismas dimensiones, en los tres años siguientes.

Como punto de vista adicional, se preguntó a los encuestados para cuándo esperaban un mayor impacto de la IA en esas dimensiones con un 50% de confianza y probabilidad. Se formularon otras preguntas, como qué agentes participaban en la obtención de determinadas oportunidades, qué obstáculos contribuían a la brecha de la IA entre países y si determinados grupos se enfrentaban a otras limitaciones a la hora de aprovechar las oportunidades de la IA, y cómo podían abordarlas.

La encuesta se realizó del 9 al 21 de agosto de 2024, y la lista de participantes se elaboró a partir de las redes de la OSET y del Órgano Asesor, e incluyó a expertos del Órgano Asesor en "inmersión en profundidad". Además, para realizar la encuesta se recurrió en gran medida a la reunión AI for Good de la Unión Internacional de Telecomunicaciones y a las redes de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo. En total, participaron más de 1.000 personas. Más de 120 personas respondieron a la encuesta, lo que aporta una perspectiva rica y diversa (también entre regiones y géneros) sobre las oportunidades de la IA.

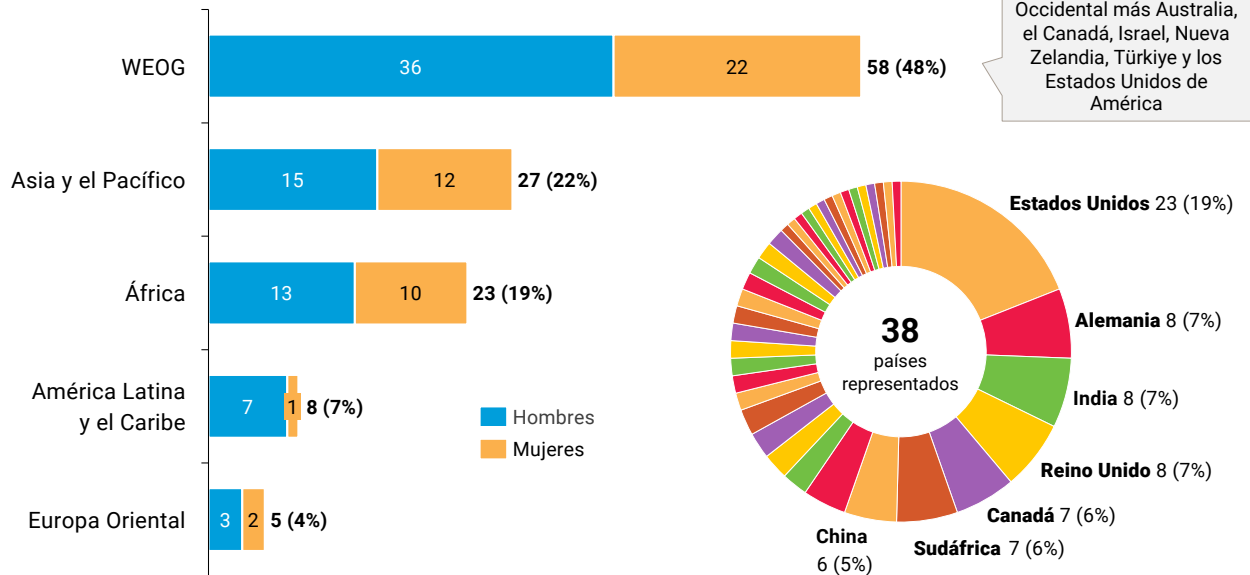
1 El ODS 8 (Trabajo decente y crecimiento económico) y el ODS 9 (Industria, innovación e infraestructuras) no se trataron por separado, dada su estrecha relación con el aumento de la actividad económica. Tampoco se preguntó específicamente por el ODS 17 (Alianzas para lograr los Objetivos).

Sinopsis de la muestra

Representación regional: sólida participación mundial

Permite comparar las respuestas entre el Grupo de los Estados de Europa Occidental y Otros Estados (WEOG) y otras regiones

Encuestados por región de nacionalidad* (n = 121)



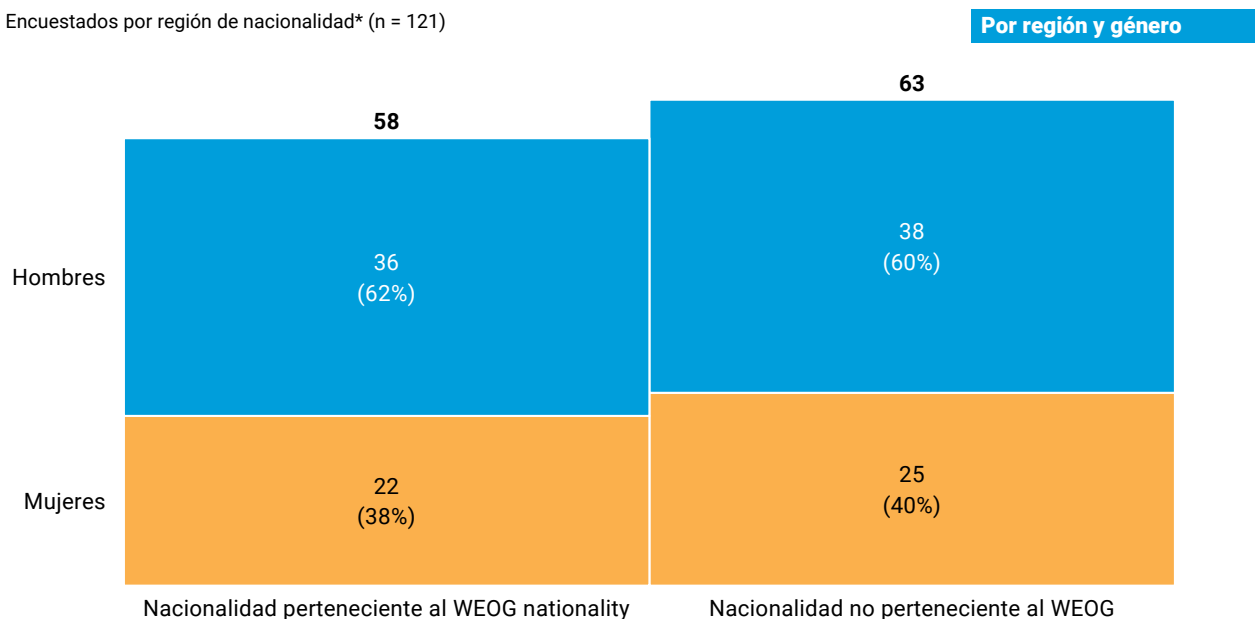
* 9 encuestados (7%) indicaron varias nacionalidades. Si los encuestados residían en uno de sus países de nacionalidad, se utilizó esa nacionalidad (8 de 9). En caso contrario, se utilizó la nacionalidad menos representada (1 de 9).

Fuente: Encuesta AI Risk Pulse Check a 21 de agosto de 2024 (n=121).

Los hombres representan aproximadamente el 60% de las muestras de población perteneciente y no perteneciente al WEOG

La coherencia significa que el análisis univariante por género y región no se contamina inmediatamente

Encuestados por región de nacionalidad* (n = 121)



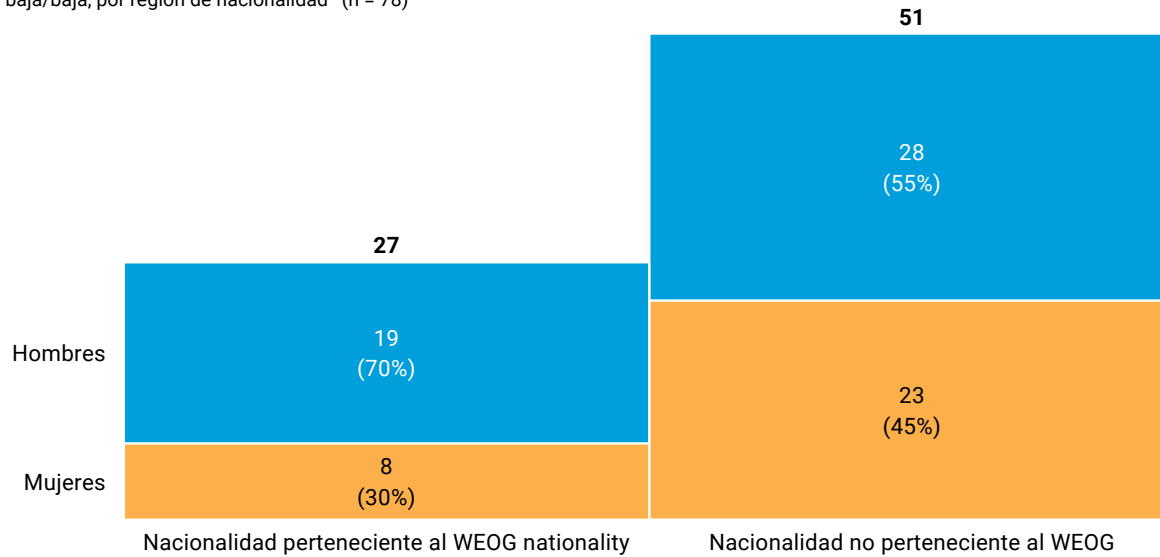
* 9 encuestados (7%) indicaron varias nacionalidades. Si los encuestados residían en uno de sus países de nacionalidad, se utilizó esa nacionalidad (8 de 9). En caso contrario, se utilizó la nacionalidad menos representada (1 de 9).

Fuente: Encuesta AI Risk Pulse Check a 21 de agosto de 2024 (n=121).

La muestra de países en desarrollo está menos equilibrada

Encuestados que declaran tener conocimientos específicos sobre contextos de países de renta media-baja/baja, por región de nacionalidad* (n = 78)

Por región y género



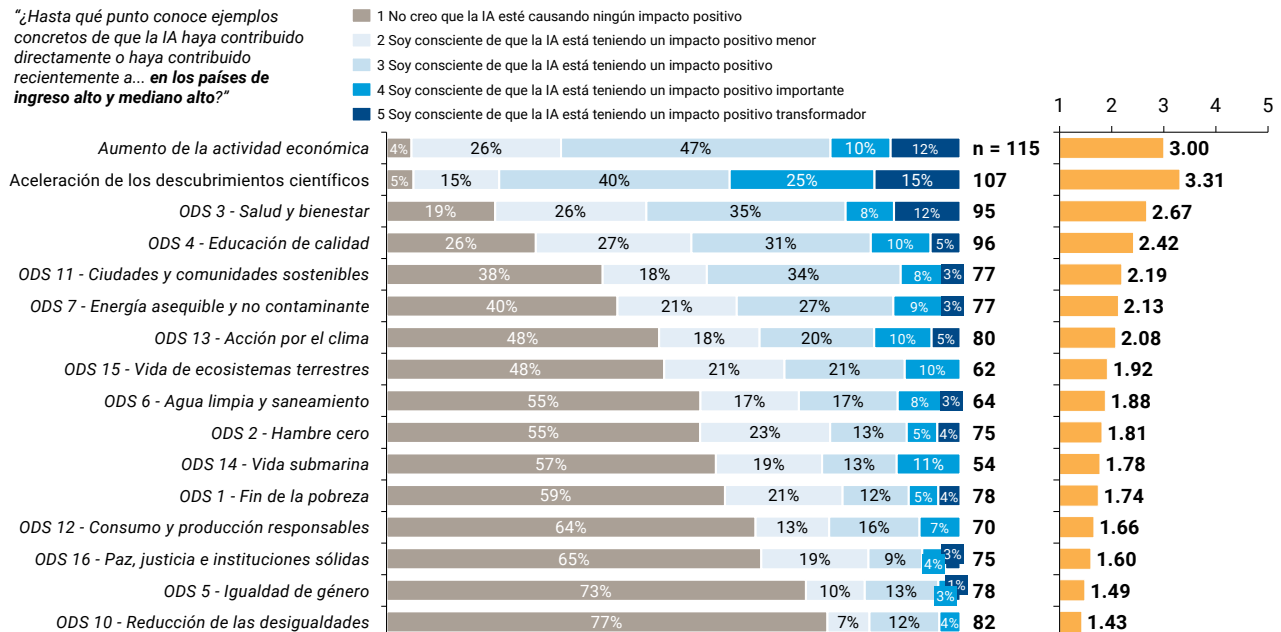
* 9 encuestados (7%) indicaron varias nacionalidades. Si los encuestados residían en uno de sus países de nacionalidad, se utilizó esa nacionalidad (8 de 9). En caso contrario, se utilizó la nacionalidad menos representada (1 de 9).
Fuente: Encuesta AI Risk Pulse Check a 21 de agosto de 2024 (n=121).

Percepciones sobre el impacto positivo de la IA hasta la fecha

Impacto positivo hasta la fecha en los ámbitos del crecimiento y la ciencia, pero menor en la mayoría de los ODS

Impacto hasta la fecha en los países de ingreso alto y mediano alto

“¿Hasta qué punto conoce ejemplos concretos de que la IA haya contribuido directamente o haya contribuido recientemente a... en los países de ingreso alto y mediano alto?”

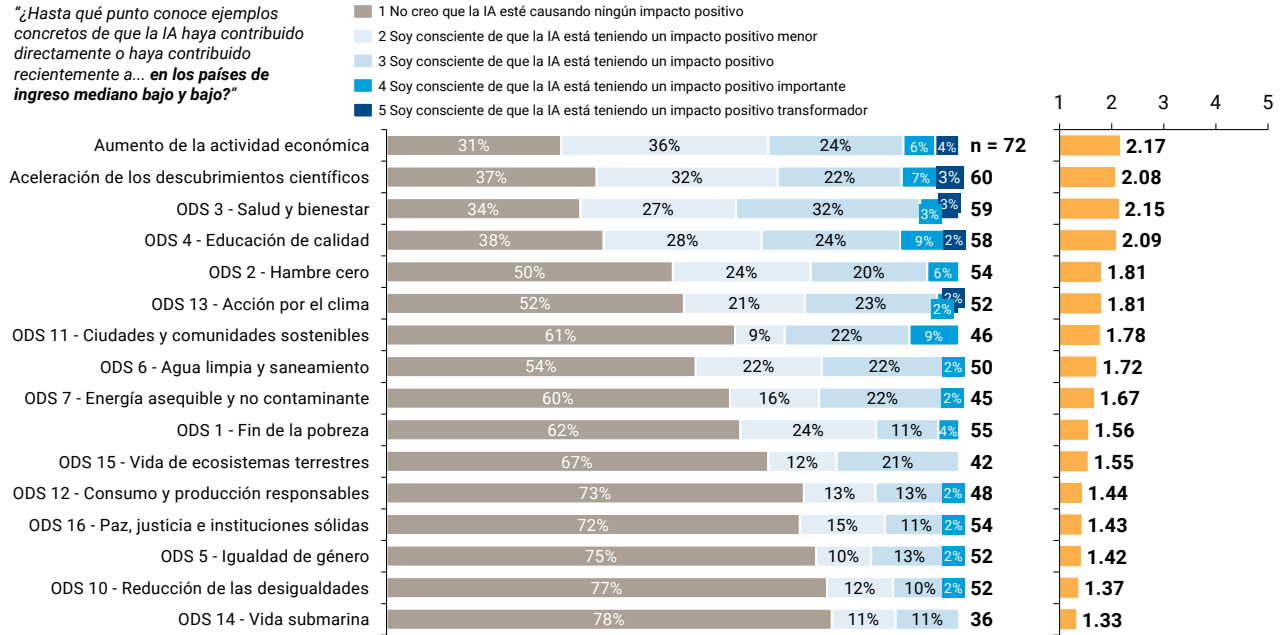


Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. El número específico de encuestados se indica en cada pregunta. No se preguntó sobre los ODS 8, 9 y 17.
Fuente: Encuesta AI Risk Pulse Check de 21 de agosto de 2024 (n=121).

Menor impacto comunicado en los países de ingreso bajo en todos los frentes

Impacto hasta la fecha en los países de ingreso mediano bajo y bajo

“¿Hasta qué punto conoce ejemplos concretos de que la IA haya contribuido directamente o haya contribuido recientemente a... en los países de ingreso mediano bajo y bajo?”

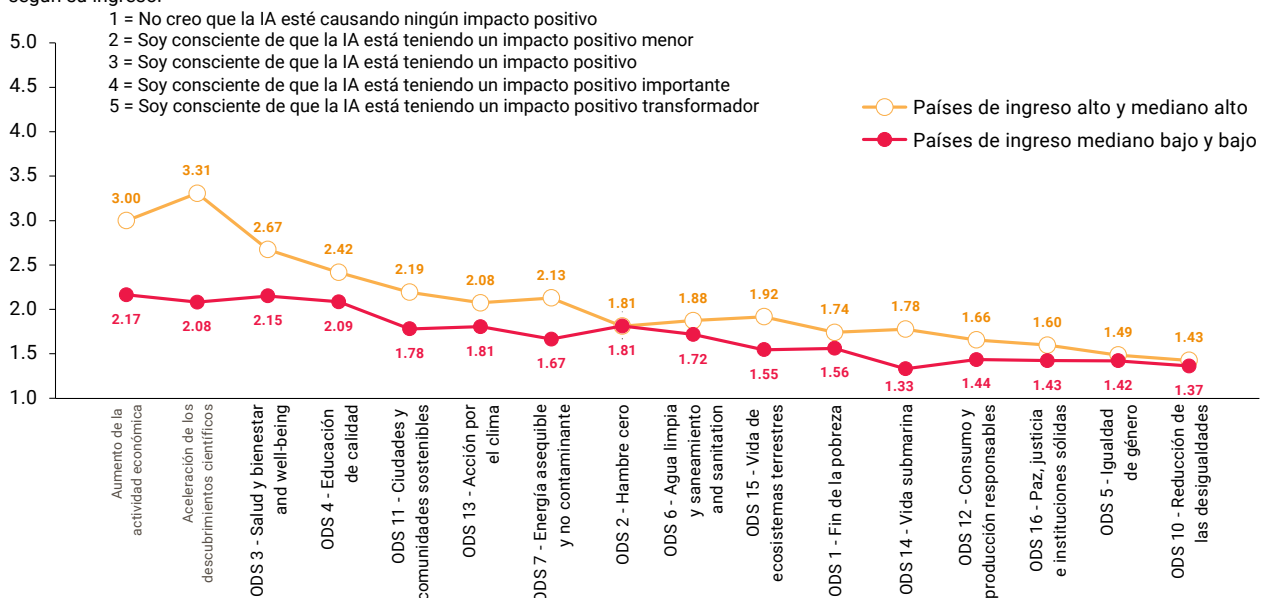


Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. El número específico de encuestados se indica en cada pregunta. No se preguntó sobre los ODS 8, 9 y 17. Fuente: Encuesta AI Risk Pulse Check de 21 de agosto de 2024 (n=121, de los cuales 78 respondieron a preguntas sobre países de ingreso mediano bajo y bajo).

Menor impacto comunicado en los países de ingreso bajo en todos los frentes

La brecha es más pronunciada en crecimiento económico y ciencia

Valoración media para “¿Hasta qué punto conoce ejemplos concretos de que la IA haya contribuido directamente o haya contribuido recientemente a ...?”, por grupo de países según su ingreso:

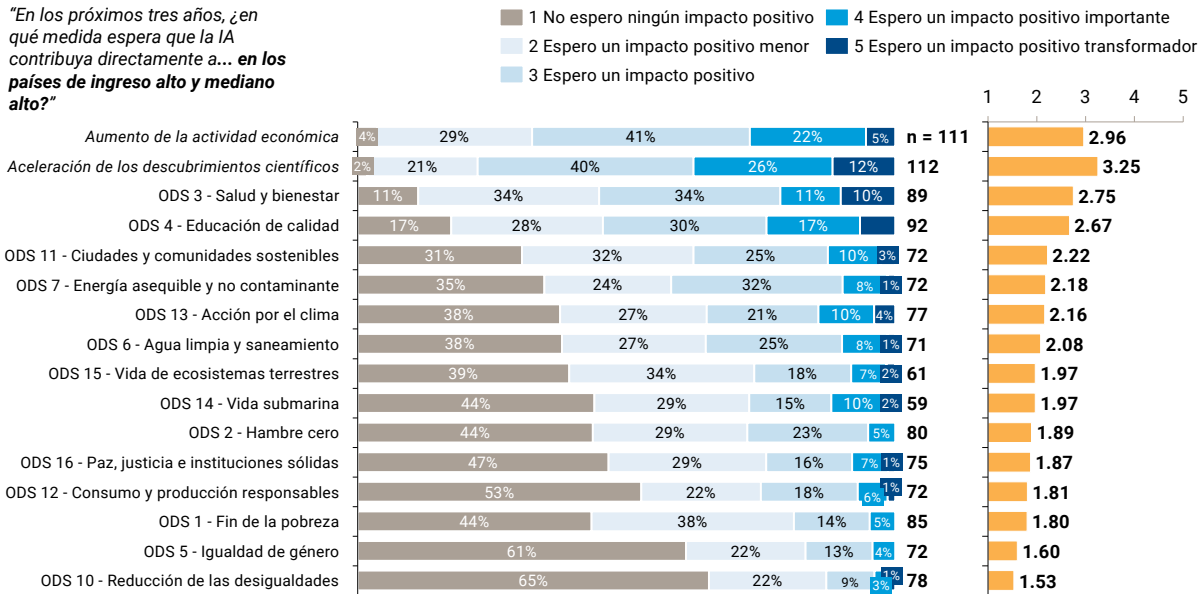


Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Sólo se preguntó sobre los países de renta media-baja/baja a los encuestados que declararon tener conocimientos relevantes. No se preguntó sobre los ODS 8, 9 y 17. Fuente: Encuesta OSET AI Opportunity Scan, 9-21 de agosto de 2024.

Percepciones sobre el impacto positivo previsto de la IA en los próximos tres años

Impacto previsto en el crecimiento, la ciencia, la salud y la educación – impacto menor en otros ámbitos

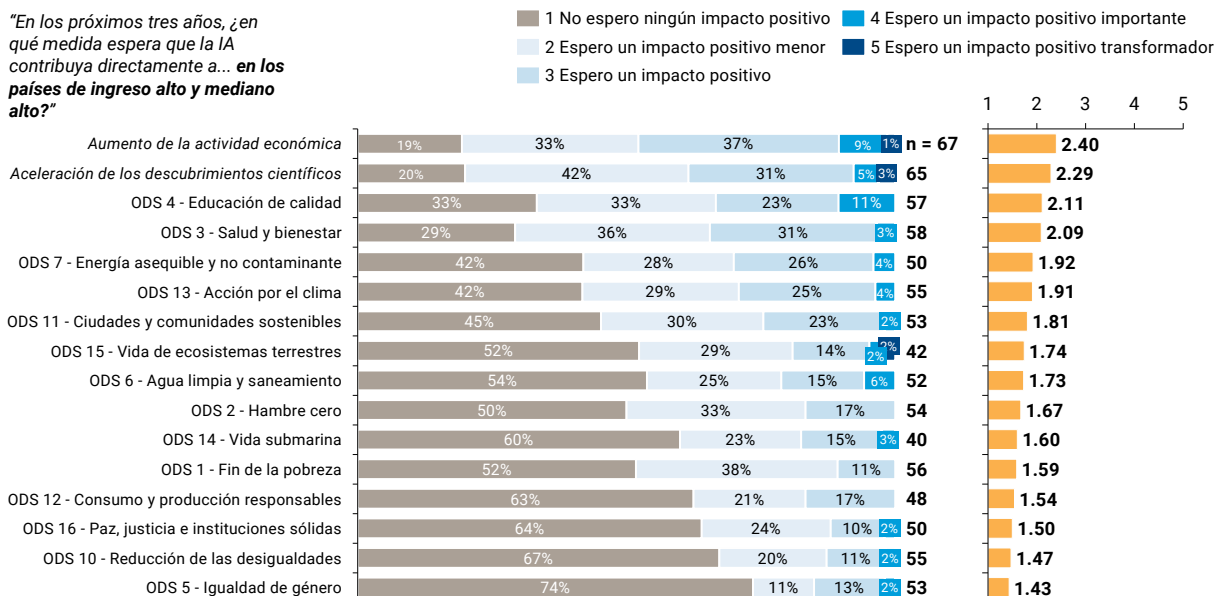
Impacto previsto en los próximos tres años en los países de ingreso alto y mediano alto



Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. El número específico de encuestados se indica en cada pregunta. No se preguntó sobre los ODS 8, 9 y 17.
Fuente: Encuesta AI Risk Pulse Check de 21 de agosto de 2024 (n=121).

Se espera cierto impacto en los países de ingreso bajo, pero de nuevo más limitado

Impacto previsto en los próximos tres años en los países de ingreso mediano bajo y bajo



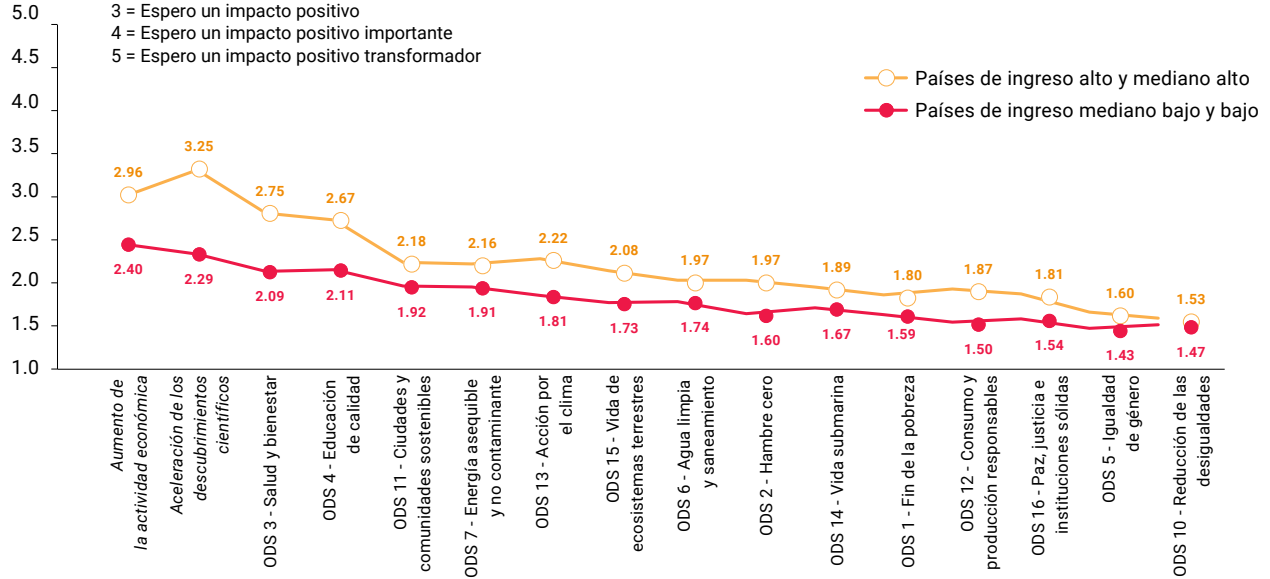
Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. El número específico de encuestados se indica en cada pregunta. No se preguntó sobre los ODS 8, 9 y 17.
Fuente: Encuesta AI Risk Pulse Check de 21 de agosto de 2024 (n=121).

Impacto previsto en los próximos tres años en los países de ingreso mediano bajo y bajo

La brecha es más pronunciada en los ámbitos de crecimiento económico, ciencia, salud y educación

Valoración media para "En los próximos tres años, ¿cuánto espera que la IA contribuya directamente a...?", por grupo de países según su ingreso, donde:

- 1 = No espero ningún impacto positivo
- 2 = Espero un impacto positivo menor
- 3 = Espero un impacto positivo
- 4 = Espero un impacto positivo importante
- 5 = Espero un impacto positivo transformador



Nota: Excluye "No sabe" / "No contesta" y las respuestas en blanco. Sólo se preguntó sobre los países de renta media-baja/baja a los encuestados que declararon tener conocimientos relevantes. No se preguntó sobre los ODS 8, 9 y 17. Fuente: Encuesta OSET AI Opportunity Scan, 9-21 de agosto de 2024. Gráficos preparados con think-cell

Anexo G: Lista de abreviaciones

ACM	Association for Computing Machinery
AG	Grupo de los Estados de África
IA	Inteligencia artificial
ANSI	Instituto Nacional Americano de Normalización
APG	Grupo de Asia y el Pacífico sobre Blanqueo de Dinero
ASEAN	Asociación de Naciones de Asia Sudoriental
BSI	Business Standards Institution
CEN	Comité Europeo de Normalización
CENELEC	Comité Europeo de Normalización Electrotécnica
CERN	Organización Europea de Investigación Nuclear
EEG	Grupo de Europa Oriental
ETSI	Instituto Europeo de Normas de Telecomunicación
FAO	Organización de las Naciones Unidas para la Alimentación y la Agricultura
GAFI	Grupo de Acción Financiera
FIMI	Manipulación e interferencia de información extranjera
FMF	Foro de modelos fronterizos
CEF	Consejo de Estabilidad Financiera
G20	Grupo de los 20
G7	Grupo de los Siete
GPU	Unidad de procesamiento gráfico
OIEA	Organismo Internacional de Energía Atómica
OACI	Organización de Aviación Civil Internacional
CEI	Comisión Electrotécnica Internacional
IEEE	Instituto de Ingenieros Electricistas y Electrónicos
OIT	Organización Internacional del Trabajo
OMI	Organización Marítima Internacional
IPCC	Grupo Intergubernamental de Expertos sobre el Cambio Climático
ISO	Organización Internacional de Normalización
UIT	Unión Internacional de Telecomunicaciones
ALC	América Latina y el Caribe
NIST	Instituto Nacional de Normas y Tecnología (Estados Unidos)
OCDE	Organización de Cooperación y Desarrollo Económicos
ACNUDH	Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos
OSET	Oficina del Enviado del Secretario General para la Tecnología
SAC	Administración China de Normalización
ODS	Objetivos de Desarrollo Sostenible
UNCTAD	Conferencia de las Naciones Unidas sobre Comercio y Desarrollo
PNUD	Programa de las Naciones Unidas para el Desarrollo
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
ACNUR	Oficina del Alto Comisionado de las Naciones Unidas para los Refugiados
OLCT	Oficina de Lucha contra el Terrorismo
WEOG	Grupo de los Estados de Europa Occidental y Otros Estados
OMS	Organización Mundial de la Salud
OMPI	Organización Mundial de la Propiedad Intelectual
WSC	World Standards Cooperation

Donantes

El Órgano Asesor agradece las contribuciones financieras y en especie de los siguientes gobiernos y socios, sin los cuales no habría podido desempeñar sus responsabilidades:

Gobierno de Alemania
Gobierno del Reino de Arabia Saudita
Gobierno de la República Checa
Gobierno de los Emiratos Árabes Unidos
Gobierno de Finlandia
Gobierno de Italia
Gobierno de Japón
Gobierno del Reino de los Países Bajos
Gobierno del Reino Unido de Gran Bretaña e Irlanda del Norte
Gobierno de Singapur
Gobierno de Suiza
Unión Europea
Omidyar Network Fund
Organización Internacional de la Francofonía

Secretaría

Simon Chesterman
Quintin Chou-Lambert
Eleonore Fournier-Tombs
Sebastian Frank
David Michael Kelly
Brian Shung Seun Lau
Andrew Morritt
Filippo Pierozzi
Mehdi Snène
Isabel de Sola
Lucia Velasco
Rebekah Hayoung Woo



United
Nations



AI
Advisory
Body