CELE
Centro de Estudios en Libertad de Expresión

Artículo de investigación

67

Gobernanza global de la IA: ¿quién regula, con qué enfoque y para quién?

Matías González Mama, Agustín Pérez Aledda y Lina Palacios

Junio de 2025

Matías González Mama, Agustín Pérez Aledda, y Lina Palacios, "Gobernanza global de la IA: ¿quién regula, con qué enfoque y para quién?", Artículo de investigación No. 67 (ESP), Centro de Estudios en Libertad de Expresión (CELE), Buenos Aires (2025)

UP
Universidad de Palermo

Facultad de Derecho

Centro de Estudios en Libertad de Expresión

Gobernanza global de la IA: ¿quién regula, con qué enfoque y para quién?

Matías González Mama

Agustín Pérez Aledda

Lina Palacios

CELE/UP

cele@palermo.edu

18 de junio de 2025

Abstract

Este informe presenta un mapeo de las principales iniciativas regulatorias a nivel internacional, regional y nacional sobre inteligencia artificial. En un escenario caracterizado por una creciente proliferación normativa y una competencia por definir modelos regulatorios en la temática, la mayoría de las iniciativas tienen puntos de coincidencia pero, en muchos casos, difieren en sus enfoques y en su naturaleza jurídica. Observamos una tendencia consolidada hacia marcos regulatorios centrados en los riesgos de desarrollo e implementación de la tecnología, donde la protección de los derechos humanos, aunque mencionada, suele quedar en un segundo plano. Aún en ausencia de un modelo regulatorio dominante, es probable que los primeros instrumentos jurídicamente vinculantes que adoptan este enfoque influyan decisivamente en futuros desarrollos normativos que los tomen como modelos, desplazando aún más el foco en derechos. Asimismo, la limitada participación de países en desarrollo en estos procesos plantea desafíos relevantes de inclusión y representatividad en la gobernanza global de la IA, con el riesgo de generar marcos regulatorios desalineados de sus contextos y necesidades locales.[1]

[1] Este informe es resultado de varios meses de debate y trabajo con el equipo del CELE. Los autores quieren agradecer Agustina Del Campo, Ramiro Álvarez Ugarte, Nicolás Zara, Vladimir Garay, Juan Marinangeli y Francesca Chocano por sus ideas y valiosos comentarios a borradores anteriores de este informe.

1. Introducción

En estos últimos años la IA ha ido cobrando una relevancia cada vez mayor. Voces de la comunidad científica, académica, técnica (incluidas las empresas[2]) así como organismos gubernamentales y de organizaciones internacionales advirtieron sobre la existencia de ciertos peligros o riesgos asociados al desarrollo e implementación de esta tecnología y propusieron respuestas regulatorias para intentar morigerar, mitigar o evitar esos riesgos.

Desde el Centro de Estudios en Libertad de Expresión (CELE), observamos que la regulación de la inteligencia artificial (IA) se ha consolidado como una temática prioritaria en la agenda regulatoria y política global. En virtud del avance acelerado de propuestas normativas —tanto vinculantes como no vinculantes— en diversas jurisdicciones, consideramos oportuno desarrollar un mapeo que funcione como una herramienta para comprender el estado del arte regulatorio en esta materia. Este ejercicio preliminar no solo permite identificar tendencias y enfoques predominantes, sino también advertir de manera temprana algunas de las posibles implicancias que estas regulaciones podrían tener sobre el ejercicio de derechos fundamentales.

El objetivo de este estudio es realizar un mapeo de iniciativas regulatorias sobre inteligencia artificial e identificar posibles modelos regulatorios de esta tecnología con el objetivo de dilucidar si algunos de estos modelos han sido replicados en América Latina[3]. Si bien el mapeo no pretende ser exhaustivo porque la presentación de iniciativas es constante, sí pretende ser lo más abarcativo posible, para lo cual se observan iniciativas de espacios o foros internacionales, surgidas dentro del sistema de Naciones Unidas, así como iniciativas regionales y nacionales. Cabe destacar que la "mayoría de las políticas y prácticas sobre inteligencia artificial se han desarrollado en y para el Norte Global, lo que ha contribuido a una creciente brecha entre los países que influyen en la gobernanza global de la IA y los que no"[4]. Regiones como América Latina y África sólo han contribuido un 5 y 2 por ciento a la producción de normas respectivamente[5].

---

[2] Kent Walker, "7 Principles for Getting AI Regulation Right," n.d., https://blog.google/outreach-initiatives/public-policy/7-principles-for-getting-ai-regulation-right/

[3] González Mama, M.; Alvarez Ugarte, R., «Modelización regulatoria. Palpitando la influencia de la Digital Services Act en América Latina», Centro de Estudios en Libertad de Expresión (CELE), Artículo de investigación No. 63, Buenos Aires, Argentina, 2025.

[4] Gordon LaForge, Robert Muggah, and Gabriella Seiler, "Bridging the AI Governance Divide" (T20, n.d.), https://www.t20brasil.org/media/documentos/arquivos/TF05_ST_05_Bridging_the_AI_gov66cdcbf06f991.pdf ; Giulia Neaher et al., "'Responsible and Safe AI: A Primer for Policymakers in the Global South.'" (Igarapé Institute, Abril 2024). Disponible en: https://igarape.org.br/wp-content/uploads/2024/04/SD_EN_Responsible-and-Safe-AI.pdf

[5] Gordon LaForge, Robert Muggah, and Gabriella Seiler, "Bridging the AI Governance Divide."

La regulación de la IA se toca con la regulación de otros temas como la desinformación, los procesos electorales, el discurso de odio, la protección de los datos personales, defensa del consumidor, propiedad intelectual, entre otros. Por esta razón, en algunas propuestas regulatorias encontramos la confluencia de distintos enfoques o prismas desde los que se aborda esta tecnología que buscan dar cuenta de esta transversalidad, aún cuando en ocasiones se reconoce que los marcos regulatorios ya existentes pueden ser aplicables y que no es necesario innovar. Además, estos enfoques buscan establecer o reflejar principios éticos, responsabilidades, estándares técnicos, medidas de promoción o desarrollo de la IA, mecanismos de prevención de riesgos, la protección de derechos, entre otros.

En relación a los enfoques basados en riesgos que encontramos en las iniciativas, es importante resaltar que se identificaron propuestas que buscan restringir, mitigar o minimizar los riesgos (en ocasiones también definidos como "sistémicos") que la tecnología puede generar o que el propio sistema de IA pueda implicar (desde su diseño hasta su uso y desarrollo durante el ciclo de vida de la IA), ya sea de manera general o sectorial. En muchos casos, las iniciativas realizan una taxonomía de riesgos o califican a los sistemas de IA según el riesgo que generan, desde aquellos cuyos usos están prohibidos[6] a aquellos permitidos[7] pero que requieren la adopción de medidas y procesos de detección y mitigación de riesgos (como evaluaciones de impacto previas al despliegue del sistema de IA). Si bien en algunos casos los riesgos mencionados están vinculados a la afectación de determinados derechos individuales o colectivos, en otros casos están relacionados con estándares técnicos, infraestructura TIC, el ambiente, o la seguridad[8].

Si bien la IA es una tecnología que lleva largo tiempo desarrollándose, la irrupción de modelos de IA generativa a disposición del público masivo ha llevado a que los gobiernos y organismos internacionales coloquen la cuestión en lo más alto de sus agendas. Como consecuencia, el debate antes centrado en la elaboración de estrategias nacionales para impulsar el desarrollo de

---

[6] Por ejemplo, en el Reglamento Europeo considerando 42 dice: "deben prohibirse las evaluaciones de riesgos realizadas con respecto a personas físicas para evaluar la probabilidad de que cometan un delito o para predecir la comisión de un delito real o potencial basándose únicamente en la elaboración de perfiles de esas personas físicas o la evaluación de los rasgos y características de su personalidad". Ver también artículo 5.

[7] En algunos casos, como la Reglamentación europea sobre la IA estos son clasificados como de alto riesgo.

[8] Por ejemplo, en el Reglamento Europeo considerando 110 dice: "Los modelos de IA de uso general pueden plantear riesgos sistémicos, por ejemplo, cualquier efecto negativo real o razonablemente previsible en relación con accidentes graves, perturbaciones de sectores críticos y consecuencias graves para la salud y la seguridad públicas, cualquier efecto negativo real o razonablemente previsible sobre los procesos democráticos y la seguridad pública y económica o la difusión de contenidos ilícitos, falsos o discriminatorios". Giovanni De Gregorio and Pietro Dunn, "The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age," *Common Market Law Review* 59, no. 2 (2022): 473–500. Disponible en: https://doi.org/10.54648/cola2022032

la IA ha girado—en los últimos años—a la cuestión de cómo regular estas tecnologías y en qué mecanismos de gobernanza establecer para su control.

Las numerosas iniciativas e instrumentos que este informe estudia pretenden marcar el rumbo de futuras regulaciones. A pesar de la fragmentación existente, muchas de éstas iniciativas utilizan un lenguaje similar y hasta coinciden en ciertos criterios comunes, pero en muchos casos difieren en las estrategias y enfoques utilizados. El siguiente relevamiento procura identificar cuáles son las tendencias presentes en los marcos regulatorios propuestos a la fecha y en los modelos de gobernanza que se encuentran en desarrollo. Uno de los objetivos principales de este trabajo es identificar si existen distintas aproximaciones o si—por el contrario—la tendencia regulatoria que de ellas emerge es más o menos homogénea. Según nuestros hallazgos existe una marcada tendencia a la adopción de un enfoque de gobernanza de la IA principalmente basado en los riesgos que ésta puede generar.

1.1. Algunas aclaraciones metodológicas

En esta investigación hemos incorporado todas aquellas regulaciones y proyectos de regulación que explícitamente buscan establecer un marco normativo para el desarrollo, investigación, uso e implementación de la inteligencia artificial. No obstante, es importante aclarar que no es sencillo definir qué es la IA o, más aún, si corresponde llamar de esa manera al fenómeno o grupo de tecnologías que se engloban bajo ese nombre.[9] Aún cuando la utilización del nombre se ha vuelto prácticamente ubicua, no existe una definición universal. No nos enfocaremos aquí en desarrollar una definición propia: esta aclaración metodológica pretende—simplemente—advertir al lector sobre las divergencias e inconsistencias que pueden existir en la definición de este grupo de tecnologías entre un instrumento regulatorio y otro. En algunos casos, las definiciones son más genéricas o vagas mientras que en otros son más específicas e incluyen la mención de tecnologías específicas como el aprendizaje automatizado (o machine learning en inglés), modelos extensos de lenguaje (LLMs por sus siglas en inglés) o las IA de propósito general,[10] y, en ciertos casos, los instrumentos se enfocan en tipos específicos de IA, como la IA generativa o predictiva. Mientras que la primera crea contenido, como textos, imágenes, voz o música, en respuesta a instrucciones humanas, la segunda analiza datos para anticipar eventos futuros y guiar decisiones en seguridad, gestión de inventarios o contratación laboral, entre otros.[11] En consecuencia los problemas asociados

[9] Evgeny Morozov, *The problem with artificial intelligence? It's neither artificial nor intelligent*. Disponible en: https://www.theguardian.com/commentisfree/2023/mar/30/artificial-intelligence-chatgpt-human-mind

[10] Observatory on Information and Democracy (2025), *Information Ecosystems and Troubled Democracy: A Global Synthesis of the State of Knowledge on News Media, AI and Data Governance*, p. 47. Disponible en:: https://observatory.informationdemocracy.org/wp-content/uploads/2024/12/rapport_forum_information_democracy_2025.pdf

[11] Arvind Narayan, Sayash Kapoor (2024), *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't,*

a cada tipo de IA también serán distintos. Algunos de estos son afectaciones a: derechos humanos (particularmente la libertad de expresión y acceso a la información), propiedad intelectual, privacidad y protección de datos personales, acceso al mercado laboral, vigilancia e inteligencia, derechos de migrantes, falta de transparencia, reproducción de estereotipos, desinformación y manipulación, por mencionar unos pocos. Las afectaciones pueden venir tanto del uso que hagan las empresas o usuarios de estas tecnologías como del uso que hagan los Estados. Narayan y Kapoor advierten que en el caso de la IA predictiva las salvaguardas deben ser aún mayores puesto que funciona mucho peor que la IA generativa. Este mal funcionamiento se combina con la potencialidad de afectar gravemente la vida de las personas, por ejemplo, al impedir el acceso a un seguro médico según las predicciones que la tecnología haga.

2. Iniciativas de bloques y foros internacionales

2.1. OCDE: Recomendación sobre Inteligencia Artificial

En 2019, el Consejo de la OCDE adoptó un conjunto de recomendaciones[12] que, si bien no son vinculantes, sentaron ciertos estándares iniciales que terminaron influenciando algunos de los procesos posteriores como el Global Partnership AI (GPAI por sus siglas en inglés), el Proceso de Hiroshima del G7 y los Principios del G20 sobre IA. El documento aplica un enfoque horizontal que incorpora principios orientados a la industria y basados en valores centrados en el ser humano.

En particular, recomienda a los actores involucrados guiarse por los siguientes principios: 1.1 Crecimiento inclusivo, desarrollo sostenible y bienestar; 1.2 Respeto por estado de derecho, los derechos humanos (entre ellos la libertad de expresión) y los valores democráticos, incluida la equidad y la privacidad[13]; 1.3 Transparencia y aplicabilidad (brindar información que proporcione un entendimiento general de los sistemas y sus resultados, que permita conocer que se interactúa con un sistema de IA e impugnar sus resultados); 1.4 Robustez y

---

and How to Tell the Difference, Princeton University Press.

[12] Organización para la Cooperación y el Desarrollo Económicos. Recomendación del Consejo sobre Inteligencia Artificial, adoptado el 21 de mayo de 2019 y enmendado el 7 de noviembre de 2023. Disponible en: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Adhirieron los 34 Estados Miembro de la OCDE y 10 Estados más, entre los cuales se encuentran Argentina, Brasil, Colombia, Costa Rica y Perú.

[13] Cabe destacar que si bien el principio 1.2a) establece que los actores de la IA deben abordar los problemas asociados con la desinformación amplificada por la IA, tienen que hacerlo respetando al mismo tiempo la libertad de expresión.

seguridad (trazabilidad y gestión de riesgos); y 1.5 Rendición de cuentas[14]. Asimismo, recomienda a los gobiernos la inversión en investigación y desarrollo, la adopción de mecanismos para la conformación de un entorno político propicio para la IA, el fomento de capacidades y de la preparación necesaria para la transformación del mercado laboral, así como también la cooperación internacional. Otro aspecto relevante es que la definición sobre IA que incluye, revisada en 2023[15], fue luego replicada con algunos cambios en otros documentos, como el Reglamento de Inteligencia Artificial de la Unión Europea[16].

Por otra parte, como mecanismos de seguimiento, encomendó al Comité de Políticas de Economía Digital de la OCDE la creación de guías de implementación y la promoción de un diálogo interdisciplinario entre las múltiples partes interesadas. Asimismo, se creó el Observatorio de Políticas de IA (OECD.AI) y una red de expertos en IA (ONE AI).

El documento de recomendación sobre inteligencia artificial fue actualizado en 2024[17], después de que el Comité de Política Digital emitiera un reporte de evaluación de las recomendaciones[18]. El reporte concluyó que era necesario actualizar la recomendación con el fin de aclarar el contenido de algunas disposiciones de modo que se facilite su aplicación y aumente su pertinencia, a la vez que refleje los avances tecnológicos importantes. Como actualizaciones específicas requeridas el reporte incluyó: i) abordar acciones respecto a la desinformación priorizando salvaguardar la integridad de la información en el contexto de la IA generativa; ii) abordar los usos no previstos, el uso indebido intencionado y el uso

[14] Ver 1.5.c). En particular deben aplicar un enfoque sistemático de gestión de riesgos en cada fase del ciclo de vida del sistema de IA de manera continua y adoptar una conducta empresarial responsable para abordar los riesgos relacionados con los sistemas de IA. Los riesgos incluyen aquellos relacionados con sesgos perjudiciales, los derechos humanos —incluida la seguridad, la protección y la privacidad—, así como los derechos laborales y de propiedad intelectual

[15] La definición enmendada estipula lo siguiente: An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

[16] Reglamento de Inteligencia Artificial de la Unión Europea, Art. 3.1. sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales.

[17] Organización para la Cooperación y el Desarrollo Económicos, Recomendación del Consejo sobre Inteligencia Artificial, adoptado el 21 de mayo de 2019 y enmendado el 3 de mayo de 2024. Disponible en: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Adhirieron los 34 Estados Miembro de la OCDE y 10 Estados más, entre los cuales se encuentran Argentina, Brasil, Colombia, Costa Rica y Perú.

[18] Organización para la Cooperación y el Desarrollo Económicos. Reporte de implementación de la recomendación de la OCDE sobre inteligencia artificial, emitido el 24 de abril de 2024. Disponible en: https://one.oecd.org/document/C/MIN(2024)17/en/pdf

indebido no intencionado. iii) clarificar qué la información deben facilitar los actores de IA sobre los sistemas para garantizar la transparencia y la divulgación responsable iii); iv) esbozar mecanismos para hacer frente a posibles daños o comportamiento no deseado a lo largo del ciclo de vida de la IA; v) promover una conducta empresarial responsable durante todo el ciclo de vida de la IA[19].

Si bien la Recomendación de la OCDE fue vista como un avance importante hacia el desarrollo y uso responsable de la IA, la falta de carácter normativo o vinculante para los Estados adherentes deja ciertos aspectos relacionados con la protección de los derechos humanos como compromisos voluntarios, cuando en efecto se trata de normas legales de cumplimiento obligatorio.[20]

2.2. G20: Principios sobre IA

En junio de 2019, los Ministros de Comercio y Economía Digital del G20 adoptaron una serie de principios no vinculantes[21] que recogen los Principios de la OCDE, aunque sin incluir una definición de IA y sin mencionar explícitamente la protección a la libertad de expresión. Cabe destacar que contaron con el apoyo de China y Rusia entre sus miembros, a diferencia del caso de la OCDE donde dichos países no participan. En la cumbre del G20 de 2023 en Nueva Delhi, los Estados reafirmaron su compromiso con tales principios y manifestaron que perseguirán un enfoque regulatorio y de gobernanza pro innovación que maximice los beneficios y tome en consideración los riesgos asociados al uso de la IA.

En noviembre de 2024, en el marco de la 10° cumbre de presidentes de los parlamentos del G20 realizada en Brasil, los jefes parlamentarios de los países del G20, a excepción de Argentina que se desvinculó del documento, adoptaron por consenso la declaración conjunta sobre "Parlamentos por un mundo justo y un planeta sostenible".

La declaración destaca el compromiso de trabajar para "promover la cooperación internacional en la creación de capacidades y seguir debatiendo sobre la IA para el desarrollo sostenible inclusivo y la reducción de las desigualdades."[22] Además resalta la necesidad de que

---

[19] Organización para la Cooperación y el Desarrollo Económicos. Reporte de implementación de la recomendación de la OCDE sobre inteligencia artificial, emitido el 24 de abril de 2024. Disponible en: https://one.oecd.org/document/C/MIN(2024)17/en/pdf

[20] En este sentido, ver Human Rights Watch, Rules for a New Surveillance Reality, 18 de noviembre de 2019. Disponible en: https://www.hrw.org/news/2019/11/18/rules-new-surveillance-reality

[21] G20 AI Principles, Anexo a la Declaración de Ministros de Comercio y Economía Digital del 9 de junio de 2019. Disponible en https://www.mofa.go.jp/files/000486596.pdf

[22] G20, Declaración Conjunta sobre Parlamentos por un mundo justo y un planeta sostenible. 10a Cumbre de Presidentes de los Parlamentos del G20 (P20). Brasilia, del 6 al 8 de noviembre de 2024. Pag 7. Disponible en: https://www.congressonacional.leg.br/documents/143854729/144591198/decesp.pdf/eaf2a720-c9bd-494c-a0f0-f7d12a8a0422

los países se preparen para la transformación digital, entendiendo la importancia de la economía digital, y, para ello, es necesario abordar la brecha digital, tecnológica y de la IA.

También anima a los gobiernos a continuar desarrollando normas internacionales que permitan el fortalecimiento de la economía digital y garantizando los derechos humanos. A la vez que los firmantes pretenden "tomar las medidas adecuadas para aprovechar una IA segura y fiable mediante un enfoque transparente, inclusivo, ético, responsable, digno de confianza y centrado en el ser humano."[23]

2.3. G7: Proceso de Hiroshima

En mayo de 2023, el G7 lanzó una iniciativa destinada a propiciar la discusión en torno a la gobernanza global de la IA, en particular, de la IA generativa. Luego de siete meses de trabajo, en diciembre de ese año fue adoptado un marco global de políticas que incluye un conjunto de principios[24] destinado a todos los actores y un código de conducta[25] para organizaciones que desarrollen sistemas avanzados de IA. Si bien los países del G7 mantienen la voluntad de promover los productos del Hiroshima Process, aún no acordaron un mecanismo de monitoreo y cumplimiento. El documento menciona que los Estados deben cumplir con sus obligaciones en virtud del derecho internacional de los derechos humanos para promover el pleno respeto y la protección de los derechos humanos, mientras que las actividades del sector privado deben alinearse con marcos internacionales como los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos y las Directrices de la OCDE para Empresas Multinacionales. Sin embargo no aborda de manera detallada cómo interactúan estas obligaciones con los principios guía que luego propone.

Entre ellos encontramos: 1) identificar, evaluar y mitigar riesgos (trazabilidad y pruebas internas y externas, como red-teaming adversarial) antes de su implementación y durante el ciclo de vida de la IA; 2) identificar y mitigar vulnerabilidades, incidentes y usos indebidos una vez implementados; 3) contar con mecanismos de reporte destinado a terceras partes y usuarios y mantener documentación sobre los riesgos e incidentes; 4) realizar reportes públicos sobre la IA avanzada e implementar medidas de transparencia que contribuyan a la rendición de cuentas y el intercambio de información sobre incidentes; 5) desarrollar políticas de gobernanza y gestión de riesgos con un enfoque basado en riesgos; 6) invertir e

---

[23] G20, Declaración Conjunta sobre Parlamentos por un mundo justo y un planeta sostenible. 10a Cumbre de Presidentes de los Parlamentos del G20 (P20). Brasilia, del 6 al 8 de noviembre de 2024. Pag 8. Disponible en: https://www.congressonacional.leg.br/documents/143854729/144591198/decesp.pdf/eaf2a720-c9bd-494c-a0f0-f7d12a8a0422

[24] G7, Hiroshima Process, International Guiding Principles for Organizations Developing Advanced AI Systems. Disponible en https://www.mofa.go.jp/files/100573471.pdf

[25] G7, Hiroshima Process, International Code of Conduct for Organizations Developing Advanced AI Systems. Disponible en https://www.mofa.go.jp/files/100573473.pdf

implementar sistemas robustos de seguridad, 7) desarrollar e implementar medidas de autenticación de contenidos, como marcas de agua u otras técnicas que permitan identificar el contenido generado por IA; 8) priorizar la investigación e inversión destinada a mitigar los riesgos; 9) priorizar el desarrollo de IA avanzadas para abordar los desafíos relacionados con la crisis climática, la salud global y la educación, 10) avanzar en la adopción de estándares técnicos internacionales; y 11) adoptar medidas para la protección de los datos personales y la propiedad intelectual.

Por su parte, el Código de Conducta avanza un paso más en relación con los principios y ofrece mayores detalles destinados a su implementación por parte de todo tipo de organizaciones que desarrollen o implementen sistemas avanzados de IA. Entre otros aspectos, propone un listado de ciertos riesgos a los que los desarrolladores se comprometerían a dedicar especial atención,[26] así como medidas mínimas de transparencia y especificaciones de ciberseguridad.

Para dar continuidad al proceso de Hiroshima, durante el año 2024, la presidencia del G7 en cabeza de Italia, abrió un espacio de reporte de informes, "con el apoyo de la OCDE y la contribución de 20 organizaciones de diez países con una verdadera perspectiva de las partes interesadas[27]."

Paralelamente al reporte, se desarrolló una caja de herramientas para la inteligencia artificial en el sector público.[28] El documento fue desarrollado con el apoyo de la OCDE y UNESCO con base en reporte de países del G7 y pasa revista sobre las estrategias nacionales de IA de Canadá, Unión Europea, Estados Unidos, Francia, Italia, Alemania, Japón y el Reino Unido. Dentro de los puntos más importantes, cabe destacar los principales desafíos identificados: la integración de infraestructuras existentes; los dilemas éticos asociados principalmente a los sesgos de los sistemas; la protección de la privacidad y seguridad de los datos; la resistencia al cambio organizacional; la falta de habilidades y adecuadas capacitaciones a los funcionarios públicos; la limitación de recursos técnicos y financieros; y, un marco regulatorio y ético que garantice la transparencia y responsabilidad.

---

[26] Riesgos químicos, biológicos, radiológicos y nucleares; capacidades ofensivas cibernéticas; riesgos a la salud o seguridad; capacidades de auto replicación o de entrenar otros modelos; riegos sociales como sesgos discriminatorios o violaciones de la privacidad; amenaza a los valores democráticos y derechos humanos, como facilitar la desinformación o dañar la privacidad; riesgos de que un evento particular pueda llevar a una reacción en cadena con efectos negativos en ciudades, actividades o comunidades enteras.

[27] G7. Presidency of the Council of Ministers. The G7 Tech and Digital Ministerial Meeting to be held on 15 October in Cernobbio. 14 de octubre de 2024. Disponible en: https://www.g7italy.it/en/the-g7-tech-and-digital-ministerial-meeting-to-be-held-on-15-october-in-cernobbio/

[28] G7 Toolkit For Artificial Intelligence In The Public Sector. Disponible en: https://www.oecd.org/en/publications/g7-toolkit-for-artificial-intelligence-in-the-public-sector_421c1244-en.html

Adicionalmente, la caja de herramientas incluye recomendaciones de buenas prácticas en el uso de IA, como el establecimiento de estructuras de gobernanza para gestionar y supervisar los sistemas de IA; la implementación de mecanismos de seguimiento de los sistemas; fomentar un entorno propicio para la implementación de IA; la incorporación de consideraciones éticas; el desarrollo de capacidades de los funcionarios públicos; establecer marcos de identificación, mitigación y gestión de riesgos; y promover la transparencia y responsabilidad en el uso de IA.

En ninguno de los documentos del G7 encontramos mención expresa a la protección del derecho a la libertad de expresión.

3. Iniciativas en el marco de la Organización de las Naciones Unidas

3.1. UNESCO: Recomendación sobre la ética de la IA

En noviembre de 2021 la Conferencia General de la UNESCO adoptó un conjunto de recomendaciones en materia de ética de la IA[29], con el propósito de brindar un marco universal de valores y principios para orientar a los Estados en la formulación de leyes y políticas que promuevan la fiabilidad e integridad de los sistemas de IA y prevenir daños durante todo su ciclo de vida así como las responsabilidad de los actores de la IA.[30] El documento adopta una definición de IA que comprende a aquellos sistemas capaces de procesar datos e información de una manera que se asemeja a un comportamiento inteligente y abarca, generalmente, aspectos de razonamiento, aprendizaje, percepción, predicción, planificación o control. Luego, propone como valores transversales a los derechos humanos, libertades fundamentales (entre ellas la libertad de expresión) y la dignidad humana; la prosperidad del medio ambiente y los ecosistemas; la diversidad y la inclusión; y la vida en sociedades pacíficas, justas e interconectadas.

A su vez, incluye los siguientes principios éticos: 1) proporcionalidad e inocuidad, 2) seguridad y protección; 3) equidad y no discriminación; 4) sostenibilidad; 5) intimidad y protección de datos; 6) supervisión y decisión humanas, 7) transparencia y explicabilidad, 8) responsabilidad y rendición de cuentas; y 9) gobernanza y colaboración adaptativas y de múltiples partes interesadas.

Entre las propuestas más destacables, el documento encomienda crear procedimientos de evaluación y mitigación de riesgos, que incorporen un análisis de proporcionalidad del uso de

[29] UNESCO, Recomendación sobre la ética de la inteligencia artificial, adoptada el 23 de noviembre de 2021. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

[30] Ibíd. 2.b) "Los actores de la IA pueden definirse como todo actor que participe en al menos una etapa del ciclo de vida del sistema de IA y pueden ser tanto personas físicas como jurídicas, por ejemplo, investigadores, programadores, ingenieros, especialistas en datos, usuarios finales, empresas, universidades y entidades públicas y privadas, entre otros".

la IA de acuerdo al contexto y basado en fundamentos científicos rigurosos. Asimismo, sostiene que cuando las decisiones tengan un impacto irreversible o difícil de revertir o que puedan implicar decisiones de vida o muerte, la decisión final siempre debería ser adoptada por un ser humano. También establece que los sistemas no deben ser utilizados con fines de calificación social o de vigilancia masiva. El documento prevé mecanismos de cooperación internacional en favor de los países menos avanzados en materia de IA; la implementación de marcos adecuados de protección y gobernanza de datos, la incorporación de evaluaciones de impacto en la privacidad; la posibilidad de que las decisiones o resultados siempre puedan ser atribuibles a una persona humana o jurídica existente; el deber de proporcionar información para que las personas tomen conocimiento de que una decisión fue tomada con base en IA y puedan solicitar explicaciones a las instituciones del sector público; y la auditabilidad y trazabilidad necesarias para la supervisión, evaluación de impacto y debida diligencia.

En una segunda sección del documento, se proponen diferentes acciones y aspectos a considerar para once ámbitos de actuación, relativos al campo de la gobernanza[31], las políticas de datos, la cultura, el empleo y la salud, entre otros. Menciona la que los Estados deberían establecer marcos de evaluación de impactos éticos y que, junto a las empresas del sector privado, deberían desarrollar mecanismos de debida diligencia para " determinar, prevenir y atenuar los riesgos y rendir cuentas de la forma en que abordan el impacto de los sistemas de IA en el respeto de los derechos humanos, el estado de derecho y las sociedades inclusivas".[32]

Al contrario de otros instrumentos, trata cuestiones como la necesidad de abordar el impacto ambiental e incentivar métodos más eficientes en el uso de los recursos; promover conjuntos de datos abiertos, que sean diversos, sólidos y de calidad; incluir una perspectiva transversal de género y el fomento de la participación de las mujeres en la investigación de la IA. También propone evitar el riesgo de concentración en el mercado cultural, promover contenidos locales e investigar las implicancias de la IA en los derechos de propiedad de intelectual; asegurar que el uso de la IA en el aprendizaje esté sujeta a requisitos estrictos, sobre todo si se utiliza para la supervisión, evaluación de capacidades o predicción de comportamientos. El documento propone que los Estados deberían utilizar los sistemas de IA para mejorar el acceso a la información y el conocimiento, con el fin de mejorar la libertad de expresión, las libertades académicas y científicas y el acceso a la información, así como aumentar la divulgación proactiva de los datos y la información oficiales. También tienen que garantizar la libertad de expresión y acceso a la información en relación con la generación, moderación y conservación automática de contenidos (evitando diseminar desinformación) y la transparencia de los operadores.

[31] Según las recomendaciones los mecanismos de gobernanza deberían ser inclusivos, transparentes, multidisciplinarios y multilaterales. Ver 54.
[32] Ibid. 50 y 51.

Como mecanismos de seguimiento y evaluación de la recomendación contempla la creación de dos herramientas: una Metodología de Evaluación del Estadio de Preparación[33] y otra Metodología de Evaluación de Impacto Ético,[34] ambas lanzadas en 2023. Posteriormente, en el 2do Foro Mundial, realizado en febrero de 2024 en Eslovenia, fueron lanzados el Observatorio Mundial de Ética de la IA y la Red UNESCO de Expertos en Ética de la IA sin Fronteras, iniciativas destinadas a funcionar como *hub* de colaboración y cooperación para la implementación de las recomendaciones. Asimismo, se acordó la realización de asistencias técnicas focalizadas para trabajar en conjunto con ciertos gobiernos en la incorporación de las recomendaciones en sus legislaciones y políticas para la IA, incluyendo en esta primera etapa a Chile y Brasil, entre otros países.

Existen otras dos iniciativas de la UNESCO que tienen como objetivo implementar y fortalecer la Recomendación sobre la ética de la IA. La primera es la plataforma Women For Ethical IA, que reúne a "17 expertas del mundo académico, la sociedad civil, el sector privado y los organismos reguladores de todo el mundo que compartirán investigaciones y contribuirán a un repositorio de buenas prácticas. La plataforma impulsará avances en algoritmos y fuentes de datos no discriminatorios, e incentivará a niñas, mujeres y grupos infrarrepresentados a participar en la IA.[35]"

La segunda es el Consejo Empresarial para la Ética de la IA que promueve la colaboración entre la UNESCO y empresas que operan en América Latina en relación a tecnologías de inteligencia artificial. "El Consejo sirve de plataforma para que las empresas se reúnan, intercambien experiencias y promuevan prácticas éticas en la industria de la IA.[36]"

A diferencia de otros instrumentos no vinculantes, las Recomendaciones de la UNESCO recogen la necesidad de adoptar un enfoque inclusivo, que no reemplace sino que construya sobre las normas existes en materia de derechos humanos, al mismo tiempo que promueva

---

[33] UNESCO, Metodología de evaluación del estadio de preparación: una herramienta de la Recomendación sobre la Ética de la Inteligencia Artificial, 2023. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000385198_spa

[34] UNESCO, Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence, 2023. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000386276?posInSet=1&queryId=24995b60-ff25-4317-bc0c-9525ea1d9431
Este instrumento está diseñado principalmente para ayudar a los funcionarios gubernamentales involucrados en la contratación de sistemas de inteligencia artificial.

[35] Unesco. 2024. Plataforma de expertas 'Women4Ethical AI' para avanzar en la igualdad de género. Ética de la IA. Disponible en: https://www.unesco.org/es/artificial-intelligence/recommendation-ethics?hub=32618

[36] Unesco. 2024. Consejo Empresarial para la Ética de la IA. Disponible en: https://www.unesco.org/es/artificial-intelligence/recommendation-ethics?hub=32618

una participación amplia en la gobernanza y evaluación de los impactos de la IA.[37] Es un producto enmarcado dentro del mandato de la UNESCO y que se concentra en los riesgos éticos dentro del ámbito de su competencia (educación, ciencia, cultura, comunicación e información). En ese sentido, la efectividad e implementación de las recomendaciones continuará dependiendo de la voluntad de los Estados por aplicarlas, así como de la voluntad de los actores del sector privado por apoyar este proceso.[38]

3.2. Secretaría General: Informe del Órgano Asesor de IA

En la Hoja de Ruta para la Cooperación Digital[39,] propuesta por el Secretario General de la Naciones Unidas en junio de 2020, fueron identificados como desafíos pendientes la falta de representación e inclusión de los países menos desarrollados en las discusiones globales en materia de IA, la falta de coordinación por fuera de los grupos tradicionales y la brecha existente para que el sector público se beneficie de la supervisión o gobernanza de estas tecnologías.

Entre los objetivos planteados se propuso la creación de un cuerpo asesor de múltiples partes interesadas[40], el que fue conformado en octubre de 2023[41], con el fin de desarrollar conocimiento a nivel global, abordar la falta de representación y promover una plataforma de liderazgo en materia de IA. En diciembre, el grupo publicó un informe provisional[42] titulado "Governing AI for Humanity", planteando los déficits en torno a la gobernanza global de la IA, reconociendo que el Sur Global ha estado en mayor medida ausente en las discusiones, e identifica los datos, la computación y el talento como habilitadores clave para su desarrollo equitativo, debiendo ser acompañados por la gobernanza como valla de seguridad.

---

[37] En este sentido, ver European Center for Non Profit Law, "The historical text incorporates critical points raised by civil society", 17 de diciembre de 2021. https://ecnl.org/news/unesco-adopts-first-ever-global-agreement-ethics-ai. Disponible en: https://ecnl.org/news/unesco-adopts-first-ever-global-agreement-ethics-ai

[38] Ocho empresas del sector privado adhirieron a los principios de la Recomendación de la UNESCO y se comprometieron a apoyar su implementación. Ver Declaración de Sector Privado, UNESCO Global Forum on the Ethics of AI, 6 de febrero de 2024. Disponible en: https://www.unesco.org/sites/default/files/medias/fichiers/2024/02/gfeai_private_sector_05022024.pdf

[39] Secretaria General de las Naciones Unidas, Hoja de ruta para la cooperación digital, A/74/821, 29 de mayo de 2020. Disponible en: https://documents.un.org/doc/undoc/gen/n20/102/54/pdf/n2010254.pdf?token=xjPsSetF0K21wVWjz0&fe=true

[40]Véase: https://www.un.org/en/ai-advisory-body

[41]Véase: https://www.un.org/sites/un2.un.org/files/organo-asesor-ia_prensa_final.pdf

[42] Advisory Body on Artificial Intelligence, Interim Report: Governing AI for Humanity, diciembre de 2023. Disponible en: https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf

El informe provisional propuso abordar los riesgos que supone la IA a través de una perspectiva de las vulnerabilidades existentes o potenciales (sobre las comunidades vulnerables y los bienes comunes), ofreciendo una categorización de aquellos riesgos identificados y señalando que un marco para la evaluación del riesgo será desarrollado en una segunda etapa.

Frente a la falta de consenso sobre cómo abordar estos riesgos, el grupo de expertos sostuvo que, al igual que con el principio precautorio en materia ambiental, la falta de certezas científicas no puede derivar en una parálisis de la gobernanza de la IA. En tanto el Órgano Asesor fue encomendado a ofrecer opciones para la gobernanza internacional de la IA, en vez de ofrecer un solo modelo, las recomendaciones preliminares se basan en proporcionar principios para la formación de nuevas instituciones y las funciones que necesitan llevarse adelante, ya sea en cabeza de una única institución o de una red de ellas.

Como principios, propone una 1) gobernanza inclusiva, por y para el beneficio de todos; 2) en función del interés público; 3) construida en conjunto con la gobernanza de datos y la promoción de datos como bien común; 4) universal, conectada y arraigada en la colaboración adaptativa entre múltiples partes interesadas; y 5) anclada en la Carta de la Naciones Unidas, el derecho internacional de los derechos humanos y otros compromisos, como los ODS.

Además, el reporte menciona siete funciones institucionales (que van de soft a hard) que deberían involucrar la gobernanza internacional de la IA. A nivel global, las organizaciones internacionales, los gobiernos y el sector privado tendrían la responsabilidad principal de estas funciones mientras que la sociedad civil, la academia y científicos independientes desempeñarían un papel fundamental en la construcción de evidencia para las políticas, la evaluación del impacto y la rendición de cuentas de los actores clave durante la implementación. Cada conjunto de funciones tendría diferentes loci de responsabilidad en los diferentes niveles de gobernanza: sector privado, gobierno y organizaciones internacionales. Las siete funciones propuestas son:1) evaluar regularmente la orientación e implicancias futuras de la IA - inclusive a través de la creación de un observatorio internacional de análisis- ; 2) reforzar la interoperabilidad de los esfuerzos de gobernanza alrededor del mundo basados en normas internacionales a través de un marco global de gobernanza respaldado por la ONU (por ejemplo, este marco podría promover la realización de evaluaciones de impacto en derechos humanos); 3) desarrollar y armonizar estándares y marcos de gestión de riesgos; 4) facilitar el desarrollo, implementación y uso de la IA para el beneficio social y económico a través de la cooperación; 5) promover la colaboración internacional en materia de desarrollo de talentos, acceso a la infraestructura computacional y la construcción de datasets diversos y de alta calidad, modelos de código abierto responsables y bienes públicos de IA para el desarrollo sostenible; 6) monitorear riesgos, reportar incidentes y coordinar respuestas de emergencia; y 7) contar con normas que aseguren el cumplimiento y la rendición de cuentas.

El Reporte Interino del Órgano Asesor recibió comentarios que alertaron sobre el peligro de caer en un enfoque que trate a la IA como un actor o entidad única, en vez de un conjunto diverso de tecnologías, y de favorecer los intereses de las grandes empresas tecnológicas a costa de sacrificar la diversidad y la innovación en la gobernanza de la IA.[43] El ex-Relator para la Libertad de Expresión, David Kaye, resaltó de manera positiva el rol propuesto para la ONU en la gobernanza global de la IA y su anclaje en las normas e instituciones DDHH del sistema pero destacó como aspectos a mejorar en el reporte final, la necesidad de analizar el riesgo de que la interoperabilidad de las regulaciones favorezca a las empresas transnacionales en perjuicio de la protección de los DDHH y así como cuestiones sobre competencia y antimonopolio, a pesar de que el reporte interino había identificado la concentración económica como una de las principales problemáticas.[44]

La versión final del informe "Gobernanza de la IA en beneficio de la humanidad[45]" fue lanzada en septiembre de 2024 y recomienda: la creación de una arquitectura global e inclusiva, basada en la cooperación internacional para la gobernanza de IA; la creación de una oficina de IA dentro de la Secretaría General, la conformación de un panel experto científico internacional; gestionar reuniones intergubernamentales sobre el futuro de la IA, que se enfoquen en discutir estándares y reglas de interoperabilidad; el establecimiento de una entidad para el intercambio de definiciones y normas aplicables a la IA; el desarrollo de un marco global de datos sobre la IA (incluyendo normas comunes, acuerdos modelo para los flujos transfronterizos y fideicomisos de datos); la creación de una red cuyo fin sea fomentar capacidades y un fondo mundial para apoyar el despliegue de la IA en países con menos recursos; implementar un enfoque ágil, inclusivo y eficaz en la gobernanza de IA, que refleje la rápida evolución tecnológica.

El informe final recoge los comentarios de la versión preliminar y reconoce a la IA como un conjunto de tecnologías con diversas funcionalidades. Destaca que estas tecnologías tienen un potencial de oportunidades que debe aprovecharse y que requiere de una regulación que permita distribuir sus beneficios de manera equitativa, ya que, la brecha digital podría limitar el abanico de oportunidades que representa. Además, reconoce que conlleva otros riesgos como los sesgos, la vigilancia, el aumento de la creación y difusión de la desinformación, daño a la integridad de la información, afectaciones a la libertad de expresión mediante burbujas de

[43] Joana Bryson, "AI Is Not a Unitary Actor: My Response to the UN Interim Report Consultation", 2 de abril de 2024. Disponible en: https://joanna-bryson.blogspot.com/2024/04/ai-is-not-unitary-actor-my-response-to.html

[44] Ver David Kaye, "Evaluating the UN AI Advisory Body Interim Report", Tech Policy Press, 14 de febrero de 2024. Disponible en https://www.techpolicy.press/evaluating-the-un-ai-advisory-body-interim-report/

[45] Órgano Asesor de Alto Nivel sobre Inteligencia Artificial. Secretaría General de Naciones Unidas (2024) Gobernanza de la Inteligencia Artificial en beneficio de la Humanidad: Informe final. Disponible en: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_es.pdf

información, la manipulación e interferencia extranjera sobre los procesos democráticos,[46] las "alucinaciones" de los grandes modelos lingüísticos, los riesgos a la paz y seguridad y el consumo de energía de los sistemas en tiempos de crisis climática. Asimismo, resalta también que las armas autónomas y los usos de la IA con fines de seguridad pública generan importantes desafíos de naturaleza jurídica y humanitaria.

En línea con los principios esbozados en su versión preliminar, el informe parte 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. Asimismo, reconoce el déficit existente relativo a la gobernanza global de la IA y la dependencia de la voluntariedad en cuanto a la implementación de estándares. En términos de representación, el informe analiza la participación de los Estados en las siete iniciativas más destacadas sobre la gobernanza global de la IA e identifica que muchas regiones del mundo quedaron fuera de las conversaciones. Son 7 los países que participaron en la totalidad de ellas, mientras que 118 países no participaron en ninguna de ellas, los cuales corresponden mayoritariamente al Sur Global.

Concluyendo con un llamado de acción, el informe destaca el rol que puede tener la Organización de Naciones Unidas como "vehículo de un nuevo contrato social para la IA que garantice la aceptación mundial de un régimen de gobernanza.[47]"

En respuesta a la publicación del informe final, Global Partners Digital (GPD) publicó un análisis[48] que resalta de manera positiva el planteamiento de un marco coherente para la gobernanza global de la inteligencia artificial (IA), abordando sus desafíos y oportunidades. No obstante, según GPD, el informe final se desvía de la aproximación original de "la forma sigue a la función" adoptado en el informe interino, centrándose en llenar vacíos existentes en lugar de analizar las funciones específicas necesarias para proporcionar una gobernanza sólida de la IA y establecer una hoja de ruta sobre cuándo y cómo avanzar en su implementación. Además, GPD advierte que la coordinación normativa, crucial para la armonización basada en derechos humanos, recibe menos énfasis en el informe final, lo que podría debilitar la rendición de cuentas en la gobernanza de la IA. Además, la centralización en una sola oficina,

[46] p. 36. Ver también:"Las ultrafalsifcaciones, 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.", p. 40.

[47] Órgano Asesor de Alto Nivel sobre Inteligencia Artificial. Secretaría General de Naciones Unidas (2024) Gobernanza de la Inteligencia Artificial en beneficio de la Humanidad: Informe final. Disponible en: https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_es.pdf

[48] Global Partners Digital (2024) The Final Report of HLAB-AI: our analysis and thoughts. Disponible en: https://www.gp-digital.org/news/the-final-report-of-hlab-ai-our-analysis-and-thoughts/

tal como se propone en el documento, podría obstaculizar el involucramiento no gubernamental, ya que un enfoque descentralizado permite más participación.

El informe destaca la necesidad de estrategias cohesivas e inclusivas, sus recomendaciones requieren una mayor alineación con los marcos de derechos humanos y una mayor refinación para asegurar una cooperación internacional efectiva y transparente.

Artículo 19 consideró que el informe final identificó "un enfoque general adecuado para la gobernanza de la IA, pero perdió varias oportunidades para implementarlo de manera efectiva, lo que pone en peligro la protección de los derechos humanos y la libertad de expresión."[49] En resumen, las organizaciones critican que si bien se menciona la importancia de los derechos humanos, no hay una debida ni suficiente adopción de principios del derecho internacional de los derechos humanos, lo cuál es problemático en el marco de un contexto (reconocido en el informe) de dependencia de la voluntad de los Estados para implementar los estándares de gobernanza.

3.3. Resolución 78/265 de la Asamblea General

El 21 de marzo de 2024, la Asamblea General de las Naciones Unidas (AGNU) aprobó la Resolución 78/265 bajo el título: "Aprovechar las oportunidades de una IA segura y confiable para un desarrollo sostenible"[50]. Además de ser la primera resolución en la materia, patrocinada por EEUU y co-patrocinada por otros 123 países, incluida China, es destacable que también fue aprobada por consenso, sin votación, por la totalidad de los Estados que forman parte de la ONU.

En primer lugar, reconoce a la IA como un ámbito en evolución y la necesidad de que los enfoques en materia de gobernanza de la IA estén basados en el derecho internacional, sean interoperables, ágiles, adaptables e inclusivos y que respondan a las necesidades de los países desarrollados y en desarrollo por igual.

Entre las medidas para promover la cooperación internacional, promueve una mayor participación de los países en desarrollo y la construcción de capacidades, intercambio de conocimientos y transferencia de tecnología. También estipula que los Estados deben abstenerse o dejar de usar sistemas de IA que sean imposibles de operar sin proteger los derechos humanos o que supongan riesgos indebidos para éstos.

Para el desarrollo de sistemas seguros, protegidos y fiables, se encomienda identificar, prevenir y mitigar las vulnerabilidades y los riesgos; incorporar mecanismos de retroalimentación;

---

[49] Article 19 (2025). UN: AI governance and safeguarding freedom of expression. Disponible en: https://www.article19.org/resources/un-ai-governance-and-safeguarding-freedom-of-expression/

[50] Asamblea General de las Naciones Unidas, Resolución A/RES/78/265, 21 de marzo de 2024. Disponible en: https://digitallibrary.un.org/record/4043244/files/A_RES_78_265-ES.pdf?ln=en

promover la alfabetización en IA; aplicar mecanismos de seguimiento, gestión de riesgos y evaluaciones de impacto, incorporar herramientas técnicas de autenticación del contenido, elaborar marcos internacionales para el entrenamiento y puesta a prueba mediante conjuntos de datos, que mitiguen los sesgos y combatan la discriminación; salvaguardar los derechos de propiedad intelectual y de autor; y proteger la privacidad y los datos personales. Asimismo, contempla la elaboración de evaluaciones de riesgos e impacto para proteger los derechos humanos.

Respecto de los sistemas de decisiones automatizadas o que apoyen decisiones que afecten a los usuarios finales, alienta a promover su transparencia, previsibilidad, fiabilidad y comprensión mediante explicaciones y herramientas de supervisión humana, y medios eficaces de reparación y rendición de cuentas.

Como elemento novedoso que no es replicado en otros documentos o instrumentos, la resolución incorpora consideraciones específicas respecto de los modelos lingüísticos de gran tamaño (LLM) y la necesidad de proteger y preservar la diversidad lingüística y cultural, teniendo en cuenta el multilingüismo en los datos que se utilizan.

Si bien desde la sociedad civil la Resolución fue recibida positivamente por su lenguaje operativo respecto de la abstención y cese del uso de sistemas de IA que vulneren los DDHH, también encontró críticas por no incorporar el uso de sistemas en el ámbito militar. Además, las implicaciones para los DDHH fueron brevemente mencionadas y los compromisos necesarios para cerrar la brecha digital tampoco fueron suficientemente contundentes. Otra crítica destacó que la resolución no fue desarrollada mediante los procesos habituales de colaboración y dentro comités de la AGNU, no refleja un verdadero modelo de múltiples partes interesadas. Como contrapartida la sociedad civil propuso desarrollar una próxima resolución sobre IA en el contexto del sistema de derechos humanos de la ONU.[51]

En julio de 2024, en el marco del septuagésimo noveno período de sesiones de la AGNU, fue publicado el informe temático sobre Inteligencia artificial y solidaridad internacional – hacia la solidaridad internacional en el ámbito de la inteligencia artificial centrada en el ser humano desde el diseño - de la Experta Independiente sobre los derechos humanos y la solidaridad internacional, Cecilia M. Bailliet.[52]

---

[51] Declaración Conjunta de OSC, "Encontrar lo bueno en la primera resolución de la Asamblea General de la ONU sobre inteligencia artificial", 10 de abril de 2024. Disponible en https://www.derechosdigitales.org/23508/encontrar-lo-bueno-en-la-primera-resolucion-de-la-asamblea-general-de-la-onu-sobre-inteligencia-artificial/

[52] Bailliet, C. Inteligencia artificial y solidaridad internacional – hacia la solidaridad internacional en el ámbito de la inteligencia artificial centrada en el ser humano desde el diseño. Informe Temático A/79/170. Disponible en: https://www.ohchr.org/es/documents/thematic-reports/a79170-artificial-intelligence-and-international-solidarity-towards

El informe resalta como criterios de solidaridad de la IA la transparencia, la alfabetización digital contra la desinformación, la equidad, no discriminación e inclusión en el diseño y aplicación de los sistemas basados en inteligencia artificial. En ese sentido, la equidad se define como la creación de algoritmos justos que eviten sesgos y discriminación, considerando las diversas necesidades de todas las partes interesadas. También incluye una serie de recomendaciones entre las que se destacan la adopción de un enfoque de múltiples partes interesadas y el deber de que los gobiernos y las empresas realicen continuas evaluaciones del impacto en los derechos humanos con "contribuciones directas de la sociedad civil, antes de recopilar, procesar o desplegar datos, y ofrecer total transparencia, en particular en el entrenamiento de la IA."[53]

3.4. Global Digital Compact

El Global Digital Compact (GDC) o Pacto Digital Global[54] fue adoptado en el 2024, como un anexo del Pacto del Futuro, celebrado en la Cumbre del Futuro, como una iniciativa de la Secretaría General de Naciones Unidas. Tiene como objetivo "abordar los desafíos de la gobernanza digital garantizando acceso equitativo, protección de derechos humanos y reducción de brechas tecnológicas."[55]

Si bien no es un documento vinculante, incluye compromisos y establece consensos relevantes respecto al uso de inteligencia artificial. El GDC destaca la necesidad de un modelo de gobernanza global para la IA, que permita llegar a acuerdos globales sobre su uso, que correspondan a un modelo de múltiples partes interesadas. Asimismo, propone la creación de un panel científico internacional independiente y el inicio de un diálogo mundial dentro de la ONU para abordar los desafíos y oportunidades de la IA.

El pacto reconoce tanto los riesgos como las oportunidades que presenta la IA. Destaca la importancia de la cooperación internacional para evaluar estos riesgos, especialmente en aplicaciones militares, aunque no establece limitaciones explícitas ni mecanismos de supervisión independientes para asegurar que la IA se utilice de manera que respete los derechos humanos. También resalta el rol que pueden tener los sistemas basados en inteligencia artificial en el desarrollo sostenible.

---

[53] Ibíd. 24/26.

[54] Global Digital Compact (2024) Disponible en: https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf

[55] Derechos Digitales (2024) Pacto Global Digital: Gobernanza tecnológica y la protección de los derechos humanos. Disponible en: https://www.derechosdigitales.org/24430/el-pacto-para-el-futuro-su-relacion-con-el-gdc-y-con-el-ambito-digital/

Entre otros aspectos, el Pacto propone adoptar un enfoque basado en riesgos y apoyar la interoperabilidad de los distintos enfoques de la gobernanza de la IA; así como también promover la transparencia, protección de los derechos humanos, rendición de cuentas y supervisión humana de los sistemas de IA. Por otra parte, incluye compromisos relativos a la elaboración de normas comunes en materia de datos para su intercambio y aprovechamiento. Asimismo, exhorta a las empresas y desarrolladores a adoptar salvaguardias en los procesos de entrenamiento de los modelos de IA, identificar el material generado por IA y certificar la autenticidad de los contenidos y su origen, así como etiquetas, marcas de agua y otras técnicas. Particularmente hace referencia a la obligación de mitigar daños potenciales, como el discurso de odio y la discriminación en contenidos generados por IA.[56]

En una carta firmada por 24 organizaciones de la sociedad civil y expertos en la materia[57], se hace una crítica al GDC por la ausencia de compromisos suficientemente robustos para evitar tecnologías incompatibles con los derechos humanos. De igual forma, el Pacto no incorpora adecuadamente una perspectiva de género interseccional, lo que puede profundizar las desigualdades. Además, los firmantes hacen un llamado a priorizar la participación de la sociedad civil, debido a que el proceso de redacción del GDC careció de transparencia y compromiso significativo con la sociedad civil. Por último, reiteran la necesidad de coherencia y colaboración efectiva dentro de la ONU para aplicar el derecho internacional de derechos humanos a Internet y tecnologías digitales.

Respecto a la omisión de una perspectiva interseccional de género, la Asociación para el Progreso de las Comunicaciones (APC) enfatizó que además de las menciones a la brecha digital de género, el pacto global no adoptó análisis profundos acerca de las múltiples formas de discriminación que enfrentan las mujeres y personas de género diverso, particularmente en el contexto digital.

Por otro lado, la APC insistió en que el Pacto "no establece limitaciones explícitas ni mecanismos de supervisión independiente que aseguren que la IA sea utilizada de manera que respete los derechos humanos.[58] Frente a este punto, resalta la necesidad de que hayan menciones expresas y directas referentes a las acciones contraterroristas y el uso de inteligencia artificial artificial en el campo militar, pues, de lo contrario, la ambigüedad puede dar lugar a usos riesgosos para los derechos humanos, que contraríen el derecho internacional de los derechos humanos.

---

[56] 36.c).

[57] Article 19 (2024). UN: Global Digital Compact must build a truly collaborative digital future. Disponible en: https://www.article19.org/resources/un-global-digital-compact-collaborative-digital-future/

[58] Association for Progressive Communications (2024) Pacto Global Digital: Gobernanza tecnológica y la protección de los derechos humanos. Disponible en: https://www.apc.org/es/blog/pacto-global-digital-gobernanza-tecnologica-y-la-proteccion-de-los-derechos-humanos

4. Instrumentos internacionales de carácter vinculante

4.1. Unión Europea: Reglamento de Inteligencia Artificial

Luego de un proceso iniciado en 2021, que incluyó negociaciones y modificaciones por parte de los diferentes órganos de la UE, el Parlamento Europeo aprobó en 2024 el Reglamento de Inteligencia Artificial (EU AI Act)[59] como primer instrumento jurídico vinculante y general en materia de IA. El Reglamento adoptó un enfoque horizontal y basado en riesgos, y a partir de un marco diferenciado que asigna obligaciones en función de si se trata de sistemas o modelos de IA de alto riesgo o no.

En primer lugar, la EU AI Act utiliza una definición abierta de IA, de modo similar a la de la OCDE, y pone el foco en los elementos de autonomía y adaptabilidad de los sistemas que, a partir de las entradas que reciben, generan resultados tales como predicciones, contenidos, recomendaciones o decisiones que afectan los ambientes físicos o virtuales. Como contrapartida, la norma no alcanza a aquellas aplicaciones que se utilicen con fines militares, de defensa o seguridad nacional; para investigación y desarrollo científico; y tampoco a los sistemas o modelos de IA antes de su puesta en servicio, a menos que realicen pruebas en entornos reales. Asimismo, ciertas obligaciones tampoco aplican a los modelos de general-purpose AI (modelos fundacionales) de código abierto.

El Reglamento establece un marco normativo de aplicación transversal que abarca a diversos agentes involucrados en el desarrollo, comercialización y utilización de sistemas de IA. Su ámbito de aplicación incluye proveedores, distribuidores, fabricantes y usuarios, independientemente de su localización geográfica, siempre que los outputs generados por dichos sistemas se empleen dentro del territorio de la UE. Asimismo, el Reglamento se articula en consonancia con el acervo jurídico comunitario en materia de protección de datos, derechos laborales, la DSA y demás legislación sectorial aplicable. Cabe destacar que los sistemas de IA de código abierto quedan excluidos de su ámbito de regulación, excepto en aquellos casos en los que sean clasificados como de alto riesgo o se encuentren comprendidos dentro de las categorías específicas sujetas a supervisión.

El artículo 3 establece definiciones clave e incorpora un extenso glosario de términos, entre los que destacan "sistema de IA", "riesgo", "riesgo sistémico" e "IA de propósito general". Resulta particularmente relevante la inclusión de conceptos vinculados a la gestión de riesgos, especialmente aquellos asociados a amenazas sistémicas, dado que estos no fueron contemplados en el marco definitorio de la Ley de Servicios Digitales (DSA). Esta omisión en

[59] Disponible en: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206

la DSA contrasta con el enfoque proactivo del Reglamento de IA, que busca anticipar y mitigar impactos adversos a gran escala derivados del uso de estas tecnologías[60].

Ciertas aplicaciones de la IA quedaron prohibidas (Art. 5).[61] Sin embargo, estas prohibiciones no son absolutas, sino que dejan espacio para su implementación cuando cumplan ciertas condiciones, como cuando tengan fines médicos o de seguridad. Por ejemplo, cuando se trata de sistemas predictivos del delito, estos son prohibidos exceptuando cuando sean "utilizados para apoyar la valoración humana de la implicación de una persona en una actividad delictiva que ya se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva."

Si bien prohíbe los sistemas de identificación biométrica en tiempo real, el reglamento contempla excepciones que permiten su uso para identificar a víctimas específicas de delitos como secuestro o trata, a personas en los casos que exista una amenaza inminente a la vida o seguridad o una amenaza de ataque terrorista; o a una persona sospechada de haber cometido un delito penado con al menos un máximo de 4 años de prisión. La redacción final del apartado incorporó una serie de salvaguardas cuando esta identificación se realice en lugares públicos como: definición del alcance temporal, geográfico y personal estrictamente necesario; obligación de registro en la base de datos de la UE; la realización de un estudio de impacto a los derechos fundamentales por parte de la autoridad garante del cumplimiento del Derecho[62]; autorización previa otorgada por orden judicial o administrativa independiente cuya decisión sea vinculante del Estado miembro en el que vaya a utilizarse dicho sistema. Sin embargo, algunos de estos requisitos pueden posponerse siempre que existan razones de urgencia justificadas[63]. La Comisión Europea elaborará y publicará informes anuales sobre la utilización de sistemas de identificación biométrica remota en tiempo real en espacios de

---

[60] Artículo 3: 2) «riesgo»: la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio; y 65) «riesgo sistémico»: un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general, que tienen unas repercusiones considerables en el mercado de la Unión debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor;

[61] Artículo 5: Son prácticas prohibidas a) el uso de técnicas subliminales o engañosas para manipular el comportamiento y la toma de decisiones, b) el aprovechamiento de vulnerabilidades por edad, discapacidad, situación social o económica, c) los sistemas de clasificación social, d) los sistemas que evalúen la probabilidad de cometer delitos, e) que creen bases de datos de reconocimiento facial a partir de imágenes de internet o CCTV, f) que infiera emociones en ambientes de trabajo o instituciones educativas, g) que infiera datos sensibles a partir de datos biométricos, y h) la identificación biométrica en tiempo real, con sus excepciones por razones de seguridad.

[62] Según el artículo 3, 45), una autoridad garante del cumplimiento del Derecho es cualquier autoridad pública u organismo habilitado por la ley de un Estado miembro con competencias para prevenir, investigar, detectar o enjuiciar delitos, o ejecutar sanciones penales. Esto incluye también la protección frente a amenazas a la seguridad pública y su prevención

[63] Artículo 5.2.b y 5.3.

acceso público, con fines vinculados a la garantía del cumplimiento del Derecho. Dichos informes se fundamentarán en datos agregados provistos por los Estados miembros.

Por otra parte, los sistemas de AI de alto riesgo serán aquellos que cumplen simultáneamente dos condiciones: que esté destinado a funcionar como componente de seguridad de un producto regulado por normativa de la Unión Europea (Anexo I) o que él mismo sea dicho producto, y que esté sujeto a una evaluación de conformidad por parte de terceros antes de ser introducido en el mercado o puesto en servicio. También lo son aquellos listados en el Anexo III, salvo cuando no planteen un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, también al no influir sustancialmente en el resultado de la toma de decisiones, y los que en un futuro se incorporen[64]. Las aplicaciones de alto riesgo deberán 1) contar con sistemas de gestión de riesgos,[65] 2) cuando los sistemas impliquen el entrenamiento de modelos de IA con datos se desarrollarán a partir de conjuntos de datos de entrenamiento, validación y prueba que cumplan criterios de calidad y se se someterán a prácticas de gobernanza y gestión de datos, 3) proporcionar documentación técnica, 4) implementar registros automáticos de eventos, 5) brindar medidas de transparencia e información para los usuarios, 6) asegurar la vigilancia humana, 7) adecuarse a niveles de precisión, solidez y ciberseguridad, y 8) contar con sistemas de control de calidad. Para ello, promueve el desarrollo de estándares, especificaciones comunes y criterios para el cumplimiento y crea una base de datos a nivel europeo en la que deberán registrarse tanto los sistemas de IA de alto riesgo como aquellos implementados en el sector público, donde habría información disponible respecto de estos sistemas.

Antes de su comercialización o uso, los sistemas de inteligencia artificial deben ser registrados en la base de datos de la UE, según el nivel de riesgo identificado. Esta obligación aplica tanto a los sistemas clasificados como de alto riesgo conforme al anexo III (exceptuando los contemplados en su punto 2) como a aquellos que, tras una evaluación conforme al artículo

[64] Artículo 7 y Anexo III: Las siguientes categorías de aplicaciones, con las especificaciones que figuran en el anexo, son consideradas de alto riesgo: biometría, infraestructura crítica, educación, empleo, servicios de salud, crediticios, de seguro y emergencias, aplicación de la ley, migración y control fronterizo, administración de justicia y proceso democrático. Artículo 7: La Comisión estará facultada para adoptar actos delegados con arreglo al artículo 97 al objeto de modificar el anexo III mediante la adición o modificación de casos de uso de sistemas de IA de alto riesgo.

[65] Según el Artículo 9, se debe implementar un sistema de gestión de riesgos continuo, debidamente documentado y actualizado a lo largo de todo su ciclo de vida. Este proceso, de carácter iterativo, abarca desde la identificación y análisis de posibles amenazas para la salud, la seguridad o los derechos fundamentales, hasta la evaluación de riesgos asociados tanto a los usos previstos como a posibles aplicaciones indebidas. Asimismo, debe incorporar la detección de nuevos riesgos que puedan surgir durante la fase de vigilancia poscomercialización, así como la aplicación de medidas específicas para su mitigación. Cabe destacar que solo se tendrán en cuenta aquellos riesgos que puedan ser minimizados de manera efectiva mediante mejoras en el diseño técnico o mediante la provisión de información adecuada a los usuarios. Las medidas de gestión de riesgos deberán tener en cuenta y aceptar los riesgos residuales asociados a cada peligro identificado, así como el riesgo residual general del sistema de IA de alto riesgo.

6.3, hayan sido considerados por el proveedor como no pertenecientes a dicha categoría. En todos los casos, el proveedor o su representante autorizado debe registrar el sistema y su identidad en dicha base de datos. En el caso de entidades públicas que desplieguen sistemas de alto riesgo, también se requiere su registro y el del uso previsto. Para ámbitos sensibles como la seguridad, migración o control fronterizo, el registro se realiza en una sección no pública y segura, con acceso restringido a la Comisión y autoridades nacionales. Los sistemas comprendidos en el punto 2 del anexo III se registran a nivel nacional.

Cabe resaltar que la decisión de si un sistema concreto es de alto riesgo o no quedó en cabeza de la empresa proveedora,[66] en lo que fue señalado como un cambio que tornó el marco de clasificación del riesgo en un ejercicio potencialmente autorregulatorio.[67] En casos de emergencia que afecten la seguridad pública, la protección de personas, el medio ambiente o infraestructuras críticas, las autoridades pueden solicitar una exención temporal del procedimiento de evaluación de conformidad que mencionamos anteriormente. La autoridad de vigilancia del mercado podrá autorizar este uso excepcional —incluso de forma retroactiva en situaciones de extrema urgencia— siempre que se cumplan los requisitos aplicables. La autorización, que debe notificarse inmediatamente a la Comisión Europea y los Estados miembros, estará sujeta a un mecanismo de supervisión.[68]

La Comisión, en consulta con el Consejo de IA emitirá directrices sobre la aplicación práctica del artículo, incluyendo ejemplos de sistemas de IA clasificados como de alto o bajo riesgo.

La EU AI Act prescribe la realización de una evaluación de impacto de los sistemas de IA de alto riesgo en los derechos fundamentales, antes de su implementación. Sin embargo, esta obligación es limitada a aquellos organismos públicos y entidades privadas que brinden servicios públicos e implementen sistemas de IA de alto riesgo (salvo aquellos de infraestructura crítica) y demás entidades privadas que implementen sistemas de calificación crediticia o de cálculo de seguros de vida. Este análisis consistirá en: 1) descripción de los procesos en los que utilizará el sistema y su finalidad 2) el período en que se implementará el sistema y la frecuencia de uso 3) categorización de personas y colectivos que puedan verse afectados 4) los riesgos de perjuicio específicos que puedan afectar a las categorías de personas físicas y colectivos identificadas 5) descripción de medidas de supervisión humana 6) las medidas que deben adoptarse en caso de que dichos riesgos se materialicen, incluidos los acuerdos de gobernanza interna y los mecanismos de reclamación Sin embargo, no existe una

[66] Artículo 6, inciso 4. " El proveedor que considere que un sistema de IA contemplado en el anexo III no es de alto riesgo documentará su evaluación antes de que dicho sistema sea introducido en el mercado o puesto en servicio". De todas maneras "Dicho proveedor estará sujeto a la obligación de registro establecida en el artículo 49".

[67] Ver Aministía Internacional y otros, "EU's AI Act fails to set gold standard for human rights, 3 de abril de 2024. Disponible en: https://www.amnesty.eu/news/eus-ai-act-fails-to-set-gold-standard-for-human-rights/

[68] Artículo 43.

obligación explícita de evaluar si los riesgos resultan aceptables a la luz del derecho internacional de los derechos humanos o de prevenirlos cuando sea posible, así como tampoco resulta aplicable para autoridades migratorias y de law enforcement.

Por su parte, para los demás sistemas de IA que no impliquen un riesgo significativo de perjuicio a la salud y a la seguridad o repercusiones negativas para los derechos fundamentales, la norma estipula obligaciones mínimas de información y se prevé la creación de códigos de conducta voluntarios.

Entre las disposiciones generales de la regulación, se establece la obligación de *labeling* o etiquetado para identificar los contenidos de audio, imagen, texto o videos producidos sintéticamente, deepfakes y textos generados con IA destinados a informar sobre asuntos de interés público, salvo, en este último caso, que haya habido control humano por parte de una persona con responsabilidad editorial por la publicación del contenido.

También fue incorporada una regulación específica para las general-purpose AI o modelos fundacionales, que comprende obligaciones consistentes en proporcionar documentación técnica, cumplir con la normativa de copyright, informar públicamente sobre el contenido utilizado para su entrenamiento y la obligación de designar un representante. Además, aquellos modelos que impliquen un riesgo sistémico por el alto impacto de sus capacidades deberán realizar evaluaciones y pruebas, reportar incidentes y asegurar niveles adecuados de ciberseguridad. La UE clasificará como de riesgo sistémico a los modelos de IA de uso general con capacidades de alto impacto (evaluadas técnicamente o por la Comisión según el Anexo XIII), presumiéndose dicho impacto según criterios técnicos[69].

Como organismos de gobernanza, la norma crea, además de la ya existente Oficina de IA de la Comisión Europea, un Comité Europeo de IA, conformado por un representante de cada Estado; un Foro Asesor, cuyos miembros serán designados por la Comisión considerando un balance entre las diferentes partes interesadas, incluyendo sociedad civil y academia; y un Panel Científico de Expertos Independientes, destinado a asesorar en el seguimiento de las general-purpose AI. A nivel nacional, los Estados deberán contar con una autoridad de supervisión, que cumplirá las funciones de vigilancia del mercado y de autoridad notificante.

Las autoridades nacionales de vigilancia tienen, entre otras competencias, la potestad de acceder al código de los sistemas de IA de alto riesgo cuando resulte necesario para determinar el cumplimiento de sus obligaciones. Las autoridades nacionales de protección de los derechos fundamentales también podrán solicitar el acceso a la documentación y, cuando ello fuera insuficiente, solicitar la realización de pruebas. La Comisión Europea conserva la competencia exclusiva para supervisar las obligaciones de los proveedores de general-purpose AI models y puede requerir información y documentación, llevar a cabo evaluaciones,

[69] Artículo 51.

requerir medidas de mitigación, restringir su disponibilidad o directamente retirarlo del mercado.

Por su parte, los usuarios tienen el derecho de realizar reclamos ante la autoridad nacional de vigilancia por incumplimiento del Reglamento y a solicitar explicaciones sobre el rol de un sistema de IA en los procedimientos de toma de decisiones que los afecten y sobre los elementos que fueron considerados para ello.

Finalmente, el Reglamento habilita la imposición de multas y sanciones[70] a aquellas empresas, proveedores de sistemas de IA y a instituciones, órganos y organismos de la Unión que no cumplan con ciertas disposiciones de la norma, así como la posibilidad de que la autoridad de vigilancia del mercado del Estado miembro correspondiente tome medidas adecuadas y proporcionales para restringir o prohibir la comercialización, recuperar o retirar sin demora del mercado un sistema de IA de alto riesgo.[71]

El texto final de la EU AI Act, al basarse en un enfoque de riesgos, deja la protección de los DDHH en un segundo plano, limitándose a la autoevaluación que realicen los proveedores o implementadores de estos sistemas y las medidas que tomen para mitigar tales riesgos. Asimismo, la prohibición de ciertas prácticas problemáticas para los DDHH fue acompañada de un número de excepciones que dejan abierta la puerta, por ejemplo, al reconocimiento de emociones y a sistemas de vigilancia masiva incompatibles con el derecho a la privacidad y la no discriminación. También permite el uso de la IA en materia de seguridad nacional y exceptúa a las autoridades migratorias, de defensa y seguridad nacional de las obligaciones de transparencia respecto sistemas de IA de alto riesgo, dando lugar a que sean implementados por fuera del conocimiento público[72]. Es notable que no hayan sido incorporados como de alto riesgo los sistemas de IA utilizados en ecosistemas informativos, como redes sociales y demás plataformas digitales, teniendo en cuenta su impacto en la circulación y amplificación de contenidos en línea.

El reglamento establece obligaciones en cabeza de los distintos actores de la cadena, así el artículo 16 describe las obligaciones de los proveedores de sistemas de IA de alto riesgo; el artículo 23 impone obligaciones a los importadores; y el artículo 24, a los distribuidores. Asimismo, el artículo 25 dispone que cualquier distribuidor, importador, responsable del despliegue o tercero será considerado proveedor de un sistema de IA de alto riesgo cuando

[70] Artículos 99, 100 y 101.
[71] Artículo 83.

[72] En este sentido, ver Amnistía Internacional y otros, “EU’s AI Act fails to set gold standard for human rights”, 3 de abril de 2024. Disponible en https://www.amnesty.eu/news/eus-ai-act-fails-to-set-gold-standard-for-human-rights/; Article 19, “EU: AI Act passed in Parliament fails to ban harmful biometric technologies”, 13 de marzo de 2024. Disponible en:https://www.article19.org/resources/eu-ai-act-passed-in-parliament-fails-to-ban-harmful-biometric-technologies/

cumpla determinadas circunstancias y estará sujeto a las obligaciones previstas en el artículo 16 s. Si bien de esta manera se extienden las obligaciones de los proveedores a los demás actores de la cadena , la norma no aborda la creciente concentración de empresas en el mercado de IA y la conformación de monopolios, aunque da ciertos incentivos y consideraciones para startups y pymes (como acceso equitativo a sandboxes regulatorios[73] y formularios simplificados).

De todos modos, la mayoría de las disposiciones requerirán reglamentos delegados, códigos de conducta, estándares técnicos y criterios de cumplimiento, que darán forma a las obligaciones específicas. También podrán ser incorporadas al listado del Anexo III otras aplicaciones de la IA que en un futuro sean consideradas como de alto riesgo por vulnerar derechos humanos. En este sentido, el Foro Asesor creado por la norma puede ser un espacio valioso de participación e involucramiento por parte de la sociedad civil y la academia.

Cabe destacar que es posible que, en virtud del conocido "Efecto Bruselas" descrito por Anu Bradford[74], esta regulación posiblemente influencie de manera significativa futuros procesos regulatorios de otros Estados por fuera de la Unión Europea, de un modo similar a lo ocurrido con su regulación en materia de protección de datos personales y la Ley de Servicios Digitales.[75]. Al convertirse en el primer instrumento vinculante, y considerando el tamaño de su economía, si bien no resulta de aplicación extraterritorial, las empresas de IA que pretendan ofrecer sus servicios en la UE deberán adaptarse a la nueva regulación. También existe la posibilidad de que el efecto Bruselas se materialice a través de la exportación de estándares a otras jurisdicciones por parte de las propias empresas proveedoras.

Si bien el reglamento entró en vigor el 1 de agosto de 2024, su aplicación y obligatoriedad fue prevista de manera progresiva. De este modo, la prohibición de determinados sistemas de IA y las obligaciones sobre alfabetización se encuentran operativas desde el 2 de febrero de 2025, mientras que las disposiciones relativas a modelos de uso general lo harán a partir del 2 de agosto de dicho año. Por último, las normas destinadas a los sistemas de IA de alto riesgo recién serán obligatorias a partir del 2 de agosto de 2027.

En septiembre de 2024 se llevó a cabo la primera sesión para el proceso de elaboración del Código de Buenas Prácticas para los modelos de IA de uso general, en el marco del EU AI

[73] Artículos 57 a 63.

[74] Anu Bradford. The Brussels Effect: How the European Union Rules the World. Oxford University Press., 2020. Disponible en: https://doi.org/10.1093/oso/9780190088583.001.0001

[75] González Mama, M.; Alvarez Ugarte, R., «Modelización regulatoria. Palpitando la influencia de la Digital Services Act en América Latina», Centro de Estudios en Libertad de Expresión (CELE), Artículo de investigación No. 63, Buenos Aires, Argentina, 2025.

Act. De acuerdo con información de la Comisión Europea[76], el Código de Buenas Prácticas tiene como objetivo la facilitación de una correcta aplicación del Reglamento de IA para los modelos de inteligencia artificial de uso general, incluidas las normas de transparencia y las relacionadas con los derechos de autor, la taxonomía del riesgo sistémico, la evaluación del riesgo y las medidas de mitigación.

El código no tendrá carácter vinculante para proveedores, pero expertos presuponen fuertes incentivos para que estos se acojan a las disposiciones acordadas[77]. Frente a este proceso, Calvet-Bademunt[78] ha identificado dos grupos de trabajo en los que existe una ventana de oportunidad para proteger la libertad de expresión, o, por el contrario, en donde la ausencia de estas protecciones puede ser problemática para la garantía del derecho.

En el grupo de trabajo sobre "identificación y evaluación de riesgos", destaca la necesidad de incluir de manera expresa la protección del derecho a la libertad de expresión, ejercicio que ha sido insuficiente por parte de la Comisión, en especial porque el Reglamento no considera la libertad de expresión en definiciones importantes como la de riesgo sistémico. Por otro lado, hace un llamado al grupo de trabajo sobre "mitigación técnica de riesgos", cuyas discusiones deben priorizar que las restricciones previstas en el código no sean desproporcionadas y obedezcan a estándares internacionales de libertad de expresión; este punto es particularmente central debido a que es posible que acudan a prácticas de censura para enfrentar fenómenos como la desinformación o los discursos de odio.

En el marco de la Cumbre de Acción sobre Inteligencia Artificial celebrada en París en febrero de 2025, la presidenta de la Comisión Europea, Ursula von der Leyen, anunció el lanzamiento de InvestAI, una ambiciosa iniciativa destinada a movilizar 200.000 millones de euros en inversiones para el desarrollo y despliegue de la inteligencia artificial y colocar a la UE como líder tecnológico y de la soberanía digital[79]. En abril de 2025 von der Leyen anunció que impulsará una implementación más "simplificada" del Reglamento de IA de la UE, con posibles revisiones a sus obligaciones de reporte, incluso antes de que entre plenamente en vigor. El objetivo es reducir la carga administrativa para facilitar la adaptación de las empresas europeas, en un contexto donde pocas grandes compañías de IA tienen base en la región.[80]

---

[76] European Commission. Industry, academia and civil society contribute to the work on Code of practice for general-purpose artificial intelligence, 30 de septiembre de 2024. Disponible en: https://digital-strategy.ec.europa.eu/en/news/kick-plenary-general-purpose-ai-code-practice-took-place-online

[77] Calvet-Bademunt, J. Safeguarding Freedom of Expression in the AI Era. Tech Policy, 2024. Disponible en: https://www.techpolicy.press/safeguarding-freedom-of-expression-in-the-ai-era/

[78] Ìbidem.

[79] EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence, 11 de febrero de 2025. Disponible en: https://luxembourg.representation.ec.europa.eu/actualites-et-evenements/actualites/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence-2025-02-11_en

[80] Virkkunen takes aim at AI Act reporting obligations, 10 de abril de 2025. Disponible en: https://www.euractiv.com/section/tech/news/virkkunen-takes-aim-at-ai-act-reporting-obligations/

4.2 Consejo de Europa: Convención Marco sobre IA

El 17 de mayo de 2024, el Comité de Ministros del Consejo de Europa aprobó el texto del primer tratado internacional en materia de IA: la Convención marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho[81]. En el proceso intervinieron representantes de los Estados miembros, Estados no miembros de la UE (como EEUU, Argentina, México y Perú, entre otros), la OECD, la UNESCO, la OSCE y representantes de 68 organizaciones de la sociedad civil en calidad de observadores[82]. El convenio fue sometido a votación y aprobado el 17 de mayo de 2024. Posteriormente, en septiembre del mismo año fue abierto para firmas. Para febrero de 2025, habían firmado Andorra, Canadá, Georgia, Islandia, Liechtenstein, Montenegro, Noruega, la República de Moldavia, San Marino, el Reino Unido, así como Israel, los Estados Unidos de América, la Unión Europea, Canadá y Japón.

Éste pretende sentar principios y normas destinadas a regular todo el ciclo de vida de los sistemas de IA para asegurarse de que sean acordes con los derechos humanos, la democracia y el estado de derecho. El texto adopta un enfoque basado en riesgos y la adopción de medidas en términos generales, ya sean específicas u horizontales, en función de la gravedad y la probabilidad de los impactos adversos.

La Convención adopta la definición de IA de la OCDE y, en cuanto a su ámbito de aplicación, alcanza a las autoridades públicas y actores privados que actúen en su nombre. Respecto de los actores privados, los Estados tendrán la obligación de "abordar" los riesgos e impactos de sus actividades y podrán elegir extender las demás obligaciones de la Convención a los privados mediante una declaración que puede ser enmendada en cualquier momento. También se excluye a los sectores de seguridad y de defensa nacional de la aplicación del tratado.

Previo a la aprobación del texto definitivo, la Asamblea Parlamentaria del Consejo de Europa emitió sus observaciones al proyecto, lamentó el enfoque diferenciado para el sector privado y propuso limitar el régimen de excepción para el sector defensa a los casos en que esas actividades se encuentren en línea con el derecho internacional de los derechos humanos, establecer prohibiciones a ciertos usos de la IA incompatibles con los DDHH, proteger a las

[81] Consejo de Europa, Convención marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho, 17 de mayo de 2024. Disponible en: https://rm.coe.int/1680afae3c

[82] Convención marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho, 05 de septiembre de 2024. Disponible en:https://rm.coe.int/1680afae3c

personas denunciantes e incluir disposiciones respecto de la protección de la salud y el ambiente[83] . Sin embargo, tales observaciones no fueron incorporadas.

La convención contiene obligaciones generales que incluyen medidas para la protección de los derechos humanos, de acuerdo a los instrumentos internacionales de derechos humanos que se citan en su preámbulo, y para asegurar la integridad de las instituciones y procesos democráticos, incluyendo el acceso a la participación en el debate público y la capacidad de formar opiniones libremente.

El capítulo III prescribe la adopción de medidas dirigidas a salvaguardar los principios de dignidad humana; transparencia y vigilancia (identificación de contenido generado por IA); rendición de cuentas y responsabilidad; igualdad y no discriminación; privacidad y protección de datos personales; confiabilidad (requisitos de calidad y seguridad); e innovación segura (entornos controlados de prueba)

En cuanto a los remedios previstos, el texto refiere al cumplimiento de las obligaciones internacionales ya existentes y agrega que las medidas al respecto deben incluir que la información relevante de sistemas de IA que puedan afectar los derechos humanos sea documentada, provista a los organismos autorizados y puesta a disposición de las personas afectadas, junto con la información suficiente para impugnar la decisión tomada por un sistema de IA y su uso. También deberán garantizar un recurso y garantías procesales efectivas para los derechos, incluyendo la notificación, de acuerdo al contexto, de que se interactúa con un sistema de IA en lugar de una persona humana.

El capítulo 5 establece el deber de adoptar medidas para identificar, evaluar, prevenir y mitigar los riesgos de impacto a los derechos humanos, de modo "graduado y diferenciado" en función de tales riesgos, considerando 1) el contexto y uso previsto, 2) la gravedad y probabilidad de posibles impactos, 3) las perspectivas de las partes interesadas relevantes y personas cuyos derechos puedan verse afectados, 4) su aplicación iterativa a través del ciclo de vida del sistema IA, 5) el monitoreo de riesgos, 6) su documentación, y 7) la realización de pruebas antes de un primer uso o modificación significativa. Como resultado de ello, los Estados deberán evaluar la necesidad de moratorias o prohibiciones cuando se considere algunos usos incompatibles con los derechos humanos.

Finalmente, el mecanismo de implementación de la convención consiste en una Conferencia de las Partes a la cual deberán reportar los Estados el cumplimiento de las obligaciones dentro de los primeros dos años de vigencia y luego periódicamente. Además, los Estados deberán establecer a nivel nacional uno o más mecanismos de monitoreo, independientes, imparciales y con la expertise y recursos suficientes.

---

[83] Asamblea Parlamentaria del Consejo de Europa, Opinion on the Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, 18 de abril de 2024. Disponible en: https://pace.coe.int/en/files/33517/html

Como primera propuesta de tratado en materia de IA, con potencial alcance global, el lenguaje utilizado resulta ciertamente vago en términos de obligaciones concretas y adopta un enfoque basado en riesgos que, si bien es similar al de la EU AI Act, le otorga a la protección de los derechos humanos un rol más central en función de la referencia explícita al cumplimiento de las normas de derecho internacional.

Entre las principales críticas a la Convención se destacan la inaplicabilidad de las disposiciones a los ámbitos de defensa y seguridad nacional y el sistema *a la carte* destinado a que los Estados puedan excluir a los actores privados del alcance de las obligaciones de la convención. Por otra parte, la metodología de evaluación de los sistemas de IA que será desarrollada por el Comité de IA tampoco será de cumplimiento obligatorio para los Estados parte ni tampoco establece líneas rojas respecto de aplicaciones específicas de IA incompatibles con los DDHH, lo cual potencialmente propiciará la proliferación y fragmentación de criterios para su implementación. Asimismo, también fue señalado que no se consideraron adecuadamente las problemáticas específicas de los países del Sur Global, quedando por fuera cuestiones como los derechos laborales de los trabajadores y la protección de la salud y el medio ambiente.[84]

5. Iniciativas en el marco de la Organización de Estados Americanos

5.1 Marco Interamericano de Gobernanza de Datos e Inteligencia Artificial (MIGDIA)

Este marco es una iniciativa de varios estados miembros de la OEA que incluye lineamientos en 12 áreas temáticas, que se presentan como hoja de ruta; un estudio sobre el estado de gobernanza de datos e IA en las Américas, acompañado de recomendaciones; y, una política modelo de gobernanza de datos e IA que oriente la generación de políticas públicas internas de los estados miembro.

La iniciativa reconoce como esencial la gobernanza de datos en los países de la región para un aprovechamiento pleno y respetuoso de los derechos humanos de las tecnologías emergentes como la inteligencia artificial. El proceso de formulación de los lineamientos se realizó

[84] En este sentido, ver Conference of INGOs, "Recommendations to PACE prior to drafting its Opinion in the Draft Framework", 12 de abril de 2024. Disponible en:https://ecnl.org/sites/default/files/2024-04/CINGO%20Recommendations%20PACE%20FC%20on%20AI%20April%202024.pdf, Global Partners Digital, The world's first treaty on AI: our thoughts and the way forward, 2 de abril de 2024. Disponible en: https://www.gp-digital.org/the-worlds-first-treaty-on-ai-our-thoughts-and-the-way-forward/; Amnistía Internacional, Recommendations to the PACE and CoM on the Draft Framework Convention on AI, 11 de abril de 2024. Disponible en: https://www.amnesty.eu/wp-content/uploads/2024/04/Amnesty-International-Recs-draft-CoECAI-11042024.pdf

alrededor de prioridades identificadas entre los países de la región durante el año 2023, y posteriormente estudiados por múltiples actores durante el 2024. Se espera que el documento sea expuesto en la Cumbre de las Américas de 2025, que tendrá lugar en diciembre de 2025, para que los Estados expresen su consenso y aprobación de los lineamientos[85].

La visión estratégica de estos lineamientos[86] destaca que la gobernanza de datos es crucial para la transformación digital, el desarrollo sostenible y la protección de derechos humanos. Busca fortalecer la toma de decisiones, la eficiencia institucional y la transparencia. Los lineamientos transversales incluyen el respeto a los derechos humanos, el diálogo multiactoral, la actualización de marcos normativos sobre privacidad y protección de datos, la participación ciudadana y la sostenibilidad ambiental.

El documento insiste en la importancia de promover la alineación de políticas nacionales con estándares internacionales, como las recomendaciones de la UNESCO sobre uso ético de la IA y los principios de datos abiertos. Esto incluye fomentar el diálogo regional para consolidar posiciones nacionales en foros internacionales. Además, plantea desarrollar marcos normativos dinámicos, definir estrategias gubernamentales claras y crear oficinas responsables de la gobernanza de datos e IA con liderazgo técnico y planes de comunicación transparentes.

Los lineamientos cubren áreas clave como: cultura de datos e IA, ética, transparencia y protección de derechos humanos, marcos de gobernanza de datos, infraestructuras de datos, dinamización de ecosistemas público-privados, maximización del impacto de la gestión de datos y IA, investigación y desarrollo, datos abiertos y ciberseguridad. En resumen, estos buscan establecer un marco de referencia para que los países adopten políticas nacionales de gobernanza de datos e IA, promoviendo la armonización regional y la colaboración entre sectores. El desarrollo de estos lineamientos involucró un proceso de co-creación con participación de actores del sector privado y la sociedad civil.

En noviembre de 2024 el Departamento para la Gestión Pública Efectiva abrió la Consulta Pública regional abierta: “lineamientos interamericanos de gobernanza de datos e IA” para la revisión del documento de lineamientos que fue presentado en la Reunión Ministerial de Gobierno Digital de la Red Interamericana de Gobierno Electrónico (Red Gealc) que tuvo lugar los días 26 y 27 de noviembre de ese año.

---

[85] Derechos Digitales (2024) Comentarios de Derechos Digitales sobre los “Lineamientos interamericanos para la gobernanza de datos y la IA” de la Organización de Estados Americanos OEA. Disponible en: https://www.derechosdigitales.org/wp-content/uploads/Sobre_lineamientos_GDIA_-OEA_postura_de_DD.pdf

[86] Organización de Estados Americanos (2024) Lineamientos Interamericanos de Gobernanza de Datos e Inteligencia Artificial. Disponible en: https://www.oas.org/ext/es/democracia/marco-interamericano-de-gobernanza-de-datos-e-inteligencia-artificial-migdia/moduleid/14735/id/1080/lang/2/controller/item/action/download

Como respuesta a la consulta, la organización Derechos Digitales publicó una carta con comentarios en la que destacan 3 ejes dentro de los cuáles considera hay puntos de mejora. Uno es el eje de maximización del impacto de la gestión de los datos y la inteligencia artificial, respecto del cual DD propone el fomento de observatorios multiactorales que investiguen y monitoreo de tecnologías aplicadas a la gobernanza de datos y de iniciativas de gobernanza nacionales y regionales. Otro es el eje de investigación y desarrollo en gobernanza de datos e inteligencia artificial, que implica la creación de marcos legales que promuevan y garanticen procesos investigativos basados en minería de datos y metadatos. Por último, el eje de ciberseguridad, para el que propone la creación de mecanismos de registro público de reportes nacionales y regionales de incidentes de ciberseguridad, así como reportes de vulnerabilidades por parte de distintos actores, a través de metodologías no intrusivas, especialmente sobre aquellos que comprometan la seguridad digital en materia de sistemas de gobernanza de datos[87].

Adicionalmente, en el marco de la VII Reunión de Ministros de Ciencia y Tecnología convocada por la OEA, la misión permanente, en alianza con el Departamento de Estados Unidos, lanzó un nuevo proyecto para desarrollar un marco de políticas de IA seguro, confiable y responsable para las Américas. El proyecto de declaración titulada "Hacia el Desarrollo y Despliegue de una Inteligencia Artificial Segura, Protegida, Confiable en las Américas"[88] resuelve urgir a los Estados Miembros a garantizar el desarrollo y despliegue seguro y confiable de la IA en la región, promoviendo la innovación y respetando los derechos humanos. Además, se compromete a definir e implementar una agenda interamericana sobre IA que fomente la innovación, proteja los derechos humanos e impulse el desarrollo integral.

También recomienda a los Estados Miembros desarrollar marcos de gobernanza modelo para abordar los desafíos de la IA, adaptándolos a cada contexto nacional. De la misma forma, el Consejo alienta a compartir información y buenas prácticas sobre políticas y marcos regulatorios de IA para construir una base común de entendimiento en la región, enfatizando en que las políticas de IA deben estar alineadas con las prioridades de desarrollo, derechos humanos y seguridad de la OEA.

---

[87] Derechos Digitales (2024) Comentarios de Derechos Digitales sobre los "Lineamientos interamericanos para la gobernanza de datos y la IA" de la Organización de Estados Americanos OEA. Disponible en: https://www.derechosdigitales.org/wp-content/uploads/Sobre_lineamientos_GDIA_-OEA_postura_de_DD.pdf

[88] Organización de Estados Americanos. Consejo Interamericano para el Desarrollo Integral (2024) Proyecto de declaración Hacia el Desarrollo y Despliegue de una Inteligencia Artificial Segura, Protegida, Confiable en las Américas. Disponible en: https://www.oas.org/ext/es/principal/calendario/evento/moduleid/14104/id/726/lang/2/controller/item/action/eventdownload

Establece que desde la OEA se deben facilitar diálogos entre gobiernos, industria y sociedad civil para desarrollar marcos de gobernanza colaborativos. Así mismo, el Consejo insta a fomentar alianzas público-privadas para impulsar la innovación y el uso responsable de la IA y recomienda establecer programas de capacitación para formuladores de políticas y ampliar la oferta educativa en IA para jóvenes.

La declaración busca fortalecer y ampliar la Red de Centros de Excelencia en Tecnologías Transformadoras y resuelve establecer dos grupos de trabajo para implementar los mandatos de la declaración y su plan de acción. Finalmente, solicita a los Estados Miembros asignar recursos técnicos, humanos y económicos para apoyar las acciones propuestas y colaborar con Estados Observadores y socios para apoyar la implementación de la declaración.

Más allá de la mención del deber de respetar el derecho internacional de los derechos humanos y los principios que rigen a la OEA, la declaración muestra una ausencia de enfoque basado en derechos humanos a la hora de concretar las iniciativas que componen el plan de acción. A pesar de que su formulación correspondió a los ministros y altas autoridades de ciencia y tecnología de los estados miembros, el proyecto de declaración no evidencia ningún análisis contextual profundo respecto al impacto que puede tener el desarrollo y uso de sistemas basados en inteligencia artificial en la región.

Además de una mención superficial acerca de la preocupación del Consejo sobre los posibles riesgos que representa la IA, a diferencia de otros instrumentos internacionales y múltiples iniciativas nacionales, la declaración no adopta un enfoque basado en riesgos, no identifica cuáles son esos riesgos ni categoriza su potencial amenaza de acuerdo a las particularidades sociales, culturales y políticas de la región.

De la misma manera, las acciones declaradas no tienen naturaleza distinta a llamados a los estados para colaborar, sin plantear una hoja de ruta concreta, clara y completa que permita la materialización de esfuerzos regionales en materia regulatoria y con perspectiva de protección y garantía de derechos humanos.

6. Regulaciones a nivel nacional

6.1 Estados Unidos: AI Bill of Rights, Orden Ejecutiva y NIST

En octubre de 2022, la Office of Science and Technology Policy de la Casa Blanca publicó un documento programático,[89] denominado AI Bill of Rights, destinado a guiar el diseño, uso y desarrollo de sistemas automatizados, basados en machine learning, estadística u otro tipo de procesamiento de datos o técnicas de inteligencia artificial por el gobierno federal.

[89] White House Office of Science and Technology Policy, Blueprint for an AI Bill of Rights, octubre de 2022. Disponible en: https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/

La propuesta contiene un conjunto de cinco principios a ser adoptados voluntariamente por todo tipo de organizaciones, públicas o privadas, y servir para el desarrollo de normas, políticas y estándares. En particular, propone los siguientes principios: 1) Sistemas seguros y eficaces (identificación y mitigación de riesgos, datos de calidad, evaluación y reportes independientes); 2) Protecciones contra la discriminación algorítmica (Evaluación de igualdad y disparidad por diseño, datos representativos); 3) Privacidad de datos (privacidad por diseño y default, protección contra la vigilancia sin controles); 4) Notificación y explicación (sobre el uso de los sistemas y las razones de una decisión automatizada), 5) Alternativas humanas, Consideración y Retroceso (opción de no utilizar el sistema, que la decisión sea considerada por una persona humana, mecanismo para recurrir ante un error o frente a los impactos de su uso).

El documento fue recibido como un paso inicial importante, llamando a que sea implementado para todas las agencias y contratistas y regulado para toda la administración federal, sin dejar por fuera a los sistemas utilizados en contextos de seguridad y seguridad nacional, en los que las poblaciones marginalizadas enfrentan obstáculos e impactos discriminatorios, como la exclusión algorítmica.[90]

Sobre la base del documento "Blueprint for an AI Bill of Rights", el 30 de octubre de 2023, el Presidente Biden dictó una orden ejecutiva[91] destinada a avanzar en el desarrollo e implementación de la IA mediante un enfoque amplio basado en la coordinación entre los diferentes departamentos y agencias del gobierno federal. A partir de ocho principios/políticas, plantea que la IA sea segura y protegida, promover la innovación y la competencia, apoyar a los trabajadores[92], que la IA sea consistente con la igualdad y los derechos civiles, proteger los intereses de EEUU, la privacidad, gestionar los riesgos del uso de la IA por parte del gobierno federal y promover el liderazgo de EEUU en la materia. Fija objetivos y acciones con plazos definidos en cabeza de diferentes departamentos y agencias y crea el White House Artificial Intelligence Council conformado por funcionarios de diversas carteras.

---

[90] ACLU, Comment on Release of White House Blueprint for an Artificial Intelligence Bill of Rights, 4 de octubre de 2022. Disponible en: https://www.aclu.org/press-releases/aclu-comment-release-white-house-blueprint-artificial-intelligence-bill-rights; Access Now, "Access Now applauds U.S. Blueprint for AI Bill of Rights, but more safeguards needed", 6 de octubre de 2022. Disponible en https://www.accessnow.org/press-release/u-s-blueprint-for-ai-bill-of-rights/; Civil Rights, Tech Groups Call on Biden to Protect Public from Harms of AI, 3 de agosto de 2023, https://cdt.org/press/civil-rights-tech-groups-call-on-biden-to-protect-public-from-harms-of-ai/

[91] White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 30 de octubre de 2023. Disponible en: https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence

[92] Esto se refiere a la posibilidad de adoptar medidas para evaluar y abordar los impactos de la IA en el mercado laboral, proteger los derechos de los trabajadores y fomentar el desarrollo de habilidades relacionadas con la IA.

La definición de IA adoptada es la del US Code[93] que comprehende a los sistemas basados en máquinas que pueden, para un conjunto dado de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o tomar decisiones que influyan en entornos reales o virtuales, así como también utilizar insumos de origen humano y de máquinas para percibir entornos reales y virtuales; abstraer dichas percepciones en modelos mediante análisis de forma automatizada; e inferir y formular opciones de información o acción. Asimismo, parte de las acciones se concentran en modelos fundacionales de doble uso, como aquellas que implican riesgos graves por su potencial de ser utilizados en relación con armas químicas, biológicas, radiológicas o nucleares, operaciones de ciberataques y evadir el control humano mediante engaños.

Entre las líneas de acción, se encuentran el desarrollo de guías, estándares y mejores prácticas en materia de seguridad de la IA (Marco de gestión de riesgos para IA generativa, entornos de prueba, reportes de empresas que desarrollen modelos de uso dual); gestión de infraestructura crítica y ciberseguridad, reducir el riesgo que supone el contenido sintético (autenticación de contenidos y rastreo de procedencia, etiquetado y marcas de agua, detección de contenidos sintéticos); promoción de la competencia (prevenir la concentración de empresas y abusos de posición dominante); promoción de la igualdad y los derechos civiles (sectorialmente, en el sistema de justicia penal, seguridad social, empleo, consumidores, pacientes y estudiantes); promover el uso de la IA en el gobierno federal (con base en la presunción de que impacta en los derechos y la seguridad y siguiendo los requisitos mínimos de gestión de riesgos planteados en la AI Bill of Rights y la IA NIST).

La orden ejecutiva de la administración de Biden fue un paso adelante en la inclusión de consideraciones específicas sobre derechos humanos en la aplicación de sistemas de IA en el sector público federal, así como también el calendario previsto para el desarrollo de instrumentos y políticas prometía mayores avances en el corto plazo. A pesar de ello, mantuvo la tendencia de excluir de la regulación a las aplicaciones de IA en seguridad nacional, actividades de vigilancia, migraciones y control fronterizo, sin establecer líneas rojas o abordar, entre otros, las implicancias de tecnologías de reconocimiento biométrico en detenciones arbitrarias en aplicación o de las tecnologías de etiquetado de contenidos para la libertad de expresión y la garantía de anonimato.[94]

---

[93] Código de Estados Unidos, artículo 9401.3. Disponible en: https://uscode.house.gov/view.xhtml?req=(title:15%20section:9401%20edition:prelim)

[94] ACLU, Statement on President Biden's Executive Order on Artificial Intelligence, 30 de octubre de 2023. Disponible en: https://www.aclu.org/press-releases/aclu-statement-on-president-bidens-executive-order-on-artificial-intelligence; Acces Now, "White House Executive Order on AI: a step forward on civil rights, but Congress must pass privacy law", 30 de octubre de 2023. Disponible en https://www.accessnow.org/press-release/white-house-executive-order-on-ai/; Center for Democracy & Technology, "CDT Welcomes White House Executive Order on AI", 30 de octubre de 2023. Disponible en https://cdt.org/press/cdt-welcomes-white-house-executive-order-on-ai/

Por su parte, el Instituto Nacional de Estándares y Tecnología (NIST) publicó en marzo de 2023 la primera versión del AI Risk Management Framework (AI RMF)[95], una herramienta voluntaria dirigida a organizaciones que diseñen, desarrollen, implementen o usen sistemas de IA y destinada a gestionar los riesgos y promover una IA responsable y confiable, de un modo que logre preservar los derechos de las personas y sea flexible y medible en su aplicación.

El modelo de análisis del AI RMF ubica en el centro a las personas y al planeta e identifica cuatro dimensiones clave (modelo, datos y entradas, tareas y salidas; y contexto de aplicación), en 6 fases del ciclo de vida, y cuatro funciones (gobernanza, mapeo, medición y gestión), divididas a su vez en 19 categorías y 72 subcategorías. Asimismo, propone que para que un sistema de IA sea confiable y seguro debe garantizarse que sea válido y fiable, seguro y resistente, responsable y transparente, explicable e interpretable, de privacidad reforzada, justo y cuyos sesgos perjudiciales hayan sido gestionados.

El AI RMF se acompaña de un manual de estrategias, que sugiere acciones para alcanzar los resultados en línea con cada una de las funciones delineadas, y un centro de recursos[96], donde se reúnen los documentos relevantes, acceso a los estándares promovidos a nivel global y métricas para asistir en la evaluación y validación de los sistemas de IA. La Executive Order on Safe, Secure and Trustworthy AI encomendó al NIST la elaboración de una versión actualizada que incorpore las particularidades relacionadas con el uso de IA generativa y también se encuentra en proceso el desarrollo de diferentes perfiles o casos de uso, adaptando la implementación del marco a diferentes sectores de aplicación.

Este marco para la gestión de riesgos refiere explícitamente a los derechos humanos y promueve la participación de la academia, sociedad civil y usuarios, personas y comunidades que potencialmente puedan verse afectadas. Al tratarse de una primera versión, se ha señalado que aún existen ciertos aspectos pendientes, tales como el establecimiento de métricas, líneas base y puntos de referencia, que estén rigurosamente apoyados en investigaciones científicas y ofrezcan diversidad lingüística, haciendo hincapié en obligaciones de transparencia y documentación.[97]

Desde el legislativo, en septiembre de 2024 el senador Edward Markey presentó en el Senado de los Estados Unidos el Artificial Intelligence Civil Rights Act of 2024. Esta propuesta de ley federal busca prevenir la discriminación prohibiendo el uso de algortimos que tengan sesgos discriminatorios; además, impone la obligación de implementar evaluaciones de impacto de

---

[95] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework, enero de 2023. Disponible en https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

[96] National Institute of Standards and Technology, Trustworthy & Responsible AI Resource Center. Mas información en https://airc.nist.gov/Home

[97] Center for Democracy & Technology, "Comments to NIST on its Assignments Under the Executive Order Concerning Artificial Intelligence", 2 de febrero de 2024. Disponible en: https://cdt.org/insights/cdt-comments-to-nist-on-its-assignments/

los sistemas de IA, junto con el deber de realizar auditorías independientes sobre los algortimos. También regula la transparencia en la contratación de sistemas basados en IA y regula estándares respecto a escenarios en los cuáles deben priorizarse las decisiones humanas sobre las algorítmicas.

En enero de 2025, el Grupo de Trabajo sobre IA y Datos Abiertos del Gobierno conformado en 2023 como iniciativa del Departamento de Comercio de los Estados Unidos, publicó el documento "Inteligencia Artificial Generativa y datos Abiertos: lineamientos y buenas prácticas[98]"

Los lineamientos están pensados para ser usados por el Departamento de Comercio, pero fue publicado con el fin de que sea un documento vivo que pueda modificarse con base en retroalimentaciones y consideraciones de diferentes actores. Su particularidad recae en que provee estándares respecto a cómo publicar datos abiertos para que sirvan para alimentar sistemas de inteligencia artificial generativa, esto significa la publicación de "datos abiertos que están formateados, estructurados y preparados de manera que faciliten su uso efectivo en sistemas de IA generativa y reduzcan los costos relacionados con la innovación en el desarrollo de sistemas de IA generativa[99]".

El inicio de la nueva gestión de Donal Trump vino acompañado de un cambio en la política de EEUU en materia de regulación de la IA Estados Unidos. El 23 de enero de 2025, el presidente dictó la Orden Ejecutiva 14179, titulada "Removing Barriers to American Leadership in Artificial Intelligence", la cual deroga las anteriores políticas "que funcionan como barrera a la innovación estadounidense en IA" y establece un plan para promover el desarrollo de la IA sin "sesgos ideológicos" ni "agendas sociales fabricadas", en la búsqueda de fortalecer el liderazgo global de Estados Unidos en la materia.[100] La orden no detalla cuáles son las políticas preexistentes que estaban generando obstáculos para el desarrollo de la IA, "pero se propone rastrear y revisar todas las políticas, directrices, regulaciones, órdenes y otras

---

[98] AI and Open Government Data Working Group (2025). Generative Artificial Intelligence and Open Data: Guidelines and Best Practices. Department of Commerce of the EEUU. Disponible en: https://www.commerce.gov/news/blog/2025/01/generative-artificial-intelligence-and-open-data-guidelines-and-best-practices

[99] AI and Open Government Data Working Group (2025). Generative Artificial Intelligence and Open Data: Guidelines and Best Practices. Department of Commerce of the EEUU. Pag 16. Disponible en: https://www.commerce.gov/news/blog/2025/01/generative-artificial-intelligence-and-open-data-guidelines-and-best-practices

[100] Reuters, "Trump orders AI action plan and more work erasing Biden's AI efforts", 23 de enero de 2025. Disponible en: https://www.reuters.com/technology/artificial-intelligence/trump-orders-ai-action-plan-more-work-erasing-bidens-ai-efforts-2025-01-23/?utm_source=chatgpt.com

acciones tomadas con base en la amplia orden ejecutiva para la IA emitida por el expresidente Joe Biden en 2023."[101]

Al respecto, la ACLU advierte sobre el desmantelamiento de protecciones clave en el uso de la IA, lo cual expone a las personas a riesgos significativos y demuestra la creciente influencia de las grandes empresas tecnológicas en la nueva administración.[102]

En materia regulatoria, el Congreso de Estados Unidos recientemente aprobó la Take It Down Act (S. 146[103]) Se trata de una norma que evidencia una tendencia que se reproduce en otros países y tiene como objeto abordar una problemática que se ha popularizado en los debates sobre los riesgos derivados de uso inteligencia artificial generativa, siendo esta la producción y reproducción de *deepfakes* de imágenes íntimas o de contenido sexual que afecta diferencialmente a mujeres.

La norma incluye medidas amplias que pueden resultar en prácticas desproporcionadas de censura y no contiene disposiciones que prevengan los abusos. Por ejemplo, plantea un sistema de"notificación y eliminación" que obliga a plataformas a retirar contenido señalado en un plazo de 48 horas y a hacer "esfuerzos razonables" para eliminar copias conocidas. La definición amplia de lo que puede ser denunciado y la ausencia de salvaguardas efectivas puede facilitar el abuso del sistema para eliminar discurso legítimo, incluyendo sátira, periodismo y comentarios políticos[104].

6.2. China: medidas administrativas para los servicios de IA generativa

En agosto de 2023 entraron en vigor una serie de medidas administrativas[105] provisorias que establecen obligaciones para los proveedores de servicios de IA generativa al público, comprendiendo a aquellos sistemas que generen texto, imágenes, audio, video u otros contenidos. Las medidas no aplicarán para los sistemas que se encuentren en las etapas de

---

[101] Los Angeles Times (2025) Trump firma decreto sobre desarrollo de inteligencia artificial 'libre de sesgo ideológico'. Disponible en: https://www.latimes.com/espanol/california/articulo/2025-01-23/trump-firma-decreto-sobre-desarrollo-de-inteligencia-artificial-libre-de-sesgo-ideologico

[102] ACLU, Trump's Efforts to Dismantle AI Protections, 11 de febrero de 2025, Disponible en: https://www.aclu.org/news/privacy-technology/trumps-efforts-to-dismantle-ai-protections-explained

[103] S.146 - TAKE IT DOWN Act. Disponible en: https://www.congress.gov/bill/119th-congress/senate-bill/146

[104] Electronic Frontier Foundation. The TAKE IT DOWN Act: A Flawed Attempt to Protect Victims That Will Lead to Censorship, 11 de febrero de 2025. Disponible en: https://www.eff.org/deeplinks/2025/02/take-it-down-act-flawed-attempt-protect-victims-will-lead-censorship

[105] China, Interim Measures for the Management of Generative Artificial Intelligence Services, 7 de octubre de 2023. Original disponible en https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm. Traducción disponible en https://www.chinalawtranslate.com/en/generative-ai-interim/

investigación, desarrollo o uso, siempre y cuando no sean ofrecidos o puestos a disposición para el público. El enfoque utilizado busca otorgar igual importancia al desarrollo y a la seguridad de los sistemas, permitiendo la innovación y la gobernanza de la IA generativa.

En función de la normativa, la provisión y el uso de la IA generativa deben respetar los valores socialistas (contenidos prohibidos, entre ellos, los que sean considerados subversivos o dañen la imagen nacional); prevenir la discriminación y respetar los derechos de propiedad intelectual y secretos comerciales, sin utilizarse monopólicamente o competir de forma desleal.

Asimismo, los proveedores deben emplear medidas efectivas de transparencia y mejorar la precisión y confiabilidad; asegurar que los datos de entrenamiento hayan sido obtenidos de fuentes legales, de calidad, y sean objetivos y diversos; implementar reglas de etiquetado. También se establece la responsabilidad de los proveedores como productores del contenido en línea y que, cuando tomen conocimiento de contenido o actividades ilegales, adopten medidas como la restricción de generación y transmisión, así como su remoción, registro y reporte a los organismos pertinentes.

Las medidas también refieren al cumplimiento de las provisiones del Consejo de Estado de China respecto de la gestión de algoritmos de recomendación[106] y sobre contenidos deepfake[107], dictadas en enero y noviembre de 2022, respectivamente.

En marzo de 2024, un grupo de expertos conformado por académicos de universidades chinas publicaron un proyecto de ley para la regulación general de los sistemas de IA[108]. El texto propuesto sugiere un enfoque pro industria, en línea con las anteriores medidas, y, si bien no hay certezas sobre si será aprobado, habría una fuerte señal en ese sentido.

Como puntos a resaltar contiene disposiciones sobre discriminación algorítmica, privacidad y etiquetado de contenido generado por IA. También establece que dicha tecnología no debería ser utilizada para analizar o evaluar comportamientos y hábitos de individuos, así como tampoco intereses o información económica, crediticia o de salud. También aborda la propiedad intelectual, estableciendo que los proveedores de IA podrían utilizar datos con copyright en conjuntos de datos de entrenamiento, siempre y cuando el uso sea diferente al propósito original de esos datos y no provoque un daño irrazonable sobre los derechos de

[106] China, Provisions on the Management of Algorithmic Recommendations in Internet Information Services, 31 de diciembre de 2021. Original disponible en http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm. Traducción disponible https://www.chinalawtranslate.com/en/algorithms/

[107] Provisions on the Administration of Deep Synthesis Internet Information Services, 25 de noviembre de 2021. Original disponible en: http://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm. Traducción disponible https://www.chinalawtranslate.com/en/deep-synthesis/

[108] China Law Society, "Ley de inteligencia artificial (Borrador sugerido por académicos", 18 de marzo de 2024. Disponible en: http://www.fxcxw.org.cn/dyna/content.php?id=26910

titular, dejando la puerta abierta en favor del desarrollo de los sistemas de IA. Por su parte, los contenidos generados por IA podrían protegerse por propiedad intelectual en función del grado de contribución del usuario sobre el resultado final, siempre y cuando se trate de una persona humana.

Si bien el proceso regulatorio de China ha consistido hasta ahora en medidas específicas respecto de tecnologías o aplicaciones concretas de la IA, se espera que el proyecto de una regulación general vea la luz en el corto plazo. Según Math Sheehan, dado el peso que tiene China en el comercio mundial y su influencia cada vez mayor en la infraestructura de los países en desarrollo[109] harán que la regulación que adopte China afecte significativamente el panorama global en materia de gobernanza de la IA[110].

6.2. Reino Unido

Unos meses antes de que en China entraran en vigor medidas administrativas sobre la IA, en marzo de 2023 en Reino Unido, la Secretaria de Estado de Ciencia, Innovación y Tecnología presentó al parlamento un white paper llamado "Una aproximación favorable a la innovación en la regulación de la IA".[111] Este borrador de política pública definía cinco principios clave para orientar a los reguladores en la gestión de los riesgos asociados con la inteligencia artificial. Estos principios abarcan la seguridad, protección y robustez; la transparencia; la justicia; la rendición de cuentas; y la capacidad de impugnación y reparación.[112]

Posteriormente, en febrero de 2024, fue publicado el informe "Una aproximación favorable a la innovación en la regulación de la IA: respuesta del gobierno[113]" que se construyó luego de

---

[109] En este sentido, ver Article 19, China: The rise of digital repression in the Indo-Pacific, 18 de abril de 2024. Disponible en: https://www.article19.org/resources/china-the-rise-of-digital-repression-in-the-indo-pacific/

[110] Matt Sheehan. "China's AI Regulations and How They Get Made", 10 de julio de 2023, Carnegie Endowment for International Peace. Disponible en: https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

[111] Secretaría de Estado de Ciencia, Innovación y Tecnología de Reino Unido, "A pro-innovation approach to AI regulation", 29 de marzo de 2023. Disponible en: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#ministerial-foreword

[112] Pascua, S. (2023) Reino Unido: El Gobierno publica una propuesta de regulación de la inteligencia artificial. Instituto Autor. Disponible en: https://institutoautor.org/reino-unido-el-gobierno-publica-la-propuesta-de-regulacion-de-la-inteligencia-artificial.

[113] Secretaría de Estado de Ciencia, Innovación y Tecnología de Reino Unido, "A pro-innovation approach to AI regulation: government response", 6 de febrero de 2024. Disponible en: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response#summary-of-consultation-evidence-and-government-response.

realizar una consulta pública en la que se recibieron numerosas contribuciones. El informe continúa la línea del white paper y se centra en fomentar la innovación mientras al mismo tiempo que propone gestionar los riesgos asociados a esta tecnología, además de ratificar los cinco principios clave: seguridad y robustez, transparencia, justicia, responsabilidad y capacidad de impugnación. También destaca los beneficios de abordar la regulación de manera flexible y adaptativa, en vez de adoptar reglamentaciones más rígidas.

A partir de dicha estrategia y dentro de las medidas destinadas a promover la transparencia de los sistemas de IA, el Gobierno de Reino Unido dispuso la obligatoriedad del Algorithmic Transparency Recording Standard[114] para todas las dependencias del gobierno central. Se trata de uno los primeros estándares de transparencia algorítmica para el sector público, destinado a que las entidades puedan brindar información clara sobre los algoritmos que utilizan y las razones que así lo justifican, a partir de un mecanismo de transparencia que permite acceder a información relevante sobre los sistemas utilizados por las agencias gubernamentales.

Recientemente, en enero de 2025, el Departamento de Ciencia, Innovación y Tecnología presentó al Parlamento un Plan de Acción sobre las Oportunidades de la IA.[115] Este plan establece como objetivo impulsar el crecimiento económico a través de los sistemas de inteligencia artificial, a partir de los siguientes pilares fundamentales: el desarrollo de la infraestructura y talento, promover la adopción de sistemas basados de inteligencia artificial en el sector público y privado y el impulso de una IA soberana y segura de desarrollo local. El plan de acción se compone de 50 recomendaciones específicas que determinan las prioridades del gobierno en materia de IA.

El plan se centra en expandir la infraestructura de IA del Reino Unido, incluyendo una inversión a largo plazo para mejorar la capacidad de investigación y desarrollo en IA. Esto incluye el establecimiento de "AI Growth Zones" para acelerar la construcción de centros de datos de IA y mejorar la capacidad computacional. Busca también impulsar la adopción de IA en todo el sector público y privado, aumentando la productividad. Además, el objetivo es que el Reino Unido se convierta en un "creador" de IA, en lugar de solo un usuario, desarrollando campeones nacionales en capas críticas de la pila de IA. Esto permitiría al país beneficiarse económicamente del avance de la IA y tener influencia en su gobernanza y seguridad.

---

[114] Gobierno de Reino Unido, "Algorithmic Transparency Recording Standard", Disponible en: https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies.

[115] Departamento de Ciencia, Innovación y Tecnología de Reino Unido, "Independent report: AI Opportunities Action Plan", 13 de junio de 2025, https://Depwww.gov.uk/government/publications/ai-opportunities-action-plan. Disponible en: https://www.gov.uk/government/publications/ai-opportunities-action-plan.

Se reconoce la necesidad de formar decenas de miles de profesionales en IA a través de canales que van desde aprendizajes hasta educación superior, y se busca aumentar la diversidad en el sector tecnológico. Aunque no establecen regulaciones específicas, enfatiza la importancia de equilibrar la innovación con medidas de seguridad y supervisión adecuadas. Menciona la necesidad de reformar las leyes de propiedad intelectual para facilitar el uso de contenido por desarrolladores de IA. Finalmente, el plan de acción se pronuncia sobre la necesidad de proporcionar acceso a conjuntos de datos públicos para apoyar a las empresas que trabajan con IA, lo que podría impulsar la innovación y el desarrollo de nuevas aplicaciones. También propone crear una Biblioteca Nacional de Datos y que el gobierno debe estar preparado para asumir algunos riesgos en un contexto de incertidumbre.

El marco regulatorio del Reino Unido, por un lado se destaca por su desarrollo en la rama ejecutiva, evitando expedir regulaciones mediante vías legislativas y absolutamente vinculantes, cuestión que ha sido abordada de manera expresa y justificada por las autoridades aludiendo la necesidad de regulaciones flexibles que permitan su fácil adaptación a los cambios tecnológicos. Por el otro, resalta de la regulación formulada el enfoque pro innovación que marca como prioridad para el país posicionarse como un líder mundial en materia de inteligencia artificial, lo que parece elevar los esfuerzos en el ámbito de desarrollo económico, más que un enfoque proteccionista de derechos humanos.

7. Hallazgos

Desde las recomendaciones y principios que surgieron en el seno de bloques y foros internacionales hasta la aprobación de las primeras regulaciones vinculantes, hoy nos encontramos en medio de una dinámica global que busca hacer frente a las problemáticas actuales y potenciales que presenta el vertiginoso desarrollo de la IA. Estamos frente a una carrera , similar a la espacial, en la que varios Estados buscan promover activamente el desarrollo de esta tecnología por sus potenciales efectos benéficos para su desarrollo y sus economías. En febrero de 2025, la presidenta de la Comisión Europea, Ursula Von der Leyen, anunció la puesta en marcha de la iniciativa InvestAI para movilizar 200.000 millones de euros en inversión en inteligencia artificial, [116] luego de que la administración de Donald Trump anunciara un plan de inversión de 500.000 millones de dólares mediante la alianza privada de los tres mayores desarrolladores de IA de Estados Unidos.[117]

---

[116] Durante el anuncio, Von der Leyen destacó que: “La IA mejorará nuestra asistencia sanitaria, estimulará nuestra investigación e innovación e impulsará nuestra competitividad. Queremos que la IA sea una fuerza para el bien y para el crecimiento”. Disponible en: https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence

[117] CNN, "Trump anuncia una inversión de US$ 500.000 millones en infraestructura de IA en EE.UU", 21 de enero de 2025. Disponible en:

La carrera por fomentar e invertir en IA se entrelaza con la necesidad de establecer marcos regulatorios que impulsen la innovación, al tiempo que garanticen un control efectivo sobre la implementación y los usos de estas tecnologías. Estos "modelos regulatorios" pueden ejercer un efecto modelizador en otros países, influyendo en el diseño de sus propias normativas. Si bien este fenómeno responde a múltiples factores, es fundamental que los países que busquen adoptar disposiciones en esta materia consideren los marcos y estándares internacionales de derechos humanos, el contexto local, la capacidad institucional para su implementación y los intereses en juego, tanto del sector público como del privado.

Si bien es muy pronto para evaluar estas iniciativas, lo cierto es que existe una "carrera global" por encontrar el marco regulatorio adecuado para este tipo de tecnologías. Resulta relevante, entonces, estudiar—como aquí se ha hecho—las diferentes propuestas. Este análisis resulta crucial para favorecer un debate informado y democrático. Por ello esperamos que este informe sirva como herramienta útil para la sociedad civil, la academia, los funcionarios públicos y el sector privado. Facilitar el acceso a información sistematizada sobre la regulación de la IA contribuye a la formulación de políticas basadas en evidencia y fortalece la capacidad de los actores involucrados para participar en la toma de decisiones de manera informada y crítica.

Las iniciativas regulatorias en la materia, a nivel global, pueden clasificarse en dos grandes grupos: los instrumentos vinculantes y los no vinculantes. Ambos tienen potenciales efectos modelizadores, pero se diferencian a grandes rasgos en los enfoques que adoptan, evidenciando así las prioridades de los entes regulatorios frente al tratamiento de la IA. A continuación, abordaremos los hallazgos respecto al predominio de determinados enfoques en los instrumentos vinculantes y no vinculantes y las relaciones influyentes de unos sobre otros.

La influencia de los principios delineados en la Recomendación de la OCDE de 2019 puede verse en las posteriores propuestas del G20 y el G7, que adoptaron un mismo enfoque no vinculante y basado en riesgos. Hasta el momento, el Proceso de Hiroshima del G7 es el único que propuso incipientes instrumentos para su implementación. Sin embargo, la protección de los derechos humanos se encuentra prácticamente ausente o únicamente mencionada como un marco cuyo cumplimento se da por sentado, sin detenerse en el reconocimiento de las posibles afectaciones concretas ni en recomendaciones al respecto. La Global Partnership on AI fue desarrollada en torno a un compromiso compartido con la Recomendación de la OCDE sobre Inteligencia Artificial. Esta iniciativa está orientada a la generación de conocimiento, como un modelo positivo de enfoque global y de múltiples partes interesadas.

En el marco de la ONU, el proceso de la UNESCO sobre ética de la IA, aunque voluntario, construye sobre las normas de DDHH y se apoya en la colaboración para la construcción de

https://cnnespanol.cnn.com/2025/01/21/economia/trump-inversion-millonaria-infraestructura-ia-oracle-openai-softbank-trax.

marcos regulatorios a nivel nacional que aborden los principios éticos que deben guiar la gobernanza de la tecnología así como los mecanismos para mitigar sus riesgos. La iniciativa del Secretario General de las Naciones Unidas para la conformación de una estrategia de gobernanza global incorpora la consideración de posibles afectaciones a los DDHH de individuos y comunidades y sugiere un anclaje en las normas del derecho internacional de los DDHH y en los órganos especializados. También es importante destacar la primera resolución de la Asamblea General sobre IA que, si bien tiene un peso mayoritariamente simbólico por resultar del consenso de la totalidad de los Estados que conforman la ONU, atiende cuestiones como la desigualdad en materia de desarrollo y la brecha digital. Sin embargo, ésta no tiene carácter vinculante ni aborda el uso de la IA en el ámbito militar. El Global Digital Compact propone un enfoque basado en riesgos así como la transparencia, protección de los derechos humanos, rendición de cuentas y supervisión humana de los sistemas de IA.

Entre los instrumentos vinculantes, es probable que el Reglamento de la UE tenga un impacto importante en la discusión global y futuras regulaciones en la materia (aún a nivel nacional), ante lo cual resultará preciso mantener presente aquellas cuestiones más problemáticas: un enfoque que ubica a los riesgos en el centro, dejando a los DDHH en un segundo lugar, y delega su mitigación a la autoevaluación que realicen los proveedores e implementadores de sistemas de IA; prohibiciones permisivas para usos en situaciones de tensión con los DDHH, incluyendo tecnologías de vigilancia biométrica; y exclusiones y excepciones del régimen para aplicaciones en el ámbito de defensa, seguridad nacional y migraciones. Es llamativo que en una regulación tan extensa y con tanto detalle no hayan mayores referencias y salvaguardas del derecho a la libertad de expresión. Las consecuencias de la implementación de este Reglamento aún son desconocidas pero sí sabemos que la propia Comisionada ha expresado que propondrá una implementación más "simplificada" de la AI ACT por sus potenciales efectos anti-innovación.

La Convención Marco del Consejo de Europa establece principios y normas destinadas a regular todo el ciclo de vida de los sistemas de IA para asegurarse de que sean acordes con los derechos humanos, la democracia y el estado de derecho. Además, adopta un enfoque basado en riesgos con obligaciones que en algunos casos requieren una mayor especificación o detalle, y deja vía libre para los sistemas de IA en los ámbitos de defensa y seguridad nacional y una válvula de escape para que los Estados decidan la implementación por parte del sector privado.

En referencia a regulaciones nacionales, de las dos grandes potencias económicas, tanto EEUU como China se encuentran inmersos en sus propios procesos regulatorios. Mientras EEUU adoptó ciertas medidas a nivel de su administración federal, con el plan de desarrollar instrumentos sectoriales a través de sus distintas agencias especializadas y un marco voluntario de gestión de riesgos para luego dar marcha atrás con ciertas acciones, China dictó medidas administrativas obligatorias en materia de algoritmos de recomendación, IA generativa y

contenidos *deepfake*, caracterizadas por un enfoque de control estatal centralizado y obligatorio para todos los actores. No obstante, los avances regulatorios de EEUU sufrieron un retroceso con la nueva administración de Trump que derogó las medidas adelantadas por el gobierno de Biden, lo que ha puesto al panorama nacional en materia de gobernanza de la IA en una situación de incertidumbre.

El Reino Unido, por el momento, ha presentado un white paper, la respuesta del gobierno y el plan de 2025 y ha decidido no implementar una regulación horizontal específica sobre inteligencia artificial.[118] En su lugar, propone un marco basado en principios que será aplicado por los reguladores sectoriales existentes según sus respectivas competencias. Este enfoque no estatutario busca ofrecer mayor flexibilidad y capacidad de adaptación frente al rápido desarrollo de la tecnología, con la finalidad de promover la innovación y posicionar al Reino Unido como una potencia en materia de desarrollo de IA. Sin embargo, el gobierno adelantó que luego de revisar durante un período la implementación de los principios, prevé introducir una obligación legal para los organismos reguladores que les exija tenerlos debidamente en cuenta.[119]

Si los países en desarrollo no logran influir en la gobernanza de la inteligencia artificial, es posible que la tecnología y la regulación asociada a ésta termine reflejando los intereses de las naciones más ricas, en ocasiones en detrimento de aquellos con menos recursos. Esto podría limitar la capacidad de desarrollar ecosistemas locales de IA que impulsen el desarrollo sostenible o generar disrupciones al adoptar marcos regulatorios ajenos a sus realidades. En este contexto, las diferencias en las prioridades de gobernanza de la IA entre las grandes jurisdicciones que establecen estándares pueden acentuar estos desafíos."[120]

El análisis de las iniciativas y regulaciones existentes revela una tendencia clara hacia la adopción de enfoques basados en riesgos para la regulación de la inteligencia artificial, tal

---

[118] White & Case, AI Watch: Global regulatory tracker - United Kingdom. Disponible en: https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-kingdom

[119] Respuesta del gobierno, Párrafo 109. Disponible en: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response

[120] Gordon LaForge, Robert Muggah, and Gabriella Seiler, "Bridging the AI Governance Divide."

como se encuentra en la DSA,[121] lo que puede derivar en efectos adversos.[122] Si bien algunos de estos marcos incorporan referencias a la protección de los derechos humanos, su orientación prioritaria parece centrarse en la facilitación del desarrollo industrial de la IA. En realidad, el problema no radica tanto en que una empresa adopte un enfoque basado en riesgos, sino en que el Estado, a través de la regulación, delegue en actores privados la responsabilidad de identificar, valorar y mitigar dichos riesgos sin establecer las salvaguardas adecuadas.

La valoración de los riesgos debería surgir de un proceso legislativo transparente y contextualizado. Sin ese marco, clasificar ciertos riesgos como aceptables o inaceptables puede generar distorsiones normativas, facilitar abusos, reproducir sesgos y producir efectos de sobreprotección o subprotección que afecten el ejercicio efectivo de los derechos humanos.

En este sentido, consideramos que es importante incorporar obligaciones específicas destinadas a garantizar el respeto y protección de los derechos humanos a lo largo del ciclo de vida de los sistemas de IA,[123] alineadas con las normas y estándares del derecho internacional de los derechos humanos. Esto permitirá identificar con mayor precisión los impactos positivos y negativos de la inteligencia artificial, según el contexto en el que se la aplique.[124] Así, los procesos regulatorios no se limitarían únicamente a una lógica de gestión de riesgos, sino que incorporarían una perspectiva centrada en la protección y el equilibrio de derechos que, en determinadas circunstancias, pueden entrar en tensión o conflicto, como la libertad de expresión y el derecho al honor. Ya en el año 2018 la Relatoría Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión de Naciones Unidas había expresado preocupación por la exclusión de los derechos a la libertad de opinión y expresión de los debates públicos y políticos sobre la IA,[125] situación que se ha evidenciado en las reglamentaciones analizadas en este informe. Como esta tecnología se utiliza para efectuar

---

[121] Martin Husovec, "The Digital Service Act's Red Line: What the Commission Can and Cannot Do About Disinformation," October 1, 2024, Husovec, Martin, The Digital Service Act's Red Line: What the Commission Can and Cannot Do About Disinformation (January 10, 2024). Disponible en SSRN: https://ssrn.com/abstract=4689926 or http://dx.doi.org/10.2139/ssrn.4689926.

[122] Del Campo, . A, Álvarez Ugarte, R., Zara, N. (2025), *Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation*, CELE Research Paper No. 64. Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5161173 (aceptado para el Journal of Intellectual Property, Information Technology and Electronic Commerce, vol. 16, No. 2, 2025, en prensa).

[123] UNESCO, Consultation paper on AI regulation: emerging approaches across the world. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000390979

[124] Margaret Mitchell, The Pillars of a Rights-Based Approach to AI Development, Disponible en: https://www.techpolicy.press/the-pillars-of-a-rightsbased-approach-to-ai-development/

[125] Asamblea General de las Naciones Unidas. A/73/348. Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, párr. 43. Disponible en: https://documents.un.org/doc/undoc/gen/n18/270/45/pdf/n1827045.pdf

moderación de contenidos en múltiples plataformas y redes sociales, existe el peligro potencial de que una regulación que imponga a los sistemas de IA silenciar, prohibir o desmonetizar determinadas expresiones o contenidos restrinja ilegítimamente la libertad de expresión. Esto debería mantenernos alerta sobre algunos de los desarrollos regulatorios que abordamos en este informe.

A continuación esbozamos algunos interrogantes que podrían guiar la discusión y que a nuestro criterio requieren mayor investigación y desarrollo: ¿Cuáles son los factores de atribución de los *outputs* generados por sistemas de IA?[126] ¿Deberían atribuirse a sus desarrolladores y programadores, a otros actores en la cadena de producción, comercialización e implementación, o a los humanos que solicitan mediante *prompts* algún tipo de respuesta de la tecnología?[127] Quizás estamos ante una gama de posibles intervenciones por parte de diferentes actores que podrían determinar su potencial atribución, caso por caso. También nos encontramos en un estado particular del desarrollo tecnológico de la todavía fallida IA o, como la han denominado algunos, Ignorancia Artificial.[128]

¿Podemos considerar que los *outputs* generados por la IA constituyen un tipo de expresión? En caso afirmativo, ¿qué protección deberían merecer? Podríamos también pensar en la dimensión colectiva del derecho a la libertad de expresión y cómo éste podría verse afectado negativamente según el grado de injerencia que los Estados ejerzan sobre estas tecnologías y los tipos de respuestas que pueden brindar. Por ejemplo, al prohibir que generen "discursos o expresiones dañinas" o desinformación. No es un dato menor que los *outputs* de la IA pueden colaborar en informar el discurso y los debates en sociedades democráticas. Por lo tanto, podríamos preguntarnos si sujetar a la gobernanza de la IA a un estándar restrictivo inferior al que, por ejemplo, está sujeto el discurso de odio en el sistema de derechos humanos, constituye desde ya una posible violación a la libertad de expresión. ¿Varía el límite entre lo permitido y lo prohibido según utilicemos un enfoque basado en riesgos o un enfoque basado en derechos?

Kogan sostiene que los *outputs* de la IA están frecuentemente protegidos por la Primera Enmienda como una forma de expresión, aún cuando puedan facilitar la comisión de delitos. Para esto recurre a la analogía con aquellos en los que la expresión "tradicional" facilita la comisión de delitos. Allí el dolo es necesario para establecer la responsabilidad penal y es un

[126] Ayres, Ian and Balkin, Jack M., The Law of AI is the Law of Risky Agents without Intentions (June 01, 2024). University of Chicago Law Review Online (forthcoming 2024), Yale Law & Economics Research Paper, Yale Law School, Public Law Research Paper, Available at SSRN: https://ssrn.com/abstract=4862025 or http://dx.doi.org/10.2139/ssrn.4862025

[127] Ilan Kogan, Artificial Intelligence, Existential Risk, and the First Amendment, 27 U. Pa. J. Const. L. 156 (2025), p. 180.

[128] Flyvbjerg, Bent, AI as Artificial Ignorance (January 31, 2025). Available at SSRN: https://ssrn.com/abstract=5119382 or http://dx.doi.org/10.2139/ssrn.5119382

requisito necesario para excluir la protección de la Primera Enmienda. De este modo, el requisito protege la libertad de expresión evitando la censura o *chilling effect* de contenidos legítimos. Para el autor, el mismo estándar debería aplicarse cuando los contenidos provienen de sistemas automatizados como los de inteligencia artificial.[129]

Ayres y Balkin parecen adoptar una postura similar al afirmar que la IA es un agente riesgoso que no tiene intenciones, a diferencia de los humanos que la desarrollan, comercializan o utilizan. Si volvemos a la pregunta sobre atribución: ¿sería correcto atribuir una responsabilidad general a los programadores o desarrolladores por el "mal uso" que se haga de la tecnología que desarrollaron? ¿Son iguales todos los "malos usos"? ¿Un mal diagnóstico médico es igual a dar indicaciones sobre cómo elaborar una bomba casera? ¿Cómo se diferencian estos casos facilitados por esta tecnología de la posibilidad de acceder a esa misma información yendo a una biblioteca o realizando una simple búsqueda en Internet?

También podríamos pensar que el código es un tipo de expresión protegida por el derecho a la libertad de expresión,[130] puesto que puede transmitir información e ideas, y ser una herramienta de comunicación. Si aceptamos esto, ¿hasta donde es aceptable la intervención estatal respecto de lo que un código puede o no expresar?

Ante la posibilidad de censura previa es necesario ejercer cautela y procurar que toda regulación que aborde estas tecnologías cumpla con los requisitos de legalidad, necesidad y proporcionalidad. Esto implica que las regulaciones que de algún modo limiten la libertad de expresión, aun indirectamente, deben ser necesarias en una sociedad democrática para perseguir un objetivo o interés imperioso, ser precisas e idóneas, no genéricas, proporcionadas para la finalidad que persiguen y no debe existir un medio alternativo menos restrictivo.[131] También deben demostrar que reducen sustantivamente la probabilidad del daño que buscan evitar en general y no sólo en la utilización de la IA.[132] Pensar en términos de derechos y no de regulación de tecnologías es fundamental para no perder esto de vista.

Por último, consideramos que el desarrollo de marcos regulatorios debe garantizar la participación activa y coordinada de los Estados, la sociedad civil, sector técnico, la academia y el sector privado, asegurando un debate inclusivo que permita diseñar normativas que resguarden los derechos humanos sin obstaculizar la innovación tecnológica.

[129] Kogan., p. 190.
[130] Ver por ejemplo el caso *Bernstein v. United States*.
[131] Marco Interamericano.
[132] Kogan, p. 192.

8. Anexo: Cuadro comparativo de las iniciativas analizadas

	Alcance	Resultado	Enfoque	Implementación
OCDE	36 miembros, 10 adherentes	Principios no vinculantes	Basado en riesgos y DDHH	Comité de OCDE, observatorio y red de expertos
G20	20 miembros	Principios no vinculantes	Basado en riesgos y DDHH	Sin mecanismos de cumplimiento
G7	7 Estados	Principios y código de conducta voluntario	Basado en riesgos	Proceso de Hiroshima. Apoyo a GPAI
UNESCO	Global	Principios éticos y metodología de evaluación de impacto	Basado en riesgos y DDHH	Observatorio y red de expertos
ONU	Global	Compromisos no vinculantes (Resolución 78/265 de la Asamblea General y Pacto Digital Global) Informe sobre Gobernanza de la IA (Secretario General)	Basado en cooperación, riesgos y DDHH	Órgano Asesor sobre IA, Panel científico internacional.
Unión Europea	Regional, 27 Estados	Regulación horizontal de carácter vinculante. Obligaciones diferenciadas según niveles de riesgo	Basado en riesgos, impacto en DDHH y estándares	Oficina y Comité de IA, Foro Asesor, panel de expertos, autoridades nacionales.
Consejo de Europa	Global	Tratado internacional de carácter vinculante	Basado en riesgos	Conferencia de las Partes, monitoreo a nivel nacional
OEA	Regional, 34 Estados	Lineamientos de gobernanza de datos	Basado en gobernanza de datos	En consulta y pendiente de aprobación.
Estados Unidos	Nacional	Principios, acciones y marco de evaluación y gestión de riesgos. Órdenes Ejecutivas	Basado en riesgos	Consejo de IA de la Casa Blanca, departamentos y agencias federales. Comisión Federal de

		de promoción de la IA. Ley federal anti deepfakes.		Comercio. Instituto Nacional de Estándares y Tecnología (NIST)
China	Nacional	Medidas administrativas de alcance general	Basado en control estatal centralizado	Administración del Ciberespacio y Ministerios
Reino Unido	Nacional	White paper y plan de acción.	Basado en principios proinnovación	Secretaria de Estado de Ciencia, Innovación y Tecnología