
Inteligencia artificial: el desafío jurídico para la protección de los derechos humanos

Artificial intelligence: the legal challenge for the protection of human rights

MA. ELIZABETH DE LOS RÍOS URIARTE

Profesora e Investigadora en la Facultad de Bioética

Universidad Anáhuac México

ORCID: 0000-0001-9600-445X

MARCIAL ORLANDO CABRERA CANTARERO

Médico Pasante de Servicio Social en la Facultad de Bioética

Universidad Anáhuac México

ORCID: 0009-0008-7610-6131

INÉS MARQUINA GURREA

Médico Pasante de Servicio Social en la Facultad de Bioética

Universidad Anáhuac México

ORCID: 0000-0003-1810-7853

Fecha de recepción: 01 marzo 2024

Fecha de aceptación: 12 mayo 2024

SUMARIO: I. Introducción. II. La necesidad de poner a la inteligencia artificial en la lupa jurídica. III. Regulaciones existentes para el buen uso de la IA 1. Principios de Asilomar sobre Inteligencia Artificial (2017). 2. Declaración de Montreal (2018). 3. *Ethics guidelines for trustworthy AI* (2019). 4. *Rome Call for Ethics* (2020). 5. Recomendación sobre el uso ético de la Inteligencia Artificial de la UNESCO (2021). 6. Declaración de Bletchley (2023). IV. Ley de Inteligencia Artificial de la Unión Europea (2023). V. El Papel de la IA en el Desarrollo Sostenible. VI. Áreas de continuo riesgo y necesidad de vigilancia. VII. Conclusiones.

RESUMEN: El avance exponencial de la IA ha fomentado la necesidad de conocer las regulaciones existentes sobre la misma, con el objetivo de promover su adecuado uso, orientado hacia principios de carácter ético y la protección de los derechos humanos. En la actualidad, muchos son los beneficios que se obtienen de sistemas de IA, sin embargo, representa un gran desafío de carácter jurídico al momento de valorar su avance y regularlo a favor de valores como la dignidad y la libertad humana. La carencia actual de normas que controlen el desarrollo de la IA ha propiciado espacios para la elaboración de nuevas leyes en países europeos.

ABSTRACT: The exponential advance of AI has fostered the need to know its existing regulations, with the aim of promoting its appropriate use, oriented towards ethical principles and the protection of human rights. Currently, there are many benefits obtained from AI systems, however, it represents a great challenge of a legal nature at the time of assessing its progress and regulating it in favour of values such as human dignity and freedom. The current lack of regulations to control the development of AI has led to the development of new laws in European countries.

PALABRAS CLAVE: *inteligencia artificial, derechos humanos, dignidad humana, libertad, tecnología.*

KEYWORDS: *artificial intelligence, human rights, human dignity, freedom, technology.*

I. INTRODUCCIÓN

Hasta la mitad del año pasado 2023, la regulación existente en materia de diseño, venta, uso y aplicaciones de la inteligencia artificial, aunque loable en sus esfuerzos, era aún porosa; a partir de junio de ese mismo año, el parlamento europeo agudizó la mirada legislativa y adoptó una postura negociadora sobre una ley que regule los usos de la inteligencia artificial con base en la estimación del riesgo existente. Lo más interesante de esta ley es

que contempla, también la inteligencia artificial generativa que, a su vez incluye, entre otros, los programas de generación de texto como *Chat GPT*.

A pesar de que esta ley aplica sólo para el espacio europeo, todos los sistemas de inteligencia artificial que se diseñen en otros países, pero sean exportados a la Unión Europea, deberán pasar por el escrutinio de dicha regulación, de la misma forma que lo que se produzca en el espacio europeo, aunque sea exportado a otros países.

De esta manera, se obtiene ya una primera clasificación de los sistemas de inteligencia artificial basados en su nivel de riesgo para los derechos humanos que, aunque aún sujeta a múltiples críticas y mejoras, brinda un marco legal que sienta precedentes a su uso indiscriminado.

En este artículo, en un primer apartado, analizaremos las regulaciones existentes a nivel mundial sobre el diseño, programación, venta, uso y aplicaciones de programas de inteligencia artificial para determinar sus propósitos, alcances y limitaciones. En un segundo momento, se describirá el así llamado *Ley de Inteligencia Artificial de la Unión Europea* que está generando antecedentes para la primera ley de Inteligencia artificial gestada desde el espacio de la Unión Europea.

En un tercer momento, se reflexionará sobre los retos que se enfrentan a medida que se van desarrollando más y mejores sistemas tecnológicos con el propósito de identificar áreas en donde su vigilancia estrecha se hace más evidente debido a sus posibles atentados contra los derechos humanos.

A modo de conclusión, se estimará el uso de la inteligencia artificial como pivote que potencie las capacidades humanas sin sustituirlas.

II. LA NECESIDAD DE PONER A LA INTELIGENCIA ARTIFICIAL EN LA LUPA JURÍDICA

Los beneficios de los sistemas y programas creados con inteligencia artificial (IA en adelante) es incuestionable. En diversas áreas, ésta ha permitido una mayor eficiencia y rapidez en la producción y predicción de procesos y diagnósticos.

Desde los servicios de salud que permiten al usuario monitorizar sus signos vitales frecuentemente y llevar un mayor control de sus niveles de glucosa, presión arterial y oxigenación, hasta los robots que realizan cirugías de alta precisión con tasas de éxito mayúsculas, así como los aparatos radiológicos y electromagnéticos capaces de diferenciar células, procesar niveles y detectar imágenes patológicas y alertar tempranamente sobre posibles complicaciones posteriores, los usos de la IA en la salud han resultado muy benéficos.

De igual modo, en la industria agropecuaria, la monitorización a distancia de cultivos y el pastoreo inteligente evitan la pérdida de cosechas y la recolección puntual de granos, frutos y hortalizas para incrementar la cadena de producción y hasta para el traslado de estos alimentos a zonas de mayor necesidad.

En educación, los beneficios con robots que llegan a zonas donde hay carencias educativas importantes por falta de maestros o bien de escuelas y sistemas escolarizados ha resultado de gran ayuda para paliar estas lagunas; así mismo, los *chatbots* que tutoran y monitorizan el aprendizaje en las aulas y coadyuvan al maestro en el proceso de atención personalizada de los alumnos proporcionan grandes beneficios para mejorar los niveles de alfabetización y las conductas de socialización y colaboración.

Hay, empero, otros desarrollos y usos de la inteligencia artificial que, al menos no a primera vista, resultan tan beneficios y, más bien, se ubican como potencialmente peligrosos, tal es el caso de las aplicaciones en la industria bélica, la automotriz, el uso

y manejo del *Big data*, etc. Desde los soldados robot que disparan a objetivos detectados como similares a aquellos con los que fueron programados hasta sistemas de reconocimiento facial que generan patrones con base en la etnia de las personas y que, después, al usarse para otros fines, pueden presentar sesgos que generen conductas discriminatorias y hasta incriminatorias; estos usos levantan serias dudas y preocupaciones fundadas que ameritan pausar y pensar en mecanismos que protejan a las personas y sus derechos de posibles violaciones generadas voluntaria o involuntariamente por los sistemas y programas de IA.

Estas prácticas hacen necesario generar regulaciones que prevengan riesgos y alerten sobre posibles atentados a los derechos humanos. Es por ello por lo que, a medida que la tecnología ha avanzado, los intentos por regular sin obstaculizar la inteligencia artificial y su desarrollo también ha ido avanzando y volviéndose más escrupulosa y de mayor alcance.

Aunque el objetivo no es cohibir y, menos aún, prohibir el uso de inteligencia artificial en la vida de las personas, las regulaciones sí tienen por finalidad que desde el inicio del diseño se cuente con una reflexión ética y antropológica que permita desarrollar una inteligencia artificial centrada en la persona humana que la tenga siempre como fin y nunca como medio. Analizaremos ahora estos avances en la regulación existente.

III. REGULACIONES EXISTENTES PARA EL BUEN USO DE LA IA

El enfoque moderno de la inteligencia artificial, como se explica en el libro de Russell y Norvig (2020), destaca la importancia de un marco ético y de regulaciones bien fundamentadas para guiar el desarrollo y aplicación de la IA. Este enfoque es necesario para asegurar que la IA beneficie a la humanidad mientras minimiza los potenciales riesgos.

Aunque se mencionó en el apartado anterior que aún no se cuenta con un solo documento internacional al que se adhieran todos los países en el desarrollo e implementación de programas y dispositivos que operan con IA sí existen algunas regulaciones que son loables.

A continuación, se mencionarán y describirán brevemente:

1. Principios de Asilomar sobre Inteligencia Artificial (2017)

Tomando como base la ciencia ficción “Círculo Vicioso” de Isaac Asimov de 1989, quien, en sus obras, ideó robots que debían comportarse según ciertos estándares que se sostenían en tres leyes (Asimov 1989) (García Manrique 2007):

- 1) Primera Ley: Un robot no debe dañar a un ser humano o, por su inacción, dejar que un ser humano sufra daño.
- 2) Segunda Ley: Un robot debe obedecer las órdenes que le son dadas por un ser humano, excepto cuando estas órdenes se oponen a la primera ley.
- 3) Tercera Ley: Un robot debe proteger su propia existencia, hasta donde esta protección no entre en conflicto con la primera o segunda leyes.

Tomando entonces como base estas tres leyes y sus alcances, se desprenden los siguientes principios (ROBOTTECHNICS 2017):

Todo desarrollo deberá tener una meta de la investigación: la investigación debe ser beneficiosa para el ser humano, a nivel individual y colectivo.

Financiación de la investigación: la inversión en IA debería ir acompañada de fondos para la investigación.

Enlace entre ciencia y política: es importante que exista un vínculo sano entre los investigadores de IA y los legisladores.

Cultura de la investigación: entre investigadores y desarrolladores de IA, debe existir una atmósfera de reciprocidad, confianza, y trabajo en equipo.

Evitar las carreras: la cooperación activa es esencial para evitar contratiempos.

Seguridad: los sistemas de IA deberían ser seguros a lo largo de su vida operativa.

Transparencia en los fallos: si un sistema de IA causa daño debería ser posible determinar la razón.

Transparencia judicial: al momento de la participación de la IA en la toma de decisiones judiciales, dicha decisión debe ser comprensible y verificada por una autoridad humana competente.

Responsabilidad: quienes diseñan y desarrollan sistemas de IA deben tomar en cuenta las repercusiones morales del uso y mal uso de los sistemas que desarrollan.

Alineación de valores: los objetivos de los sistemas de IA deben ser coherentes con los valores humanos.

Valores humanos: los sistemas de IA deben ser compatibles con la dignidad humana, derechos, libertades y diversidad cultural.

Privacidad personal: las personas deben tener el derecho controlar la propia información, a eliminarlos o administrarlos según sus preferencias.

Libertad y privacidad: los sistemas de IA no deben atentar bajo ninguna circunstancia, en contra de la libertad personal o privacidad.

Beneficio compartido: los sistemas de IA deben beneficiar a la mayor cantidad posible de personas.

Prosperidad compartida: el beneficio económico generado por la IA debe ser compartido con todos.

Control humano: los sistemas de IA deben ser controlados por los seres humanos.

Sin subversión: los sistemas de IA no deben ir en contra de los procesos sociales y cívicos de los que depende la salud de la sociedad.

Carrera armamentística: cualquier tipo de carrera armamentística de armas autónomas letales debe ser evitada a toda costa.

Capacidad de precaución: los límites de la IA en el futuro son inciertos, por lo que es recomendable evitar hacer suposiciones. Es mejor adoptar una postura prudente al respecto.

Importancia: los futuros alcances de la IA avanzada pueden llegar a cambiar la historia, por lo que es importante planificar su desarrollo con cautela.

Riesgos: es imprescindible identificar los posibles riesgos asociados al desarrollo de la IA y buscar medidas de mitigación.

Automejora recursiva: los sistemas de IA tienen una función de automejora continua, por lo que deben estar sujetos a medidas específicas que controlen dicho crecimiento o mejora.

Bien común: los beneficios de la IA no deben ser reservados para un Estado u organización, en cambio, deben ser para el bienestar de toda la humanidad.

Aunque amplia, esta declaración contempla que, desde la misma idea de producto o servicio que se va a generar ya haya una intención de que sea algo beneficioso para la persona y la familia humana y no algo que sólo busque el beneficio individual o de unos cuantos.

2. Declaración de Montreal (2018)

Por iniciativa de la Universidad de Montreal y firmada por más de 70 instituciones educativas, gubernamentales, ONGS, etc. esta declaración contiene los siguientes principios:

Principio de bienestar: El desarrollo y uso de sistemas de inteligencia artificial (AIs por sus siglas en inglés) debe permitir el crecimiento y el bienestar de todos los seres “conscientes”.

Principio de respeto a la autonomía: Los sistemas de inteligencia artificial deben desarrollarse y utilizarse respetando la autonomía de las personas, con el objetivo de aumentar el control de las personas sobre sus vidas y su entorno.

Principio de protección a la privacidad y la intimidad: La privacidad y la intimidad deben protegerse de la intrusión de sistemas de inteligencia artificial y de los sistemas de adquisición y archivo de datos (DAAS por sus siglas en inglés).

Principio de solidaridad: El desarrollo de sistemas de inteligencia artificial debe ser compatible con el mantenimiento de los lazos de solidaridad entre las personas y las generaciones.

Principio de participación democrática: Los sistemas de inteligencia artificial deben cumplir con los criterios de inteligibilidad, justificación y accesibilidad, y deben estar sujetos al escrutinio democrático, el debate y el control.

Principio de equidad: El desarrollo y uso de sistemas de inteligencia artificial debe contribuir a la creación de una sociedad justa y equitativa.

Principio de inclusión y diversidad: El desarrollo y uso de sistemas de inteligencia artificial debe ser compatible con el mantenimiento de la diversidad social y cultural y no debe restringir el alcance de las elecciones de estilo de vida o experiencias personales.

Principio de prudencia: Toda persona involucrada en el desarrollo de IA debe tener precaución anticipando, en la medida de lo posible, las consecuencias adversas del uso de sistemas de inteligencia artificial y tomando las medidas apropiadas para evitarlas.

Principio de responsabilidad: El desarrollo y uso de sistemas de inteligencia artificial no debe contribuir a disminuir la responsabilidad de los seres humanos en la toma de decisiones.

Principio de desarrollo sostenible: El desarrollo y uso de sistemas de inteligencia artificial debe llevarse a cabo para garantizar una fuerte sostenibilidad ambiental del planeta.

Aunque cuestionable por el uso de palabras como *sintientes* carentes de una definición precisa y que, por cuestiones de espacio no se analizarán en este artículo, esta declaración hace palpable la necesidad de que todo uso de IA vaya orientado a buscar el bien de las personas y, de las creaturas en general (suponiendo que por *sintiente* se refiera a toda la creación), además es digno de resaltar el principio de prudencia que invita a anticipar los posibles riesgos de los productos creados con inteligencia artificial y detener aquellos en donde no se puedan anticipar las consecuencias hasta que el panorama resulte más claro, así como el de responsabilidad por el que se deben asumir las consecuencias que se desprendan del uso de la IA.

Huelga decir que, aunque no quede explícito en el principio mencionado, esta responsabilidad se hace extensiva a todos los involucrados: desde el diseñador, el programador, el vendedor y el usuario final del producto.

3. *Ethics guidelines for trustworthy AI (2019)*

El Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial creado por la Comisión Europea fue el principal protagonista de estos siete lineamientos:

Apropiamiento y empoderamiento humano: que los programas de IA fortalezcan las capacidades humanas y que se orienten a procesos de aprendizaje y autonomía de las personas.

Solidez tecnológica y seguridad: que ofrezcan explicaciones sólidas y serias sobre los modelos de su desarrollo y que no dejen nada al azar ni a lo accidental para que pueda ser analizado y corregido todo error que se pudiera presentar.

Privacidad y gobernabilidad: que se cuiden estrictamente los datos personales sobre todo en los programas de IA que se alimentan de ellos y que exista la cooperación internacional para que los sistemas puedan ser utilizados para beneficio social y el fortalecimiento de instituciones.

Transparencia: se exige que cada paso del desarrollo de la IA quede debidamente documentado y que esta información esté disponible en todo momento para consulta o bien para facilitar a otros el proceso de sus propios desarrollos.

No discriminación, diversidad y justicia. Existe la obligación de evitar todo acto de real o posible discriminación en los algoritmos con los que las operaciones de IA son programadas y bajo ninguna condición deberán de priorizarse actos de injusticia o marginación. La IA jamás deberá favorecer la brecha entre ricos y pobres y, para ello, deberá cuidar que los costos de acceso a dichas tecnologías sean accesibles para todos.

Bienestar ambiental y social: se contempla el equilibrio y cuidado ambiental en el diseño y producción de IA así como el alcance social de los beneficios ofrecidos.

Confiabilidad: la transparencia en su producción fomentará su confiabilidad; no obstante, cualquier sistema de IA está sujeto a sufrir fallas y cometer errores por lo que se hace aún más necesario hacer los sistemas más confiables y validarlos en sus funciones finales.

4. *Rome Call for Ethics de la Renaissance Foundation (2020)*

Esta declaración es, quizá, una de las más trascendentes en la materia debido a que fue firmada por el presidente de la Academia Pontificia para la Vida, Monseñor Vincenzo Paglia, el CEO de Microsoft, el Dr. Brad Smith, el CEO de IBM el Dr. John Kelly III, director de la FAO, el Dr. Dongyu Qu, la ministra italiana de innovación, la Dra. Paola Pisano y el Papa Francisco.

La relevancia queda implícita en tanto que fue firmada por al menos dos de los más grandes desarrolladores de la tecnología y de la IA como Microsoft e IBM, además de representar una adhesión al bien común a través de la firma del director de la FAO en una intención de que la IA sea utilizada para bien.

Este llamamiento fue ratificado y actualizado en enero de 2023 también por un representante del judaísmo, el rabino Eliezer Simha Weisz y otro del islam, el Jeque Abdallah bin Bayyah, convirtiéndose en el primer documento interreligioso de esta índole.

El documento contempla tres áreas transversales para las que la inteligencia artificial debería servir y ser desarrollada:

- 1) *Educación*: considera la necesidad de asumir el compromiso de cambiar el mundo a través de la IA con y para las generaciones más jóvenes.
- 2) *Ética*: considera el preámbulo de la Declaración Universal de los DH's sobre que todos los seres humanos nacen libres e iguales en dignidad y derechos.
- 3) *Derechos humanos*: exhorta a que el desarrollo de la IA sirva para proteger a las personas, especialmente a los más pobres y débiles y al medio ambiente.

Sobre estas áreas se desglosan 6 principios:

- 1) *Transparencia*: todos los procesos de desarrollo de inteligencia artificial deben ser de fácil entendimiento para todos los actores involucrados y, por ende, replicables.
 - 2) *Inclusión*: no deben discriminar a nadie porque todos tenemos la misma dignidad.
 - 3) *Responsabilidad*: siempre debe haber alguien que asuma la responsabilidad de lo que las máquinas hacen o pueden hacer.
 - 4) *Imparcialidad*: no deben contener sesgos o conflictos de interés.
 - 5) *Confiabilidad*: deben ser confiables para hacer aquello para lo que se les programa y no desviarse a otras actividades no previstas.
 - 6) *Seguridad y privacidad*: se deberá respetar siempre la privacidad de los usuarios y mantener estándares de seguridad que no expongan datos personales.
5. *Recomendación sobre el uso ético de la inteligencia artificial de la UNESCO (2021)*:

Fue adoptada por los 193 estados miembros de la UNESCO. En ella se promueven valores que deben inspirar el diseño y producción, y principios que deben orientar los usos y aplicaciones de tecnología con inteligencia artificial. Entre los valores están:

- 1) Promoción y defensa de la dignidad y derechos humanos.
- 2) Promoción y defensa de los ecosistemas y medio ambiente.
- 3) Promoción y respeto por la diversidad y la inclusión.

Generación de sociedades justas y pacíficas que estén interconectadas y con una visión de bien común.

Es de reconocerse que este documento va más allá de la simple enumeración de principios y propone una reflexión ética desde los valores que debe estar a la base del seguimiento de los principios y, de igual modo, es importante señalar que hay una mención explícita al cuidado del medio ambiente y a la paz como anhelo común de la humanidad.

Respecto de los principios se mencionan diez:

- 1) Proporcionalidad e inocuidad.
- 2) Seguridad y protección.
- 3) Equidad y no discriminación.
- 4) Sostenibilidad.
- 5) Privacidad y protección de datos.
- 6) Supervisión y decisión humanas.
- 7) Transparencia y explicabilidad.
- 8) Responsabilidad y rendición de cuentas.
- 9) Sensibilización y educación.
- 10) Gobernanza mundial y colaboración.

Se pretende que estos principios puedan ser aplicados en 11 diferentes ámbitos de acción:

- 1) Evaluación del impacto ético
- 2) Gobernanza y administración ética
- 3) Política de datos
- 4) Desarrollo y cooperación internacional

- 5) Medio ambiente y ecosistemas
- 6) Género
- 7) Cultura
- 8) Educación e investigación
- 9) Comunicación e información
- 10) Economía y trabajo
- 11) Salud y bienestar social

Estos once ámbitos de acción pueden considerarse las herramientas necesarias para que los principios se ejecuten y los valores permeen las prácticas y los usos, incluso se relacionan con los derechos humanos en tanto que proporcionan estrategias integrales donde la IA puede jugar un papel protagónico en la protección y promoción de la persona humana y sus derechos y, por ende, facilitar y sumar esfuerzos para lograr la paz.

Otro aspecto muy valioso de esta declaración es su firma y adhesión por parte de 197 países que, aunque su carácter no sea vinculante, en el ámbito del derecho internacional, a la firma de tantos países puede ser un factor de presión para que se avance en la misma dirección.

Es digno de mencionarse que esta regulación marca también la necesidad de dar seguimiento y hacer evaluación de la incorporación y adherencia a estos valores y principios en las legislaciones nacionales.

6. Declaración de Bletchley (2023)

Firmada en Inglaterra en el histórico parque Bletchley donde Alan Turing y otros científicos descifraron el código enigma del que se habló en el marco de la “Cumbre por la Seguridad de la Inteli-

gencia Artificial”. Un total de 28 países con representantes de los cinco continentes la firmaron, dándole con ello un carácter internacional fuerte al igual que la anterior sólo que de fecha más reciente y, por ende, más actualizada.

Contempla 5 principios que son:

- 1) *Innovación*: exhorta a garantizar la seguridad en las regulaciones sin diluir ni alejar la innovación.
- 2) *Inclusión*: destaca la importancia de que la IA sea utilizada para reducir la brecha digital existente.
- 3) *Adecuación*: plantea la necesidad de desarrollar estrategias para medir y mitigar los posibles riesgos de aquellas aplicaciones de IA potencialmente peligrosas.
- 4) *Compresión*: buscar inteligibilidad en todas las etapas del ciclo de vida de la IA con el fin de dotarla de transparencia.
- 5) *Cooperación*: trabajar en conjunto para desarrollar una IA centrada en la persona, segura y confiable.

El documento deja ver una enorme conciencia al afirmar su carácter provisional, es decir, considera que el desarrollo de la IA seguirá aumentando y perfeccionándose por lo que la reflexión ética deberá continuar profundizándose también.

IV. LEY DE INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA (2023)

En junio 2023 el parlamento europeo votó por un proyecto que abría la discusión en torno a la permisividad o prohibición del desarrollo y uso de sistemas y programas creados con inteligencia artificial. Dicha discusión ha tenido como eje rector la estimación del riesgo que dichos programas conllevan para la protección de los derechos humanos.

Aunque objeto de múltiples críticas aún, el proyecto ha considerado que, en lo general, la inteligencia artificial es beneficiosa en áreas como la medicina, la educación y la agricultura, sin embargo, no está exenta de usos que impliquen vulnerar derechos humanos, es por ello que resulta preciso clasificar éstos según el alcance de sus riesgos.

Esta regulación es la única en el mundo hasta la actualidad ya que contempla sanciones y multas a diferencia de las que se han expuesto en páginas anteriores que adoptan un estilo a modo de recomendación más que de obligación.

Uno de los problemas fundamentales a los que se enfrentó esta propuesta es la definición de inteligencia artificial ya que no existe una única adoptada de forma unánime por todos los países, sin embargo, tomaron como base la siguiente: *“software that is developed with [specific] techniques and approaches [listed in Annex 1] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with”*.

Al adoptar esta definición, se parte de los “sistemas” de IA lo que comprende a los softwares desarrollados con base en *machine learning*, los sistemas de lógica y conocimiento, así como los cálculos estadísticos.

Por su parte, la estimación de los riesgos de estos sistemas se clasifica en cuatro:

- 1) Riesgo inaceptable
- 2) Riesgo alto
- 3) Riesgo limitado
- 4) Riesgo bajo o mínimo

Las clasificadas como riesgos inaceptables quedan estrictamente prohibidas tanto en su desarrollo, venta, uso y/o aplicaciones e incluyen los siguientes:

- 1) Sistemas dañinos o manipulativos de técnicas subliminales.
- 2) Sistemas que exploten grupos vulnerables (física o mentalmente)
- 3) Sistemas usados por autoridades públicas para clasificar sujetos
- 4) Sistemas de identificación biométrica en lugares de acceso público con fines judiciales.

La segunda clasificación, de alto riesgo, se divide, a su vez, en sistemas que incluyen partes de seguridad de otros productos o bien que pertenecen a la lista de productos de salud y seguridad oficial según la legislación para tal propósito y los sistemas que se abarquen dentro de las siguientes áreas:

- 1) Identificación biométrica y categorización de personas
- 2) Manejo y operación de infraestructura crítica
- 3) Entrenamiento con fines educativos y vocacionales
- 4) Empleo y seguridad en el trabajo
- 5) Acceso y disfrute de servicios y beneficios privados y públicos
- 6) Seguridad y cumplimiento de la ley
- 7) Migración, asilo y control de fronteras
- 8) Administración de justicia y procesos democráticos

Estos sistemas clasificados como de alto riesgo deberán registrarse en una base de datos europea creada para estos fines antes de ser puestos en el mercado y utilizados. Además, deberán ser sometidos a controles estrictos de pruebas, entrenamiento, seguridad, transparencia y ciberseguridad.

La tercera clasificación, de riesgo limitado, se refiere a aquellos sistemas de IA que interactúan con los seres humanos como *chatbots*, sistemas de reconocimiento de emociones, sistemas de categorización basados en datos biométricos, contenidos de audio o video considerados “*deep fakes*” únicamente deberán cumplir con obligaciones referentes a la transparencia en sus usos y propósitos.

Por último, los sistemas considerados como de bajo riesgo estarán libres de requisitos y prohibiciones; no obstante, se considera la elaboración de un código de conducta que sea recomendable para los desarrolladores y usuarios de este tipo de IA.

La ley también contempla la conformación de un Consejo Europeo de Inteligencia artificial que supervise que se cumplan las regulaciones necesarias, así como un conjunto de medidas correctivas, sanciones y hasta multas que pueden ascender a los 30 millones de euros para aquellos que desarrollen, vendan o usen sistemas de IA sin cumplir con las exigencias requeridas.

Un aspecto relevante de esta ley es que es muy consciente de la necesidad de innovar y, para ello, de generar un espacio seguro para probar y entrenar los sistemas de IA antes de lanzarlos al mercado.

Ahora bien, esta ley europea, como todo documento legislativo, está sujeta a críticas y propuestas de mejora. Dentro de éstas se encuentra la insistencia de algunas de las partes interesadas en afinar la definición propuesta por considerarla demasiado amplia. Otra propuesta versa sobre las estimaciones de riesgos y niveles de responsabilidad pues, en un sistema de IA se involucran muchos roles y funciones y no todos deberán asumir el mismo nivel de obligaciones ni todas las acciones que se realicen en el desarrollo de un sistema de IA conllevan el mismo nivel de riesgo; así también, se propone que la estimación se haga mediante un análisis de cada caso en particular sin caer en generalizaciones que pueden afectar los beneficios de un determinado sistema o programa. En este aspecto, es de la mayor consideración la petición de estimar, también,

los riesgos de dichos sistemas de IA sobre el medio ambiente y no sólo sobre los derechos humanos.

Otra de las preocupaciones externadas por algunas partes es el cálculo de los costos que tendría adherirse a las regulaciones solicitadas por parte de las industrias y empresas que desarrollan estos sistemas y programas. En una estimación del Centro de Datos para la Innovación, se calcula un costo de 31 billones de euros para los próximos 5 años, lo que detendría la inversión de las empresas dedicada al desarrollo tecnológico.

Otro aspecto muy importante dentro de las propuestas de mejora es el hecho de que las personas, en lo individual, cuenten con mecanismos, instancias y procesos para denunciar daños sufridos o vulneraciones a sus derechos por programas de IA.

Esto es de especial importancia debido a que, aunque inicialmente los programas se aprueben y cumplan con los requisitos, por las características mismas de la IA que se alimenta del entorno, al momento de su uso, puede producir sesgos discriminatorios y vulnerar derechos humanos, por ello, se aconseja que los usuarios cuenten con un canal de denuncias que atienda estos casos.

A pesar de los esfuerzos para regular el uso de la inteligencia artificial y sus aplicaciones con el fin de proteger los derechos humanos, las regulaciones actuales pueden quedarse cortas frente a la apresurada evolución tecnológica. Estas normativas, aunque son esenciales, no siempre abarcan todas las áreas de desarrollo ni anticipan las consecuencias de tecnologías emergentes, lo que podría dejar vacíos críticos en la protección ante nuevos avances.

Por ejemplo, desarrollos recientes en inteligencia artificial como los sistemas autónomos avanzados y la inteligencia artificial generativa presentan desafíos que pueden no estar completamente contemplados por las regulaciones que ya existen. Además, la velocidad con la que se desarrollan estas tecnologías requiere un marco

regulativo que no solo se establezca, sino que también se adapte continuamente, permitiendo actualizaciones regulares y la integración de nuevas comprensiones éticas y técnicas. Por lo tanto, es crucial reconocer que las regulaciones actuales son parte de un proceso evolutivo y deben ser revisadas y adaptadas continuamente para asegurar que la implementación de la IA se organice de manera efectiva y ética con los principios de progresividad de los derechos humanos, dignidad y libertad.

V. EL PAPEL DE LA IA EN EL DESARROLLO SOSTENIBLE

En el estudio titulado “*The role of artificial intelligence in achieving the Sustainable Development Goals*”, realizado por Vinuesa y colaboradores en 2020, se analiza detalladamente cómo la IA puede contribuir al cumplimiento de los ODS establecidos por las Naciones Unidas. Se destaca que la IA tiene el potencial de impulsar la innovación y la eficiencia en áreas cruciales como la salud, la educación, la energía y el medio ambiente.

La IA puede ser una herramienta poderosa para la identificación de patrones en grandes bases de datos, lo que permite mejorar la precisión en el diagnóstico médico, optimizar la gestión de recursos naturales y facilitar la toma de decisiones informadas en políticas públicas. Además, la automatización de procesos mediante la IA puede aumentar la productividad y reducir los costos en sectores clave, promoviendo así un desarrollo económico sostenible.

Es importante reconocer que el crecimiento de la IA también plantea desafíos éticos y regulatorios significativos. Entre estos desafíos se encuentran la privacidad de los datos, el sesgo algorítmico, la seguridad cibernética y el impacto en el empleo y la desigualdad social. La utilización de algoritmos puede generar decisiones discriminatorias o injustas si no se diseñan y utilizan de manera responsable y transparente.

Además, la rápida evolución de estas tecnologías plantea mucha duda sobre la capacidad de los marcos regulatorios existentes para adaptarse y mantener el ritmo de los avances tecnológicos. Se requiere una colaboración interdisciplinaria entre los gobiernos, la industria, la sociedad civil y la academia para desarrollar políticas y normativas que promuevan el uso ético y equitativo de la IA, garantizando al mismo tiempo la protección de los derechos humanos y el medio ambiente.

En este aspecto, podemos concluir que la inteligencia artificial tiene el potencial de desempeñar un papel crucial en la consecución de los Objetivos de Desarrollo Sostenible al ofrecer soluciones innovadoras y eficientes para los desafíos globales. Sin embargo, es fundamental abordar los dilemas éticos y regulatorios asociados con su uso, asegurando que se utilice de manera responsable y en beneficio de toda la humanidad.

VI. ÁREAS DE CONTINUO RIESGO Y NECESIDAD DE VIGILANCIA

Si bien la ley de la Unión Europea representa un gran avance que demuestra un modelo adecuado de regulación de los programas creados con inteligencia artificial, éste presenta dos grandes desventajas: la primera es que sólo es válido dentro del espacio de la Unión Europea y la segunda que es una desventaja compartida por todas las demás regulaciones que se presentaron en el primer apartado, es la rapidez con la que estos sistemas avanzan y se mejoran de tal suerte que las estimaciones de riesgo y sus alcances cambian drásticamente en cuestión de meses o hasta de semanas.

Esta fue la razón que llevó a los directores de las empresas más importantes de desarrollo e innovación tecnológica a proponer en marzo de 2023 una moratoria para detener los avances de la IA hasta no poder contar con mecanismos de protección confiables

que pudieran garantizar la seguridad en el uso y aplicaciones de los sistemas y programas creados con IA.

Han transcurrido ya los seis meses solicitados y aún no se cuenta con un sistema seguro y transparente a nivel mundial y, por el contrario, se cuenta con múltiples avances en el mejoramiento y sofisticación de la IA especialmente de la de tipo generativo, tal es el caso de *Chat GPT 4* que aún ni si quiera cuenta con sistemas de verificación que puedan definir si el texto presentado fue o no creado mediante inteligencia artificial.

Ante esto, el Índice de Inteligencia Artificial 2023 (Maslej *et al.*, 2023) advierte de algunos ejemplos que ya están suscitando serias reflexiones éticas debido a las violaciones a los derechos humanos que representan. Algunos de ellos son:

- 1) Escaneo y monitoreo de teléfonos celulares en determinados grupos poblacionales como los prisioneros o los migrantes en centros de detención, lo que es una clara invasión a la privacidad de las personas.
- 2) Sistemas de reconocimiento facial en centros de trabajo y escuelas. Independientemente de los propósitos para los que se implementen, estos representan un peligro en materia de bioseguridad y pueden generar sesgos algorítmicos que, a su vez, generen marginación. También, se encuentra latente el peligro de que estos sistemas puedan ser hackeados para cometer fraudes financieros o robos de identidad. A este respecto, también hay que mencionar la posibilidad de que se asocien los rasgos faciales con comportamientos delictivos y se acusen como delincuentes a personas inocentes sólo por el hecho de asemejarse a un patrón de rasgos específico.
- 3) La propagación de videos y noticias falsas que generan confusión y propagan información incorrecta o inexacta usada para realizar afirmaciones falsas o hasta difamaciones y dañar seriamente

la reputación o imagen de una persona, incluso, en algunos casos, generar inestabilidad política a nivel mundial.

- 4) La generación de textos, imágenes o diseños sin citar las fuentes de donde son extraídos ni solicitar los permisos necesarios para su reproducción, lo que genera problemas de derechos. Este es un problema frecuente con el uso de *Chat GPT*.

Además de estas posibles dificultades para el respeto de los derechos humanos, la inteligencia artificial presenta otros desafíos en cuanto a la igualdad y equidad en el acceso a servicios en materia de salud, vida digna, salario justo, vivienda, alimentación, educación, trabajo digno y seguridad laboral, acceso a internet y banda ancha.

Ante la posibilidad del robo de identidades y la comisión de fraudes, se pueden vulnerar los derechos, entre otros, de la privacidad y confidencialidad, la presunción de inocencia y la personalidad jurídica. Sobre los sistemas desarrollados en la industria armamentística y los sistemas nacionales de defensa y seguridad pública, los derechos a la paz y a la no tortura pueden ser obviados.

En riesgo pueden también ponerse el derecho a la libertad de pensamiento, creencias, circulación, de expresión, derecho a un medio ambiente saludable, a la seguridad jurídica, derecho a ser oído públicamente, a la inclusión y no discriminación, vida libre de violencia, trato igual; derecho a la no detención arbitraria, etc.

Las áreas donde se analiza que es preciso tener especial precaución y una vigilancia más estrecha, por ende, son las relacionadas con la salud y la obtención de datos biométricos y su consecuente tratamiento confidencial y privado, las que versan sobre temas de seguridad que pueden llevar a inculpar personas inocentes, detenciones arbitrarias, privación de libertad de manera injustificada, actos de violencia hacia grupos y poblaciones vulnerables, etc.

De igual manera, la vulneración al derecho a un medio ambiente saludable es un área muy sensible en la actualidad debido al uso de materiales no biodegradables y de procesos no sustentables que dañan considerablemente el medio ambiente y generan contaminación y desprendimiento de gases de efecto invernadero que contribuyen al aumento de la temperatura global. El cuidado y procuración de entornos sustentables y sostenibles es, hoy, una prioridad que es necesario tener presente.

Por último, se considera necesario desarrollar mecanismos de gobernanza mundial y de políticas y esfuerzos multilaterales para que todos los espacios geográficos -y no sólo la Unión Europea- unifiquen criterios y cierren filas para la protección de los derechos humanos universales.

Estos ejemplos son sólo algunos de los muchos desafíos que la inteligencia artificial presenta para la legislación.

Adicionalmente, los dirigentes del G-7 han desarrollado los “Principios Rectores Internacionales del Proceso de Hiroshima para un sistema avanzado de IA” (2023). Los mismos son:

- 1) Tomar medidas para la detección de riesgos mediante sistemas de prueba que detecten cualquier posible eventualidad en el ciclo de vida de la IA y adoptar acciones para su mitigación
- 2) Adoptar sistemas de seguimiento a posibles eventualidades una vez que se han introducido al mercado dichos sistemas. Se deberá, por tanto, contar con mecanismos accesibles para el reporte de fallas o riesgos sobre el uso de la IA y su debido seguimiento para incorporar acciones que los mitiguen.
- 3) Hacer públicas las capacidades, limitaciones y áreas de uso de los sistemas de IA con el fin de asegurar la transparencia y confiabilidad.

- 4) Promover el intercambio responsable de información, así como el reporte de riesgos entre organizaciones incluidas la industria, gobierno, sociedad civil y la academia.
- 5) Implementar y hacer públicas políticas de gobernanza y manejo de riesgo de sistemas avanzados de IA. Se deben incluir políticas de privacidad y mitigación de riesgos.
- 6) Invertir e implementar controles estrictos de seguridad, incluida la seguridad física y la ciberseguridad, así como de posibles amenazas en cualquier parte del ciclo de la IA
- 7) Desarrollar mecanismos de autenticación de contenidos que faciliten identificar aquellos desarrollados mediante IA. Esto incluye declaraciones de origen de contenidos generados.
- 8) Priorizar e invertir en la investigación que permita mitigar riesgos sociales y de seguridad.
- 9) Priorizar el desarrollo de sistemas de IA para enfrentar los retos mundiales y las necesidades sociales tales como el cambio climático, salud global y educación.
- 10) Favorecer y, cuando sea pertinente, adoptar estándares técnicos internacionales.
- 11) Implementar medidas de alimentación de datos, así como protección de la privacidad de datos personales y de la propiedad intelectual.

Estos principios iniciales tienen por objetivo desarrollar un código de conducta que sea universal en el desarrollo de sistemas de Inteligencia artificial para todas las partes involucradas en su ciclo de vida.

VII. CONCLUSIONES

La inteligencia artificial ha tenido un ascenso vertiginoso tanto en su desarrollo como en su implementación en diversos ámbitos industriales y de la vida cotidiana. Sin embargo, a pesar de estos avances, las discusiones en torno a sus posibles riesgos y consecuencias éticas no han alcanzado el mismo ritmo.

La variada y débil regulación existente a nivel mundial refleja este desafío jurídico, que oscila entre la prohibición y la permisividad, en un intento por no obstaculizar el progreso tecnológico, pero sin descuidar los valores fundamentales como la libertad y dignidad humana frente a la visión utilitarista que insiste más en los fines que en los medios.

Los beneficios que aporta el desarrollo de la IA son innegables, especialmente en términos de exactitud y rapidez que ayudan y facilitan acciones que, de otra manera, se realizan de forma lenta e inexacta perdiendo, muchas veces, su efectividad y disminuyendo sus alcances. Por esto, siempre se deberá favorecer que la ciencia avance en su conocimiento y desarrolle herramientas que sirvan para mejorar la calidad de vida de las personas y de las comunidades.

La IA, al integrarse en herramientas y plataformas de aprendizaje y desarrollo personal, está transformando la forma en la que las personas adquieren habilidades y conocimientos, potencializando su capacidad de innovación y resolución de problemas.

Es crucial reconocer que también conlleva riesgos inherentes, como el sesgo del algoritmo y la pérdida de control sobre los sistemas. Por esto, la responsabilidad recae en las personas y empresas que la desarrollan y utilizan, quienes deben estar comprometidas en prevenir y mitigar posibles riesgos y eventualidades que de dichos sistemas se puedan desprender al momento de su puesta en práctica.

A pesar de la importancia de abordar estos desafíos éticos y regulatorios, es evidente que queda mucho por hacer en términos de generar conciencia y promover el debate público al respecto. Es necesario pues, que se sigan realizando foros y espacios públicos para debatir los avances de la IA en cuanto a su alcance ético pues, sólo en la medida en que la sociedad civil, la academia, la industria y el gobierno puedan colocar su voz y externar sus inquietudes, podremos generar una alianza internacional con colaboración multidisciplinaria que permita, sin prohibir, regular y encauzar los usos de la IA en beneficio del bien común, de la justicia, de la equidad y de la dignidad humana.

La necesidad de una regulación más sólida y coherente se vuelve más evidente cuando se considera el rápido ritmo de avance tecnológico en este campo. La IA está en constante evolución lo que significa que los marcos regulatorios deben adaptarse rápidamente para abordar los nuevos desafíos y riesgos que surgen. Esta adaptabilidad es necesaria y crucial para garantizar que la inteligencia artificial se use de manera ética y responsable en beneficio del bien común y del progreso humano.

BIBLIOGRAFÍA

Asimov, Isaac. (1989): *Círculo vicioso*, en Los robots, Domingo Santos, Martínez Roca (trad.), Barcelona.

Comisión Europea (2023): *Policy and Legislation: Principios rectores internacionales del proceso de Hiroshima para un sistema avanzado de IA*, en página web oficial de la Comisión Europea. Disponible en: «<https://digital-strategy.ec.europa.eu/es/library/hiroshima-process-international-guiding-principles-advanced-ai-system>» [Consultado el 12 de febrero de 2024].

Department for Science, Innovation & Technology; Foreign, Commonwealth & Development Office; Prime Minister's Office (2023): *Declaración de Bletchley de los países que asisten a la Cumbre de seguridad de la IA, 1 y 2 de noviembre de 2023*, en página web del Gobierno de Reino Unido. Disponible en: «<https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/dbc58681-1b68-47e0-8e3f-f91435fdf8ce>» [Consultado el 9 de enero de 2024].

European Parliamentary Research Service (2023): “Artificial intelligence act”, en *Briefing: EU Legislation in Progress*. Disponible en: «[https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)» [Consultado el 7 de enero de 2024].

García Manrique, Ricardo (2007): “Yo, robot, o las tribulaciones de un robot kantiano” en página web de *Observatori de Bioètica i Dret Universitat de Barcelona*, Barcelona. Disponible en: «<https://www.bioeticayderecho.ub.edu/es/yo-robot-o-las-tribulaciones-de-un-robot-kantiano>» [Consultado el 10 de enero de 2024].

High-Level Expert Group on Artificial Intelligence of European Commission (2019): *Ethics guidelines for trustworthy AI*, en página web oficial de la Comisión Europea. Disponible en: «<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>» [Consultado el 9 de enero de 2024].

Maslej, Nestor *et al.* (2023): *Artificial Intelligence Index Report 2023*, Stanford, California.

RenAIssance Foundation (2020): “The call | Rome Call”, en *Rome Call | What is the Matter with Ai Ethics?* Disponible en: «<https://www.romecall.org/the-call/>» [Consultado el 9 de enero de 2024].

Robotechnics (2017): “Principios de Asilomar de la Inteligencia Artificial” en página web de *Robotechnics Robótica Educativa*, Alcalá de Henares. Disponible en: «<https://www.robotechnics.es/asilomar/>» [Consultado el 10 de enero de 2024].

Russell, Stuart y Norvig, Peter (2020): *Artificial Intelligence: A Modern Approach*. Prentice Hall.

UNESCO (2021): “Recomendación sobre la ética de la inteligencia artificial” en página web oficial de UNESCO, París. Disponible en: «<https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial>» [Consultado el 10 de febrero de 2024].

Universidad de Montreal (2018): *Declaración de Montreal para un desarrollo responsable de la inteligencia artificial*. Disponible en: «https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/ES-UdeM_Decl-IA-Resp_LA-Declaration_v4.pdf» [Consultado el 9 de enero de 2024].

Vinuesa, Ricardo *et al.* (2020): “The role of artificial intelligence in achieving the Sustainable Development Goals”, en *Nat Commun.*, vol. 11, núm. 233, 1-10.