

REVISTA INTERNACIONAL & COMPARADA DE DERECHOS HUMANOS

Human control and discrimination: responding to the challenge of drone-bombing by organized criminal groups in Mexico.

Ana Martín Beringola

Prolegómenos de inteligencia artificial en la impartición de justicia. Implicaciones prácticas y nuevos desafíos para los derechos humanos.

Alejandro Vilchis Robles

Trazando el camino hacia una regulación efectiva: Breve análisis del Reglamento Europeo de Inteligencia Artificial de 2024.

José Rafael Grijalva Eternod



I | D | H
ACADEMIA
INTERAMERICANA
DE DERECHOS HUMANOS

REVISTA INTERNACIONAL & COMPARADA DE DERECHOS HUMANOS, Año 2024, Vol. 7, Núm. 1, Enero-Junio 2024; pp. 239; 24 cm; Semestral.

I. ARTÍCULOS DOCTRINALES, II. ENSAYOS, III. NOTAS, IV. RESEÑAS BIBLIOGRÁFICAS.



REVISTA INTERNACIONAL & COMPARADA DE DERECHOS HUMANOS, Año 7, Núm. 1, Enero-Junio 2024, es una publicación semestral editada por la Academia Interamericana de Derechos Humanos. Carretera 57 km. 13. Ciudad Universitaria. Arteaga, Coahuila. C.P. 25350 Tel: +52 (844) 4 11 14 29, <https://www.academiaidh.org.mx/revista-icdh>, revista.icdh@academiaidh.org.mx. Editor responsable: Irene Spigno. Reserva de Derechos al Uso Exclusivo No. 04-2022-011415483600-102 otorgado por el Instituto Nacional del Derecho de Autor. ISSN 2992-8613. Responsable de la última actualización de este número, Centro de Estudios Constitucionales Comparados de la Academia Interamericana de Derechos Humanos, Dra. Irene Spigno, Carretera 57 km.13. Ciudad Universitaria. Arteaga, Coahuila, C.P. 25350, fecha de última modificación 25 de junio de 2024.

El contenido de los documentos que conforman esta obra es responsabilidad exclusiva de los autores y no representan en forma alguna la opinión institucional de la Academia Interamericana de Derechos Humanos.

Esta obra está sujeta a la licencia Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional de Creative Commons.

Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-sa/4.0/>.



REVISTA INTERNACIONAL & COMPARADA
— **DE DERECHOS HUMANOS** —

REVISTA INTERNACIONAL & COMPARADA
DE DERECHOS HUMANOS

INTERNATIONAL & COMPARATIVE
JOURNAL OF HUMAN RIGHTS

Directorio / Directory

Vol. VII Núm. I

Luis Efrén Ríos Vega

José Rafael Grijalva Eternod

Irene Spigno

Editor científico

Dirección

Scientific Editor

Direction

Irene Spigno

Juan Antonio Corral Reséndiz

Editora en jefe

Coordinación

Editor in Chief

Coordination

María Gpe. Imormino De Haro

Santiago D. Sánchez Juárez

Coordinación editorial

Victor Manuel Vera García

Editorial Coordination

Formateo de textos

Text Formatting

Sandra Elizabeth Martínez Torres

Diseño editorial y maquetación

Editorial Design & Layout

Ana Daniela García Hernández

Diseño de portada

Cover Design

Contacto / Contact:

revista.icdh@academiaidh.org.mx

<https://www.academiaidh.org.mx/revista-icdh>

Consejo editorial

Jaime Fernando Cárdenas Gracia, Jorge Castellanos Claramunt,
Eleonora Ceccherini, José Ramón Cossío Díaz, Sergio Díaz Rendón,
Eduardo Ferrer Mac-Gregor, Tania Groppi, Marco Olivetti,
José de Jesús Orozco, Ilenia Ruggiu, Irene Sobrino Guijarro.

Comité evaluador

Wilma Arellano Toledo	Rubiela Gaspar Clavo
Michelle Azuaje-Pirela	Herón Gómez Eddy
José Rafael Belandria García	Diego Mauri
Ricardo Corona	Luis Alberto Reyes Figueroa
Myrna Elia García Barrera	Jaime Iván Rodríguez Lozano

Santiago Daniel Sánchez Juárez

Equipo editorial

José Antonio Estrada Marún
María Guadalupe Imormino de Haro
Carlos E. Zamora Valadez

ARTÍCULOS DOCTRINALES

- 11 **Inteligencia artificial: el desafío jurídico para la protección de los derechos humanos.**
Ma. Elizabeth de los Ríos Uriarte
Marcial Orlando Cabrera Cantarero
Inés Marquina Gurrea
- 41 **Human control and discrimination: responding to the challenge of drone-bombing by organized criminal groups in Mexico.**
Ana Martín Beringola
- 81 **Prolegómenos de inteligencia artificial en la impartición de justicia. Implicaciones prácticas y nuevos desafíos para los derechos humanos.**
Alejandro Vilchis Robles
- 123 **Trazando el camino hacia una regulación efectiva: Breve análisis del Reglamento Europeo de Inteligencia Artificial de 2024.**
José Rafael Grijalva Eternod
- 161 **Inteligencia artificial y derecho en México. Un dilema: “El mito del vehículo automatizado en México”.**
Juan Carlos Gámez Sagrero

ENSAYOS

- 197 | Educación para la paz en tiempos de algoritmos y gadgets.
Mónica Adriana Mendoza González

NOTAS

- 213 | La Inteligencia artificial como objeto y recurso en la educación jurídica.
Moramay Guerra García

RESEÑAS BIBLIOGRÁFICAS

- 225 | Spigno, Irene; Scotti, Valentina Rita y Penalva da Silva, Janaína Lima, *The Rights of Women in Comparative Constitutional Law*, Nueva York, Routledge, 2023.
Santiago Daniel Sánchez Juárez
- 231 | Reseña. Ferri, Delia; Palermo, Francesco y Martinico, Guisepppe, *Federalism and the rights of persons with disabilities. The implementation of the CRPD in federal systems and its implications*, Oxford, Hart Publishing, 2023.
Daniela Monserrat Hernández Ávila

ARTÍCULOS DOCTRINALES



Esta sección conforma el apartado distintivo de la Revista Internacional & Comparada de Derechos Humanos. Contiene los artículos doctrinales académicamente consistentes que constituyen los temas prioritarios y específicos de la publicación y cuyo enfoque es el estudio de los derechos humanos en perspectiva internacional y comparada. Para decidir sobre su publicación los textos recibidos son sometidos a un sistema de evaluación por pares de doble ciego conforme a estrictos estándares académicos.

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 beneficiosos, 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 adherencia 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 *chat-bots*, 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.

Human control and discrimination: responding to the challenge of drone-bombing by organized criminal groups in Mexico

*Control humano y discriminación: respondiendo
al desafío de los drones explosivos utilizados
por grupos de crimen organizado en México*

ANA MARTÍN BERINGOLA

Universidad Internacional de Valencia

ORCID: 0000-0002-4102-5410

Fecha de recepción: 01 marzo 2024

Fecha de aceptación: 07 junio 2024

*“[A]ll predictions agree that if man
does not master technology, but allows it to
master him, he will be destroyed by technology”
(Icrc, Commentary to Art. 37 Api, para. 1476).*

SUMMARY: I. Introduction. II. Delimiting the scope of violence. 1. Drone bombing by organized criminal groups. 2. The humanitarian nature of violations. III. Challenges posed by artificial intelligence in armed conflict. 1. Human control. 2. Machine bias. 3. Responsibility. IV. Suggestions for a domestic legal framework. 1. Integrating meaningful human control. 2. Addressing bias. V. Conclusion.

ABSTRACT: This article addresses the challenge of drone bombing by organized criminal groups in Mexico as a surging but unregulated phenomenon. It is argued that, while the Government is using a law enforcement approach, drone bombing, which intensifies the nature of confrontations, falls under certain circumstances under the regime of international humanitarian law. This results in the need to consider the difficulties that autonomous weapons systems pose in situations of armed conflict. Accordingly, the article examines *human control* and *discrimination* as key

issues in the international debate around autonomous weapons systems. Based on this discussion, the article brings forward some proposals to inform a domestic legal framework applicable to drone bombing in Mexico –and autonomous weapons at large– in consistency with the State’s international obligations and the domestic policy framework.

RESUMEN: Este artículo aborda el desafío que plantean los drones explosivos utilizados por los grupos de crimen organizado en México como un fenómeno de creciente importancia pero no regulado. Se argumenta que, si bien el Gobierno utiliza un enfoque policial frente a estos ataques, los drones explosivos, al intensificar la naturaleza de la violencia, recaen bajo ciertas circunstancias en el ámbito del derecho internacional humanitario, lo cual hace necesario considerar las dificultades que los sistemas de armas autónomas plantean en situaciones de conflicto armado. En consecuencia, el artículo examina el *control humano* y la *discriminación* como cuestiones clave que plantea el debate internacional sobre sistemas de armas autónomas. Teniendo en cuenta este debate, el artículo realiza algunas propuestas que deberían informar el marco legal aplicable a los drones explosivos en México –y a las armas autónomas en general– coherente con las obligaciones internacionales del Estado y con su política interna.

KEYWORDS: *international humanitarian law, human rights, discrimination, human control, autonomous weapons systems, drones.*

PALABRAS CLAVE: *derecho internacional humanitario, derechos humanos, discriminación, control humano, sistemas de armas autónomas, drones.*

I. INTRODUCTION

The use of Artificial Intelligence (AI) has rapidly expanded as evidenced by its multiple applications, not only in daily life but, also, in exceptional situations like armed confrontations. This is taking place in Mexico where organized criminal groups are increasingly using unmanned armed drones to combat the State or other criminal organizations. The phenomenon, which commenced in 2017, is surging and even becoming a daily reality in some States (Sullivan 2021: 867; Bunker and Sullivan 2021; Plaw

et al. 2020: 101). Drones –aircrafts remotely piloted– are autonomous weapons systems (AWS). This implies that, after initial activation by a person, AWS are able to “select and apply force to targets without human intervention [...] in response to information from the environment received through sensors and on the basis of a generalized ‘target profile’”(ICRC 2021)¹.

Artificial intelligence (AI), which involves the use of computer systems [algorithms] to carry out tasks that replicate human cognition, planning or reasoning (ICRC 2023) is being integrated in the military worldwide for its many benefits –intelligence, autonomous weapons and vehicles, resource management, logistics, surveillance, recognition– (Chandler 2021). AI is featured by its capacity to act independently and efficiency (Mainzer 2020: 3), hence, the generalized interest to incorporate it in the field of State security interests. When AI is embedded in AWS, these weapons allow obtaining a more complete analysis of conflict environment with capacity to advance critical functions in armed conflict such as targeting and engagement in the use of force. Their benefits include precision selecting targets –face, voice recognition, etc.–, reduction in the number of casualties, greater protection of civilian infrastructure (Salgado Espinoza 2020: 65) and, even indication of the time and location of crimes helping the process of accountability (Chandler 2021: 21)². However, dealing with *targeting*, AI-powered weapons pose a huge challenge. The limited human control over AWS “the user does not choose, or even know, the specific target and the precise timing and/or location of the resulting applica-

¹ International Committee of the Red Cross (ICRC).

² This is not to speak of the many benefits that artificial intelligence has in armed conflict, for instance, the identification of patterns of crime through satellite image, establishing evidence of incitement to commit crime through the use of algorithms in the social media, locating mass graves helping investigations, crime prevention and reparations. See, for instance Håkansson Schmidt (2022), see references; Porup, J.M. (2017), “Hunting for Mexico’s mass graves with machine learning” in *BIZ & IT*. Available at: «<https://arstechnica.com/information-technology/2017/04/hunting-for-mexicos-mass-graves-with-machine-learning/>» [Consulted on 28 February, 2024].

tion of force” (ICRC 2021) triggers ethical and legal debates around the issues of *control* and *responsibility*³. On the one hand, handing out *human control* to Aws for critical target functions would imply leaving live or death decisions in the hands of machines. On the other hand, the use of Aws raises the question of who is criminally responsible for machine failures if a human agent has no meaningful control –no agency *ergo* no intent– over the self-initiation of the machine in targeting and engagement (Asaro 2016: 190-192).

This paper addresses the legal gap surrounding the use of explosive drones by organized criminal groups in Mexico with the aim of suggesting a way forward to inform the domestic legal framework. In this way, the article paves the way for the future use of Aws by non-State and State actors. The paper highlights that Aws raise the intensity of violence resulting in the need to rethink the nature of armed violence in Mexico and the potential application of international humanitarian law (IHL), currently not applied in the State’s confrontation with cartels. As it is argued, under *certain* conditions, the legal requirements triggering the application of IHL exist in Mexico. Accordingly, the article delves on the circumstances that make the application of IHL a complicated task at the international and national levels in that State. One reason hampering the application of IHL is the lack of international legal standards on Aws. Another reason is official reluctance to recognize the potential application of IHL to national security issues. Discrimination is identified as another issue requiring consideration in a future domestic legal framework in Mexico. Studies reveal that AI tends to *generalize* patterns or features of conflict environment resulting in discrimination in the coding of algorithms. Consequently, machine bias is often behind failures in targeting and engagement leading to the violation of IHL principles, including distinction, proportionality, non-discrimination that result

³ Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): “Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems”, CCW/GGE.1/2022/WP.5.

in unnecessary harm (Chandler 2021; Houghton 2022; Håkansson Schmidt 2022). Discrimination in the use of AI is rampant in Mexico (Spitx *et al.* 2023) and it is especially detrimental for the most vulnerable groups such as women, girls and indigenous peoples (Casados *et al.* 2020). Moreover, as structural discrimination is exacerbated in armed conflict, machine bias is expected to aggravate these inequalities (Red de Investigación Feminista en Inteligencia Artificial *et al.* 2023). This fact weakens the official strategy under the National Plan of Peace and Security to end violence by tackling its root causes of discrimination and promoting human rights.

From this backdrop, the article's methodology first focuses on the international legal framework applicable to drone attacks in Mexico and, more broadly, to Aws. Then, with this insight, it engages on a discussion of *de lege ferenda* at the domestic level in Mexico to address the phenomenon of drone bombing. Accordingly, the article is divided in three sections. The first section discusses the phenomenon of drone bombing in the context of confrontations between criminal organizations and the State in Mexico providing arguments for the application of IHL in certain situations. Then, the article addresses the ethical challenges posed by Aws making their regulation complicated in IHL, paying special attention to the issues of *human control*, *machine bias*, and responsibility. Following this discussion, the article finally suggests the tenets of a responsible domestic legal framework in line with Mexico's international obligations and national policies. It is hoped that, in doing so, Mexico will set an example of good practice which might help moving the global discussion ahead in that direction.

II. DELIMITING THE SCOPE OF VIOLENCE

1. Drone bombing by organized criminal groups

Over the past years the intensity of violence between Mexican authorities and criminal groups, and between criminal groups

in Mexico, has drastically surged⁴. This fact questions the effectiveness of the official strategy which shifted from waging war against drug cartels towards addressing the structural causes of violence underpinned by gross inequalities and poverty (Plan Nacional 2018-2024: 2,7). Complex reasons explain the difficulty for halting organized criminal groups in Mexico, including the deterioration of living standards, the growth of the illegal drug market and a failed public strategy against organized crime (Padin 2023: 94). Growing militarization in Mexico, whose military budget increased by 52% during the term of President Andrés Manuel López Obrador (2018-2024), has resulted in the country advancing 11 positions within the most powerful armies from 2022 to 2023. However, militarization has not reversed trends in violence. Instead, both militarization and violence have gone in tandem questioning the official goal of caring for human rights to defeat drug-related violence (González 2023)⁵. Interestingly, what has been almost unnoticed is that the socialization of technology among cartels –especially the use of AI– has been a major driver of the rising levels of organized crime and violence in Mexico (Plaw *et al.* 2020: 95).

Aircrafts without a pilot and remotely controlled, commonly known as drones, are responsible for a large amount of the new forms of violence involving Mexican cartels. While the use of drones by criminal groups is not new, the constant technological improvements of AI are now reverberating into more intense confrontations. Cartels have resorted to drones strategically to gain an advantage over rivals in four main ways (Bunker and Sullivan 2021). From 2010, drones were employed to smuggle narcotics due to their considerable capacity and the difficulty of radars

⁴ Over the past 15 years, more than 35.000 intentional killings have been registered per year and the number of persons disappeared exceeds 100.000. See, WOLA (2022): “México profundiza la militarización. Los hechos muestran que es una estrategia fallida” in WOLA. Available at: «<https://www.wola.org/es/analisis/mexico-profundiza-militarizacion-hechos-muestran-estrategia-fallida/>» [Consulted on 16 February, 2024].

⁵ Popularly known in Spanish as “política de abrazos, no balazos”.

to detect them (Plaw *et al.* 2020: 101). A few years later drones served for intelligence, surveillance and reconnaissance purposes, including night vision combat. From 2017, they started to be used as weapons for targeting purposes. In 2020, from the beginning of the pandemic, they have served propaganda purposes competing against the State for political support. Whilst all of these functionalities make up the cartels criminal enterprise, the *weaponization* of drones is directly linked with an increase in violence and harm in armed confrontations in Mexico.

Over 600 attacks with explosive-equipped drones have been attributed to criminal groups in Mexico, the majority in the states of Guerrero, Tamaulipas and Michoacán (Mexico Now 2023). In the latter, drone-bombing is a nearly daily occurrence, with 260 of such attacks recorded in 2023 (CBS 2023). Drones are used by non-state actors as weapons and as weapon platforms and are becoming increasingly sophisticated –they can be single-use– and, often, multi-use platforms (Bunker and Sullivan 2021). They often incorporate homemade weapons such as explosives, nails and pellets due to the online training of technological skills by cartel members. Further, drones are operated by specialist *droneros* –drone operators–. Certain criminal groups, as the Cartel Jalisco Nueva Generación (CJNG), are training groups of drone operators to produce explosives and operate drones (Mondragón Toledo 2023). Critically, a consequence of using AWS like drones is the increasing asymmetry of warfare between the parties and the lower threshold of escalation, which is resulting in more severe harms⁶. For instance, to respond to drone firepower, the Mexican Government is acquiring helicopters equipped with rotating electric machine guns as, allegedly, the only way to overpower the increasing drone bombing capacity of the Jalisco cartel (Stevenson 2021).

⁶ Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): “Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems”, CCW/GGE.1/2022/WP.5.

The escalation of violence due to the use of explosive-equipped drones can be observed in the evolution of incidents. In Tecate (Baja California, July 2018), in a single-use incident, a police commander found a weaponized drone with two grenades and a note threatening not to mess with the local cartels. In Aguililla (Michoacán), the Jalisco and Viagras cartels regularly fight each other with bomb-dropping drones resulting in serious damage to civilian infrastructure and violence. This has triggered the use of dangerous counter-weapons by the Government. For example, the public authorities have resorted to Blackhawk helicopters for targeting drones, also called *indiscriminate blanket fire* (firing 6000 rounds per minute), which are prohibited in most civil conflicts (Stevenson 2021). Another example of escalation is violence in Nuevo Caracol (Guerrero, May 2023). While drone attacks were usually carried out between cartels, these attacks are now spreading to the rural community where they are taking their toll: civilian killings, damage of civilian infrastructure and the spread of terror have led to the forced displacement of an estimated number of 600 people in 2023, which has resulted in 67 communities demanding protection to the local and federal authorities (Mondragón Toledo 2023).

2. *The humanitarian nature of violations*

There is an alarming gap in Mexico's law and policy to address the use of AI in armed confrontations. Neither the National Plan on Peace and Security (National Plan 2018-2024) nor the National Agenda on Artificial Intelligence consider the implications of using AI in the military and security sectors in spite of the seriousness of the extant evidence and its trend to spread and become more sophisticated (Bunker and Sullivan: 2021). Yet, the high levels of armed violence in Mexico involving drones and the loophole around the use of AWS bear the question: What is the nature of violations resulting from the use of AWS by Mexican cartels? Answering to this question is of the utmost importance for minimizing legal uncertainty, arbitrary responses and, instead, clarifying the standards that would provide suitable protection against

violations as the use of Aws by non-State actors continues to increase. However, answering to this query is challenging for two reasons. One is the lack of binding international standards to guide a domestic legal framework in Mexico. Another is the difficulty of qualifying the nature of the conflict between the cartels and the State, and between cartels –i.e. armed or civilian in nature– whose discussion is viewed as opening Pandora’s box (Padin 2023).

Whether armed hostilities involving Mexican cartels constitute an armed conflict or internal disorders has been considerably debated. At stake is the application of either the law of armed conflict (LOAC), IHL or, alternatively, a law enforcement regime. The former would imply applying binding international standards on Mexico, namely, the Geneva Conventions (1949), Additional Protocol I⁷ and customary international law. Conversely, the later would entail political freedom to regulate the conflict as an internal issue whilst taking a positive note of relevant soft law standards⁸ and ensuring compliance with International Human Rights Law (IHRL). It is worth noting that, although Mexico has not ratified Additional Protocol II⁹ concerning non-international armed conflicts –an indicator that it does not consent to international interference in domestic confrontations– Article 3 Common to the Geneva Conventions applicable to internal armed conflicts is legally binding. Similarly, Mexico must abide by customary international humanitarian law, applicable at all times (ICRC 2005), and international jurisprudence. It is also the case that an eventual qualification of certain situations

⁷ (ICRC), Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1125 UNTS 3 (8 June 1977).

⁸ Refer to “The Code of Conduct for Law Enforcement Officials”, adopted by General Assembly resolution 34/169 (1979), and to the “Basic Principles on the Use of Force and Firearms by Law Enforcement Officials of 1990”, adopted by the Eighth United Nations Congress on the Prevention of Crime and the Treatment of Offenders, Cuba, 1990.

⁹ ICRC, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non- International Armed Conflicts (Protocol II), 1125 UNTS 609, 8 June 1977.

in Mexico as armed conflict would not exclude the application of IHRL, whose protection does not cease in armed conflict. IHRL would apply in a complementary way allowing greater protection in situations not covered by the *lex specialis* or IHL (ICJ 2004: para. 106), for instance, acting as a safeguard of fair treatment and judicial guarantees against members of organized criminal groups.

While law enforcement, characterized by police arrest powers¹⁰ is the official approach to address organized crime in Mexico, the reality is a *de facto* militarization. The Government has consistently recalled the application of IHRL when its armed forces act in security operations. This official position, complemented by application of law enforcement standards and the lack of ratification of Additional Protocol II to the Geneva Conventions, are good indicators of the official view that internal hostilities do not attain the threshold of Common Article 3 which would trigger the application of IHL in Mexico (Padin 2023: 792). In this regard, the National Peace and Security Plan (2018-2024) establishes the aim to pacify the country through the explicit goal of national peace-building and reconciliation, which is to be achieved by securing respect for human rights. In this regard, the problem of national insecurity and violence is officially recognized as a one of structural inequalities which must be addressed from a human rights-based approach (Chanona 2019: 6). To accomplish this aim, the National Peace and Security Plan articulates a proposal for transitioning from a public security model reliant on the military forces to one of enhanced police forces, as shown by the creation of the National Guard (Guardia Nacional), a police body with larger powers. However, this official position is contradicted by the means of implementation. The National Guard is largely made up of military staff and dependent upon the National Defense Secretary (SEDENA),

¹⁰ The term “police” refers to officers who exert the powers of arrest or detention to implement the law and order, which may imply the need to resort to the use of force. See General Assembly of the United Nations (2014): “Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns”, para. 22, A/HRC/26/36.

which has resulted in an increasing militarization of public security issues. The official approach is further complicated by a lack of clear protocols on the use of force under the National Peace and Security Plan, which is critically silent on the standards applicable to the key substantive issue of the use of force against organized criminal groups (Chanona 2019: 7-8).

Regarding the nature of the conflict involving State and non-state actors, international standards consistently establish the existence of an internal armed conflict when two concurrent conditions are present: a degree of organization and protracted violence¹¹. On the one hand, international jurisprudence considers the *organization* criterion fulfilled in the presence of indicators such as the existence of a command structure, the ability to gain access to weapons, military equipment and training, the ability to plan, coordinate and to carry out military operations¹². On the other hand, international precedents declare that violence is protracted based, not as much on duration, as on the intensity of hostilities appreciated by non-exhaustive factors, including the existence of armed clashes, the mobilization and distribution of weapons, the types of weapons and military equipment, the number of casualties due to shelling and fighting, and the quantity of troops deployed¹³.

Whether the criteria of organized and protracted violence determining the existence of an armed conflict and triggering the application of IHL exist in Mexico is highly debatable. Critics argue that the main goal of cartels is to obtain economic gain and, consequently, they adapt their organizational structure not to confront

¹¹ International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Duško Tadić*, 02 October 1995: para. 70.

¹² International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Ramush Haradinaj, Idriz Balaj and Lahi Brahimaj*, 3 April 2008: para. 60. International Criminal Court, *The Prosecutor v. Thomas Lubanga Dyilo*, 14 March 2012: para. 537.

¹³ International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Ljube Bošković and Johan Tarčulovski*, 10 July 2008: para 177. *The Prosecutor v. Thomas Lubanga Dyilo*: para. 538.

the Government *per se* but to protect their *plazas*. Accordingly, as cartels diversify their profitable activities, their structure becomes more decentralized (Padin 2023: 780-781; Muggah 2023: 570). Furthermore, despite the surge in violence in Mexico from 2007, studies show that only a small number of homicides is related to violence between State officials and illegal armed groups. This pattern is mainly attributed to situations combining drug consumption and poverty as pulling factors that affect the most vulnerable populations, therefore, confirming that economic (not political) factors are the key drivers of most murders directly related to organized crime. Accordingly, what extant research reveals is that *most* of the violence involving organized criminal groups in Mexico does not reach the threshold of armed conflict and, therefore, it is not suitable for the IHL paradigm (Padin 2023: 786).

These observations are, however, not conclusive. Despite the increasing fragmentation of Mexican cartels and their fighting for lucrative aims (Bellal 2017: 86), the latest ICRC's interpretation of armed conflict supports the notion of *aggregated intensity*. According to this view, when a coalition of armed actors has sufficient level of coordination, the intensity of hostilities "between each of them and an opposing party may be aggregated when considering whether the threshold of intensity has been reached" (D' Cunha *et al.* 2024:3). According to this view, the lack of *traditional* hierarchical structure of Mexican cartels would not prevent from considering the aggregation of violence of each of these criminal *coalitions* for the purpose of establishing the intensity requirement of an armed conflict. Additionally, even if cartels' criminal activities mainly rely on economic objectives, international jurisprudence considers that this fact does not prevent them from becoming parties to an armed conflict. In this regard, the International Criminal Tribunal for the Former Yugoslavia (ICTY) has noted that the existence of an armed conflict "is based solely on two criteria: the intensity of the conflict and organization of the parties,

the purpose of the armed forces to engage in acts of violence or also achieve some further objective is, therefore, irrelevant”¹⁴.

According to experts, the violence involving *certain* criminal groups in Mexico amounts to a non-international armed conflict in terms of organization and intensity. It is considered that, there are at least two internal armed conflicts taking place between the Government and the CJNG and the Sinaloa Cartel (CDS) (Bellal 2017: 87). Additionally, it has been noted that violence between the CDS and the CJNG reaches the threshold of an internal armed conflict as well (RULAC 2022). The CDS consists of a coalition of leaders of criminal organizations that acts as a federation relying on corruption and internal alliances. For more than ten years, the CDS has been involved in violent turf wars against the Mexican armed forces and, more recently, against other cartels for the control of drug trafficking routes (Bellal 2017: 87; RULAC 2022). In 2022, violence was particularly intense, although only one incident – a prison riot with 11 people killed, mostly civilians – has been attributable with certainty to the CDS (RULAC 2022). As for the CJNG, which emerged in 2010 as a splinter of the CDS, it has expanded its operations fast to many federal states. Considered one of the most aggressive groups, it uses highly sophisticated weapons in high-intensity attacks both against the State security forces and against other cartels, especially the CDS, to control illicit drug routes (Bellal 2017: 87).

The increasing use of drone bombing by powerful Mexican cartels reinforces the argument that criminal groups, such as the CDS and CJNG, are parties to an internal armed conflict that would trigger the application of IHL. On the one hand, the training of specialized drone operators, as is the case of the CJNG (Mondragón Toledo 2023), reflects enhanced organizational capacity in terms of access to weapons, military training and equipment, ability to plan, coordinate and carry out armed attacks, even if its orga-

¹⁴ International Criminal Tribunal for the Former Yugoslavia, *The Prosecutor v. Limaj et al.*, 30 November 2005.

nizational structure has become more loosely connected. On the other hand, incidents, like those mentioned above, where drone bombing is repelled by battle-grade force such as Blackhawk helicopters and rocket launchers by the State security forces (Shuldiner 2022), prove that escalation is a direct consequence of the cartels' increasing reliance on explosive-equipped drones. While violence between the Government and cartels, and between cartels, remains high, it becomes increasingly difficult to attribute violent incidents to specific armed groups, which has led some observers to declassify these armed conflicts due to a lack of certainty (RULAC 2022). A declassification –coming to an end– of an internal armed conflict takes place when one of its two conditions ceases to exist (D' - Cunha *et al.* 2024). However, in the case of drone bombing, there is abundant evidence –as described by this section– that cartels, including the CDs and CJNG, are increasingly using this technique, and that this fact is aggravating the scale and consequences of armed confrontations, even if complaints of armed attacks by the affected communities are often not followed-up by the authorities (Janetsky 2024; Investigator.Mx 2024). As it becomes plausible that IHL applies in certain contexts involving drone bombing attacks in Mexico –and this possibility increases in the future as drones and Aws become more commonplace– a dilemma arises: There are no international legal standards on the use of Aws in armed conflict to guide policy-making in Mexico. Accordingly, it is necessary, as will be examined in the next section, to review the main challenges that are preventing from reaching an agreement on the international standards applicable to Aws and, consequently, that are making it difficult address the phenomenon of drone bombing in Mexico.

III. CHALLENGES POSED BY ARTIFICIAL INTELLIGENCE IN ARMED CONFLICT

Aws pose two major legal issues which are unregulated: human control regarding critical combat functions and discrimination or bias in the programming of algorithms embedded in Aws. This

section addresses these legal gaps in primary norms of international law and the resulting challenge they give rise to in secondary norms in terms of establishing international responsibility for these loopholes.

1. *Human control*

The autonomy of Aws constitutes their curse. Ai allows the autonomous implementation of highly complex tasks resulting in a reduced ability of human agents to understand, predict, or mitigate their impacts (Asaro 2016: 190). One thing is decisive: the greater the autonomy of weapons, the lesser the ability of humans to control the machine. Actually, the degree of human-machine interaction can be programmed to enable humans exercise greater or lesser agency. This translates into different degrees of control. Human operators may be either *in the loop* –the machine identifies a target but a human must validate it before the machine initiates engagement– or *on the loop* –the machine identifies the target and commences engagement unless the human agent overrides it– (Coco 2023:3-4). Accordingly, Aws give rise to a critical ethical issue: a *control* problem regarding the critical functions of selecting the target and engagement in the use of force, hence, the difficulty of attributing responsibility for the harms caused by Aws when human control cannot be adequately established. The more autonomous weapons are, the more limited human control –and agency– over them, even in situations of negligence, resulting in the dilemma of who is responsible which risks turning upside down the paradigm of accountability for gross violations in armed conflict. Likewise, the human control problem may compromise other responsibility regimes. This may apply to the international responsibility of States for gross violations committed by its agents, and it may also interfere with domestic regulations sanctioning Aws, including the use weapons and explosives from the airspace. Observers have suggested that a taxonomy of the use of increasingly autonomous Aws in critical targeting and engagement functions may include the following levels of autonomy:

“1) A human engages with targets, selects targets and initiates attack; 2) A program suggests alternative targets and a human decides which to attack; 3) A machine selects targets which a human must validate before engagement; 4) A machine selects and engages targets under supervision of a human who may override the machine’s choice and halt the attack; 5) a machine selects targets and initiates attack according to the mission’s goals defined at the planning/activation stage without additional human intervention” (Amoroso and Tamburrini 2020: 191).

Even though all these levels, except level 1, involve a degree of unpredictability in critical functions, the degree of autonomy and corresponding human control over Aws clearly varies (Acquaviva 2021: 7). Only levels 1 to 3 allow human agents to be *on the loop* and, therefore, the possibility of retaining a *significant* degree of human control. Levels 4 to 6 allow no significant human control, and control is completely absent in stage 5. Furthermore, the scenario complicates if Aws failures are combined with situations where human operators do not exercise control properly such as situations of automation bias –excessive human trust in the machine’s determinations despite contradicting information from other sources– and complacency –excessive trust in the machine’s determination resulting in diminished attention from human operators– (Coco 2023: 1). Who is under control in these situations: humans, robots, or both?¹⁵ Who violates the laws of armed conflict? Can we conceive IHL without significant human intervention resulting in warfare being waged by robots instead of by humans?¹⁶ Situations where human control is not significant or absent are the object of diplomatic debate and, still, are far from reaching the necessary consensus for their regulation necessary for a consistent and principled application of IHL.

¹⁵ For an excellent discussion on the topic see, Acquaviva (2021).

¹⁶ In the same sense, consider the ethical and legal concerns on lethal Aws raised by Special Rapporteur: i) compliance with IHL, ii) attribution of responsibility, iii) violation of human dignity, and iv) risk for peace and stability. See General Assembly of the United Nations (2013): “Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns”, A/HRC/23/47.

As a warfare weapon, the legality of Aws must be considered within IHL's general framework, where it is a highly divisive matter. A first consideration is that States have the duty to assess whether the employment of new weapons would be prohibited by IHL¹⁷. Thus, it impinges on each Government a preliminary duty to create internal procedures to elucidate the legality of a new Aws before putting it to use (ICRC 1987). The academic debate on the legality of Aws moves between the waters of a full ban and regulation¹⁸. Those in favor of a complete prohibition of Aws posit their incapacity to ensure respect for IHL because they lack the moral qualities of awareness and judgment which only humans have (Amoroso and Tamburrini 2020: 189). In this sense, it is observed that interactions among Aws in unstructured warfare contexts are highly unpredictable to the extent of rendering combat *beyond* human control, which would amount to accepting the possibility of removing ethics from warfare. Furthermore, it is argued that Aws make wars easier to wage due to the possibility of saving more military lives in one's own camp –as more militaries operate remotely– which would pose increasing risks for the maintenance of international peace and security (Sharkey 2012)¹⁹. Conversely, arguments in favor of a controlled development of Aws and their regulation claim numerous benefits. These include regulating transparency, increasing control over State's use of these weapons, and potential

¹⁷ While Additional Protocol I refers to international armed conflicts only, consider that other IHL norms potentially applicable to Aws do apply to internal armed conflict. For instance, consider the prohibition to use indiscriminate weapons, which includes those whose effects “cannot be limited” which may be the case of Aws without human control. Additional Protocol I 1977: Article 36.

¹⁸ Consider also international responses. For instance, the International Committee of the Red Cross (ICRC) supports banning unpredictable Aws (as does the UN Secretary-General) and autonomous target recognition, but it endorses the regulation and use of Aws. The Human Rights Commissioner for Human Rights has called for a moratorium on the sale and use of Aws until safeguards ensure compliance with IHRL standards (see Chandler 2021: 25).

¹⁹ General Assembly of the United Nations (2013): “Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns”, A/HRC/23/47.

to save human lives due to the accuracy of their data. It has even been argued that *ethically restrained* Aws would abide by the laws of war better than humans, who are subject to emotional reactions in the battlefield that do not affect robots (Arkin 2009). Moreover, prohibiting Aws would stifle innovation, economic growth and development, and it would involve renouncing to the possibility of devising alternatives for more proper management of AI in fields where it can provide considerable advantages (Asaro 2016).

Diplomatic responses have failed to provide an international legal framework on Aws so far, a core reason being the objection amongst powerful States to limit their technological lead (Amoroso and Tamburrini 2020: 192)²⁰. Nonetheless, the international community is engaged, although limitedly, in the process of regulating lethal Aws²¹. The most relevant forum, and most concrete proposal, are the Guiding Principles on Lethal Autonomous Weapons Systems (LAWS), adopted by consensus by the Group of Governmental Experts (GGE) in the context of the Convention on Certain Conventional Weapons (Ccw)²². These recommendations –where Mexico has played a leading role– aim to establish the basis for a “Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapons Systems”²³. They have merit for adopting a middle way position between the opposite approaches, i.e.,

²⁰ Critical detractors of regulation include the Us and the Russian Federation.

²¹ The debate on Aws began in the Human Rights Council. Then, the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions called for a high-level panel to articulate an international policy in its 2013 report. From 2013, the Group of Governmental Experts within the Ccw has taken up the issue and made substantive recommendations after eight years of discussions.

²² The aim of the Ccw is to restrict and ban the use of weapons deemed to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately.

²³ Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): “Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems”, CCW/GGE.1/2022/WP.5.

a complete ban and the regulation of AWS, in the spirit of garnering enough consensus to agree on legally binding standards.

The Guiding Principles' middle way position between a full ban *versus* the regulation of AWS is rooted in the claim that human agency must prevail and underpin all prohibitions and regulations of AWS. Accordingly, the Guiding Principles suggest a straightforward ban on those AWS that remove human control, do not respect IHL principles, and whose effects cannot be sufficiently understood. From this position, which ensures respect for existing international law, the Guiding Principles leave space for negotiating a regulatory framework. In this regard, the Guiding Principles suggest two key regulations: ensuring *sufficient* human control –that will require a definition– in the critical functions of targeting and the use of force –while recognizing that the scope of human control may vary depending on the context environment– and establishing individual criminal responsibility for violations committed using AWS. Accordingly, this international initiative makes room for lethal AWS in IHL but always within the framework of sufficient human control, respect for IHL and international responsibility.

The notion of *sufficient* human control undergirding the GGE Principles emerges as a key concept in international negotiations on AWS. This term matches what observers consider the *meaningful* human control of AWS which, accordingly, may provide helpful guidance establishing the boundaries of what this degree of human control entails. Meaningful human control has been defined by a threefold threshold: i) humans act as a fail-safe actor –the machine design prevents/mitigates the endangering of life or property in situations concerning direct attacks against civilians or excessive damage–; ii) human accountability –individual criminal responsibility for AWS violations–, and iii) respect for human dignity –moral agency– so that decisions relative to harm caused to people and property are not left to artificial agents (Amoroso and Tam-

burrini 2020: 192)²⁴. Internationally agreed standards on the scope of meaningful/sufficient human control seems, therefore, of the utmost importance as a bulwark that mitigates the ethical dilemma posed by the autonomy of Aws. Another advantage of the notion of sufficient/meaningful human control under the Guiding Principles –which may be also of great help in future negotiations– is that it makes room for a “jointly differentiated and prudential” approach (Amoroso and Tamburrini 2020: 192), thereby, avoiding a one-size-fits- all approach to Aws and allowing States a margin of framed autonomy for decision-making in critical targeting and engagement functions.

2. *Machine bias*

Machines, like the humans who create them, are not infallible and make mistakes. Differently from humans though, machine failures are not underpinned by agency and moral judgment but by codes of algorithms which, in fact, make robots as good as their data (Farrés 2021: 3). The problem with algorithms, as those informing Aws, is that they are frequently informed by bias or discrimination. Gender-based discrimination is a good example. According to studies, gender discrimination is pervasive across systems of AI in warfare. Recognition tools are 70% more likely to recognize the voice of a man than of a woman, and relief systems tend to overlook the specific needs of women and girls (Houghton 2022: 2). Women are frequently stereotyped as submissive and sexualized whereas men tend to be, by default, associated with combatants, resulting in the filtering of content that exacerbates preexisting inequalities (Un Håkansson Schmidt 2022: 1-3). Likewise, gender intersecting with race discrimination is embedded in military codes. While facial recognition software makes 1% of error for light-skin women, this figure elevates to 35% for dark skin women (Håkansson Schmidt 2022: 1). Images from the most populated countries

²⁴ Note the links between this conceptualization of human control and the conclusions of the Special Rapporteur on extrajudicial summary or arbitrary executions (Heyns 2014).

–as China and India– only represent 3% of the total images of AI systems. Accordingly, while a white woman wearing a white dress is recognized as a bride, a North Indian woman wearing a sari is labelled as an artist (Chandler 2021: 15). Overall, there is a tremendous problem of representation informing the algorithms of Aws resulting in that bias underpins the machine’s critical functions in armed conflict.

Aws failures in selecting targets and engagement recurrently violate core IHL norms such as the principles of distinction –between civilians and combatants– and proportionality –refraining from causing excessive civilian loss in comparison with concrete and direct military advantage–²⁵. Interestingly, observations from the field reveal that drone bombing is the autonomous weapon *par excellence* responsible for causing these violations. This might be due to algorithm bias combined with the features of Aws, which are deprived of sensory systems for separating civilians from combatants and lack situational awareness for making proportionality decisions (Sharkey 2012: 789-790). In Afghanistan, drones for signature strikes –taking place outside traditional battlefields– were programmed to consider any military-age male a military target. This gross generalization of the target profile led to a weaponization of civilians in Afghanistan by turning civilian men into military targets. This discriminatory coding resulted in a gross breach of the principle of distinction and it had a disproportionate impact on the number of civilian casualties that combatants must spare above all (Farrés 2021: 2). In Libya, *kamikaze* Aws –selecting and attacking human targets based on proximity to an object– have been labelled *battlefield experiments* and are considered key contributors of the serious humanitarian crisis in that country (Chandler 2020: 17-18). Similarly, research on the Syrian conflict suggests that selecting aerial bombardment as a method of warfare has a drastic gender discriminatory impact. According to estimates, these methods have caused a disproportionate number of women and girls being killed in the conflict –37% women v 23% men– when com-

²⁵ Additional Protocol I, Arts. 48 and 51.

pared to other methods of warfare such as shooting –13% women v 23% men– (Farrés 2021: 6).

Avoiding bias in Aws requires the adoption of anti-approaches which are now absent in the coding industry, including a conflict analysis of preexisting discrimination and the mainstreaming of non-discrimination through the coding cycle. A conflict analysis of discrimination should consider factors such as the conflict dynamics, culture, and existing inequalities that help understand the nature and effects of violence. Then, these should be integrated in the coding of algorithms so that Aws can factor this sensitive information in. In this regard, Chandler notes, “[i]f AI [artificial intelligence] is to replicate human intelligence, a narrow understanding of what is human must be replaced by a more complex model that includes the range of bodies, abilities and emotions that are all part of human intelligence” (Chandler 2020: 9). If, as studies have shown, algorithm bias is structural, its elimination requires the mainstreaming of non-discriminatory approaches throughout the coding cycle, including the manufacturing, developing, use and control of Aws. Some key non-exhaustive measures include: inclusive gender programming –women represent 24% of computer scientists–; hiring experts to sensibly inform algorithm data (Chandler 2020: 9); a context-specific approach taking note of preexisting inequalities affecting the most vulnerable groups in peacetime which are likely to exacerbate with technologies (Spitx *et al.* 2023: 6-7), and reflecting the views of underrepresented groups, mainly women, children and other marginalized groups at a higher risk of having experienced the harms of machine failures (Håkansson Schmidt 2022: 2). Above all, informing Aws with an inclusive non-discriminatory approach to conflict analysis demands what feminist scholars have called a *human and differentiated impact* which takes note of the gender and other intersecting inequalities affecting marginalised groups who are the most affected by armed conflict (Red de Investigación Feminista en Inteligencia Artificial *et al.* 2023: 3). Overall, if Aws bias is to be avoided, it makes sense that algorithms are specifically programmed to avoid discrimination.

3. Responsibility

IHL is built upon the paradigm that only humans can be accountable for the agency and choices resulting in violations in armed conflict. Accordingly, “no one shall be convicted of an offence except on the basis of individual criminal responsibility”²⁶. This paradigm is challenged by Aws where human agency –hence intent– typically preceding targeting and the use of force is blurred by the autonomy of these weapons. The Rome Statute (Rs) establishing the International Criminal Court (Icc), and requiring knowledge and intent to establish liability²⁷, illustrates well that Aws challenge the paradigm of individual criminal responsibility. As Coco convincingly notes, when Aws make targeting choices – regardless of whether humans are *in the loop* or *on the loop*), international criminal law (ICL) is ill-equipped to establish the *mens rea* or subjective elements (Coco 2023: 11-12). Considering the war crime of intentionally directing attacks against the civilian population²⁸, even if a human must validate or supervise the machine’s targeting choices, human failure to recognize a wrong choice by the machine would likely deny their knowledge of the civilian status of victims and, consequently, also their intent. The outcome would be similar for the war crime of intentionally directing an attack causing excessive incidental death, injury or damage²⁹. The perpetrator would not be blamed for not being aware of the machine’s failure to evaluate that the attack would cause excessive damage. Likewise, the responsibility of commanders and superiors³⁰ would not be engaged if they did not know or should have known that the machine would fail in critical functions of targeting and engagement. Even if such knowledge existed, robots –lacking human

²⁶ Additional Protocol I, Art. 75(4)(b); Additional Protocol II, Art. 6(2)(b)

²⁷ UN General Assembly, Rome Statute of the International Criminal Court (last amended 2010), ISBN No. 92-9227-227-6, UN General Assembly (17 July 1998), Article 30.

²⁸ Rs, Art. 8.2, section b, fraction i.

²⁹ *Ibidem*, fraction iv.

³⁰ Rs, Article 28.

agency– could not be considered subordinates who committed crimes for which superiors are responsible in the current state of international law (Coco 2023: 15-16).

The *control* problem brings about the immediate dilemma of who is internationally responsible for IHL violations committed as a result of Aws failures during targeting and the use of force. If humans cannot be held criminally responsible, who then? Scholars are putting their legal imagination to work to answer this question due to the increasing use of Aws –among which critically drone-bombing– in conflict settings³¹. Acquaviva (2021) suggests various alternatives of how a responsibility framework might work. A first option would consider that only humans are responsible for Aws mishaps because IHL and its sanctioning through ICL are built on that premise. This solution may require some adaptations though. For instance, the Rs could be amended to introduce individual criminal responsibility for the mere use of Aws and, therefore, exceptionally require a *dolus eventualis* –a lower threshold than intent– for accepting the risk of using Aws in critical functions. Also, superiors could be held responsible based on the possibility that they infer knowledge and intent from situations where they “became aware of the crimes, accepted them and [thus] came to intend them”³². A second option for the attribution of responsibility would consider Aws legal persons subject to liability. Even if this scenario seems unpalatable today because robots have no agency and, therefore, cannot be accountable for their actions, this might change in the future. In this line, scholars have suggested establishing different types of agencies that distinguish between the purposes of algorithms and those of individuals who use them,

³¹ For instance, see Paton Walsh, Nick (2023): “Inside the unit deploying small explosive drones in Ukraine’s cash-strapped war” in *CNN*. Available at: «<https://edition.cnn.com/2023/12/12/europe/ukraine-kherson-explosive-drones-intl/index.html>» [Consulted on 28 February, 2024].

³² International Criminal Tribunal for the Former Yugoslavia, *Prosecutor v. Momcilo Krajisnik (Trial Judgment)*, 27 September 2006: paras. 1096 ff; International Criminal Tribunal for the Former Yugoslavia, *Prosecutor v. Momcilo Krajisnik (Appeals Chamber)*, 17 March 2009.

and attribute responsibility accordingly (Asaro 2016: 193-194). A third alternative which is receiving attention would envisage a hybrid system of human and machine responsibility (Acquaviva 2021: 37). When humans retain meaningful control, there would be no questioning that ICL applies to them. Yet, in the absence of meaningful human control, criminal punishment may be foregone and there might be room to consider other forms of responsibility. These could include: State responsibility for allowing the deployment of unlawful/unpredictable weapons; civil remedies against the companies who created and procured Aws, and against the individuals who programmed these weapons wrongly. Overall, the view is that the harms caused by Aws cannot go unpunished and responsibility should be assumed by the different societal actors involved.

IV. SUGGESTIONS FOR A DOMESTIC LEGAL FRAMEWORK

Mexico's regulatory gap dealing with Aws is inconsistent with the State's leadership on this matter, especially in view of the escalation of violence involving explosive-equipped drones. On the one hand, Mexico's proactiveness within the Ccw (submitting jointly with Brazil and Chile) a draft proposal for a legally binding treaty on Aws attests to strong views on the IHL standards that should apply to these weapons. Essentially, Mexico's claim, also underpinning the Ccw Guiding Principles, is that an ethical approach to regulate Aws is necessary to ensure respect for the principles of humanity and dignity underpinning IHL and IHRL. Accordingly, respecting these principles requires, as core tenets, retaining the meaningful human control of Aws and ensuring individual accountability for the failures of Aws resulting in IHL and IHRL violations. On the other hand, Mexico's domestic legal system does not address the increasing use of Aws, especially in confrontations with cartels, from the ethical lens of human control and responsibility that it sponsors internationally. Even though the National Agenda on Artificial Intelligence is underpinned by an

ethical pillar that enshrines human rights and non-discrimination –especially on gender and indigenous grounds–, it is silent on the role of the military and the police as key security issues affected by AI (Casados *et al.* 2020: 47-75). Notwithstanding these inconsistent approaches, the existing policy framework contains the necessary elements to articulate a sensitive regulation of Aww and give an adequate solution to the phenomenon of armed drones. Based on the discussion of the previous section, it is argued that a domestic legal framework of Aww in Mexico should consider two relevant issues –human control and discrimination– to which the article now turns.

1. Integrating meaningful human control

In November 2023, the Congress backed a presidential proposal to regulate the use of drones for criminal purposes which, at the time of writing, is being debated by the authorities and it is likely to make its way into the Federal Penal Code and the Federal Law of Fireweapons and Explosives. The proposal criminalizes various scenarios: i) *whoever* uses drones to throw any explosive object or weapons as well as chemical substances over persons or goods –10 to 20 years of imprisonment, increased to 30 years if the military or security forces are affected–; ii) *whoever* imports, manufactures, arms or adapts drones for the transport of explosives –5 to 10 years of imprisonment–, and iii) using drones for the surveillance of public authorities to evade or attack them –3 to 10 years of prison–. For the purpose of this discussion, it is the use of drones for critical target functions that matters.

The initiative to criminalize the use of explosive drones is progressive in various regards. It sets a leading example regulating an issue which is absent in many jurisdictions around the world and for which there are no international or regional standards. Also, the proposal encompasses responsibility for a broad range of situations such as negligence, automated bias or complacency. These situations are uncovered by current ICL standards under

the Rs. This requires intent of first or second degree³³, despite scholars have considered the inclusion of lower thresholds than intent desirable in domestic regulations, given the nature of Aws (Coco 2023: 14). In addition, the presidential proposal establishes the responsibility of manufacturers and others who contribute to the process of making the use of robots for criminal purposes possible. This promotes a policy of realistic accountability and inclusiveness of the different societal actors responsible for the harms caused by Aws (Acquaviva 2020: 37; Asaro 2016: 192).

The Congress' proposal to criminalize explosive drones is, however, affected by certain weaknesses in the light of Mexico's international obligations and policies. One shortcoming is that the initiative does *not address* the issue of human control. Hence, it is contradictory with Mexico's position under the GGE-CCW sponsored framework where the human control of Aws is considered the key principle of the proposed regulation³⁴. The presidential initiative criminalizes anyone *who throws* –explosives from a drone–. Therefore, it merely covers situations that involve an active human subject making decisions remotely and, therefore, which involve significant human control. However, it does not address the kern of the ethical debate around Aws which focuses on the lack of human control. In view of the fact that explosive drones, AI-powered weapons affected by the control problem, are increasingly used in confrontations with cartels, it would be reasonable to expect that this point would not be missed out in domestic regulations.

Another limitation of the presidential initiative is that it establishes a *blanket* prohibition of drone bombing which may contradict Mexico's IHL obligations as well as the CCW Guiding

³³ Rs, Articles 25 and 28.

³⁴ Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System (2022): "Elements for a Legally Binding Instrument to Address the Challenges Posed by Autonomy in Weapon Systems", CCW/GGE.1/2022/WP.5.

Principles on Aws. Indeed, IHL allows the use of legal weapons between the warring parties provided that its principles are respected. Moreover, States have so far not fully banned Aws in armed conflict but, as the GGE Guiding Principles show, they are paving the way towards the regulation of Aws³⁵. The proposal being discussed in Congress thus seems suitable for law enforcement situations where domestic regulations apply and attacks should be measured by *more stringent* IHRL requirements on the lethal use of force³⁶. Yet, as of today, IHL obligations do not ban drone bombing if human control is retained and humanitarian principles are respected. A future domestic framework could therefore consider the possibility of distinguishing between situations where law enforcement and IHL apply as different scenarios and with different legal implications.

The above considerations allow suggesting in this article that a future regulation of explosive drones aligned with Mexico's IHL obligations would consider prohibiting: Whoever uses or contributes to the use of drones, *with or without* meaningful human control, to throw any explosive artifacts or weapons harming persons or objects in violation of international obligations under IHRL and IHL. This formulation seems appropriate for two reasons. Firstly, it would allow attributing individual responsibility when drone bombing was carried out without meaningful human intervention, even against those highly responsible in the chain of command, not clearly covered by the present initiative. Secondly, the suggested provision would allow considering both situations of law enforcement and armed conflict, potentially applicable to armed confrontations in Mexico where the State has different legal obligations under IHL and IHRL. Accordingly, faced with a situation of drone bombing, public officials may ask: Is it a law enforcement

³⁵ *Ibid.*

³⁶ These include sufficient legal basis, necessity and proportionality. See General Assembly of the United Nations (2014): "Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns", paras. 86-92 and 136, A/HRC/26/36.

or a IHL situation? If it falls under law enforcement, a blanket prohibition on the use of drone bombing may apply –in line with the presidential initiative–. If it falls under IHL, this regime is silent about the use of Aws and the only limitations on the use of weapons require respect for the law of armed conflict. However, in line with Mexico’s foreign policy under the GGE-CCW framework, drone bombing attacks falling under IHL should be prohibited whenever IHL is violated *and* whenever the absence of meaningful human control of Aws results in IHL violations that demand accountability.

2. Addressing bias

The presidential initiative criminalizes the activities involved in the process leading to the use of explosive drones for critical functions, including the manufacturing, arming, acquisition or adaptation of drones to transport explosives. Yet, if the aim is to hold social actors responsible for the harm caused by Aws, there is still a margin of appreciation for more inclusive regulation. As explained above, an important reason why Aws make mistakes during targeting and engagement is discrimination imbedded in the process of coding. If, as explained above, bias is structural, there is considerable potential in establishing a duty of non-discrimination throughout the coding cycle which only adequate regulation can bring into effect.

Discrimination in AI is a matter of great concern and which severely affects Mexico. Not only because inequalities pre-existing confrontations are likely to exacerbate during armed hostilities (Red de Investigación Feminista en Inteligencia Artificial *et al.* 2023: 4). Also, because algorithm bias further aggravates these inequalities (Spitx *et al* 2023: 6-7). Discrimination in Mexico is intimately connected with organized crime. Studies reveal that AI in Mexico, and technology in general, is designed for men resulting in the aggravation of gender inequalities and human rights violations in realms such as violence against women, the labor market, education and health, among others. To halt these discriminatory processes, studies are recommending Mexican policymakers

to adopt an inclusive anti-discriminatory approach considering gender and other intersecting grounds in the coding of AI so that robots can capture the complexity of societal needs (Spitx et al 2023: 4-7; Casados *et al.* 2020: 62-63).

Mexico's national policies recognize the *rationale* linking discrimination to violent crime. Accordingly, there is a good reason to consider this connection in a domestic regulation of Aws as well. The National Agenda on Artificial Intelligence enshrines respect for human rights and non-discrimination as an ethical axis across the agenda. It highlights the need to identify the most vulnerable groups –noting women and indigenous peoples– and discriminatory practices in order to obtain representative data for algorithm coding. Actually, the aim of the National Strategy is to use the potential of AI to revert discrimination in Mexico by means of inclusive data that improves decision-making and promotes human rights (Casado et al 2020: 60-62). This goal is also linked with the essence of the National Plan on Peace and Security. The latter recognizes that social inequalities are the root cause of violent crime, hence, it declares that the way out of violence is to promote human rights and non-discrimination that will bring peace (Ch-
anona 2019: 6).

To benefit from the advantages of AI improving data accuracy and promoting respect for human rights in conflict, scholars have suggested the regulation of a non-discriminatory approach dealing with Aws. A duty of non-discrimination should cover those involved in the chain preceding the use of Aws and allow the attribution of criminal and civil responsibility for the damages caused by bias embedded in Aws coding (Acquaviva 2021: 37-38; Asaro 2016: 192). This approach suits well Mexico's domestic law and policy as it would strengthen the aims of the National Agenda on Artificial Intelligence and the National Plan on Peace and Security underpinned by human rights and non-discrimination. Thus, establishing a duty of non-discrimination in a future regulation of Aws would bring in alignment domestic policies dealing with

AI and security around the promotion of equality. Furthermore, a duty of non-discrimination squarely fits the approach of Federal Law to Prevent and Eliminate Discrimination (2003) which allows liability for omissions and actions *without* the intention of discrimination but with a discriminatory effect³⁷. This is the case of manufacturers, developers, traders, and others involved in the process of AI, who may act based on cultural prejudices but without intending to harm. Accordingly, a non-discriminatory approach to Aws in Mexico may consider extending liability to those contributing –even if unwillingly– to biased algorithms that result in machine failures and, thus, IHL and IHRL violations that could have been prevented with due diligence.

As outlined above, a future regulation of Aws in Mexico may provide for: A duty of non-discrimination for those involved in the process of manufacturing, developing, deploying and controlling Aws, establishing their criminal or/and civil responsibility, as corresponds. In this sense, faced with bias dealing with military applications of Aws, public officials may ask questions such as (Chandler 2021: 23; Houghton 2022: 3-4): Was the attack committed due to discrimination or did it result in discrimination on one or intersecting grounds due to bias in the coding of Aws? Was the participation and representation of vulnerable groups – women and children, indigenous, etc.– considered in the processes of algorithm coding and decision-making? Was conflict prevention considered in the system design? Did the model used consider the protection of human rights, including reparations? Did these protections apply equally to women, girls, men and boys without discrimination? If ending armed confrontations in Mexico largely depends on promoting human rights and equality, ensuring that Aws data is not discriminatory becomes a critical ethical and legal component of the national strategy.

³⁷ Ley Federal para Prevenir y Eliminar la Discriminación, publicada en el Diario Oficial de la Federación el 11 de junio de 2003. Article 1.

V. CONCLUSION

This article addressed the legal gap on the use of Aws in armed confrontations in Mexico, especially drone bombing, dealing with organized criminal groups. It argued the need for an IHL approach in certain instances of violence involving Aws which is sensitive to the issues of human control, discrimination and individual criminal responsibility. Two issues which are hampering the application of this perspective were identified. Firstly, Mexico's armed hostilities are addressed from a law enforcement perspective due to the absence of political will to apply IHL obligations. Secondly, there is a lack of international standards on Aws due to ethical concerns around the issues of human control and international criminal responsibility for the violations caused by Aws. These issues raised a major concern: If Aws have autonomy to carry out the critical functions of targeting and engagement, is there room for human accountability? It was argued that, as a key supporter crafting the Guiding Principles in the framework of the Ccw, Mexico holds a strong international position on the IHL standards that should apply to Aws. These are underpinned by the principles of significant human control and individual criminal responsibility. Accordingly, the same standards should underpin Mexico's domestic legal framework to address the increasing escalation of violence involving drone bombing by powerful criminal groups.

Based on a discussion of the key legal issues posed by Aws, the article suggested the tenets of a domestic legal framework aligned with Mexico's IHL obligations and national policies underpinned by non-discrimination and human rights. First, it was suggested that a domestic framework should expressly address the control problem establishing the illegality of Aws without meaningful human control as well as those Aws that violate IHL. A more blanket criminalization –i.e. the 2023 presidential initiative on drone bombing– would be suitable for law enforcement situations which fail to reach the threshold of armed conflict. Second, it was recommended that a domestic legal framework of Aws

should establish a duty of non-discrimination for those involved in putting Aws to work –from developers to the military– holding them responsible for machine bias resulting in IHL and IHRL violations. This broad form of social accountability is necessary for prevention purposes and to ensure reparations. Moreover, a duty of non-discrimination regarding Aws is consistent with and enhances the implementation of core Mexican policies in relation with Ai and security, namely, the National Agenda on Artificial Intelligence and the National Plan on Peace and Security, both underpinned by human rights and non-discrimination standards.

Protecting human control and non-discrimination are the core ethical reasons for which establishing a regulatory framework of Aws is critical for Mexico and for any State. Meaningful human control ensures that human agency –judgment, not automation– is behind targeting and use of force. Therefore, it involves a progressive choice to promote the rights to dignity –decisions affecting human life should be taken by humans, not by robots– and reparations in armed confrontations. As for the duty of non-discrimination in the coding of Aws, this standard shows commitment to put technology to work to improve human rights in armed conflict, promoting a culture of peace. Mexico’s support for these standards is commendable. Establishing a domestic regulation along these lines can set a compelling example to advance this urgent topic in armed confrontations where the autonomy of weapons is becoming increasingly harmful.

BIBLIOGRAPHY

Acquaviva, Guido (2021): “Autonomous weapons systems controlled by Artificial Intelligence: a conceptual roadmap for international criminal responsibility”, in *SSRN*. Available at «<https://ssrn.com/abstract=4070447>» [Consulted 31 January, 2024].

- Amoroso, Daniele and Tamburini, Guglielmo (2020): “Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues”, in *Current Robotics Report*, Vol. 1, 187-194.
- Arkin Ronald C. (2009): “*Governing Lethal Behavior in Autonomous Robots*”, Chapman and Hall/CRC, New York.
- Asaro, Peter M. (2016): “The Liability Problem for Autonomous Artificial Agents”, in *Association for the Advancement of Artificial Intelligence*, Spring Symposium, Num. 4, 190-194.
- Bellal, Annyssa (2017): “The War Report: Armed Conflicts in 2017” in *Geneva Academy*. Available at: «https://www.geneva-academy.ch/joomlatools-files/docman_files/The%20War%20Report%202017.pdf» [Consulted on 23 February, 2024].
- Bunker, Robert J. and Sullivan, John P. (2021): “Mexican Cartels are Embracing Aerial Drones and They’re Spreading” in *War on the Rocks*. Available at: «<https://warontherocks.com/2021/11/mexican-cartels-are-embracing-aerial-drones-and-theyre-spreading/>» [Consulted on 31 January, 2024].
- Casados, Daniel *et al.* (2020): “Agenda Nacional Mexicana de Inteligencia Artificial”, in *Coalición IA2030Mx*. Available at: «<https://www.ia2030.mx/agenda2020>» [Consulted on 27 February, 2024].
- CBS (2023): “Drug cartels are sharply increasing use of bomb-dropping drones, Mexican army says”, in *CBS News*. Available at: «<https://www.cbsnews.com/news/drug-cartels-more-bomb-dropping-drones-mexico-army/>» [Consulted on 01 February, 2024].
- Chandler Katherine (2021): “Does Military AI Have Gender? Understanding Bias and Promoting Ethical Approaches in Military Applications of AI”, in *UNDIR Research*
-

Report. Available at: «<https://unidir.org/publication/does-military-ai-have-gender-understanding-bias-and-promoting-ethical-approaches-in-military-applications-of-ai/>» [Consulted on 27 February, 2024].

Chanona, Alejandro (2019): “México: la reforma del sector seguridad y la estrategia contra la delincuencia organizada de Andrés Manuel López Obrador”, in *Real Instituto Elcano ARI/3/2018*. Available at: «<https://media.realinstitutoelcano.org/wp-content/uploads/2021/11/ari3-2018-chanona-mexico-reforma-seguridad-estrategia-contra-delincuencia-organizada-lopez-obrador.pdf>» [Consulted on 05 February, 2024].

Coco, Antonio (2023): “Exploring the Impact of Automation Bias and Complacency on Individual Criminal Responsibility for War Crimes”, in *Journal of International Criminal Justice*, Vol. 21, Issue 05, 1077–1096.

D’Cunha, Samit *et al.* (2024): “Defining armed conflict: some clarity in the fog of war”, in *Humanitarian Law & Policy*. Available on: «<https://blogs.icrc.org/law-and-policy/2024/05/02/defining-armed-conflict-some-clarity-in-the-fog-of-war/>» [Consulted on 23 May, 2024].

Farrés Jiménez, Andrea (2021): “Embedding Gender in International Humanitarian Law: Is Artificial Intelligence Up to the Task?”, in *Just Security*. Available at: «<https://www.justsecurity.org/77970/embedding-gender-in-international-humanitarian-law-is-artificial-intelligence-up-to-the-task/>» [Consulted on 27 February, 2024].

González, Isabella (2023): “El Ejército mexicano incrementa su poderío militar: escala 11 posiciones en el ranking mundial de las Fuerzas Armadas”, in *El País*. Available at: «<https://elpais.com/mexico/2023-03-15/el-ejercito-mexicano-incrementa-su-poderio-militar-escala-11-posi->

ciones-en-el-ranking-mundial-de-fuerzas-armadas.html» [Consulted on 27 February, 2024].

Håkansson Schmidt, Alexandra (2022): “Action Brief-From Warfare to Peace building: Employing Artificial Intelligence for Women, Peace and Security”, in *UN Women*. Available at: «<https://asiapacific.unwomen.org/en/digital-library/publications/2022/04/from-warfare-to-peacebuilding>» [Consulted 13 February, 2024].

Houghton, Isabelle (2022): “The Implications of Artificial Intelligence in Women, Peace and Security: Part 2”, in *Our Secure Future*. Available at: «<https://oursecurefuture.org/our-secure-future/news/implications-artificial-intelligence-women-peace-and-security-part-2>» [Consulted on 27 February, 2024].

ICJ (2004): “*Advisory opinion of the International Court of Justice on the Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory*”, in *United Nations*. Available at: «<https://www.un.org/unispal/document/auto-insert-178825/>» [Consulted on 28 February, 2024].

ICRC (2023): “What you need to know about artificial intelligence in armed conflict” in *International Committee of the Red Cross*, United Kingdom. Available at: «<https://www.icrc.org/en/document/what-you-need-know-about-artificial-intelligence-armed-conflict>» [Consulted on 20 May, 2024].

ICRC (2021): “Position on Autonomous Weapon Systems: ICRC Position and Background Paper” in *International Review of the Red Cross*, Geneva, 1-12.

ICRC (2005): “International Humanitarian Law Databases, Rules”, in *ICRC*. Available at: «<https://ihl-databases.icrc.org/en/customary-ihl/v1>» [Consulted on 20 February, 2024].

ICRC (1987): “Commentary of 1987, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the

Protection of Victims of International Armed Conflict (Protocol I), 8 June 1977”, in *ICRC*. Available at: «<https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-36/commentary/1987?activeTab=undefined>» [Consulted on 28 February, 2024].

Informador.Mx (2024): “Todo sobre la emboscada del CNG que dejó 3 lesionados al sur de Jalisco”, in *Informador.Mx*. Available at: «<https://www.informador.mx/jalisco/CJNG-TO-DO-sobre-la-emboscada-que-dejo-3-lesionados-al-sur-de-Jalisco-20240426-0065.html>» [Consulted on 29 May, 2024].

Janetsky, Megan (2024): “A drug cartel has attacked a remot Mexican community with drones and gunmen, rights group says”, in *Associated Press*. Available at: «<https://apnews.com/article/mexico-cartel-attack-guerrero-drones-helidoro-castillo-ed795770ce9a09b1c063d91913268b99>» [Consulted on 29 May, 2024].

López Obrador Andrés Manuel (2018): “Plan Nacional de Paz y Seguridad 2018-2024”, in *AMLO* Available at: «https://lopezobrador.org.mx/wp-content/uploads/2018/11/PLAN-DE-PAZ-Y-SEGURIDAD_ANEXO.pdf» [Consulted on 15 February, 2024].

Mainzer, Klaus (2020): *Artificial Intelligence: When do Machines Take Over?*, Springer, Berlin.

Mexico Now (2023): “Congress approves penalties for drone misuse”, in *Mexico Now*. Available at: «<https://mexico-now.com/congress-approves-penalties-for-drone-misuse/>» [Consulted on 28 February, 2024].

Mondragón Toledo, Gabriel (2023): “NarcoDrones Have Become a Growing Scare Tactic in Mexico’s Drug Wars”, in *Inkstick*. Available at: «<https://inkstickmedia.com/narcodrones->

have-become-a-growing-scare-tactic-in-mexicos-drug-wars/»
[Consulted on 01 February, 2024].

- Muggah, Robert (2023): “Organized crime in armed conflicts and other situations of violence”, in *International Review of the Red Cross*, Vol. 105, Num. 923, 569-574.
- Padin, Juan Francisco (2023): “Opening Pandora’s box: The case of Mexico and the threshold of non-international armed conflicts”, in *International Review of the Red Cross*, Vol. 106, Num. 923, 772-794.
- Paton Walsh, Nick (2023): “Inside the unit deploying small explosive drones in Ukraine’s cash-strapped war”, in *CNN*. Available at: «<https://edition.cnn.com/2023/12/12/europe/ukraine-kherson-explosive-drones-intl/index.html>» [Consulted on 28 February, 2024].
- Plaw, Avery *et al.* (2020): “The process of technologization of the drug war in Mexico”, in *The Politics of Technology in Latin America* (Volume 1), Plaw Avery *et al.* (eds.), Taylor & Francis, United Kingdom, 95-108.
- Porup, J.M. (2017), “Hunting for Mexico’s mass graves with machine learning”, in *BIZ & IT*. Available at: «<https://arstechnica.com/information-technology/2017/04/hunting-for-mexicos-mass-graves-with-machine-learning/>» [Consulted on 28 February, 2024].
- Red de Investigación Feminista en Inteligencia Artificial *et al* (2023): “Aportes para el llamado de la ANUDH sobre los desafíos de la tecnologías nuevas y emergentes en el ámbito militar”, in OHCHR. Available at «<https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/advisorycommittee/techmilitarydomain/submissions/3-academia-red-investigacion-feminista-ia.pdf>» [Consulted on 27 February, 2024].

- RULAC (2022): “Non-International Armed Conflicts in Mexico”, Geneva Academy of International Humanitarian Law and Human Rights, Geneva, 30 May 2021. Available at «<https://www.rulac.org/browse/conflicts/non-international-armed-conflict-in-mexico>» [Consulted on 28 February, 2024].
- Salgado Espinoza, Raúl (2020): “Ethical controversies about Lethal Autonomous Weapons Systems, Views of small South American states” in *The Politics of Technology in Latin America (Volume 1)*, Plaw Avery et al. (eds.), Routledge, London, 65-79.
- Sharkey Noel E (2012): “The evitability of autonomous robot warfare”, in *International Review of the Red Cross*. Available at: «<https://international-review.icrc.org/articles/evitability-autonomous-robot-warfare>» [Consulted on 27 February, 2024].
- Shuldiner, Henry, (2022): “Military-Grade Weapons in Michoacán Put Mexican Forces on War Footing”, in *Insight Crime*. Available at: «<https://insightcrime.org/news/war-type-weapons-in-michoacan-put-mexican-forces-on-battle-footing/>» [Consulted on 05 February, 2024].
- Spitx, Sarah et al. (2023): “A Gender Perspective on Artificial Intelligence (AI) and the Tech-Sector. Position Paper for the Multi-Stakeholder Dialogue in Mexico”, in *Alexander von Humboldt Institute for Internet and Society*. Available at: «<https://www.hiig.de/publication/gender-ai-techsector-mexico/>» [Consulted on 27 February, 2024].
- Stevenson, Mark (2021): “Mexico’s drug war uses drones, human shields, gunships”, in *AP*. Available at: «<https://apnews.com/article/business-mexico-army-gangs-f63d6aac1d8057c8024ae7688c465696>» [Consulted on 01 February, 2024].

- Sullivan, John P (2021): “Non-International Armed Conflict: Mexico and Colombia”, in *Revista do Ministério Público Militar*, Vol. XLVI, Num. 35. Available at: «www.academia.edu/62379204/Non_International_Armed_Conflict_Mexico_and_Colombia» [Consulted on 05 February, 2024].
- WOLA (2022): “México profundiza la militarización. Los hechos muestran que es una estrategia fallida” in WOLA. Available at: «<https://www.wola.org/es/analisis/mexico-profundiza-militarizacion-hechos-muestran-estrategia-fallida/>» [Consulted on 16 February, 2024].

Prolegómenos de inteligencia artificial en la impartición de justicia. Implicaciones prácticas y nuevos desafíos para los derechos humanos

*Prolegomena of artificial intelligence
in the administration of justice. Practical implications
and new challenges for human rights*

ALEJANDRO VILCHIS ROBLES
Poder Judicial de la Federación
ORCID: 0009-0002-9781-0677

Fecha de recepción: 20 marzo 2024

Fecha de aceptación: 10 junio 2024

*El éxito en la creación de IA sería
el evento más grande en la historia de la humanidad.
Desafortunadamente,
también podría ser el último,
a menos que aprendamos a evitar los riesgos
(Stephen Hawking)**

SUMARIO: I. Nota introductoria. II. Una aproximación a la IA: conceptos clave. III. Herramientas de IA en el derecho y en la impartición de justicia. IV. Función jurisdiccional y toma de decisiones judiciales. V. Aplicaciones prácticas de la IA a la judicatura: una visión en prospectiva. 1. Juicios o procedimientos automáticos. 2. Admisión de demandas y recursos. 3. Admisión y valoración de pruebas. 4. Audiencias y comparecencias. 5. Sentencias. VI. Nuevos desafíos para los derechos humanos con la implementación de IA en la función jurisdiccional. 1. Juez natural. 2. Derecho

* Independent (2014): "Stephen Hawking: 'Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?'". Disponible en: «<https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html>» [Consultado el 1 de junio de 2024].

a la igualdad y no discriminación. 3. Imparcialidad judicial. 4. Legalidad y seguridad jurídica. 5. Justicia pronta. 6. Derecho de defensa. 7. Presunción de inocencia. 8. Derecho a la privacidad y protección de datos. 9. Derecho a recurrir el fallo. VII. Conclusiones.

RESUMEN: Las modernas tecnologías están transformando nuestra realidad y con ello marcan una nueva era en la historia de la humanidad. En este contexto de innovación tecnológica surge la inteligencia artificial (IA), que tiene la posibilidad de imitar distintas actividades humanas a partir del procesamiento de información y mediante la operación de algoritmos inteligentes. La IA ha tenido un auge importante en los últimos tiempos y debido a su creciente desarrollo puede aplicarse en diversos ámbitos, entre ellos, el derecho y particularmente la impartición de justicia. En ese sentido, los sistemas de IA pueden auxiliar a la judicatura para desarrollar eficazmente sus funciones ordinarias y para abatir distintos problemas que la caracterizan. Por ello, es preciso conocer cuáles son las principales implicaciones prácticas de estos sistemas de IA en la judicatura y qué desafíos representarán para los derechos humanos.

ABSTRACT: Modern technologies are transforming our reality, marking a new era in human history. In this context of technological innovation, artificial intelligence (AI) emerges, with the ability to imitate distinct human activities through information processing and intelligent algorithms. AI has experienced significant growth recently, and due to its ongoing development, it can be applied in various fields, including law and particularly in the administration of justice. In this sense, AI systems can assist the judiciary in effectively carrying out its ordinary functions and addressing various challenges it faces. Therefore, it is essential to understand the main practical implications of these AI systems in the judiciary and the potential impact on human rights.

PALABRAS CLAVE: *Inteligencia artificial, algoritmos, judicatura, sistemas expertos, derechos humanos.*

KEYWORDS: *Artificial intelligence, algorithms, judiciary, expert systems, human rights.*

I. NOTA INTRODUCTORIA

En la actualidad las modernas tecnologías están transformando nuestra realidad. Se encuentran presentes en todas —o casi todas— las actividades humanas. Facilitan nuestras tareas cotidianas o por lo menos a las personas que tenemos acceso a la tecnología. En pleno siglo XXI vivimos, pues, en una revolución tecnológica que avanza deprisa y que, por lo mismo, está marcando una nueva era en la historia de la humanidad. En este contexto, surge la Inteligencia artificial [en adelante IA], como un universo de innovación tecnológica, en el que máquinas, robots y sistemas computacionales, por medio de algoritmos inteligentes, están imitando cada vez más actividades humanas a partir del procesamiento de una gran cantidad de información.

El avance de la IA es sorprendente. Actualmente, la IA tiene grandes capacidades y aplicaciones: crea y replica voces, imágenes, videos, música, textos e incluso obras de arte; genera asistentes virtuales inteligentes; programa vehículos autónomos; optimiza cadenas de suministros; también se utiliza en la elaboración de diagnósticos médicos; contribuye en las tareas de seguridad y prevención del delito; auxilia en las labores de la judicatura, entre otras funciones.

Todo este avance nos conduce a pensar que la IA puede tener efectos positivos para los seres humanos, en la medida en que se ha concebido como un medio para facilitar —y reemplazar— una gran cantidad de actividades que las personas realizamos cotidianamente. Sin embargo, no debemos desconocer que la IA también comportará múltiples consecuencias de carácter negativo, entre ellas, el desplazamiento laboral, la existencia de discriminación algorítmica, la intromisión en la privacidad y seguridad de los datos personales, la dependencia de las personas por el uso de la tecnología, el impacto ético y social de la IA, la posibilidad de su uso con fines maliciosos, así como la afectación que puede generar la IA en los derechos humanos (Equipo de Expertos en Ciencia y Tecnología 2024; Ríos 2023: 41-43).

Ante esta serie de implicaciones es fundamental abordar los desafíos de la IA de manera cuidadosa para maximizar sus beneficios y aminorar el impacto de sus riesgos. Por esta razón, el derecho debe centrarse en las repercusiones jurídicas que pueden derivarse por el uso IA, no solo porque está llamado a regularlas sino también porque debe interactuar con la IA.

En ese sentido, algunos documentos internacionales dan cuenta sobre la necesidad apremiante de regular la IA, como la Recomendación sobre la ética de la inteligencia artificial (UNESCO 2022), Inteligencia artificial promesas y amenazas (UNESCO 2018), Declaración de Santiago —para promover una inteligencia artificial ética en América Latina y el Caribe— (CMAAAL 2023)¹ y la Carta Iberoamericana de inteligencia artificial en la Administración Pública (CLAD 2023)². Debido a que el impacto de la IA se trata de un tema que genera preocupación a nivel mundial, algunos países, incluido México³, han presentado iniciativas legislativas para regular el control, desarrollo, impacto, tráfico y uso en general de la IA.

Por lo que concierne a la interacción entre IA y derecho, surge la implementación de sistemas de IA en la judicatura. Se trata de un campo propicio para llevar a cabo la aplicación de estos sistemas, debido a que en la impartición de justicia existen una serie de problemáticas que pueden resolverse con la ayuda de IA. Precisamente, “el auge de la IA experimentado en los últimos años coloca a la ad-

¹ Cumbre Ministerial y de Altas Autoridades de América Latina y el Caribe (CMAAALC).

² Centro Latinoamericano de Administración para el Desarrollo (CLAD).

³ Concretamente en México se han presentado tres iniciativas. La primera es de 23 de mayo de 2023 y se dio a conocer en la Cámara de Diputados, se propone expedir la *Ley de Regulación Ética de la Inteligencia Artificial y la Robótica* (Cámara de Diputados 2023). La segunda se presentó en el Senado el 5 de septiembre de 2023 y en ella se propone reformar la fracción XVII del artículo 73 de la Constitución, con el objetivo de facultar al Congreso de la Unión para emitir las normas necesarias que regulen la investigación, desarrollo y aplicaciones de la inteligencia artificial (Senado de la República 2023). La tercera también se presentó en el Senado el 3 de abril de 2024 y consiste en la creación de la *Ley Federal que regula la Inteligencia Artificial* (Monreal 2024).

ministración de justicia en un nuevo contexto tecnológico marcado por la implementación de los sistemas inteligentes” (San Miguel Caso 2021: 287).

En este aspecto, la IA puede contribuir fundamentalmente a combatir el *rezago jurisdiccional* y, con ello, agilizar las funciones ordinarias de los juzgados y tribunales, que, en gran parte, son mecánicas. Otros usos de la IA se relacionan con el perfilado de jueces y magistrados, y los destinados a anticipar el sentido de las resoluciones judiciales (Simón Castellano 2021a: 286). Aunque en realidad “la tecnología parece actuar principalmente como una herramienta de ayuda a la resolución de litigios, más que como un sistema autónomo que pueda realmente procesar, adjudicar o resolver litigios de forma independiente” (Morison y Harkens 2019: 8). Sin embargo, no debemos desconocer que también existen diversos riesgos que debemos enfrentar para contar con una justicia moderna y tecnológica. Por este motivo se dice que los usos de la IA “plantean notables retos para la administración de justicia desde el punto de vista jurídico, así como amenazas significativas a la calidad de nuestros sistemas de justicia, a la protección de los derechos fundamentales y al Estado de derecho” (Simón Castellano 2021b: 8).

En ese sentido, este trabajo tiene el objetivo de ilustrar, desde un enfoque exploratorio y propositivo, que la IA puede traer diversas ventajas para la judicatura, en particular la optimización de las labores jurisdiccionales y el ahorro de tiempo y recursos que permitan asegurar una justicia eficiente y pronta, pero además, que estas ventajas deben abordar y conciliar los desafíos que pueden generarse en los derechos humanos. Para tal efecto, seguiremos la siguiente metodología: en primer lugar, abordaremos algunos conceptos clave relacionados con la IA; después, exploraremos algunos sistemas expertos que operan en la impartición de justicia en la actualidad; en tercer orden, analizaremos el modo en que opera la judicatura y especialmente la forma en que las y los juzgadores toman decisiones; posteriormente nos centraremos en determinar

cuáles serán las aplicaciones prácticas de la IA a la judicatura y los retos de la IA en los derechos humanos; para finalizar expondremos las conclusiones que se derivan de este trabajo.

II. UNA APROXIMACIÓN A LA IA: CONCEPTOS CLAVE

Con frecuencia se identifica a la inteligencia artificial como uno de los avances tecnológicos más importantes de la *Cuarta Revolución Industrial*⁴, no solo por su profundo impacto en la sociedad y la economía, sino también por su potencial y constante desarrollo en la actualidad. En mayor o menor medida la IA va redefiniendo la vida en sociedad. En términos generales, la IA está destinada a igualar —o incluso superar— determinadas tareas que son realizadas ordinariamente por personas y que, por ello, requieren la aplicación de capacidades cognitivas. Esto se logra a partir del procesamiento efectivo de una gran cantidad de información y mediante la operación de algoritmos inteligentes. De esta manera, la IA tiene la capacidad de resolver problemas y tomar decisiones de acuerdo con su modelo de programación. Pero eso no es todo. La IA tiene un amplio almacenamiento de información y cuenta con modelos de aprendizaje automatizado.

Pero a todo esto, ¿qué entendemos por inteligencia artificial? Por principio de cuentas habremos de advertir que no existe una definición única y aceptada sobre IA —lo cual, dicho sea de paso, es uno de los principales desafíos frente a una eventual regula-

⁴ De acuerdo con Klaus Schwab “La Cuarta Revolución Industrial, no obstante, no solo consiste en máquinas y sistemas inteligentes y conectados. Su alcance es más amplio. Al mismo tiempo, se producen oleadas de más avances en ámbitos que van desde la secuenciación genética hasta la nanotecnología, y de las energías renovables a la computación cuántica. Es la fusión de estas tecnologías y su interacción a través de los dominios físicos, digitales y biológicos lo que hace que la Cuarta Revolución Industrial sea fundamentalmente diferente de las anteriores” (Schwab 2016: 21).

ción—. No obstante, citaremos algunas definiciones acuñadas en la comunidad académica para identificar sus notas características.

De acuerdo con Jordi Nieva Fenoll, la IA “describe la posibilidad de que las máquinas, en alguna medida, piensen, o más bien imiten el pensamiento humano a base de aprender y utilizar las generalizaciones que las personas usamos para tomar nuestras decisiones habituales” (Nieva Fenoll 2018: 20). Se trata, pues, de “la disciplina que estudia los sistemas tecnológicos que reproducen algunos procesos, operaciones y tareas del cerebro humano y de los sistemas mismos” (Campuzano Gallegos 2019: 9).

También se ha ubicado a la IA como “una rama de las ciencias computacionales que se encarga de estudiar modelos de cómputo capaces de realizar actividades propias de los seres humanos con base en el razonamiento y la conducta”; o bien, se le identifica como “la posibilidad de que sistemas y máquinas lleguen a realizar acciones complejas que tradicionalmente se consideraban propias de los seres humanos, de manera incluso más eficiente y rápida” (Cárdenas Rincón y Molano Martínez 2021: 2).

En nuestro concepto, podríamos decir que es el procedimiento mediante el cual las computadoras o sistemas, por medio del procesamiento eficiente de datos e información, están en posibilidad de realizar tareas que, cuando las realizan los seres humanos, implican una forma de razonamiento.

En la actualidad existe una distinción básica —aunque bastante controvertida— entre inteligencia artificial *fuerte* y *débil*, tomando como criterio de referencia el grado de réplica del pensamiento humano que una y otra adquieren, así como la autonomía inteligente que se pueda desprender de ambas (San Miguel Caso 2021: 289). De forma tal que la IA débil únicamente puede desarrollar actividades o tareas específicas para las cuales fue programada, por lo que no tiene el alcance de replicar determinados estados cognitivos y, por ende, su capacidad de aprendizaje es limitada. La IA fuerte, en cambio, tiene la posibilidad de imitar el pensamiento humano

y, por tanto, realizar una gran cantidad de tareas o cualquier actividad que se le indique, además de que cuenta con aprendizaje automatizado autónomo. Es importante resaltar que, en nuestros tiempos, predomina la IA que denominamos débil o estrecha.

Otro concepto clave para hablar de IA es el *algoritmo*, el cual desde un enfoque gramatical se concibe como un “conjunto ordenado y finito de operaciones que permite hallar la solución de un problema” (RAE 2024)⁵. Jordi Nieva Fenoll lo concibe como “el esquema ejecutivo de la máquina almacenando todas las opciones de decisión en función de los datos que se vayan conociendo” (Nieva Fenoll 2018: 21).

Podemos convenir en que los algoritmos son la base de la IA y se identifican como una serie de pasos o instrucciones ordenadas y secuenciales, que se utilizan para desarrollar una o varias tareas específicas. Los algoritmos funcionan por medio de sistemas de aprendizaje automatizado y ejecutan las instrucciones ordenadas por estos sistemas. Los algoritmos son, en términos simples, una secuencia de órdenes ejecutivas para realiza una actividad particular ordenada por un sistema.

Según hemos dicho la IA destaca por su amplitud en el almacenamiento de información. Pero ¿qué sería de este cúmulo de información sin una estrategia de análisis y de procesamiento de datos? Esta interrogante nos lleva al siguiente concepto clave *Machine learning* o aprendizaje automatizado. Se trata de una rama de la IA y la informática que se centra en el uso de datos y algoritmos para imitar la forma en la que aprenden los seres humanos, con una mejora gradual de su precisión (IBM 2024).

La cantidad de datos que generamos en la actualidad es exponencial y no deja de aumentar. Esta información se acumula en grandes bases digitales de datos, que conocemos como *Big data*. Tan solo para tener una idea general de esto, en el año 2020 una persona

⁵ Real Academia Española (RAE).

generó, en promedio, 1.7. *Megabytes* por segundo (Universidad Veracruzana 2024). La información se crea por la gran variedad de tecnologías y dispositivos a nuestro alcance, que producen una cantidad enorme de datos. Pero de nada sirve a la IA contar con un cúmulo infinito de información sin un tratamiento adecuado de datos. Precisamente el *Machine learning* permite utilizar y hacer funcional todo este conjunto de datos, pues hace posible que las máquinas, ordenadores o sistemas puedan extraer información y de esta manera aprender de ellos de forma autónoma, tal como lo hacen los seres humanos.

A través del *Machine learning* las máquinas pueden aprender y evolucionar a partir de la búsqueda de grandes bases de datos —*data mining*— y del procesamiento adecuado de información, con la finalidad de resolver problemas. Por lo demás, el aprendizaje automatizado adiestra los algoritmos para hacer clasificaciones o predicciones que pueden extraerse de *Big data* y con ello lograr información más exacta.

Una vez que hemos expuesto algunos conceptos clave relacionados con la IA, corresponde ahora analizar cómo se han ido implementando en distintas latitudes del mundo los sistemas de IA en el derecho y particularmente en la judicatura. Lo anterior servirá de preámbulo para ilustrar por qué la IA puede representar enormes ventajas para mejorar la función jurisdiccional y a su vez nos permitirá advertir cuáles serían los principales riesgos en los derechos humanos.

III. HERRAMIENTAS DE IA EN EL DERECHO Y EN LA IMPARTICIÓN DE JUSTICIA

Sin duda la IA puede permear en todas las áreas y ciencias del conocimiento humano. Hemos destacado que se trata de “una herramienta que cada vez se encuentra más presente en el desarrollo de las actividades cotidianas realizadas por los seres humanos

y, en esta medida, se busca aplicarla en diferentes sectores, como lo es el campo del derecho” (Cárdenas Rincón y Molano Martínez 2021: 1). Esto obedece a que el derecho es un objeto de aplicación tecnológica que puede tener frutos muy interesantes, pero también problemas profundos (Suárez Manrique y De León Vargas 2019: 73). En el campo legal, la IA ha tenido una aplicación importante en las últimas décadas, debido a su constante evolución e implementación en distintos sectores del ámbito jurídico, principalmente en la impartición de justicia.

La IA puede tener distintas aplicaciones en la ciencia jurídica, pero las más comunes giran en torno a la modernización de la justicia y con ello la adopción de herramientas tecnológicas para facilitar las labores jurisdiccionales, aplicación de sistemas inteligentes para la investigación de los delitos, creación de sistemas expertos que colaboren en el diseño de estrategias procesales idóneas ante los tribunales y cálculos estadísticos sobre la posibilidad de éxito en un caso concreto, así como la amenaza sobre los derechos humanos con motivo de la implementación de sistemas de IA.

En algunos países ya se encuentran en funcionamiento diversas herramientas de IA y derecho. En China, por ejemplo, se han instalado juzgados virtuales, que son cyber-cortes especializadas en conflictos de internet, comercio electrónico, privacidad y nombres de dominio, en cuyas controversias no es necesaria la presencia de las partes. En Canadá se analiza la iniciativa de utilizar sistemas inteligentes con el fin de resolver reclamaciones que no excedan de 5,000 dólares y los conflictos sobre propiedad y arrendamientos (Campuzano Gallegos 2019: 172). En el caso de Estonia “se ha propuesto que sea la inteligencia artificial la encargada de analizar gran cantidad de documentos legales y demás información relevante para la toma de decisiones en un determinado caso” (Cárdenas Rincón y Molano Martínez 2021: 13). Además, en este país se ha trabajado en un proyecto para crear un sistema de *jueces-robot* (Segura 2023: 50) basado en aprendizaje automatizado que planea resolver de-

mandas de hasta 7,000 euros o menos, con lo cual la IA permitiría acelerar decenas de casos atrasados (The Technolawgist 2019).

Desde la década pasada comenzaron a operar en distintas latitudes del mundo *sistemas expertos jurídicos* para apoyar fundamentalmente en labores de los despachos jurídicos, la investigación de los delitos y la impartición de justicia.

Un *sistema experto* es “un conjunto de procedimientos lógicos de inteligencia artificial que permiten que los computadores apoyen a los humanos en la tarea de analizar situaciones y adoptar decisiones” (Cárdenas Rincón y Molano Martínez 2021: 6), de modo que su objetivo principal es ayudar a encontrar la solución óptima a un problema específico. En otras palabras, es un subconjunto de la IA que emplea conocimiento humano capturado en una computadora para resolver problemas que normalmente requieren de expertos humanos (Badaró *et al.* 2013: 351). Aun cuando estos sistemas entran en el concepto de IA no son los únicos ni los que más se utilizan en la actualidad, pero sí destacan por su enorme capacidad para colaborar en distintas tareas que habitualmente realizan las personas, entre ellas las de carácter jurídico. De ahí reciben el nombre de sistemas expertos jurídicos. Si bien existe una multiplicidad de ellos, a continuación, presentamos los que, en nuestro concepto, son los más conocidos:

País	Función
España	<p>Es una herramienta de analítica jurisprudencial predictiva destinada a ayudar a definir la estrategia procesal idónea con respecto a un caso concreto, que se basa en el análisis cognitivo de millones de decisiones judiciales para determinar las posibilidades de éxito de un caso.</p> <p>Sistematiza y extrae de forma exhaustiva la inteligencia que reside en un conjunto de más de 10 millones de resoluciones judiciales, procedentes de todas las instancias y órdenes jurisdiccionales de España, a las que se incorporan medio millón de nuevas resoluciones cada año. De la misma manera, se procesan, actualizan, enriquecen e integran de forma continua todos los parámetros de la estadística judicial de la totalidad de juzgados y tribunales del país, incluyendo información sobre la duración, congestión, resolución, pendencia y litigiosidad del sistema jurídico (Jurimetría 2024).</p>
España	<p>Se trata de un sistema experto que genera una serie de gráficos interactivos, obtiene conexiones legales y jurisprudenciales, conceptos clave, mapas conceptuales e informes sobre las posibilidades de éxito en un caso, con la finalidad de ayudar a las y los operadores jurídicos a tomar las mejores decisiones en los litigios.</p>
Francia	<p>Es un programa desarrollado para la recuperación de documentación jurídica que forma parte del derecho francés. Proporciona información jurídica en tiempo real, personalizada y lo más contextual posible, lo que permite a las personas usuarias, abogadas preferentemente, seguir las principales tendencias jurídicas y defender de la mejor manera posible a sus clientes. Ayuda a buscar decisiones en casación, textos de ley, comentarios de sentencias y fichas jurídicas que son actualizadas diariamente (Galindo Ayuda 2019: 44).</p>

Prolegómenos de inteligencia artificial en la impartición de...

País	Función
Estados Unidos de América	Es una plataforma de investigación jurídica que funciona utilizando los recursos de IA en la consulta del derecho de Estados Unidos. Esta tecnología es capaz de determinar con precisión, en segundos, respuestas actualizadas, acordes con el derecho vigente —nuevas decisiones judiciales—, a cuestiones hechas por los usuarios sobre documentación jurídica y que puede auxiliar principalmente en la formulación de la fundamentación de casos concretos (Galindo Ayuda 2019: 45). De hecho, esta empresa diseñó un robot-litigante, que es capaz no solo de construir la solución del caso, sino de argumentar.
Estados Unidos de América	Se trata de un programa que, a través de la recopilación de datos, realiza cálculos sobre el perfil criminológico, riesgo de reincidencia y posibilidades de rehabilitación de personas que han cometido delitos. La herramienta toma datos de registros similares y les añade la evaluación estadística de una entrevista con la persona y un autoinforme, lo que equivale a un análisis de las circunstancias y necesidades socioeconómicas y psicológicas de la persona acusada (Morison y Harkens 2019: 14).
Argentina	Un sistema que puede predecir la solución de expedientes jurídicos simples, cuyas propuestas son calificadas por personal humano. Examina la información del expediente, considerando ciertos patrones que descubre a partir del examen de dictámenes emitidos, realiza algunas preguntas al operador y emite su propuesta de resolución.

Colombia	Es un sistema de IA que busca mejorar el proceso de selección de tutelas en la Corte Constitucional. Todas las tutelas del país llegan a la Corte para su eventual revisión. Esto implica un largo proceso para los funcionarios encargados de seleccionar los casos que la Corte revisará con el fin de emitir una sentencia. PretorIA facilita la identificación del contenido de las sentencias de tutela en el proceso de selección, cuyo resultado es la decisión de cuáles serán revisadas por la Corte (Saavedra y Upegui 2021: 36).
Colombia	Asiste a las y los jueces en la valoración de las pruebas judiciales. Este sistema emplea, a través de la IA, una metodología, procedimiento o algoritmo, que asiste a la persona juzgadora en la construcción ordenada de inferencias o argumentos probatorios. Lo anterior desde la prueba de los hechos hasta las hipótesis o pretensiones que son consideradas en el proceso judicial y que son el fundamento para la motivación de la sentencia que decide dicho proceso (Universidad de Medellín 2024). Tiene como principales usuarios a los estudiantes y profesores de derecho, abogados en ejercicio, fiscales y jueces.
México	Su finalidad es ayudar a la toma de decisiones y a la homogenización del conocimiento colectivo de la comunidad judicial, en el dominio del juicio especial de alimentos (Cáceres Nieto 2023: 69). Los módulos que tiene el sistema son tres: a) modelo tutorial, que es donde se expone o se presenta el consejo de los expertos; b) módulo inferencial, que se encarga de realizar las inferencias a partir de un esquema de argumentación dialógica y derrotable, es decir, se toman en cuenta los argumentos que están enfrentados y la forma de determinar cuál de ellos derrota al otro; y c) módulo financiero, determina la pensión provisional cuando la ley nos dice que se fijará conforme a las necesidades de una parte y las necesidades de la otra, sin tener ningún otro criterio (Martínez Bahena 2012: 841-842).

Fuente: elaboración propia

Existen, sin embargo, otros programas informáticos o asistentes virtuales de IA que, sin ser sistemas expertos legales, pueden auxiliar en tareas jurídicas y en las labores de la judicatura. Tenemos, pues, sistemas de IA generativa como *Bard* de Google o *Copilot* de Microsoft, pero el ejemplo más representativo es *ChatGPT*, que es un asistente virtual de chat operado por inteligencia artificial GPT-3 o GPT-4. Funciona a través de procesamiento de lenguaje natural y puede servir de asistente en diversas tareas y aplicaciones. Permite entablar diálogos con las y los usuarios, quienes por medio de indicadores —*prompts*— obtienen respuestas útiles y coherentes en formato texto. Además, utiliza aprendizaje automatizado, de modo que, mientras mayor sea el acceso a información, entonces más precisas serán tanto búsquedas como respuestas.

ChatGPT ha demostrado ser de gran utilidad en las labores legales. Mediante la ejecución de comandos determinados puede elaborar modelos de demandas y contratos acordes a la legislación y jurisprudencia vigente en cada país, colaborar en el diseño de estrategias jurídicas, buscar precedentes de derecho comparado, establecer la probabilidad de que un determinado hecho se encuentre probado a través del empleo de inferencias probatorias, incluso ha sido utilizado como apoyo en la decisión de controversias judiciales. Sobre este último punto, destacan tres resoluciones en las que se utilizó *ChatGPT* para abordar o resolver la cuestión planteada.

En el primer caso, el juzgado 1 Laboral del Circuito de Cartagena, Colombia, resolvió un asunto sobre el derecho a la salud de un niño autista con ayuda de *ChatGPT*, siendo la primera sentencia en redactarse a partir de inteligencia artificial en el país. La sentencia de 30 de enero de 2023 resolvió el pedido de una madre para que su hijo autista fuera exonerado del pago de citas médicas, terapias y transporte hacia los centros hospitalarios pues la familia no contaba con recursos económicos⁶.

⁶ La sentencia puede consultarse en la siguiente liga: «<https://img.lpderecho.pe/wp-content/uploads/2023/02/Sentencia-032-Colombia-LPDerecho.pdf>»

El segundo caso se trata de la sentencia dictada por el Juzgado Civil Transitorio de San Juan de Miraflores del Poder Judicial de Perú, con competencia en asuntos de familia, en el Expediente 00052-2022-18-3002-JP-FC-01, sobre proceso de alimentos. En este caso, mediante el aplicativo *ChatGPT* se logró fijar la pensión alimenticia que debía percibir una niña⁷. Si bien en esta controversia la aportación de la IA fue para realizar un cálculo matemático, lo que destaca es que la persona juzgadora confió al aplicativo la determinación del monto de los alimentos.

El tercer caso aconteció en nuestro país. Se trata de la sesión pública de la Sala Superior del Tribunal Electoral del Poder Judicial de la Federación (TEPJF) realizada el 29 de marzo de 2023, en donde se analizó el expediente SUP-JE-21/2023, que consistía en un medio de impugnación interpuesto en contra de una sentencia que trataba sobre actos anticipados de precampaña y campaña. El magistrado presidente de la Sala indicó que, al consultar *ChatGPT*, le cuestionó sobre si sabía a quién se refería la expresión “ya sabes quién” —con la finalidad de ilustrar si esa frase pudiera constituir uno de esos actos— y la respuesta del aplicativo fue que era al presidente de México y que esa referencia se popularizó en su campaña electoral de 2018 (TEPJF 2023)⁸.

En este apartado hemos observado que la interacción entre IA y derecho es una realidad y que existen aplicaciones importantes en el campo jurídico en general y en la judicatura en lo particular. Quizá por esto se piensa que la IA debe utilizarse con mayor amplitud en el ámbito judicial, pues de esta manera sería posible mejorar la toma de decisiones para hacerlas más eficientes y precisas. Pero ¿esto realmente es cierto?

⁷ La sentencia puede consultarse en la siguiente liga: «<https://img.lpderecho.pe/wp-content/uploads/2023/03/Expediente-00052-2022-18-3002-JP-FC-01-LPDerecho.pdf>»

⁸ Al analizar el caso, un diverso magistrado del tribunal externó su preocupación de que las resoluciones se tomaran a partir de lo que indicaba *ChatGPT*, pues expresó que, de ser así, se debería cuestionar el rol de los tribunales y si en un futuro serían necesarios.

IV. FUNCIÓN JURISDICCIONAL Y TOMA DECISIONES JUDICIALES

Ahora bien, para hablar sobre las ventajas y los posibles retos en los derechos que enfrentará la impartición de justicia con la adopción de sistemas de IA, es necesario, en primer lugar, comprender el papel que desempeñan las y los juzgadores, cuáles son sus debilidades y fortalezas, pero, sobre todo, la manera en que toman decisiones. Esto es relevante, porque con la IA se pretende replicar o realizar distintas tareas que, ordinariamente, realiza el funcionario judicial.

En línea de principio, habrá que precisar que la labor de juzgar se integra por un conjunto de acciones secuenciales que culminan con una solución a un caso concreto. Para dictar una sentencia es necesario “leer la información del expediente, precisar las pretensiones de las partes, identificar los hechos relevantes, construir ciertos patrones de análisis, buscar legislación, precedentes y doctrina, revisar información técnica, construir argumentos y escenarios y otros procesos internos que determinan la toma de decisión” (Campuzano Gallegos 2019: 162). Después de todo, juzgar no es una actividad única, con un método fundamental que no cambia en ningún contexto. Es necesario distinguir entre los litigios de menor importancia, los asuntos importantes y aquéllos en los que la ley o los hechos son complejos (Morison y Harkens 2019: 22), pues en función de éstas y otras circunstancias se puede tomar la decisión para resolver el caso concreto.

Pese a estas características, no es desconocido que la labor de los órganos jurisdiccionales suele ser bastante mecánica. Situaciones como el sobrecargo de asuntos, los plazos legales y la falta de recursos humanos han orillado a la judicatura a adoptar soluciones de paso para agilizar y optimizar sus funciones —por ejemplo, la utilización de modelos de resolución o formatos—. La mecanización de las decisiones judiciales se ha convertido, más de lo que sería deseable, en una práctica bastante habitual, puesto que las juezas y jueces, para facilitar la emisión de sentencias, clasifican

o sistematizan expedientes y emplean analogías, esto es, copian la fundamentación y motivación de casos similares previamente resueltos, para después adaptarlas al caso concreto (Nieva Fenoll 2018: 24ss; Campuzano Gallegos 2019: 162).

En este ejercicio de toma de decisiones las personas juzgadoras emplean procesos mentales en donde tienen en cuenta elementos subjetivos y objetivos, y se basan en su propio criterio en conjunto con el análisis de la ley, los hechos, las pruebas, los criterios jurisprudenciales, doctrinantes, entre otros (Cárdenas Rincón y Molano Martínez 2021: 3).

La forma en que resuelven las y los jueces obedece con frecuencia a ciertos *heurísticos* que están presentes en las personas juzgadoras. Un heurístico es como una directriz genérica o un atajo mental que pueden seguir las personas para tomar una decisión (Nieva Fenoll 2018: 45). Esto es, cualquier persona, a la hora de decidir, acostumbra a recordar una situación análoga en la que obró de determinada forma, por lo cual se sigue el patrón de conducta más frecuente y la repite para un caso similar. A esto debe sumarse que toda jueza o juez “como cualquier persona humana, tienes vivencias que marcan su ciclo vital, y que desde luego le influyen a la hora de juzgar” (Nieva Fenoll 2013: 172).

También existen otros factores que pueden influir en una decisión judicial, como el clima, la alimentación o la fatiga. Este tipo de situaciones se puede ilustrar con un experimento interesante realizado en Estados Unidos en 2011 (Danziger *et al* 2011; Instituto TsJCDMX 2018), que consistía en saber si las y los juzgadores tenían una diferencia en la toma de decisiones frente a casos idénticos o similares a lo largo del día. El estudio reveló que las personas juzgadoras, durante la mañana, tienden a poner mayor atención en los casos que están juzgando. Pero, conforme pasan las horas, ante casos similares, se van desgastando y toman las decisiones a través de heurísticos, y con ello las resoluciones no son tan apagadas a derecho ni tan racionales.

El mismo estudio expuso que una vez que las y los jueces hacían un pequeño descanso y consumían algún almuerzo o refrigerio retomaban la objetividad que habían mostrado al inicio del día, pero nuevamente con el paso de las horas y frente a casos similares fallaban de una manera menos racional cuanto más cansancio tenían. Esto implicaba que en las primeras horas del día y en momentos posteriores a breves descansos, las personas juzgadas dictaban una mayor cantidad de decisiones favorables y, en contraste, en la tarde/noche y cuanto más agotamiento mental tenían dictaban fallos desfavorables. Este estudio, de alguna manera, centró la cuestión en que la justicia quizá dependía en gran medida de lo que había desayunado la persona juzgada o del descanso que había tenido.

El actuar automático o recurrente de las personas juzgadas también puede tener cabida en la forma en que valoran los medios de prueba. Si bien en este ejercicio se involucran, por una parte, experiencias, ideologías, aspectos subjetivos y, por otra, juicios *lógicos* e inferencias, en realidad la valoración es una actividad que, en mayor o menor medida, tiene un cierto paralelismo o similitud en gran parte de las controversias jurisdiccionales. De modo que si la IA pretende replicar este ejercicio debe comprender cómo valoran las pruebas las y los jueces, pues solo de esta forma sería posible “articular una fórmula algorítmica que emule o replique su proceso lógico y psicológico de valoración probatoria” (Simón Castellano 2021a: 288).

A pesar de esta *mecanicidad* que existe en la toma de decisiones, debemos reconocer que no todas las controversias judiciales se resuelven de la misma forma, existen casos difíciles o paradigmáticos que, por sus propias particularidades, ameritan un pronunciamiento y análisis excepcional por parte de los tribunales. En este punto es importante destacar que, entre las características más valiosas que tienen las y los juzgadores está su creatividad, su criterio jurídico y su sensibilidad ante los problemas sociales. En ese sentido, las y los juzgadores no se limitan —en todos los casos— a aplicar

la norma al caso concreto, sino que tienen la posibilidad de interpretar la ley para adaptarla a la realidad social y, con ello, lograr que el sistema jurídico continúe evolucionando.

Todas estas virtudes son las que han permitido, por un lado, que el orden jurídico encuentre nuevas y mejores soluciones y, por otro, que se protejan progresivamente los derechos humanos. Además, las y los funcionarios judiciales tienen la facultad de someter sus decisiones a una nueva reflexión y, de ser el caso, modificar o superar sus criterios previamente establecidos. Y entonces la pregunta es ¿todo esto lo podrá desarrollar un sistema de IA?

V. APLICACIONES PRÁCTICAS DE LA IA A LA JUDICATURA: UNA VISIÓN EN PROSPECTIVA

En una primera aproximación, podemos afirmar que, en seguimiento a la hipótesis que hemos planteado, la IA será de gran ayuda para facilitar las labores ordinarias de los órganos jurisdiccionales. La IA presenta una utilidad “en todas aquellas labores que se realizan de manera casi automática, lo que implicaría atajar tres de los problemas más acuciantes del sistema judicial: la excesiva burocratización de la justicia, el tiempo en la tramitación y la escasez de recursos” (San Miguel Caso 2021: 291).

Entre los aportes más importantes de la IA en la función jurisdiccional está “la celeridad para la resolución de los procesos judiciales. Lo cual implica descongestión judicial, tutela judicial efectiva y convivencia pacífica” (Suárez Manrique y De León Vargas 2019: 79). De esta manera, los órganos jurisdiccionales pueden optimizar tiempo y trabajo, lo que conduciría a lograr una justicia no solo más ágil sino también más eficiente.

No obstante, más allá de los adelantos que pueda aportar la IA, la impartición de justicia requiere de una profunda revisión y si lo que se busca es una *reingeniería* procesal y judicial (Segura 2023:

64) basada en sistemas de IA, previamente se deben cuestionar distintos aspectos: la estructura procedimental; el sentido y objetivo del diseño formal existente; la razón de ser que subyace a las prácticas jurídicas; y el perfil de profesionales que integran a la judicatura. Todo ello con miras a generar una justicia abierta, transparente y al servicio de la población.

En ese sentido, pudieran programarse sistemas, máquinas u ordenadores para colaborar en las tareas ordinarias que realizan las y los funcionarios judiciales en los órganos jurisdiccionales, pero también para combatir los resabios que existen en este ámbito. Así pues, es posible programar algoritmos inteligentes para capturar información general de las controversias; formar expedientes; analizar las demandas y documentos; elaborar acuerdos de trámite; organizar los expedientes por tipo de asunto⁹; agendar los plazos legales; realizar notificaciones electrónicas, entre otras. También es posible que los sistemas de IA puedan colaborar en actividades más complicadas como la valoración de las pruebas, la selección de la norma o jurisprudencia aplicable al caso o, incluso, en el dictado de determinadas sentencias.

Si un sistema de IA realizara todas estas tareas ahorraría tiempo y recursos tanto materiales como humanos y combatiría el rezago jurisdiccional, al tiempo que garantizaría una justicia pronta y expedita. En ese contexto, la aplicación práctica de sistemas de IA en la judicatura pudiera tener incidencia principalmente en los siguientes rubros:

1. Juicios o procedimientos automáticos

Es posible que la adopción de la IA por parte de la judicatura implique la selección de determinadas controversias o procedimientos que puedan sustanciarse de forma [semi]automática y sin intervención de las partes —o al menos con una intervención re-

⁹ Estos modelos de IA “podrían adaptarse de forma útil como método para clasificar los casos en un esfuerzo por hacer frente a la creciente carga de trabajo” (Morison y Harkens 2019: 28).

ducida—. En estos casos bastaría que la demanda, contestación, pruebas y alegatos se pudieran plantear a través de un sistema o plataforma, y que una vez verificadas estas fases se emitiera el fallo que resuelva el caso. También esto sería posible en procedimientos no contenciosos o controversias que no impliquen una problemática compleja. De esta manera, los juicios o procedimientos se tramitarían en plazos cortos y con ello permitirían descongestionar el sistema judicial.

En el caso de nuestro país, pudiera aplicarse este modelo para determinadas controversias, como son los juicios ejecutivos mercantiles de cuantía menor; divorcios unilaterales o bilaterales, jurisdicciones voluntarias o procedimientos no contenciosos en materias civil y familiar; procedimientos paraprocesales en materia laboral; determinación de medidas cautelares en materia penal; o en cualquier materia cuando la solución verse sobre figuras exploradas como la cosa juzgada, litispendencia o prescripción. Sin duda es posible diseñar algoritmos inteligentes que puedan sustanciar y, en su caso, resolver estos asuntos. Sin embargo, desde este momento es preciso aclarar que todas las actividades realizadas con sistemas de IA deben estar bajo la supervisión de personal humano.

2. Admisión de demandas y recursos

Tanto el trámite de las demandas como de los recursos suelen tener un alto grado de automatización en la práctica jurisdiccional.

En el caso de la demanda, el escrito relativo puede incorporarse a un sistema o plataforma virtual, incluidos los documentos que en su caso deban anexarse. Una vez hecho, el sistema pudiera revisar si la demanda cumple con todos los requisitos legales y dictar el acuerdo de trámite que corresponda: admisión, prevención o desechamiento. Lo mismo puede acontecer con la contestación de demanda, que puede recibirse por medio de la propia plataforma. En ambos casos, los emplazamientos o notificaciones pudieran realizarse electrónicamente.

El trámite de recursos suele compartir un tratamiento similar que en el caso de la demanda, aunque con menores requisitos. Una serie de algoritmos bien podrían identificar la parte recurrente, el acuerdo o resolución que se impugna, así como cuestiones de temporalidad y si, conforme a la legislación correspondiente el recurso es procedente contra la decisión impugnada. Verificados estos supuestos, el propio sistema estaría en aptitud de disponer la admisión o el desechamiento del recurso, así como el trámite para ponerlo en estado de resolución, incluidas las notificaciones que procedan. Incluso, si el medio de impugnación es de resolución sencilla, el sistema bien podría generar una propuesta de solución.

3. Admisión y valoración de pruebas

Ciertamente, la admisión y valoración de pruebas suelen tener un determinado grado de automatismo, por ende, es posible que un sistema de IA pueda auxiliar en estas labores. En la fase de admisión, la IA puede verificar si las pruebas cumplen con los requisitos establecidos en la ley procesal y, de ser así, disponer su admisión y preparación o, en su caso, requerir a la parte promovente para que ajuste el ofrecimiento a las exigencias legales. La admisión es, pues, una fase previsible y, por ende, el sistema no tendría mayor inconveniente en verificar que cualquier prueba ofrecida —documental, testimonial, declaración de partes, confesional, pericial, etcétera— se ajuste a las exigencias de la norma procesal. Quizá el problema se encuentre a la hora de ponderar si la prueba es idónea, pertinente y conducente —en los casos en que la legislación lo exija como requisito previo a la admisión—, pues estas características, por lo general, comportan una serie de razonamientos y justificación por parte del operador judicial.

En la valoración probatoria la IA puede desempeñar un papel significativo en los procesos judiciales. En el caso de la documental no existe mayor problema, porque un sistema puede detectar el tipo de documento, identificar la información relevante en segundos y resumirla en caso de ser necesario, con lo cual se puede prescindir de lecturas innecesarias. También puede detectar erro-

res e inconsistencias en el contenido del documento. Desde luego que, la revisión final de la documental debe quedar a cargo de las y los funcionarios judiciales, más aún cuando involucre aspectos de interpretación.

Tratándose de la declaración de personas o prueba testimonial, un algoritmo puede evaluar la declaración de un testigo y su coherencia, es decir, que esté libre de contradicciones; la capacidad de la persona interrogada de recuperar la descripción del escenario donde sucedieron los hechos; la existencia o no de corroboraciones en otros medios de prueba de lo declarado por la parte deponente (Nieva Fenoll 2018: 84). Incluso, es posible diseñar un algoritmo que involucre pautas de psicología del testimonio para evaluar la veracidad del testimonio. A esto se suma que la IA puede auxiliar en la admisión y calificación de preguntas que formulen las partes tanto de manera escrita como oral, con la finalidad de que únicamente se tomen en cuenta aquellas interrogantes que cumplan con la fórmula establecida en la legislación procesal.

En el supuesto de la prueba pericial, un sistema de IA puede colaborar en la corroboración tanto de la metodología como de las conclusiones de la ciencia, arte, técnica u oficio de la prueba. Es decir, puede emplearse un programa para validar no solo que el procedimiento seguido por la persona experta sea el indicado —principalmente tratándose de ciencias—, sino también que los resultados de las pruebas se encuentren validados por la comunidad científica, o que sean acordes con las reglas del arte, técnica u oficio respectivo, además de que el propio sistema pudiera establecer el grado de fiabilidad de la prueba. Sin duda esto apoyaría a las personas juzgadoras en la valoración de las periciales, pues el famoso apotegma de que *el juez es perito de peritos* es una utopía.

En otro orden de ideas, es posible que la IA pueda asumir labores de valoración conjunta de las pruebas y emplear inferencias probatorias —en función de algoritmos diseñados para tal efecto—, para finalmente plantear una propuesta de valoración probatoria. En estos casos, habría que poner atención en que la IA se ajuste

al sistema de valoración de pruebas establecido en la legislación procesal. En ese contexto, una herramienta de IA que esté programada para valorar pruebas puede ser compatible con un sistema legal o tasado, pues conforme a este esquema la ley establece el peso probatorio que debe tener cada prueba. La cuestión se complica, en cambio, cuando se tratan de sistemas de íntima convicción o sana crítica, pues éstos permiten cierto margen de arbitrio judicial en el proceso de valoración. En el caso del sistema de sana crítica es posible que un sistema de IA sea compatible con el empleo de las reglas de la lógica y los conocimientos científicos, más puede resultar complicado que se ajuste a las máximas de la experiencia, por tratarse de juicios hipotéticos que derivan de la experiencia o el sentido común de la vida en sociedad.

4. Audiencias y comparecencias

Con las facilidades que proporciona la IA se piensa en la posibilidad de privilegiar las comparecencias telemáticas, pues en la *praxis* se ha demostrado que no es del todo necesaria la presencia física de las partes en los órganos jurisdiccionales, debido a que pueden comparecer o rendir declaraciones por medio de videoconferencia, desde cualquier parte en que se encuentren. El principal reto radicaría, sin embargo, en compatibilizar estos mecanismos digitales con los principios procesales establecidos en algunas materias, como sucede, por ejemplo, con el principio de intermediación, que exige la presencia física de la persona juzgadora y las partes procesales en la sala de audiencias.

Incluso, si se adoptara la modalidad virtual para audiencias y comparecencias, pudiera prescindirse de la necesidad de contar con órganos jurisdiccionales en todo el territorio del país, pues estos pudieran concentrarse en unas pocas sedes. Esto además pudiera tener impacto en la “abolición de las normas de competencia territorial, en beneficio de normas de reparto distributivas del trabajo jurisdiccional por criterios objetivos” (Nieva Fenoll 2018: 34).

5. Sentencias

Posiblemente la implementación de IA en la elaboración de sentencias sea uno de los temas más polémicos, porque irremediablemente sugiere pensar en el escenario en que un *juez robot* reemplaza a la persona juzgadora y dicta la sentencia que define la situación legal de las partes en la controversia.

En ese sentido, se ha debatido mucho sobre el juez artificial, la posibilidad de que sea un programa, y no una persona humana, quien dicte sentencia. Lo cual no es una apreciación del todo objetiva¹⁰, pues este tipo de programas sobre el *juez autómatas*, no se refieren, especialmente, a que sea una máquina la que remplace la labor de la persona juzgadora, sino que se refiere a que existan programas o sistemas que ayuden o que hagan más eficiente la toma de decisiones judiciales (Suárez Manrique y De León Vargas 2019: 73).

En ese contexto, pues, la IA puede colaborar en múltiples tareas relacionadas con el dictado de sentencias. Es posible que a partir de la revisión integral del expediente judicial el sistema precise las pretensiones de las partes y las pruebas conducentes para acreditarlas; fije los antecedentes que dieron origen a la controversia; genere esquemas de valoración probatoria que pueden concluir con una posible solución del caso, es decir, cuál de las posturas en conflicto se encuentra mayormente acreditada en función de las pruebas desahogadas en un proceso; colaborar en la identificación de la ley, jurisprudencia o precedentes —nacionales o internacionales— aplicables para resolver el caso sometido a la potestad

¹⁰ Para que los algoritmos impregnaran por completo el entorno socio jurídico, las tecnologías de IA tendrían que llegar al punto en el que la persona juzgadora, o al menos algún tipo de sistema judicial sustituto, pudiera producirse de forma semi o totalmente autónoma utilizando análisis de *big data*, y que éste pudiera actuar de forma independiente. Para esto, sería necesario instrumentalizar por completo todos los aspectos sociales del Derecho para que pudieran calcularse y predecirse tecnológicamente (Morison y Harkens 2019: 24-26). Lo cual sería una labor casi imposible.

del órgano judicial; presentar una propuesta de motivación de la sentencia de acuerdo al sentido del fallo, esto es, formular argumentos que sustenten el criterio que será asumido en la sentencia; inclusive, el sistema de IA puede presentar una o varias propuestas de solución a un caso concreto a partir del uso de patrones de predicción y estadística, con la acotación de que la decisión final de un asunto siempre debe asumirse por una persona juzgadora, debido a los intereses que se encuentran en conflicto y a los posibles márgenes de error que puede tener el sistema de IA.

A partir de este brevísimo esbozo sobre las implicaciones prácticas de la IA en la judicatura, podemos concluir preliminarmente que las aportaciones de la IA pueden contribuir a optimizar distintos sectores de la justicia y al mismo tiempo combatir otros problemas presentes en esta área, principalmente el rezago jurisdiccional que se ha vuelto una nota distintiva de las judicaturas a nivel mundial. De modo que las ventajas de la IA permiten asegurar una justicia eficiente y pronta.

No obstante, debe tenerse presente que,

“cualquiera que sea la tarea que realicen los sistemas de IA en el sistema de impartición de justicia y específicamente en el proceso judicial, su función es de auxilio y de complemento a las tareas de las y los operadores judiciales, de modo que, en todos los casos, los resultados arrojados por aquéllos deben ser validados y revisados por humanos” (Campuzano Gallegos 2019: 167).

Por ello, la discusión acerca de si seguirán existiendo personas juzgadoras en el futuro no es una cuestión que deba preocuparnos por el momento, más bien la interrogante debe girar en torno a si las y los jueces sobrevivirán a la tecnología y, en ese sentido, si estarán en posibilidad de incorporar a sus funciones ordinarias las herramientas de IA que les auxilien en el trámite y resolución de las controversias sometidas a su potestad. Con todo, no se vislumbra que la IA persiga el objetivo de sustituir a las personas juzgadoras o si este fuera el objeto está muy lejos de consolidarse.

De hecho, “los enfoques de la IA aún no pueden, y probablemente nunca puedan, desarrollar la complejidad necesaria para reproducir la actividad esencialmente social de impartir justicia” (Morison y Harkens 2019: 2). En realidad, se pretende que existan programas y máquinas que ayuden a que la toma de decisiones judiciales sea más eficiente y expedita.

VI. NUEVOS DESAFÍOS PARA LOS DERECHOS HUMANOS CON LA IMPLEMENTACIÓN DE IA EN LA FUNCIÓN JURISDICCIONAL

Abordada la cuestión de las principales implicaciones y tendencias de la IA en el proceso judicial, corresponde ahora centrarnos en el impacto que pueden tener estas nuevas tecnologías en los derechos humanos, lo cual no es un tema baladí, sino de la mayor importancia, pues si una de las funciones del Poder Judicial es tutelar derechos, entonces sería un contrasentido que los sistemas de IA —que se implementen para facilitar las labores judiciales— vulneren prerrogativas fundamentales.

En nuestro concepto, es inevitable que la IA comporte afectaciones a derechos humanos, por la sencilla razón de que diversos avances tecnológicos aún no se encuentran regulados. Por ello, es importante conocer cuáles serían las implicaciones más importantes en los derechos, con la finalidad de centrar el debate en las consecuencias tanto positivas como negativas que puedan relacionarse con los derechos fundamentales por el uso de IA en la judicatura y, en todo caso, analizar la posibilidad de conciliar las ventajas que hemos destacado con los desafíos que pueden generarse en las prerrogativas fundamentales de las personas.

1. Juez natural

Este derecho puede afectarse seriamente si la judicatura privilegiara la automatización de decisiones judiciales o, en el peor

de los casos, si un sistema de IA asumiera totalmente las funciones de juzgamiento, porque de esta manera no se garantizaría que las personas sean juzgadas por tribunales ordinarios, a quienes la ley les hubiera encomendado esa tarea y conforme a los procedimientos legalmente establecidos. Mientras la función jurisdiccional esté reservada a jueces, siempre necesitaremos contar con la participación decisiva de la persona juzgadora, quien, en todo caso, debe ser el ser humano al frente del sistema experto de IA (Simón Castellano 2021a: 290).

Este posible desafío pudiera prevenirse si se adopta la premisa de que IA debe operar en términos de colaboración y no sustitución de la persona juzgadora. No es necesario —e incluso es poco probable— que la justicia sea totalmente automatizada, es preferible adoptar un esquema de semiautomatización, como un método para contrarrestar los problemas asociados con el análisis algorítmico (Morison y Harkens 2019: 28)

2. Derecho a la igualdad y no discriminación

La afectación a estos derechos puede advertirse desde dos enfoques. Primero, existiría vulneración al derecho a la igualdad en sentido amplio cuando la implementación de sistemas de IA favoreciera únicamente a las personas que tienen acceso al uso de tecnologías o que cuentan con recursos económicos, además de que esto propiciaría invariablemente la violación a otros derechos como el acceso a la justicia y el debido proceso. Sin duda, el uso de IA puede generar una brecha digital o un analfabetismo digital cuando sólo un sector de la población pueda utilizarla en su vida cotidiana, mientras que otros grupos no tengan acceso a la IA o, aunque lo tuvieran, no pudieran utilizarla (Icaza-Álvarez *et al* 2019: 402). A esto se suma que los procesos de digitalización pudieran afectar en mayor proporción a las personas en situación de pobreza por cuanto a la dificultad de acceder a su utilización en aquellas controversias en que sean parte. Desde esta perspectiva, la IA en la judicatura representa un enorme desafío para la satisfacción del derecho a la igualdad.

El segundo enfoque se refiere a la discriminación o mejor dicho a la *discriminación algorítmica*. Es innegable que los sistemas de IA son programados por seres humanos, por lo cual en esa labor de programación se pueden heredar determinados sesgos o patrones de discriminación en los algoritmos. De modo que, si estos sistemas asumen decisiones o realizan tareas basadas en patrones discriminatorios, pueden generar desigualdad y, con ello, afectar a determinados grupos sociales o a personas con determinadas características.

Pese a los desafíos a estos derechos, consideramos que las ventajas que ofrece la implementación de IA en las labores judiciales son mayores, pues representan una optimización y una celeridad en la justicia. No obstante, se debe buscar un esquema de implementación en el que se garantice a todas las personas, en igualdad de condiciones y sin discriminación alguna, el acceso y uso a la justicia tecnológica.

3. *Imparcialidad judicial*

Se trata de un derecho destinado a garantizar el recto ejercicio de la función jurisdiccional a efecto de que no exista algún interés personal en la toma de decisiones. Tal como hemos anunciado, en las personas juzgadoras concurren una serie de factores externos e internos —sesgos, heurísticos, ideologías, emociones, problemas personales, fatiga, ignorancia, etcétera— que pueden incidir en el criterio que habrán de adoptar para resolver un caso concreto. De ahí que la imparcialidad judicial ha sido una institución bastante cuestionable.

Así, con la adopción de sistemas de IA se concibe la posibilidad de que se garantice eficazmente la imparcialidad en las decisiones judiciales, debido a que la IA sería más *objetiva y neutra* en la resolución de controversias solo en la medida en que se logre despojar al sistema de determinados inconvenientes como los sesgos discriminatorios. Esto obedece a que, en los casos en que sea necesario el auxilio de un sistema inteligente, se limitaría a aplicar la ley a los

casos concretos para los que fue programado, por lo tanto, no habría posibilidad de resolver con base en emociones o sensaciones personales, pues se piensa que los sistemas o máquinas no pueden desarrollar estos estados subjetivos de las personas. Desde este enfoque, pareciera que más que un desafío, hablaríamos de algo positivo, empero debe analizarse la cuestión con suma cautela, porque resulta bastante riesgoso dejar en manos de una máquina la decisión de un asunto, debido a que se encuentran en juego valores importantes de las personas.

4. Legalidad y seguridad jurídica

El uso de sistemas inteligentes en las decisiones judiciales implicaría una mayor eficacia de estos derechos y, con ello, es plausible que aumente la confianza de la ciudadanía en la judicatura, debido a que “un caso similar a otro se decidirá en sentido análogo o teniendo en cuenta los criterios tomados por el primero” (Cárdenas Rincón y Molano Martínez 2021: 7). La lógica de un sistema de IA en estos términos correspondería con la aplicación de la ley al caso concreto, de manera similar a un ejercicio de subsunción normativa, garantizándose además que existe uniformidad de criterios. De forma tal que con esta dinámica las partes pudieran tener mayor certeza jurídica sobre la forma en que resolverán los tribunales.

5. Justicia pronta

Este es otro de los derechos que lejos de afectarse, en realidad se maximizaría con el uso de sistemas inteligentes en el ámbito de la justicia. Hemos insistido en que uno de los problemas más frecuentes en la judicatura es que la justicia es lenta y esto obedece, en gran medida, a la enorme cantidad de asuntos y a los escasos recursos materiales y humanos que se destinan para abatir esas cargas, con lo que se da paso al rezago jurisdiccional. Estamos seguros de que los sistemas de IA contribuirán a lograr el descongestionamiento jurisdiccional, por un lado y, por otro, lograrán que la impartición de justicia sea más ágil, de forma tal que las controversias serán resueltas en un tiempo breve. Es precisamente en este rubro

donde existe una mayor articulación entre las ventajas que supone el uso de IA en la labor jurisdiccional y los derechos humanos de las personas usuarias de la justicia.

6. *Derecho de defensa*

Es posible que, con la adopción de sistemas de IA, este derecho se vea afectado por dos razones fundamentales: la primera, que el desconocimiento del funcionamiento de los algoritmos, derivado de su falta de transparencia, obstaculizaría que las y los abogados pudieran realizar los actos de defensa necesarios en favor de los intereses de sus clientes. La segunda razón obedece a que, con la incorporación paulatina de herramientas de IA en la impartición de justicia, es probable que las partes puedan aportar la información que requiera el tribunal *virtual* y, por ende, es posible que prescindan de los servicios jurídicos que prestan las y los abogados, pues los actos de defensa serían limitados y en algunos casos innecesarios.

Estimamos que estos desafíos al derecho de defensa pueden solucionarse a través de la adopción del principio de máxima transparencia en el acceso a la configuración de los algoritmos y al funcionamiento mismo del sistema de IA. Por otro lado, a pesar de que la adopción de sistemas tecnológicos supone automatización en los servicios legales, las personas deben contar, en todos los casos, con la asistencia de un profesional del derecho que colabore no sólo en la supervisión y los trámites del proceso sino también en el diseño de la estrategia procesal. Los sistemas tecnológicos no son la panacea a los males de la justicia, son un mecanismo que, en todo caso, optimizará la eficacia de los procesos judiciales. Por este motivo, tal como sucede con las personas juzgadoras, las y los abogados no debieran ser sustituidos por sistemas o robots.

7. *Presunción de inocencia*

Con el uso cada vez más frecuente de sistemas de IA en el campo de la justicia, el derecho a la presunción de inocencia puede

ser de los más afectados. En principio, porque si en la valoración probatoria el algoritmo comete algún error, esto puede incidir inevitablemente en una condena injusta en materia penal. Lo mismo sucedería si el sistema de IA no garantiza la satisfacción del estándar probatorio necesario para emitir un fallo condenatorio.

Por ello, para contrarrestar estos desafíos a la presunción de inocencia, sería fundamental que las máquinas no asumieran la toma de decisiones de forma automatizada y total; pero si por alguna razón se regulara la posibilidad de resolución para determinadas controversias, entonces tales resoluciones siempre deben ser validadas por personas juzgadoras, pues sólo de esta forma se evitará que las sentencias sean arbitrarias, injustas o violatorias de derechos.

8. Derecho a la privacidad y protección de datos

El derecho a la privacidad es aquél que “todo individuo tiene a separar aspectos de su vida privada del escrutinio público” (García Ricci 2009: 189). Se trata de una esfera de la vida que está fuera de la intromisión del Estado y, por ende, protegida del escrutinio de los otros, porque sólo atiende a la persona en lo individual (Pescharde 2017: 362). De cara a la implementación de sistemas de IA en la justicia, el derecho a la privacidad puede ser afectado con motivo de la recopilación de los datos personales y sensibles de las personas que formen parte de las controversias jurisdiccionales. Por esta razón, se impone a la judicatura el deber de garantizar que las herramientas de IA protegerán la información confidencial de las personas para evitar que se haga un mal uso de ella. Este es el principal desafío que deberá enfrentarse. Para atenderlo es suficiente que la ley tutele estos derechos frente al uso de la tecnología y que el propio sistema de IA avale la privacidad y protección de datos. Esta serie de ideas nos permite reflexionar acerca de los enormes desafíos que la IA impone a los derechos humanos cuando interactúa con la función jurisdiccional. Sea en términos de afectación o maximización, los derechos son el centro de atención, constituyen el límite de la actuación del Estado y, en estos

tiempos, también sirven como medio de contención para proteger a las personas frente el uso de las nuevas tecnologías.

Con todo, frente a las ventajas de la implementación de la IA en la judicatura, que conllevan a la optimización de las labores jurisdiccionales y al ahorro de tiempo y recursos para ofrecer una justicia eficiente y pronta, lo cierto es que, a partir del breve análisis que hemos planteado es posible establecer que los desafíos a los derechos humanos pueden ser controlados y mitigados. Desde esta perspectiva, es indispensable que, antes de incorporar la IA a las labores de la judicatura, se cuente con una legislación que contemple un esquema de protección de derechos fundamentales, pues de nada serviría contar con una justicia moderna y tecnológica si no se garantiza la progresividad de los derechos.

9. Derecho a recurrir el fallo

En nuestro concepto, debe existir la posibilidad de que cualquier decisión judicial en la que intervenga algún sistema de IA sea revisada por un tribunal superior. En tal virtud, si la judicatura incorpora IA en los procesos judiciales, entonces debe, a su vez, garantizar que tanto las actuaciones procesales como las resoluciones finales puedan ser revisadas por una instancia superior, para corregir posibles errores o decisiones ilegales e injustas.

Esta serie de ideas nos permite reflexionar acerca de los enormes desafíos que la IA impone a los Derechos Humanos cuando interactúa con la función jurisdiccional. Sea en términos de afectación o maximización, los derechos son el centro de atención, constituyen el límite de la actuación del Estado y, en estos tiempos, también sirven como medio de contención para proteger a las personas frente el uso de las nuevas tecnologías. Desde esta perspectiva, es indispensable que, antes de incorporar la IA a las labores de la judicatura, se cuente con una legislación que contemple un esquema de protección de derechos fundamentales, pues de nada serviría contar con una justicia moderna y tecnológica si no se garantiza la progresividad de los derechos.

VII. CONCLUSIONES

La IA es una innovación tecnológica cuyo objetivo es imitar distintas actividades humanas a través del procesamiento de grandes cantidades de información y por medio de la operación de algoritmos inteligentes. El constante y potencial desarrollo de la IA va redefiniendo la vida en sociedad, por lo cual es necesario advertir sus beneficios, pero también sus consecuencias negativas.

En la actualidad la IA ha tenido una aplicación muy importante en el campo del derecho, muestra de ello son los distintos sistemas expertos jurídicos que existen en el mundo y que se utilizan fundamentalmente para labores de despachos jurídicos, investigación de delitos y la impartición de justicia. También existen otros programas o sistemas que pueden auxiliar en distintas tareas legales.

La IA puede aportar diversas ventajas cuando se traslada al campo de la impartición de justicia, en particular puede contribuir a la optimización de las labores jurisdiccionales, así como al ahorro de tiempo y recursos, lo cual permitiría asegurar una justicia moderna, eficiente y pronta. Sin embargo, todas las tareas que realicen los sistemas de IA en la impartición de justicia deben ser supervisadas y validadas por seres humanos.

Al lado de las implicaciones prácticas de la IA en la función jurisdiccional, no podemos desconocer que los sistemas inteligentes pueden incidir en los derechos humanos. No obstante, hemos planteado que es posible conciliar las ventajas de la IA con los desafíos que pueden generarse en los derechos, por lo que no advertimos que sean mayores los riesgos que las bondades del uso de la IA en la impartición de justicia.

BIBLIOGRAFÍA

- Badaró, Sebastián *et al.* (2013): “Sistemas expertos: fundamentos, metodologías y aplicaciones”, en *Ciencia y Tecnología*, Vol. 13, 349-364.
- Cáceres Nieto, Enrique (2023): “La Inteligencia Artificial aplicada al derecho como una nueva rama de la teoría jurídica”, en *Anales de la Cátedra Francisco Suárez*, Vol. 57, 63-89.
- Cámara de Diputados (2023): “Impulsan iniciativa para expedir la Ley de Regulación Ética de la Inteligencia Artificial y la Robótica”, en *Comunicación Social de la Cámara de Diputados*. Disponible en: «<https://comunicacionsocial.diputados.gob.mx/index.php/boletines/impulsan-iniciativa-para-expedir-la-ley-de-regulacion-tica-de-la-inteligencia-artificial-y-la-robotica#:~:text=El%20diputado%20Ignacio%20Loyola%20Vera,y%20financieros%2C%20con%20apego%20a>» [Consultado el 2 de junio de 2024].
- Campuzano Gallegos, Adriana (2019): “*Inteligencia artificial para abogados. Ya es tiempo...*”, Thomson Reuters, México.
- Cárdenas Rincón, Erick y Molano Martínez, Valeria (2021): “Un estudio sobre la posibilidad de aplicar la inteligencia artificial en las decisiones judiciales”, en *Revista DIREITO GV*, Vol. 17, Núm. 1.
- CLAD (2023): “Carta Iberoamericana de inteligencia artificial en la Administración Pública” Disponible en: «<https://clad.org/wp-content/uploads/2023/08/Borrador-CI-IA-V2-ES-08-2023.pdf>» [Consultado el 01 de junio de 2024].
- CMAAAL (2023): “Declaración de Santiago”, en *Minciencia*. Disponible en: «https://media.datacenterdynamics.com/media/documents/Declaracion_de_Santiago_Inteligencia_Artificial.pdf» [Consultado el 01 de junio de 2024].

- Danziger, S. *et al.* (2011): “Extraneous factors in judicial decisions”, en *PNAS*. Disponible en: «<https://doi.org/10.1073/pnas.1018033108>» [Consultado el 10 de febrero de 2024].
- Equipo de Expertos en Ciencia y Tecnología (2024): “Inteligencia artificial, ventajas y desventajas”, en *Universidad Internacional de Valencia*. Disponible en: «<https://www.universidadviu.com/es/actualidad/nuestros-expertos/inteligencia-artificial-ventajas-y-desventajas>» [Consultado el 01 de junio de 2024].
- Galindo Ayuda, Fernando (2019): “¿Inteligencia Artificial y Derecho? Sí, pero ¿cómo?”, en *Revista Democracia Digital e Gobierno Electrónico, Florianópolis*, Vol. 2, Núm. 18, 36-57.
- García Ricci, Diego (2009): “El derecho a la privacidad en las redes sociales en internet”, en *Derechos Humanos México. Revista del Centro Nacional de Derechos Humanos*, Núm. 12, 183-200.
- IBM (2024): “¿Qué es Machine Learning?”, en *IBM*. Disponible en: «<https://www.ibm.com/es-es/topics/machine-learning>» [Consultado el 1 de marzo de 2024].
- Icaza-Álvarez, Daniel Orlando *et al* (2019): “El analfabetismo tecnológico o digital”, en *Polo del conocimiento*, Vol.4, Núm. 2, 396-406.
- Independent (2014): “Stephen Hawking: ‘Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?’”, en *Independent*. Disponible en: «<https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html>» [Consultado el 1 de junio de 2024].
-

- Instituto TSJCDMX (2018): “El Cerebro de la Toga: Procesos Cognitivos y Toma de Decisiones Judiciales”, en *You Tube*. Disponible en: «<https://www.youtube.com/watch?v=ykjj-tVvmD5I&t=598s>» [Consultado el 15 de enero de 2024].
- Jurimetría (2024): “¿Cómo funciona?”, en *Jurimetría*. Disponible en: «<https://jurimetria.laleynext.es/content/ComoFunciona.aspx>» [Consultado el 02 de marzo de 2024].
- Martínez Bahena, Goretty Carolina (2012): “La inteligencia artificial y su aplicación al campo del Derecho”, en *Alegatos*, Vol.82, 827-846.
- Monreal, Ricardo (2024): “Ley Federal que Regula la Inteligencia Artificial. Exposición de motivos para publicación”, en *Ricardo Monreal*. Disponible en: «<https://ricardomonrealavila.com/ley-federal-que-regula-la-inteligencia-artificial-exposicion-de-motivos-para-publicacion/>» [Consultado el 1 de junio de 2024].
- Morison, John y Harkens, Adam (2019): “Re-Engineering Justice? Robot Judges, Computerised Courts and (Semi) Automated Legal Decision-Making”, en *Legal Studies*, Disponible en: «<https://ssrn.com/abstract=3369530>» [Consultado el 03 de junio de 2024].
- Nieva Fenoll, Jordi (2018): “*Inteligencia artificial y proceso judicial*”, Marcial Pons, Madrid.
- Nieva Fenoll, Jordi (2013): “*Proceso judicial y Neurociencia: una revisión conceptual del derecho procesal*”, Taruffo, Michele y Nieva Fenoll, Jordi (dirs.), Marcial Pons, Madrid.
- Peschard, Jacqueline (2017): “Cien años del derecho a la privacidad en la Constitución”, en *Cien ensayos para el centenario. Constitución Política de los Estados Unidos Mexicanos*, Esquivel, Gerardo *et al.* (coords.), Universidad Nacional Au-

tónoma de México, Instituto de Investigaciones Jurídicas, Instituto Belisario Domínguez, 361-378.

RAE (2024): “Algoritmo”, en *RAE*. Disponible en: «<https://www.rae.es/diccionario-estudiante/algoritmo>» [Consultado el 3 de junio de 2024].

Ríos, Sebastián (2023): “Los riesgos de la Inteligencia artificial”, en *Revista Mensaje*. Disponible en: «<https://www.dii.uchile.cl/wp-content/uploads/2023/05/02-REVISTA-MENSAGE-Los-riesgos-de-la-Inteligencia-Artificial.pdf>» [Consultado el 2 de junio de 2024].

Saavedra, Víctor y Upegui, Juan Carlos (2021): “*PretorIA y la automatización del procesamiento de causas de Derechos Humanos*”, *Derechos Digitales-IDRC-CRDI*, Colombia.

San Miguel Caso, Cristina (2021): “La aplicación de la Inteligencia Artificial en el proceso: ¿un nuevo reto para las garantías procesales?”, en *IUS ET SCIENTIA*, Vol.7, Núm.1, 286-203.

Schwab, Klaus (2016): “*La cuarta revolución industrial*”, Penguin Random House Grupo Editorial, México.

Segura, Romina (2023): “Inteligencia artificial y administración de justicia: desafíos derivados del contexto latinoamericano”, en *Revista de Bioética y Derecho*, Vol. 58, 45-72.

Senado de la República (2023): “Presentan iniciativa que faculta al Congreso para legislar en materia de Inteligencia Artificial”, en *Comunicación Social del Senado de la República*. Disponible en: «<https://comunicacionsocial.senado.gob.mx/informacion/comunicados/6661-presentan-iniciativa-que-faculta-al-congreso-para-legislar-en-materia-de-inteligencia-artificial/>» [Consultado el 1 de junio de 2024].

Simón Castellano, Pere (2021a): “Inteligencia artificial y valoración de la prueba: las garantías jurídico-constitucionales del órgano de control”, en *THEMIS-Revista de Derecho*, Núm.79, 283-297.

Simón Castellano, Pere (2021b): “Inteligencia artificial y Administración de Justicia: Quo vadis, justitia?”, en *IDP. Revista de Internet, Derecho y Política*, Núm. 33, 1-15.

Suárez Manrique, Wilson Yesid y De León Vargas, Georgina Isabel (2019): “Inteligencia artificial y su aplicación en la administración de justicia”, en *Revista Jurídica Mario Alario D’Filippo*, Vol. 11, Núm. 21, 71-83.

TEPJF (2013): “Sesión Pública – miércoles 29 marzo 2023 - TEPJF”, en *YouTube*. Disponible en: «<https://www.youtube.com/live/OwaZg3quyls?t=3679s>» [Consultado el 3 de junio de 2024].

The Technolawgist (2019): “Estonia se prepara para tener “jueces robot” basados en Inteligencia Artificial”, en *The Technolawgist*. Disponible en: «<https://www.thetechnolawgist.com/2019/06/12/estonia-se-prepara-para-tener-jueces-robot-basados-en-inteligencia-artificial/>» [Consultado el 01 de marzo de 2024].

UNESCO (2022): “Recomendación sobre la ética de la inteligencia artificial”, en *UNESCO*. Disponible en: «<https://www.unesco.org/es/articulos/recomendacion-sobre-la-etica-de-la-inteligencia-artificial>» [Consultado el 1 de junio de 2024].

UNESCO (2018): “Inteligencia artificial promesas y amenazas”, en *El Correo de la UNESCO*. Disponible en: «https://unesdoc.unesco.org/ark:/48223/pf0000265211_spa» [Consultado el 1 de junio de 2024].

Universidad de Medellín (2024): “Juez inteligente”, en *Universidad de Medellín*. Disponible en: «<https://cide.udemedellin.edu.co/tecnologias/juez-inteligente-sistema-experto-que-asis-te-al-juez-en-la-valoracion-de-la-prueba-judicial/>» [Consultado el 10 de febrero de 2024].

Universidad Veracruzana (2024): “Conocimientos generales: ¿Sabes cuántos datos se generan en un minuto?”, en *Universidad Veracruzana*. Disponible en «https://www.uv.mx/infosegura/general/conocimientos_datos/» [Consultado el 14 de febrero de 2024].

Trazando el camino hacia una regulación efectiva: Breve análisis del Reglamento Europeo de Inteligencia Artificial de 2024

Charting the path to effective regulation: Brief analysis on the 2024 European Artificial Intelligence Act

JOSÉ RAFAEL GRIJALVA ETERNOD

Academia Interamericana de Derechos Humanos

Universidad Autónoma de Coahuila

ORCID: 0000-0001-6496-7016

Fecha de recepción: 02 mayo 2024

Fecha de aceptación: 10 junio 2024

SUMARIO: I. Introducción. II. Antecedentes. III. Estructura y contenido del AIA. 1. Considerandos. 2. Disposiciones Generales. 3. Prácticas de Inteligencia Artificial. 4. Sistemas de IA de alto riesgo. 5. Obligaciones de transparencia para proveedores e implantadores de determinados sistemas de IA y modelos GPAI. 6. Modelos de IA de propósito general. 7. Medidas de apoyo a la innovación. 8. Gobernanza. 9. Base de datos de la UE para sistemas de IA de alto riesgo. 10. Control posterior a la comercialización, intercambio de información y vigilancia del mercado. 11. Código de conducta y directrices. 12. Delegación de poderes y funcionamiento de los comités. 13. Confidencialidad y sanciones. 14. Disposiciones finales. IV. Conclusiones.

RESUMEN: El presente artículo ofrece un análisis exploratorio y descriptivo del Acta Europea de Inteligencia Artificial de 2024. Dada la novedad de este marco regulatorio el estudio se centra en proporcionar una visión general sobre la estructura y alcance del Acta. Para ello, se examinan las disposiciones clave y sus posibles impactos, intentando esbozar un marco comprensible para cualquier persona interesada en adentrarse en la materia. Además, se busca identificar áreas donde la regulación puede influir significativamente en el desarrollo y la implementación de tecnologías de inteligencia artificial, con un enfoque particular en la

protección de derechos y la responsabilidad de los actores involucrados. Finalmente, el trabajo refleja la perspectiva del autor sobre el contenido del Acta, buscando ilustrar cómo esta nueva regulación puede moldear el futuro de la inteligencia artificial tanto en la Unión Europea como en otras latitudes.

ABSTRACT: This article offers an exploratory and descriptive analysis of the 2024 European Artificial Intelligence Act. Given the recent nature of this regulatory framework, the study focuses on providing an overview of the structure and scope of the Act. To this end, the relevant regulations and its possible impacts are examined, attempting to outline a comprehensible framework for anyone interested in delving into this subject. Additionally, the study aims to identify areas where regulation may significantly influence the development and implementation of artificial intelligence technologies, with a particular focus on the rights protection and responsibility of the involved actors. Finally, the work reflects the author's perspective on the content of the Act, seeking to illustrate how this new regulation might shape the future of artificial intelligence both within the European Union and beyond.

PALABRAS CLAVE: *Inteligencia Artificial, Unión Europea, desarrollo, derechos humanos, tecnologías emergentes.*

KEYWORDS: *Artificial Intelligence, European Union, development, human rights, emerging technologies.*

I. INTRODUCCIÓN

La inteligencia artificial (IA) ha experimentado un crecimiento exponencial en su aplicación en diversos ámbitos, llevando consigo tanto oportunidades como desafíos significativos. La regulación de este fenómeno en la Unión Europea (UE) ha constituido un esfuerzo constante, dinámico y fundamental para abordar jurídicamente los avances tecnológicos y sus implicaciones en múltiples esferas de la vida cotidiana. A medida que la IA se ha instalado en los sistemas económicos, sociales y políticos de esta región, se ha consolidado también la necesidad de concebir un marco regula-

torio que no solo continúe fomentando la innovación y el desarrollo tecnológico, sino que también establezca un potente marco de protección de los derechos fundamentales de los ciudadanos europeos para asegurar un uso ético y responsable de estas tecnologías (García 2022: 305).

Si bien existen otros proyectos relevantes en materia de regulación de IA en el ámbito regional y universal (Drnas de Clément 2022), la UE ha sido pionera en el establecimiento de un enfoque regulatorio para la misma. En este sentido, la Comisión Europea, consciente de los desafíos y oportunidades que la IA presenta, publicó en abril de 2021 una propuesta de regulación de la inteligencia artificial, conocida como el Reglamento Europeo de Inteligencia Artificial (AIA por sus siglas en inglés) la cual, además de ser la primera de su tipo a nivel global, buscaba generar un ambiente de confianza que garantizara la seguridad y la dignidad del ser humano mientras se promovía la adopción y el desarrollo de la IA en toda la región (Gill 2019).

Este enfoque de *hard law*¹ adoptado por la UE para regular la IA destacó, desde sus inicios, la importancia de establecer, mediante regulaciones explícitas, los confines y las responsabilidades asociadas con las tecnologías de inteligencia artificial, donde la participación de gobiernos y organizaciones supranacionales es fundamental. Asimismo, promovió la consolidación de un marco ético dirigido a determinar si las acciones y decisiones realizadas en el desarrollo y aplicación de la inteligencia artificial serían *adecuadas* o *inadecuadas* (Piedra 2023: 6).

De esta manera, la iniciativa de la UE para regular la inteligencia artificial constituyó el inicio de un camino crucial hacia la creación

¹ Además de este enfoque de *hard law*, es posible adoptar una perspectiva más flexible o de *soft law*, centrada en recomendaciones y sugerencias no obligatorias, manteniendo así un diálogo continuo sobre la regulación de la inteligencia artificial. Este método ha sido implementado por entidades como la Organización para la Cooperación y el Desarrollo Económico (OCDE), que en 2019 emitió su Recomendación del Consejo sobre la IA (Morandín-Ahuerma 2023: 96).

de un marco ético y legal que rigiera la evolución de estas tecnologías disruptivas el cual ha culminado con la redacción del AIA. Con ella, la UE busca equilibrar los beneficios de la IA con la protección de sus ciudadanos, estableciendo un importante precedente internacional que, sin duda alguna, influirá en la manera en la que otros países y regiones abordarán la regulación de la IA en el futuro.

Como respuesta a este escenario surge el presente artículo, el cual pretende configurarse como un estudio exploratorio y descriptivo cuyo objetivo inicial es ofrecer una panorámica general de esta nueva legislación. Dada la novedad de la temática² y la carencia de un aparato crítico consolidado que analice el AIA, este trabajo no pretende ofrecer un análisis exhaustivo sino más bien trazar y perfilar las principales disposiciones del Acta y su posible impacto en el escenario tecnológico y social de la UE. En consecuencia, los comentarios al contenido del Acta, breves como son, reflejan únicamente la perspectiva del autor y no se apoyan en un marco teórico contrastado, dado que el desarrollo teórico y crítico sobre esta nueva regulación de la IA en el espacio europeo aún está en sus etapas iniciales.

De igual forma, debe señalarse que la inclusión de cada capítulo y artículo del AIA en este trabajo responde a la intención de ofrecer una estructura clara y accesible que facilite la comprensión y el acceso al contenido del acta, garantizando con esta metodología que todas las personas interesadas en esta regulación europea, sin importar su conocimiento previo del documento, puedan identificar y comprender las disposiciones que la integran sin dificultades.

En resumen, este trabajo se propone como una herramienta de primer contacto para aquellas personas que buscan adentrarse en los fundamentos y especificidades del marco jurídico que la

² El Acta de Inteligencia Artificial de la Unión Europea fue aprobada el pasado 13 de marzo de 2024, es decir, un mes antes de iniciar la redacción del presente trabajo.

Unión Europea ha establecido para regular el dinámico campo de la inteligencia artificial.

II. ANTECEDENTES

Como ya se ha apuntado, el surgimiento y la evolución de la IA han planteado desafíos sin precedentes en el panorama tecnológico global, obligando a los Estados, las empresas y los organismos internacionales a reconsiderar los enfoques tradicionales hacia la gobernanza tecnológica. En este contexto, la UE ha emergido como una líder proactiva en el establecimiento de un marco normativo para la IA. Este esfuerzo regulatorio no solo refleja las preocupaciones sobre la privacidad, la seguridad y la ética, sino que también destaca el compromiso de los países de la región con la protección de los derechos fundamentales en la era digital. Pero ¿cómo se fue gestando este enfoque pionero?

Los antecedentes de la regulación de la IA en la UE pueden rastrearse hasta los primeros debates que se realizaron en el seno de dicha organización sobre las implicaciones éticas, sociales y legales de la tecnología digital y la automatización (Drnas de Clément 2022: 13). A medida que la IA comenzó a integrarse en varios aspectos de la vida económica y social, la Comisión Europea reconoció la necesidad de un diálogo estructurado sobre cómo estas tecnologías deberían ser gestionadas. En este sentido, el establecimiento en 2018 del Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial (AIHLEG por sus siglas en inglés) por parte de la Comisión Europea fue uno de los primeros pasos significativos. Desde su creación este grupo formado por personas expertas de diversos sectores –incluyendo la academia, la industria y la sociedad civil–, asumió la función de brindar asesorías y consultorías sobre la implementación de políticas y estrategias de IA, así como de elaborar recomendaciones éticas para el desarrollo y uso de la IA (Antonov 2022: 42 y ss). En consecuencia, el AIHLEG asumió el mandato de ofrecer consejos sobre cómo la UE podía aprovechar

al máximo los beneficios de la IA, a la vez que abordara los desafíos y riesgos asociados con esta tecnología.

Una de las contribuciones más significativas del AIHLEG fue la elaboración de las *Directrices Éticas para una IA Confiable*³ las cuales se centran en asegurar que el desarrollo y uso de la IA en Europa se realice bajo principios éticos y siguiendo valores fundamentales. Bajo esta idea, la Inteligencia Artificial Confiable parte del principio de que la confianza es fundamental para las sociedades, las economías y el desarrollo sostenible por lo que resulta crucial que individuos, organizaciones y sociedades puedan establecer confianza en el desarrollo, implementación y uso de la IA para explotar completamente su potencial (Thiebes *et al.* 2021: 448 y ss).

Asimismo, este documento señala la necesidad de que el desarrollo e implementación de IA se realice bajo un enfoque centrado en el ser humano. Sobre este rubro, la estrategia europea de IA señala que la confianza es esencial para asegurar una aproximación humanocéntrica a la IA la cual no debe considerarse un fin en sí misma, sino más bien una herramienta al servicio de las personas, pues su propósito primordial siempre debe ser mejorar el bienestar humano. Para alcanzar este objetivo, dice el AIHLEG, es crucial asegurar la fiabilidad de la IA, integrando completamente valores fundamentales aceptados por los países miembros de la UE, y reflejados en la Carta de Derechos Fundamentales de la UE⁴, como el respeto a la dignidad humana, la libertad, la democracia, la igualdad, el Estado de derecho, el pluralismo, la no discriminación, la tolerancia, la justicia, la solidaridad y la igualdad de condiciones⁵. En resumen, las Directrices elaboradas por el AIHLEG resaltan

³ Dichas directrices así como una compilación de comentarios recibidos por medio de una consulta abierta pueden revisarse en «<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.» Último acceso 30 de abril de 2024.

⁴ Ver el *Preámbulo* de la Carta de Derechos Fundamentales de la Unión Europea. Disponible en «<https://www.derechoshumanos.net/normativa/normas/europa/CDFUE/CartaDerechosFundamentalesUnionEuropea-v2007.htm>.» Último acceso 30 de abril de 2024.

⁵ European Commission, *Com (2019)168*, 8 abril 2019.

la necesidad de desarrollar y sostener un enfoque proactivo y considerado hacia el desarrollo y la implementación de la IA, alineado con los principios éticos y el contenido de los derechos humanos.

Junto a este enfoque de confiabilidad humanocéntrica de la IA desarrollado por el AIHLEG, en abril de 2018 la Comisión Europea presentó un documento titulado *Inteligencia Artificial para Europa* el cual destacaba la importancia de fomentar el desarrollo tecnológico garantizando un alto estándar ético. Así, dicha comunicación proponía un enfoque equilibrado que promoviera la innovación y el uso de la IA bajo un marco ético riguroso, respetando los valores y normativas fundamentales de la UE. Los propósitos de la estrategia descrita en este documento incluyen: fomentar el desarrollo tecnológico e industrial de la UE así como la incorporación de la IA en los sectores económicos, tanto público como privado; adaptarse a las transformaciones socioeconómicas que la IA introduce; y, finalmente, establecer un entorno ético y legal apropiado (Gamero Casado 2021: 271 ss).

En diciembre de 2018, la Comisión Europea junto con los Estados miembros divulgaron un *Plan Coordinado sobre la Inteligencia Artificial*⁶ que versaba sobre el avance de la IA en la UE. Este plan fue complementado en enero de 2019 por la *Carta ética europea sobre el uso de la IA en los Sistemas Judiciales*⁷, que estipula cinco principios destinados a orientar la aplicación ética de la IA, especialmente en el ámbito judicial, entre los que se encuentran el respeto por los derechos fundamentales, la no discriminación, la calidad y seguridad de los datos –asegurando que las fuentes sean certificadas e intangibles y que los modelos sean diseñados de manera multidisciplinar y segura tecnológicamente–, la transparencia, la imparcialidad y la equidad, y el control del usuario

⁶ Disponible en «<https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>.» Última revisión 30 de abril de 2024.

⁷ Disponible en «<https://protecciondata.es/wp-content/uploads/2021/12/Carta-Etica-Europea-sobre-el-uso-de-la-Inteligencia-Artificial-en-los-sistemas-judiciales-y-su-entorno.pdf>.» Última revisión 30 de abril de 2024.

–permitiendo que los usuarios estén bien informados y mantengan control sobre sus decisiones–. Durante el mismo año, la Comisión Europea proclamó la *Comunicación General Confianza en la Inteligencia Artificial centrada en el ser humano*⁸ junto con las *Directrices éticas para una Inteligencia Artificial Confiable*⁹. Estos documentos refinaron y extendieron las directrices sobre ética y confiabilidad establecidas previamente (Drnas de Clément 2022: 14).

Un año más tarde, en febrero de 2020, la UE publicó el *Libro Blanco sobre la Inteligencia Artificial*¹⁰ el cual constituyó el avance más significativo en la regulación de la IA hasta ese momento al proponer diversas medidas para la creación de un ecosistema de excelencia y confianza, y para delinear un marco regulador que clasificara las aplicaciones de IA según el nivel de riesgo que presentaban. En dicho documento se resalta la capacidad de las nuevas tecnologías para transformar varios aspectos de la vida cotidiana al tener la posibilidad de mejorar significativamente la asistencia médica –mediante diagnósticos más precisos y una prevención más efectiva de enfermedades–, la eficiencia en la agricultura –optimizando el uso del agua, detectando plagas o previendo el comportamiento de consumidores–, los esfuerzos de mitigación y adaptación al cambio climático –generando patrones climáticos de largo plazo, optimizando el uso de energía, acelerando el descubrimiento de materiales más sostenibles–, aumentando la seguridad de las y los ciudadanos –mejorando rutas para vehículos de emergencia, detectando amenazas cibernéticas, analizando datos históricos y tendencias para predecir dónde y cuándo es más probable que ocurran delitos–, entre otros aspectos. No obstante, también se reconoce que la IA presenta riesgos potenciales, incluyendo decisiones opacas –debido a la complejidad y la falta de transparencia

⁸ Disponible en «<https://eur-lex.europa.eu/legal-content/es/ALL/?uri=CELEX%3A52019DC0168>.» Última revisión 30 de abril de 2024.

⁹ Disponible en «<https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.» Última revisión 30 de abril de 2024.

¹⁰ Disponible en «https://commission.europa.eu/document/d2ec4039-c5be-423a-81ef-b9e44e79825b_es.» Última revisión 30 de abril de 20204.

en cómo los modelos de IA procesan los datos y llegan a conclusiones¹¹, discriminación –si la IA es entrenada con datos sesgados y con subrepresentación en los mismos–, invasiones a la privacidad –con el uso de cámaras de reconocimiento facial o recopilando datos personales de comportamiento– y su posible uso en actividades delictivas –ciberataques, evasión de sistemas de seguridad–, entre otros peligros (Drnas de Clément 2022: 14).

Para contrarrestar los riesgos identificados y maximizar los beneficios anticipados, el *Libro Blanco* de la UE recomienda un enfoque dual. Esto implica crear un ambiente propicio para la excelencia en la IA, así como establecer un marco de confianza. La excelencia se logra fortaleciendo las capacidades de la UE en materia de IA, colaborando estrechamente con los Estados miembros y centrándose en la investigación e innovación, además de fomentar la implementación de la IA en el tejido económico y administrativo europeo, con una atención particular hacia el impulso de las pequeñas y medianas empresas y el sector privado. Por otro lado, el entorno de confianza se orienta hacia la creación de un marco regulatorio sólido, imponiendo obligaciones legales a las aplicaciones de IA consideradas de alto riesgo, como las utilizadas en el campo médico o en la seguridad pública, para garantizar su seguridad, fiabilidad y adhesión a los derechos humanos. Además, se sugiere la introducción de un sistema de etiquetado voluntario para las aplicaciones de IA de menor riesgo, con el fin de verificar su conformidad con las regulaciones existentes. De este modo, este enfoque dual aspiró a guiar el desarrollo de la IA de forma alineada con los valores europeos, estableciendo a la UE como un referente en el campo de las tecnologías digitales con perspectiva ética y de confianza (Colcelli y Burzagli 2021: 3 ss).

¹¹ La opacidad en la IA puede llevar a desafíos significativos, especialmente en contextos críticos como la medicina, la justicia y la seguridad pública, donde las decisiones deben ser justificables y transparentes. Esto ha llevado a un creciente interés en el campo de la IA explicable (XAI), que busca desarrollar métodos y técnicas para hacer que las decisiones y operaciones de la IA sean más comprensibles para los humanos (Vaquero de Miguel 2020).

Siguiendo las recomendaciones del *Libro Blanco*, en abril de 2021 la Comisión Europea dio un paso audaz al presentar la primera propuesta legislativa a nivel mundial específica para regular la IA conocida como *Acta de Inteligencia Artificial de la Unión Europea*¹² –también llamada Ley de Inteligencia Artificial de la Unión Europea–, la cual respondió, *inter alia*, a la demanda del Parlamento Europeo de contar con un marco normativo que garantizara un mercado interno equilibrado para los sistemas de IA, atendiendo a sus riesgos y beneficios. Por ello, uno de los objetivos de esta propuesta fue proporcionar claridad legal, fomentar la inversión y la innovación, y asegurar que los sistemas de IA en el mercado fueran seguros, respetuosos con los derechos fundamentales y con una gobernanza efectiva. De esta manera, la iniciativa buscaba prevenir la fragmentación del mercado y promover un enfoque basado en el riesgo, todo ello dentro de un marco que favoreciera la confianza y la legalidad¹³.

Finalmente, en marzo de 2024, después de diversas revisiones que derivaron en más de 800 enmiendas al Acta propuesta por la Comisión Europea, el Parlamento Europeo aprobó el AIA marcando, con ello, un hito regional y global en la regulación de esta tecnología. Si bien la implementación completa de esta normativa está prevista para 2026, con un período transitorio para la prohibición de sistemas de IA específicos, su aprobación representa indudablemente un avance significativo para la supervisión de la IA, resaltando la importancia de un equilibrio entre avance tecnológico y derechos individuales en el contexto europeo.

III. ESTRUCTURA Y CONTENIDO DEL AIA

El AIA de la Unión Europea se encuentra sistematizada y diseñada como un marco regulatorio funcional para abordar las múltiples

¹² Disponible en «<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.» Última revisión 30 de abril de 2024.

¹³ European Commission, *COM (2021)206*, 21 abril 2021: 2 ss.

facetas del desarrollo y aplicación de la inteligencia artificial dentro de los Estados miembros de la UE. Esta normativa se divide en diversos títulos y capítulos, cada uno enfocado en regulaciones específicas que garantizan la implementación segura y ética de la IA. Dentro del contenido de sus principales títulos podemos encontrar desde regulaciones detalladas para los sistemas de IA considerados de alto riesgo hasta la obligación de los proveedores de sistemas de IA a divulgar información clave sobre el funcionamiento y los propósitos de sus aplicaciones, permitiendo así que los usuarios y reguladores tengan un entendimiento claro y accesible de las tecnologías empleadas. De esta manera, cada título de este documento parece haber sido diseñado con el propósito de cubrir exhaustivamente los distintos aspectos regulatorios, desde la protección de datos hasta la interacción ética con sistemas autónomos, asegurando así que la tecnología de IA se desarrolle y aplique de una manera que beneficie a la sociedad, promueva la innovación y proteja los derechos fundamentales de todas las personas dentro de la Unión Europea.

La estructura del AIA está compuesta por un nutrido apartado de considerandos y 13 capítulos principales conformados por diversos artículos enfocados en distintos aspectos regulatorios. Su organización es la siguiente:

1. Considerandos

El AIA incluye 180 razones de derecho que proporcionan fundamento y contexto sobre cómo deben interpretarse y aplicarse los artículos del Acta. Estos considerandos son esenciales porque ofrecen justificaciones detalladas sobre la necesidad de regulación y las intenciones legislativas detrás de cada artículo del Acta abarcando una gama de temas, desde la necesidad de establecer un marco jurídico coherente en toda la UE para evitar la fragmentación del mercado interno, hasta cuestiones específicas de derechos humanos, ética y seguridad. Destacan la importancia de crear un entorno de IA confiable y seguro, garantizando que el desarrollo

y la implementación de sistemas de IA respeten los derechos fundamentales y la dignidad humana.

Además, explican la categorización de los sistemas de IA según su riesgo potencial y la imposición de requisitos más estrictos para aquellos sistemas considerados de alto riesgo. Por ejemplo, se establecen criterios detallados para la transparencia y la supervisión de la IA, incluyendo la obligación de informar cuando se interactúa con sistemas automáticos, como los *chatbots* o sistemas generativos, y medidas para prevenir usos abusivos como la manipulación comportamental o el reconocimiento facial indebido en espacios públicos.

Estos considerandos también reflejan el compromiso de la UE con la innovación responsable, mencionando la creación de espacios de prueba regulados para permitir el desarrollo seguro de nuevas tecnologías antes de su introducción en el mercado. La profundidad y amplitud de estos considerandos muestran cómo la UE intenta equilibrar la promoción de la tecnología y la innovación con la protección de los ciudadanos y los valores éticos.

2. Disposiciones Generales

Artículos que integran el capítulo 1 del AIA.

Artículo 1: Objeto.

Artículo 2: Ámbito de aplicación.

Artículo 3: Definiciones.

Artículo 4: Alfabetización en IA.

En este capítulo se establece que la finalidad del AIA es mejorar el mercado interior y fomentar la adopción de una IA confiable y centrada en el ser humano que ofrezca un elevado nivel de protección al Estado de derecho, al medio ambiente y a los derechos humanos resaltando, de esta forma, la importancia de equilibrar el progreso tecnológico con la protección integral de valores fundamentales aceptados por todos los miembros de la UE. Asimismo, establece que todos los actores involucrados en la creación, comer-

cialización y uso de sistemas de IA en la UE –proveedores, distribuidores, fabricantes, entre otras– deberán ceñirse a lo dispuesto por este marco regulatorio. No obstante, el Acta señala que estas disposiciones no serán aplicables a los sistemas de IA utilizados exclusivamente para fines militares, de defensa o de seguridad nacional basándose en el respeto a las competencias de los Estados Miembros en materia de seguridad nacional y a la soberanía de cada país en gestionar sus propios asuntos de defensa sin la interferencia del derecho comunitario.

Este apartado ofrece también una larga lista de términos en la cual, *inter alia*, se determina lo que deberá entenderse por sistema de IA, por datos biométricos y sensibles y por riesgo, enfocándose en la probabilidad y gravedad de perjuicios potenciales. De igual forma, señala la obligación de que proveedores y responsables de sistemas de IA garanticen que quienes implementan y operan sistemas de IA posean conocimientos adecuados sobre la tecnología –*alfabetización en IA*–. Lo anterior busca asegurar un uso responsable, considerando la experiencia y formación de los involucrados, así como el contexto particular de aplicación de la IA.

3. *Prácticas de Inteligencia Artificial Prohibidas*

Artículos que integran el capítulo 2 del AIA.

Artículo 5: Prácticas de Inteligencia Artificial prohibidas.

Este capítulo constituye un componente crucial para la protección de los derechos humanos en la era de la inteligencia artificial. Su contenido prohíbe explícitamente el uso de sistemas de IA que manipulen subliminalmente a las personas o exploten vulnerabilidades específicas, como la edad, la discapacidad, la condición social o económica, entre otras, para alterar el comportamiento de manera perjudicial por medio de manipulaciones o engaños derivados de sistemas de IA dirigidos a alterar de manera significativa el comportamiento de las personas afectadas, lo que podría llevarlas a tomar decisiones que no tomarían de otro modo.

Estas prohibiciones aseguran que la IA no se utilice para socavar la autonomía individual o la capacidad de tomar decisiones informadas, poniendo un especial énfasis en el deber de proteger a los individuos más vulnerables de prácticas que puedan abusar de sus circunstancias particulares. Este enfoque preventivo refleja un compromiso serio con la ética y la protección contra el abuso tecnológico, priorizando la dignidad y los derechos fundamentales sobre las capacidades técnicas de los sistemas de IA.

4. Sistemas de IA de alto riesgo

Artículos que integran el capítulo 3 del AIA.

Sección 1: Clasificación de los sistemas de IA como de alto riesgo.

- Artículo 6: Normas de clasificación de los sistemas de IA de alto riesgo.
- Artículo 7: Modificaciones del anexo III.

Sección 2: Requisitos de los sistemas de IA de alto riesgo

- Artículo 8: Cumplimiento de los requisitos.
- Artículo 9: Sistema de gestión de riesgos.
- Artículo 10: Datos y gobernanza de datos.
- Artículo 11: Documentación técnica.
- Artículo 12: Mantenimiento de registros.
- Artículo 13: Transparencia y suministro de información a los empresarios.
- Artículo 14: Supervisión humana.
- Artículo 15: Precisión, robustez y ciberseguridad.

Sección 3: Obligaciones de los proveedores e implantadores de sistemas de IA de alto riesgo y otras partes interesadas.

- Artículo 16: Obligaciones de los proveedores de sistemas de IA de alto riesgo.
- Artículo 17: Sistema de gestión de la calidad Sistema de gestión de la calidad.
- Artículo 18: Conservación de la documentación.
- Artículo 19: Registros generados automáticamente.
- Artículo 20: Acciones correctoras y deber de información.
- Artículo 21: Cooperación con las autoridades competentes.
- Artículo 22: Representantes autorizados de los proveedores de sistemas de IA de alto riesgo.
- Artículo 23: Obligaciones de los importadores.
- Artículo 24: Obligaciones de los distribuidores.
- Artículo 25: Responsabilidades a lo largo de la cadena de valor de la IA.
- Artículo 26: Obligaciones de los implantadores de sistemas de IA de alto riesgo.
- Artículo 27: Evaluación de impacto sobre los derechos fundamentales de los sistemas de IA de alto riesgo.

Artículos que integran el capítulo 3 del AIA.

Sección 4: Autoridades de notificación y organismos notificados.

- Artículo 28: Autoridades de notificación.
 - Artículo 29: Solicitud de notificación de un organismo de evaluación de la conformidad.
 - Artículo 30: Procedimiento de notificación.
 - Artículo 31: Requisitos relativos a los organismos notificados.
 - Artículo 32: Presunción de conformidad con los requisitos relativos a los organismos notificados.
 - Artículo 33: Filiales y subcontratación de los organismos notificados.
 - Artículo 34: Obligaciones operativas de los organismos notificados.
 - Artículo 35: Números de identificación y listas de organismos notificados designados en virtud del presente Reglamento.
 - Artículo 36: Modificaciones de las notificaciones.
 - Artículo 37: Impugnación de la competencia de los organismos notificados.
 - Artículo 38: Coordinación de los organismos notificados.
 - Artículo 39. Organismos de evaluación de la conformidad de terceros países.
-

Sección 5: Normas, evaluación de la conformidad, certificados, registro

- Artículo 40: Normas armonizadas y productos de normalización.
 - Artículo 41. Especificaciones comunes.
 - Artículo 42. Presunción de conformidad con determinados requisitos.
 - Artículo 43. Evaluación de la conformidad.
 - Artículo 44. Certificados.
 - Artículo 45: Obligaciones de información de los organismos notificados.
 - Artículo 46: Excepción al procedimiento de evaluación de la conformidad.
 - Artículo 47 Declaración de conformidad de la UE.
 - Artículo 48: Marcado CE.
 - Artículo 49. Registro.
-

Este apartado revisa y regula detalladamente los sistemas de IA denominados *de alto riesgo* –dispositivos médicos, sistemas para determinar el acceso a instituciones educativas, sistemas utilizados en los ámbitos de la aplicación de la ley o control de fronteras, entre otros– enfocándose en su clasificación, requisitos regulatorios, y medidas de gestión. Asimismo, el capítulo señala los distintos procedimientos que el ordenamiento contempla para asegurar que estos sistemas cumplan con los estándares establecidos, incluyendo la supervisión de su desempeño y la gestión de riesgos a lo largo de su ciclo de vida. Sobre este punto el AIA resalta la impor-

tancia de la transparencia y la trazabilidad de los sistemas de IA por lo que consagra la obligación de proporcionar documentación y registros detallados que faciliten la evaluación de dichos sistemas de conformidad con los parámetros establecidos en la propia Acta –y en otros documentos normativos de la UE como los reglamentos (UE) 2016/679 y (UE) 2018/1725 y la directiva (UE) 2016/680–.

De manera general, en esta sección también se instauran una serie de obligaciones específicas para los proveedores, importadores y distribuidores, asegurando que todos los actores en la cadena de suministro asuman el compromiso de mantener los altos estándares de seguridad y eficacia de los sistemas de IA de alto riesgo. La responsabilidad jurídica que deriva del desarrollo y uso de este tipo IA, y que en este apartado se detalla, puede apreciarse en la regulación que se establece sobre la gobernanza de datos, la transparencia en las comunicaciones y la evaluación del impacto sobre los derechos fundamentales de los sistemas de IA de alto riesgo.

En relación con la gobernanza de datos, el artículo 10 del AIA contempla diversas directrices fundamentales para la gestión ética y responsable de los datos utilizados en la inteligencia artificial asegurando que los procesos asociados a la recolección y salvaguarda de estos sean comprensibles y accesibles para todas las partes. Lo anterior resulta fundamental para garantizar el derecho de las personas a entender cómo se utilizan sus datos y cómo se toman decisiones que los afectan, permitiéndoles ejercer sus derechos de manera efectiva. Este enfoque resulta esencial no solo para la eficacia técnica de los sistemas de IA, sino también para asegurar la protección y el respeto de los derechos humanos fundamentales frente al desarrollo y uso de IA de alto riesgo. En orden de alcanzar este objetivo, el Acta contempla una serie de condiciones fundamentales para la gestión adecuada de este tipo de IA. Entre ellas se pueden mencionar las siguientes:

- 1) *Representatividad de los datos*: los datos deben ser pertinentes, suficientemente representativos y libres de errores pues ello permite evitar los sesgos que pueden llevar a decisiones discrimi-

natorias o injustas por parte de sistemas de IA, especialmente en contextos donde los sistemas de IA pueden influir significativamente en las vidas de personas que, por sus condiciones particulares, se encuentren en un grado especial de vulnerabilidad.

- 2) *Privacidad y seguridad de los datos*: el AIA busca asegurar que cualquier dato personal sea recogido y tratado de manera que respete la privacidad y la dignidad de las personas. Esto incluye limitaciones en la reutilización de los datos personales y medidas avanzadas de protección de la privacidad como la seudonimización. En este punto el Acta se armoniza con las disposiciones establecidas en Reglamento (UE) 2016/679 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD)¹⁴.
- 3) *Prevención y Mitigación de Sesgos*: de acuerdo con el AIA, la detección y mitigación de sesgos en los datos es fundamental para garantizar que los sistemas de IA operen de manera justa y no perjudiquen la salud, la seguridad, o los derechos fundamentales de las personas. De esta forma, la identificación de posibles impactos negativos y la implementación de correcciones son pasos vitales para salvaguardar estos derechos.

Con respecto al tema de transparencia y comunicaciones, el artículo 13 del AIA contempla una serie de medidas relevantes para garantizar la transparencia en el uso y manejo de sistemas de IA de alto riesgo con el propósito de garantizar que los operadores de estos sistemas, así como los usuarios finales, tengan acceso a información completa y comprensible sobre cómo estos funcionan, cuáles son sus capacidades y limitaciones y cómo deben ser manejados correctamente para prevenir malos usos o efectos no deseados. Para ello, el Acta promueve diversos estándares clave

¹⁴ El texto de dicho reglamento europeo se encuentra disponible en «<https://eur-lex.europa.eu/legal-content/ES/TXT/?qid=1532348683434&uri=CELEX%3A02016R0679-20160504>.» Última revisión 30 de abril de 2024.

para un uso ético y responsable de la IA, entre los que se pueden mencionar los siguientes:

- 1) *Impulso a una transparencia eficaz*: que los sistemas de IA sean diseñados para garantizar un nivel adecuado de transparencia que permita a los usuarios entender cómo se toman las decisiones automatizadas, especialmente en términos de explicabilidad y justicia.
- 2) *Documentación comprensible*: que las instrucciones de uso sean claras y accesibles pues ello ayuda a garantizar que los responsables de operar la IA los hagan de manera segura y eficaz, reduciendo así el riesgo de errores que podrían comprometer la seguridad personal o los derechos fundamentales de los individuos afectados.
- 3) *Derecho a la información*: el numeral respalda el derecho a la información al garantizar que los usuarios y las partes interesadas comprendan plenamente cómo los datos personales son utilizados y gestionados.

Finalmente, la evaluación de impacto relativa a los derechos fundamentales para los sistemas de IA de alto riesgo se regula en el artículo 27 del AIA desde un enfoque preventivo y reflexivo para garantizar que la implementación de este tipo de tecnología no solo cumpla con los requisitos técnicos y de seguridad, sino que también respete y proteja los derechos humanos y libertades fundamentales de las personas involucradas. Para ello el artículo contempla la posibilidad de requerir evaluaciones de impacto antes de la implementación de un sistema de IA pues, basándose en un principio precautorio, el AIA busca prevenir violaciones o afectaciones de derechos antes de que ocurran, en lugar de tratar de rectificarlas posteriormente. Algunos elementos clave de este numeral son los siguientes:

- 1) *Evaluación previa obligatoria*: se debe realizar una evaluación de impacto que explique cómo los sistemas de IA podrían afectar

Trazando el camino hacia una regulación efectiva: Breve...

los derechos fundamentales de personas y grupos específicos. Esta evaluación debe incluir una descripción detallada del uso previsto del sistema, las personas afectadas y los posibles daños.

- 2) *Responsabilidad de entes públicos y privados*: el contenido del artículo vincula tanto a instituciones públicas como a entidades privadas que prestan servicios públicos bajo el argumento de que quien maneje infraestructuras que puedan generar un impacto significativo en la sociedad, tienen un deber especial con la protección de los derechos fundamentales.
 - 3) *Interacción con otras evaluaciones de impacto*: el artículo también menciona que la evaluación de impacto sobre los derechos fundamentales puede complementar otras evaluaciones, como las relativas a la protección de datos personales bajo el RGPD, lo que muestra una intención de integrar los riesgos asociados a la IA pues estos pueden ser multifacéticos y afectar diversos aspectos de la vida de las personas.
5. *Capítulo 4. Obligaciones de transparencia para proveedores e implantadores de determinados sistemas de IA y modelos GPAI*

Artículos que integran el capítulo 4 del AIA.

Artículo 50: Obligaciones de transparencia para proveedores e implantadores de determinados sistemas de IA y modelos GPAI.

El contenido de este artículo va dirigido a garantizar la transparencia en la interacción entre sistemas de IA y personas físicas. Con este propósito, en sus disposiciones se reconoce que los sistemas de IA que interactúan directamente con personas deben ser diseñados de tal manera que los usuarios sean conscientes de que están interactuando con un sistema de IA pues ello es indispensable para asegurar que las decisiones que tomen estén adecuadamente informadas, protegiendo así su autonomía y evitando engaños. Asimismo, se busca garantizar que al momento de interactuar con una IA las personas usuarias puedan distinguir entre los contenidos

generados por humanos y aquellos desarrollados por máquinas, por lo que se exige un etiquetado de contenidos generados por IA para advertir sobre su origen artificial.

En esta misma línea, el numeral dispone que tratándose de sistemas de reconocimiento de emociones y de categorización biométrica, los usuarios deberán ser oportuna y suficientemente informados sobre cómo funcionan y cómo se manejan sus datos personales, alineándose esta obligación con diversos reglamentos de protección de datos de la UE (v.gr.: reglamentos (UE) 2016/679 y (UE) 2018/1725 y directiva (UE) 2016/680). No obstante, debe resaltarse que el AIA señala que esta obligación no será aplicable cuando los sistemas de IA se utilicen bajo autorización legal para detectar, prevenir, investigar o enjuiciar delitos, siempre que se mantengan las garantías adecuadas para los derechos y libertades de terceros.

6. Modelos de IA de propósito general

Artículos que integran el capítulo 5 del AIA.

Sección 1: Normas de clasificación.

- Artículo 51: Clasificación de los modelos de IA de propósito general con riesgo sistémico.
- Artículo 52: Procedimiento.

Sección 2: Obligaciones de los proveedores de modelos de IA de propósito general.

- Artículo 53. Obligaciones de los proveedores de modelos de IA de propósito general.
- Artículo 54: Representante autorizado.

Sección 3: Obligaciones de los proveedores de modelos de IA de propósito general con riesgo sistémico.

- Artículo 55: Obligaciones de los proveedores de modelos de IA de propósito general con riesgo sistémico.
- Artículo 56: Códigos de buenas prácticas.

La regulación de este capítulo va dirigida a normar y clasificar modelos de IA de uso general, destacando particularmente aquellos

identificados como poseedores de un riesgo sistémico¹⁵ ya sea porque estos cuentan con una capacidad de gran impacto (significativo en contextos críticos), o porque la Comisión Europea, basándose en evaluaciones técnicas o recomendaciones de grupos de expertos decide incluirlos dentro de esta clasificación. No obstante, en apego al principio de seguridad jurídica, el AIA contempla la posibilidad de que cualquier proveedor inicie un procedimiento de evaluación cuando considere que su modelo de IA no debe incluirse dentro de la lista de riesgo sistémico.

De esta forma, el numeral contempla una serie de obligaciones diferenciadas que deberán cumplir los proveedores de modelos de IA con base en si estos son considerados de uso general o si son identificados con riesgos sistémicos. Entre estas obligaciones encontramos mantener una documentación detallada y accesible sobre los modelos que se utilizan –incluyendo su diseño, entrenamiento, evaluación, entre otras– y la realización de pruebas de adversarios para identificar y mitigar riesgos potenciales.

El contenido regulatorio de este capítulo evidencia una intención proactiva y detallada al considerar los riesgos que pueden asociarse al uso de la IA en general, con especial atención a aquellos modelos que pueden generar consecuencias sistémicas. Por ello, con estas disposiciones el AIA busca mitigar los riesgos potenciales que estos modelos pueden representar para la sociedad buscando garantizar un equilibrio entre la innovación tecnológica y la protección del bienestar público.

¹⁵ El AIA señala, en su artículo 3.65, que a efectos de este reglamento por daño sistémico deberá entenderse "... 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".

7. Medidas de apoyo a la innovación

Artículos que integran el capítulo 6 del AIA.

Artículo 57: Espacios aislados de regulación de la IA.

Artículo 58: Disposiciones detalladas y funcionamiento de los espacios aislados de regulación de la IA.

Artículo 59: Tratamiento posterior de datos personales para el desarrollo de determinados sistemas de IA de interés público en el espacio aislado de regulación de la IA.

Artículo 60: Pruebas de sistemas de IA de alto riesgo en condiciones del mundo real fuera de los espacios aislados de regulación de la IA.

Artículo 61: Consentimiento informado para participar en pruebas en condiciones del mundo real fuera de los espacios aislados de regulación de la IA.

Artículo 62: Medidas para proveedores e implantadores, en particular las PyMES, incluidas las empresas de nueva creación.

Artículo 63: Excepciones para operadores específicos.

Este apartado va dirigido a establecer medidas que incentiven el desarrollo seguro de sistemas de IA. Este enfoque evidencia la intención de equilibrar la pujante innovación en esta área con la necesidad de su regulación lo que supone una preocupación central del AIA en la gobernanza de las tecnologías inteligentes. De esta forma, el articulado de este capítulo señala, *inter alia*, que los Estados miembro de la UE deberán desarrollar, por lo menos, un espacio controlado de pruebas para los sistemas de IA¹⁶ con el objetivo de proporcionar un entorno regulado, pero flexible en términos regulatorios, para que las empresas puedan desarrollar y probar tecnologías de IA, lo que permite realizar una evaluación precautoria de los potenciales impactos, positivos y negativos, de un sistema de IA en una diversidad de contextos.

¹⁶ Los espacios de prueba controlados para IA –*regulatory sandboxes*– constituyen un entorno regulatorio y/o experimental diseñado para probar y evaluar tecnologías de IA de manera segura y controlada antes de su implementación plena

Trazando el camino hacia una regulación efectiva: Breve...

Estos espacios, dicta este apartado, no solo deberán promover la innovación, sino que deberán hacerlo garantizando un mínimo de estándares éticos y de seguridad, especialmente en el tratamiento de datos personales, por lo que su diseño deberá ser inclusivo y accesible. Esto cobra mayor relevancia en las pruebas de IA de alto riesgo las cuales, si bien podrán realizar fuera de los entornos controlados para evaluarse en entornos reales, deberán contar con estrictas condiciones de supervisión y de consentimiento informado.

Por ende, el análisis integral de este capítulo refleja un marco jurídico robusto orientado al establecimiento de normas que fomenten un ambiente de creatividad en el desarrollo de tecnologías de IA, dentro de un entorno seguro y ético. Este planteamiento integral se dirige a equilibrar la innovación con la responsabilidad y la supervisión normativa.

8. *Gobernanza*

Artículos que integran el capítulo 7 del AIA.

Sección 1: Gobernanza a escala de la Unión.

- Artículo 64: Oficina de IA.
- Artículo 65: Creación y estructura del Consejo Europeo de Inteligencia Artificial.
- Artículo 66: Funciones del Consejo.
- Artículo 67: Foro consultivo.
- Artículo 68: Grupo científico de expertos independientes.
- Artículo 69: Acceso de los Estados miembros al grupo de expertos.

Sección 2: Autoridades nacionales competentes.

- Artículo 70: Designación de las autoridades nacionales competentes y punto de contacto único.

en situaciones reales. Ello permite a los desarrolladores, investigadores y reguladores estudiar cómo los sistemas de IA operan en escenarios simulados o bajo ciertas condiciones controladas, para identificar potenciales problemas, riesgos o desafíos operativos y éticos (Bustos Moreno 2022).

Este apartado del AIA insta un marco jurídico-organizativo de gobernanza dedicado a la supervisión del uso adecuado y efectivo de los sistemas de IA dentro de la UE. Para lograr ello, contempla una regulación dedicada a la gobernanza dentro de la Unión y una regulación enfocada a las responsabilidades de las autoridades nacionales competentes.

En relación con la gobernanza a escala de la UE, el AIA contempla la creación de un *Comité Europeo de Inteligencia Artificial* compuesto por un representante de cada Estado miembro –designado por su Estado por un período de 3 años con posibilidad de renovación única– y el Supervisor Europeo de Protección de Datos¹⁷ –quien participará en calidad de observador–. Entre las funciones principales de este órganos se encuentran: el asesoramiento y la asistencia técnica tanto a la Comisión Europea como a los Estados miembros en formas para la aplicación coherente y eficaz del AIA; facilitar la coordinación y comunicación entre autoridades nacionales; recopilar conocimientos técnicos y mejores prácticas entre los Estados miembro; coadyuvar en la armonización de las prácticas administrativas relacionadas con la IA; emitir recomendaciones y dictámenes relacionados con cualquier asunto relacionado con la ejecución y aplicación del AIA; promover la alfabetización en IA para la mejor comprensión de riesgos, ventajas, obligaciones y derechos en el uso de sistemas de IA; contribuir al desarrollo de parámetros de referencia para la mejor comprensión entre operadores del mercado y autoridades competentes de los conceptos incluidos en el AIA; entre otros.

¹⁷ El Supervisor Europeo de Protección de Datos (SEPD) es una autoridad independiente encargada de garantizar que las instituciones y órganos de la Unión Europea respeten el derecho a la protección de datos personales cuando procesan la información de individuos. El SEPD también asesora sobre políticas y legislación que afectan a la privacidad y cooperan con autoridades similares para asegurar la coherencia en la protección de datos en toda Europa. Ver Reglamento (UE) 2018/1725 del Parlamento Europeo y del Consejo de 23 de octubre de 2018. Disponible en «<https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A32018R1725>.» Última revisión, 30 de abril de 2024.

Trazando el camino hacia una regulación efectiva: Breve...

Por lo tanto, el Comité Europeo de Inteligencia Artificial se constituye como una pieza clave dentro de la estructura de gobernanza de la IA en la UE, al asumir el mandato de asegurar una aplicación uniforme y efectiva de las normativas de IA y de actuar como un órgano central de conocimiento y coordinación entre los Estados miembros y la Comisión.

En este mismo rubro, el capítulo VII del AIA contempla la creación de un Foro Consultivo y de un Grupo de Expertos Científicos Independientes. El primero tiene como propósito principal proporcionar conocimientos técnicos y asesoramiento al Comité Europeo de Inteligencia Artificial así como facilitar la participación de representantes de la industria, la sociedad civil y el sector académico en el proceso de gobernanza de la IA en la UE. El segundo, conformado por personas especialistas en el ámbito de la IA, tiene como funciones principales la de alertar sobre posibles riesgos sistémicos y contribuir al desarrollo de herramientas y metodologías para evaluar estos sistemas.

Respecto a la gobernanza enfocada a las autoridades nacionales competentes, el AIA establece el deber de cada Estado miembro de designar autoridades que actúen de manera independiente e imparcial para asegurar el cumplimiento del contenido normativo del Acta. Para ello, dichas autoridades tendrán la responsabilidad de mantener un nivel óptimo de recursos y conocimientos técnicos para supervisar efectivamente la IA, lo que incluye competencias en tecnologías de IA, protección de datos, ciberseguridad, y conocimiento de las normativas legales aplicables.

9. Base de datos de la UE para sistemas de IA de alto riesgo

Artículos que integran el capítulo 8 del AIA.

Artículo 71: Base de datos de la UE para los sistemas de IA de alto riesgo.

El artículo que conforma este apartado insta a un marco regulatorio sobre la gestión de información relacionada con los sistemas

de inteligencia artificial. De esta forma, el numeral establece el mandato, dirigido a la Comisión y a los Estados miembro, de conformar una base de datos que distinga entre los sistemas de IA de alto riesgo y los sistemas que no se consideren así. Dicho banco de información debe ser accesible, abierto al público y construirse bajo el principio de estricta necesidad, por lo que únicamente deberán almacenarse los datos personales que sean necesarios para cumplir con los objetivos del AIA. Esto subraya la importancia de la privacidad y la protección de datos dentro del marco regulatorio de la IA en la UE.

10. Control posterior a la comercialización, intercambio de información y vigilancia del mercado

Artículos que integran el capítulo 9 del AIA.

Sección 1: Seguimiento postcomercialización.

- Artículo 72: Seguimiento postcomercialización por parte de los proveedores y plan de seguimiento postcomercialización para sistemas de IA de alto riesgo.

Sección 2: Intercambio de información sobre incidentes graves.

- Artículo 73. Notificación de incidentes graves.

Sección 3: Ejecución.

- Artículo 74: Vigilancia del mercado y control de los sistemas de IA en el mercado de la Unión.

- Artículo 75: Asistencia mutua, vigilancia del mercado y control de los sistemas de IA de propósito general.

- Artículo 76: Supervisión de las pruebas en condiciones reales por las autoridades de vigilancia del mercado.

- Artículo 77: Competencias de las autoridades de protección de los derechos fundamentales.

- Artículo 78. Confidencialidad.

- Artículo 79: Procedimiento para tratar los sistemas de IA que presenten un riesgo a nivel nacional.

- Artículo 80: Procedimiento para tratar los sistemas de IA clasificados por el proveedor como de riesgo no elevado en aplicación del Anexo III.

- Artículo 81: Procedimiento de salvaguardia de la Unión.

- Artículo 82: Sistemas de IA conformes que presentan un riesgo.

- Artículo 83. Incumplimiento formal Incumplimiento formal.

- Artículo 84: Estructuras de apoyo a las pruebas de IA de la Unión.

Artículos que integran el capítulo 9 del AIA.

Sección 4: Recursos.

- Artículo 85: Derecho a presentar una reclamación ante una autoridad de vigilancia del mercado.
 - Artículo 86: Derecho a la explicación de las decisiones individuales.
 - Artículo 87: Notificación de infracciones y protección de los denunciantes.
-

Sección 5: Supervisión, investigación, aplicación y control de los proveedores de modelos de IA de propósito general.

- Artículo 88: Cumplimiento de las obligaciones de los proveedores de modelos de IA de propósito general.
 - Artículo 89: Acciones de seguimiento.
 - Artículo 90: Alertas de riesgos sistémicos por la Comisión técnica científica.
 - Artículo 91: Facultad de solicitar documentación e información.
 - Artículo 92: Facultad de realizar evaluaciones.
 - Artículo 93: Facultad de solicitar medidas.
 - Artículo 94: Derechos procesales de los operadores económicos del modelo de IA de propósito general.
-

Este capítulo del AIA va dirigido a confeccionar un marco regulatorio para la supervisión de sistemas de IA de alto riesgo una vez que estos están en el mercado. De su contenido, variado y robusto, se pueden resaltar las siguientes disposiciones:

- Artículos 72 y 73: contemplan la obligación de los proveedores de sistemas de IA de alto riesgo de constituir un sistema de vigilancia postcomercialización que recopile, documente y analice permanentemente datos relacionados al rendimiento de los sistemas de IA para asegurar que estos continúan ajustándose a los estándares regulatorios establecidos en el AIA. Asimismo, se establece la obligación de notificar a las autoridades de vigilancia del mercado de los Estados miembro cualquier incidente grave inmediatamente después de que el proveedor haya vinculado el incidente con el sistema de IA. Este proceso de notificación rápida y detallada es vital para mitigar posibles daños y para que las autoridades pertinentes puedan tomar acciones correctivas de manera oportuna. En este punto,

la cooperación y las investigaciones resultantes son esenciales para entender y rectificar las fallas en los sistemas de IA.

- Artículos 74 a 84: En esta sección el AIA se centra en garantizar el cumplimiento y supervisión de los sistemas de IA en el mercado de la Unión. Para ello, se dispone que el Reglamento (UE) 2019/1020 se adaptará en todo lo que sea necesario para que pueda aplicarse específicamente a los sistemas de IA. Por lo tanto, la referencia a *operador económico* que realiza el reglamento de 2019 debe incluir a los proveedores de sistemas de IA y la referencia a *producto* incluirá los sistemas de IA. De esta manera se armoniza y consolida el marco legal aplicable a la IA en la UE permitiendo adaptar estructuras y procesos ya existentes a las particularidades de la IA. Asimismo, se especifica que las autoridades de vigilancia del mercado tendrán la obligación informar anualmente a la Comisión y a las autoridades nacionales competentes sobre cualquier información recabada que, a su juicio, consideren que puede influir en la aplicación de las normas de competencia de la UE. Sobre ello, se resalta que para ciertos sistemas de IA de alto riesgo asociados a productos ya regulados previamente por legislación de la Unión, se designarán autoridades específicas para su vigilancia, asegurando una supervisión coherente, armónica y efectiva.

En esta misma línea, esta sección señala que en el caso de sistemas de IA de uso general estos serán supervisados por las autoridades de vigilancia del mercado y la Oficina de IA del Comité Europeo de Inteligencia Artificial, estableciendo para ello un marco jurídico para la acción en caso de que un sistema no cumpla con los requisitos reglamentarios. A continuación, se resalta la importancia de proteger los derechos fundamentales en el contexto de la IA, por lo que el AIA reconoce que las autoridades nacionales ostentan el poder de solicitar y acceder a cualquier documentación relacionada con sistemas de IA de alto riesgo para asegurar que su uso no infrinja los derechos fundamentales. Esto resalta el compromiso de esta normativa con los principios éticos y los derechos humanos, aspectos esenciales dado el impacto potencial de la IA en áreas

Trazando el camino hacia una regulación efectiva: Breve...

sensibles como la privacidad, la justicia, la seguridad, el medio ambiente y la no discriminación.

Finalmente, en los artículos señalados se contemplan procedimientos aplicables a los sistemas de IA que presenten riesgos, estableciendo un marco detallado para la evaluación, el control, y la toma de medidas correctivas, incluyendo la retirada de productos del mercado cuando sea necesario. Lo anterior resulta fundamental para responder de manera efectiva a cualquier riesgo que los sistemas de IA puedan presentar para la salud, la seguridad o los derechos fundamentales.

- Artículos 85 a 94: En las dos secciones finales de este capítulo se abordan aspectos relevantes sobre la supervisión, el cumplimiento, la protección de los derechos y la transparencia en relación con el uso de sistemas de IA. En consecuencia, los numerales mencionados prescriben un marco normativo detallado para asegurar que los sistemas de IA operen de manera segura y conforme a las normativas, además de establecer vías de recurso y protección para los individuos y operadores afectados.

Del contenido de estos artículos se colige que este capítulo ofrece herramientas y procedimientos esenciales para garantizar que los sistemas de IA de alto riesgo sean seguros, confiables y merecedores de la confianza tanto del público como de los entes reguladores. Esta orientación proactiva hacia la vigilancia después de la comercialización y el intercambio de información resulta fundamental para manejar eficazmente los riesgos que la IA presenta en la sociedad.

11. Códigos de conducta y directrices

Artículos que integran el capítulo 10 del AIA.

Artículo 95: Códigos de conducta para la aplicación voluntaria de requisitos específicos.

Artículo 96: Directrices de la Comisión sobre la aplicación del presente Reglamento.

Este apartado da cuenta de la obligación del Comité Europeo de Inteligencia Artificial y de los Estados miembros de fomentar y facilitar la elaboración de códigos de conducta que estimulen la adopción voluntaria de estas normas para mejorar la confianza y la responsabilidad en el desarrollo y uso de sistemas de IA, especialmente aquellos que no se clasifican como de alto riesgo. Este enfoque no solo busca la adaptación de los sistemas de IA a las normativas vigentes de manera flexible, sino que también apunta a integrar aspectos éticos y de sostenibilidad medioambiental en el diseño y operación de IA. Además, se enfatiza la importancia de la alfabetización en IA y la inclusión, asegurando que los sistemas de IA tengan perspectiva de género y sean accesibles y equitativos para grupos diversos, incluidas personas en situación de vulnerabilidad y personas con discapacidad.

En este capítulo también se establece que la Comisión Europea desarrollará una serie de directrices para facilitar la aplicación práctica del reglamento. Estas directrices, que resultan fundamentales al actuar como un puente entre la legislación y su aplicación efectiva, deberán contemplar aspectos desde la transparencia y las prácticas prohibidas hasta la coherencia con otras regulaciones de la UE y la definición técnica de los sistemas de IA. La intención de las mismas es clarificar cómo se deben interpretar y aplicar las disposiciones del AIA en el contexto práctico, lo que es especialmente crucial para las pequeñas y medianas empresas (PYMES), las startups y las autoridades públicas locales pues, de acuerdo con la presente Acta, son estos actores los que tienen más probabilidad de verse afectados por las disposiciones del AIA.

De esta forma, los numerales que integran este apartado ofrecen una guía para la implementación voluntaria de estándares para el desarrollo y uso de sistemas de IA, asegurando que se promuevan prácticas éticas y sostenibles, y destaca las necesidades específicas de diferentes actores en el ecosistema de la IA. Ello, asegura que la regulación no solo sea aplicable y pertinente en el momento de su

Trazando el camino hacia una regulación efectiva: Breve...

implementación, sino que también permanezca efectiva y relevante ante los desarrollos futuros en la tecnología de IA.

12. Delegación de poderes y funcionamiento de los comités

Artículos que integran el capítulo 11 del AIA.

Artículo 97: Ejercicio de la delegación.

Artículo 98: Procedimiento de comité.

Este capítulo se integra por dos artículos que detallan los mecanismos mediante los cuales la Comisión Europea puede adoptar actos delegados y cómo estos interactúan con los procedimientos del Comité Europeo de Inteligencia Artificial, garantizando así un enfoque estructurado y coherente en la regulación de la IA. Los actos delegados permiten a la Comisión Europea, bajo ciertas condiciones y por un periodo de cinco años a partir de la entrada en vigor del AIA, modificar o añadir disposiciones reglamentarias sin necesidad de cumplimentar los requisitos propios de un proceso de modificación legislativa con el objetivo de adaptar el contenido del Acta a los avances tecnológicos y a las circunstancias emergentes propias del campo de la IA.

No obstante, en este apartado se contempla también la facultad del Parlamento Europeo y del Consejo de revocar esta facultad de la Comisión Europea cuando consideren que existe la necesidad de reestablecer un balance democrático en el caso concreto. El procedimiento para adoptar un acto delegado incluye realizar consultas a expertos de cada Estado miembro y requiere que éste sea notificado al Parlamento Europeo y al Consejo inmediatamente tras su adopción. De esta forma, se busca asegurar que los actos delegados solo entren en vigor si no hay objeciones por parte de estas instituciones dentro de un periodo de tres meses, con posibilidad de extensión por tres meses más, lo que proporciona un mecanismo de control adicional.

13. *Confidencialidad y sanciones*

Artículos que integran el capítulo 12 del AIA.

Artículo 99. Sanciones.

Artículo 100: Multas administrativas a las instituciones, agencias y organismos de la Unión.

Artículo 101: Multas para proveedores de modelos de Ia de propósito general.

Este capítulo indica cuál es la naturaleza, magnitud y forma en la que deberán instaurarse y aplicarse las medidas correctivas necesarias para garantizar que el contenido del AIA no sea solo mandatorio sino que su incumplimiento derive en responsabilidades diferenciadas con base en la condición y gravedad de las infracciones.

En este sentido, el articulado que integra este apartado establece un marco sancionatorio aplicable a las acciones y omisiones que infrinjan el Acta el cual deberá ser equilibrado, proporcionado, justo y con perspectiva de derechos, pues ello resulta decisivo para no disuadir el desarrollo y la innovación tecnológica al imponer cargas que podrían obstaculizar la viabilidad económica de las empresas, especialmente las PYMES y startups. Asimismo, la perspectiva adoptada por el AIA en este numeral deriva en un reconocimiento de los riesgos potenciales para los derechos humanos que podrían ocasionarse por operaciones irresponsables de sistemas de IA, incluyendo la invasión de la privacidad, la discriminación y el daño económico.

En términos específicos, las multas contempladas en el Acta pueden alcanzar hasta un monto de 35 millones de euros o el 7% del volumen de negocios mundial total. Para decidir la imposición de la multa así como la cuantía de esta los Estados miembro deberán tomar en consideración todas las circunstancias que consideren relacionadas al caso concreto, así como la naturaleza, gravedad y consecuencia de la infracción. Asimismo, deberán analizar el nú-

mero de personas afectadas, el nivel de daño generado, si existen factores agravantes o atenuantes, la intencionalidad o negligencia y si otras autoridades han impuesto previamente multas administrativas al operador por la misma acción u omisión que haya infringido el AIA. La consideración de todos estos elementos busca asegurar que las decisiones sean justas y consideren todos los aspectos relevantes de cada caso, lo cual promueve un entorno de respeto y cumplimiento de las disposiciones del Acta –y de la normativa europea relacionada a la misma–, además de proteger los derechos humanos al considerar el impacto de las infracciones en las personas afectadas.

Junto a las sanciones dirigidas a operadores de sistemas de IA, el capítulo contempla la constitución de un marco sancionatorio dirigido a instituciones de la UE las cuales no están exentas de incumplir las regulaciones establecidas en el AIA. En consecuencia, el Supervisor Europeo de Protección de Datos asume el mandato de imponer multas administrativas a órganos y organismos de la UE el cual, al momento de determinar la imposición y cuantía de las multas, deberá considerar, *inter alia*, la naturaleza, gravedad, duración y consecuencias de la infracción, el grado de responsabilidad de la institución, órgano u organismo de la UE, los esfuerzos que se hayan instaurado para mitigar los daños generados y la disposición a colaborar con el Supervisor Europeo, quien tendrá la obligación de informar anualmente a la Comisión Europea el número y cuantía de las multas impuestas y de los litigios relacionados con el objetivo de garantizar la transparencia y la rendición de cuentas dentro del sistema regional generado para regular la IA.

Finalmente, este capítulo contempla también la posibilidad de aplicar multas a proveedores de modelos de IA de uso general cuando estos infrinjan las disposiciones del AIA, omitan dar respuesta a las solicitudes de información, no cumplan con lo ordenado en alguna directiva impuesta directamente a ellos, impidan el acceso a la Comisión para que pueda realizar evaluaciones independientes de los modelos de IA, entre otras.

Debe señalarse que en apego estricto a los principios de seguridad jurídica, debido proceso, entre otros, la Comisión Europea deberá ofrecer al infractor la oportunidad de ser oído, garantizando así un proceso justo y una defensa adecuada para las entidades involucradas. Asimismo, este marco regulatorio contempla la posibilidad de instaurar un control judicial sobre las decisiones administrativas que al respecto toma la Comisión Europea. Éste recae sobre el Tribunal de Justicia de la Unión Europea quien es el órgano competente para revisar las decisiones de la Comisión sobre el tipo y el monto de las sanciones impuestas, pudiendo modificar tanto la forma como la cuantía de éstas.

En conclusión, el capítulo XII del AIA instaura un modelo sancionador que, sin disuadir la innovación y desarrollo en materia de IA, implementa una regulación sólida para prevenir y reparar comportamientos contrarios al AIA. Lo anterior no solo busca reforzar la eficacia del Acta sino que también pretende proteger derechos al garantizar que las entidades que operan con IA actúen de manera responsable y transparente, priorizando la seguridad, la responsabilidad y la equidad en el ecosistema tecnológico emergente.

14. Disposiciones finales

Artículos que integran el capítulo 13 del AIA.

Artículo 102: Modificación del Reglamento (CE) núm. 300/2008.

Artículo 103: Modificación del Reglamento (UE) núm. 167/2013.

Artículo 104: Modificación del Reglamento (UE) núm. 168/2013.

Artículo 105: Modificación de la Directiva 2014/90/UE.

Artículo 106: Modificación de la Directiva (UE) 2016/797.

Artículo 107: Modificación del Reglamento (UE) 2018/858.

Artículo 108: Modificación del Reglamento (UE) 2018/1139.

Artículo 109: Modificación del Reglamento (UE) 2019/2144.

Artículo 110: Modificación de la Directiva (UE) 2020/1828.

Artículo 111: Sistemas de IA ya comercializados o puestos en servicio.

Artículo 112: Evaluación y revisión

Artículo 113. Entrada en vigor y aplicación Entrada en vigor y aplicación.

Trazando el camino hacia una regulación efectiva: Breve...

Finalmente, el último capítulo de esta normativa contempla la necesidad de modificar el contenido de diversos reglamentos y directivas de la UE para armonizar dichos documentos con el contenido del AIA. De esta manera, el Acta busca garantizar que las disposiciones vigentes en la legislación de la UE se alineen con los nuevos estándares europeos establecidos para la inteligencia artificial intentado así realizar una incorporación adecuada que permita mantener un marco regulatorio integral, coherente y actualizado que abarque las innovaciones tecnológicas en IA, asegurando que éstas mantengan altos estándares de seguridad y fiabilidad en su aplicación dentro de los Estados miembros de la UE.

IV. CONCLUSIONES

La entrada en vigor del AIA de la UE constituye un avance significativo en el desarrollo de una regulación global dirigida a gestionar adecuadamente el crecimiento e implementación de tecnologías que están redefiniendo la industria –revolución 4.0– y reconfigurando diversos aspectos de la interacción social –trabajo, medicina, seguridad, medio ambiente, entre otros–. Bajo esta consideración, el Acta europea no solo evidencia una comprensión avanzada de las implicaciones técnicas de la IA, sino que también subraya un fuerte compromiso con la protección de los derechos fundamentales y la promoción de un desarrollo tecnológico ético y responsable.

Con este propósito es que su contenido muestra una constante intención por equilibrar la promoción a la innovación tecnológica con la necesidad de establecer un marco que brinde seguridad y certeza en el uso de sistemas de IA, especialmente cuando estos sean considerados de alto riesgo. Para ello, el Acta contempla un mandato robusto en materia de transparencia y responsabilidad ajustado a las operaciones de IA, dirigido a la construcción de confianza, comprensión y aceptación de las tecnologías emergentes. Además, el Acta de IA resalta la importancia de la transparencia y la *accountability* en las operaciones de IA, mandatos que ayu-

dan a construir confianza en las tecnologías emergentes y facilitan una mayor aceptación pública. Estos elementos resultan de la mayor relevancia en un momento en el que la IA ha mostrado el enorme potencial que tiene para beneficiar, pero también para perjudicar, a la humanidad.

No obstante, si tomamos en cuenta la enorme carga que el AIA deposita en los Estados miembros, su implementación puede generar también incertidumbre y preocupación en el entorno de la UE, pues si tomamos en cuenta las notables diferencias sociales y económicas que existen entre ellos podemos cuestionar si a futuro estos contarán con la capacidad suficiente para aplicar efectivamente esta normativa y para mantener el ritmo de evolución de la tecnología de IA. Indudablemente, el éxito en este punto dependerá de la educación y la capacitación en IA, tanto para los reguladores como para el público en general, como de la cooperación internacional y el intercambio de mejores prácticas, no solo para mitigar estos desafíos, sino también para adaptar continuamente el nuevo marco regulador europeo a los subsecuentes desarrollos tecnológicos.

En conclusión, el AIA constituye un notorio avance en el camino de la regulación efectiva de la IA pues no solo configura una regulación para las tecnologías emergentes dentro de sus fronteras sino que además marca la dirección que debe seguir la regulación del desarrollo tecnológico. En este sentido, el contenido de este reglamento puede representar un excelente punto de partida para que otras regiones y países avancen adecuadamente en el desarrollo de sus propias políticas y normativas. Para ello, el Acta deja clara la relevancia de fomentar, en la era de la inteligencia artificial, la innovación tecnológica con transparencia, responsabilidad y perspectiva de derechos humanos.

BIBLIOGRAFÍA

- Antonov, Alexander (2022): “Gestionar la complejidad: la contribución de la UE a la gobernanza de la inteligencia artificial”, en *Revista CIDOB d’Afers Internacionals*, núm. 131, 41-68. Disponible en: «<https://doi.org/10.24241/rcai.2022.131.2.41>»
- Bustos Moreno, Yolanda B. (2022): “La responsabilidad civil en los espacios controlados de pruebas (regulatory sandboxes) sobre movilidad aérea urbana y la futura Ley de movilidad sostenible”, en *Cuadernos de Derecho Privado*, vol. 2, núm. 2, 8-49.
- Colcelli, Valentina y Burzagli, Laura (2021): “Elementos para una cultura europea de desarrollo de herramientas de inteligencia artificial: el Libro blanco sobre la inteligencia artificial y las Directrices éticas para una IA fiable”, en *Revista Justicia & Derecho*, vol. 4, núm. 2, 1-12. Disponible en: «<https://doi.org/10.32457/rjyd.v4i2.1349>»
- Drnas de Clément, Zlata (2022): “Inteligencia Artificial en el Derecho Internacional, Naciones Unidas y Unión Europea”, en *Revista Estudios Jurídicos*, núm. 22, 1-28. Disponible en: «<https://doi.org/10.17561/rej.n22.7524>»
- Gamero Casado, Eduardo (2021): “El Enfoque Europeo de Inteligencia Artificial”, en *Revista de Derecho Administrativo*, núm. 20, 268-289.
- García, Sara (2022): “Una aproximación a la futura regulación de la inteligencia artificial en la Unión Europea”, en *Revista de Estudios Europeos*, vol. 79, enero-junio, 304-323. Disponible en: «<https://doi.org/10.24197/ree.79.2022.304-323>»
- Gill, Amandeep (2019): “Artificial Intelligence and International Security: The Long View”, en *Ethics & International Affairs*, vol. 33, núm. 2, 169-179. Disponible en: «<https://doi:10.1017/S0892679419000145>»
-

- Morandín-Ahuerma, Fabio (2023): “Recomendación del Consejo sobre inteligencia artificial de la OCDE: desigualdad e inclusión”, en *Principios normativos para una ética de la inteligencia artificial*, Morandín-Ahuerma, Fabio (coord.), CONCYTEP, Puebla, México, 95-102.
- Piedra Alegría, Jonathan (2023): “Anotaciones iniciales para una reflexión ética sobre la regulación de la Inteligencia Artificial en la Unión Europea”, en *Revista de Derecho*, Universidad núm. 28, 1-34. Disponible en: «<https://doi.org/10.22235/rd28.3264>»
- Thiebes, Scott *et al.* (2021): “Trustworthy artificial intelligence”, en *Electron Markets*, núm. 31, 447–464. Disponible en: «<https://doi.org/10.1007/s12525-020-00441-4>»
- Vaquero de Miguel, Guillermo (2020): *Aproximaciones a la explicación de decisiones algorítmicas: Inteligencia Artificial Explicable*, Trabajo de Grado, Universidad Politécnica de Madrid, España.

Inteligencia artificial y derecho en México. Un dilema: El mito del vehículo automatizado en México

*Artificial intelligence and law in Mexico. A dilemma:
“The myth of the automated vehicle in Mexico”*

JUAN CARLOS GÁMEZ SAGRERO

Escuela Libre de Derecho

ORCID: 0009-0009-2150-4316

Fecha de recepción: 01 marzo 2024

Fecha de aceptación: 17 junio 2024

Sumario: I. Introducción. II. ¿Qué es la IA en un vehículo de conducción automatizada? III. Sesgos y fallas de la IA. IV. Normatividad de derecho positivo mexicano que da respuesta al caso hipotético planteado. 1. El concepto jurídico de persona. 2. Las personas jurídicas. 3. Temperamentos legales. 4. Repensar el concepto de persona moral o jurídica. 5. Un punto medio. 6. El aparato legal mexicano. 7. La perspectiva del derecho civil mexicano. 8. La perspectiva del derecho penal mexicano. 9. La responsabilidad penal de la persona moral o jurídica. V. La teoría de la imputación objetiva. VI. Posibles causas que puede tener el caso bajo el enfoque penal por la integración del elemento IA en la respuesta legal. VII. Un breve ejercicio en materia penal. VIII. Un posible caso de atipicidad por ausencia de una adecuada regulación administrativa sobre la IA. IX. Conclusiones y reflexiones.

RESUMEN: El presente artículo propone una visión panorámica de derecho penal y civil mexicano, aplicable ante el hipotético caso en que un vehículo de conducción automatizada por medio de inteligencia artificial esté involucrado en un evento que genere muerte, lesiones o daños. Se abordan diversos conceptos de la tónica jurídica mexicana aplicable, principalmente de fuente normativa y doctrinal, así como su interacción con el elemento de inteligencia artificial. También presentamos algunas

reflexiones sobre el impacto que implica la ausencia de regulación de carácter administrativo en materia de inteligencia artificial. Por último, hacemos un somero ejercicio legal donde evidenciamos la metodología en la cadena de responsabilidad, y concluimos quien podría ser responsable por un evento dañoso derivado de la intervención de un vehículo automotor de conducción automatizada, por medio de la inteligencia artificial en México.

ABSTRACT: This article presents an overview of the Mexican criminal and civil law applicable to the hypothetical case in which an artificial intelligence-driven vehicle participates in an event that causes death, injury, or damage. The discussion covers various concepts of the applicable Mexican legal topic, primarily from normative and doctrinal sources and their interaction with the artificial intelligence element. We also present some reflections on the impact of the absence of administrative regulation on artificial intelligence. We then conduct a brief legal exercise to showcase the methodology in the chain of responsibility for identifying the party responsible for a harmful event caused by the intervention of an automated motor vehicle using artificial intelligence in Mexico.

PALABRAS CLAVE: *Inteligencia artificial, vehículo de conducción automatizada, derecho penal mexicano, derecho civil mexicano, regulación administrativa, persona moral o jurídica, responsabilidad penal de la persona moral o jurídica, imputación objetiva, atipicidad, responsabilidad penal.*

KEYWORDS: *Artificial intelligence, artificial intelligence-driven vehicle, Mexican criminal law, Mexican civil law, administrative regulation, legal entity, legal entity's criminal liability, objective imputation, atypicality, criminal liability.*

I.- INTRODUCCIÓN

A raíz de la existencia en el mercado mexicano de los vehículos con modalidad de *conducción automatizada*, es decir, con la capacidad de conducirse a través de los sistemas y programas que emplean inteligencia artificial (IA), prácticamente sin la intervención del conductor humano, surgen incógnitas sobre las consecuencias legales en caso de ocurrir algún siniestro causante de muerte, lesiones o daños materiales.

Ante la novedad del avance tecnológico que representa la operación de software de inteligencia artificial (IA) y la carencia de información divulgada, se ha generalizado un entorno social de desconocimiento y desconfianza ante el fenómeno.

En el presente trabajo nos cuestionamos, bajo la óptica preponderante de las leyes y doctrinas mexicanas, sobre la IA en un vehículo de conducción automatizada, tratando de conceptualizar los elementos fundantes como *persona moral* o *jurídica*, y si la IA encuadra en ellos. Hacemos patente la existencia de los llamados sesgos y fallas de la IA, así como la trascendencia que pueden tener en el devenir de la tecnología. También esbozaremos la normatividad de derecho positivo mexicano, que da respuesta al caso hipotético planteado, principalmente bajo el enfoque penal y civil, con la estrecha interacción del derecho administrativo mexicano. Especularemos sobre los posibles causes legales en materia penal que puede tener un caso, por la integración del elemento IA en la respuesta legal, y finalmente, plantaremos nuestras conclusiones y reflexiones.

II. ¿QUÉ ES LA IA EN UN VEHÍCULO DE CONDUCCIÓN AUTOMATIZADA?

La noción de inteligencia artificial tiene distintas acepciones, depende del enfoque con que se aborde. Bajo la perspectiva tecnológica, tal y como lo conceptualiza Rouhiainein, se refiere a “la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano” (Rouhiainein 2018: 17), se debe considerar siempre la evolución del campo de la ciencia y tecnología para contextualizar adecuadamente.

Sobre las características de la inteligencia artificial (IA), la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) refiere:

“Los sistemas de la Inteligencia Artificial (IA) están diseñados para funcionar con diferentes grados de autonomía, mediante la modelización y representación del conocimiento y la explotación de datos y el cálculo de correlaciones. Pueden incluir varios métodos, como, por ejemplo, aunque no exclusivamente: i) el aprendizaje automático, incluido el aprendizaje profundo y de refuerzo; y ii) el razonamiento automático, incluidas la planificación, la programación, la representación del conocimiento y el razonamiento, la búsqueda y la optimización” (UNESCO 2021: 10).

La doctrina generalizada en el tema¹ distingue entre la *Soft IA* y la *Hard IA*. La *soft IA*, se encuentra presente en los modelos comerciales que conocemos actualmente como las aplicaciones de mapas navegadores, los *autobots*, los *chatbox*, y muchas de las aplicaciones de asistencia administrativa que analizan y procesan millones de datos en segundos; en las comunicaciones y de manera más compleja, contamos a las aplicaciones comerciales de conducción asistida o autónoma de los vehículos automotores (que es la que nos interesa en el presente artículo).

En cambio, la llamada *Hard IA*, se caracteriza por la pretensión de reproducir las funciones cognitivas de la mente humana, con los grandes retos que conlleva tratar de definir, entender y maquetar la mente de los humanos y las derivaciones éticas que implica.

Resulta oportuno recordar la distinción básica entre el llamado *Software* y *Hardware*. El *Hardware*, son los medios tangibles donde puede reposar cualquier programa tecnológico, es decir el receptáculo del robot; el *software* se refiere a los intangibles de la programación que se basa en comandos e instrucciones binarias como punto de partida y va elevando la complejidad de sus algoritmos y funciones.

Cuando nos referimos al concepto de IA, aplicado a un vehículo de conducción automatizada en México, lo encuadramos dentro de los modelos de la *Soft IA*.

¹ Véase Minsky 2016, Simon 1969 y McCarty.

III. SEGOS Y FALLAS DE LA IA

Como toda creación humana que incide en la revolución científica, encontramos que la IA enfrenta grandes desafíos en su desarrollo, comenzando por el propio devenir tecnológico y falsación de sus principios. También se presentan retos en el uso práctico social.

Estos retos en la aplicación social de la IA, involucra una narrativa contemporánea que se enfoca en las implicaciones éticas del uso de la IA. No es posible cuantificar el impacto real que tendrá el uso de la IA en áreas económicas, laborales, ambientales, bélicas, de seguridad pública, seguridad humana, por mencionar algunas, pero se debe hacer un gran esfuerzo global para mitigar lo que a primera vista pudiese representar un riesgo a la sociedad.

En este sentido, se advierte un entorno legal desfasado en México, ya que las fuentes que informan el orden jurídico doméstico y dan directrices, son fuentes del derecho internacional.

Muchos de los esfuerzos regulatorios sobre la IA, hacen énfasis en los llamados *usos éticos* de la misma, y establecen una serie de principios que servirán como base del desarrollo jurídico subsecuente. A manera de ejemplo destaca lo declarado por la Unión Europea que señala siete requisitos esenciales para que la IA pueda ser considerada fiable:

- “1.- Intervención y supervisión humanas
- 2.- Solidez y seguridad tecnológica
- 3.- Privacidad y gestión de datos
- 4.- Transparencia
- 5.- Diversidad, no discriminación y equidad
- 6.- Bienestar social y medioambiental
- 7.- Rendición de cuentas” (Comisión Europea sobre IA 2019: 4).

Todas las directrices deben adoptar un enfoque basado en el impacto social que tiene la IA en el campo específico, pues se debe ponderar las áreas de aplicación para hacer énfasis de la legislación, pues no resulta de igual trascendencia la salud o las funciones de seguridad nacional, que las funciones de ocio y entretenimiento. En el rubro atinente a la intervención y supervisión humana, resalta que “los sistemas de IA deben ayudar a las personas a elegir mejor y con más conocimiento de causa en función de sus objetivos” (Comisión Europea sobre IA 2019: 4). Se deben garantizar medidas de control adecuadas como la adaptabilidad, exactitud y explicabilidad de los sistemas de inteligencia artificial.

La Unión Europea señala que “la supervisión debe lograrse a través de mecanismos de gobernanza, tales como el enfoque de la participación humana (*human in the loop*), la supervisión humana (*human on the loop*), o el control humano (*human in comand*)” (Comisión Europea sobre IA 2019: 5). Se debe de garantizar siempre que las autoridades públicas tengan la capacidad para ejercer sus funciones de supervisión.

Siguiendo a la Unión Europea en el tema de seguridad, encontramos que “además, los mecanismos de IA deben integrar mecanismos de seguridad desde su diseño” (Comisión Europea sobre IA 2019: 5) para garantizar su desempeño seguro, salvaguardando la seguridad física y psicológica de los usuarios, esto incluye la minimización y reversibilidad de las consecuencias no deseadas (Comisión Europea sobre IA 2019).

Bajo el tamiz jurídico, resulta útil identificar las fuentes del derecho con relación a la IA, pero adelantamos que, en el caso de México, es prácticamente nula la regulación *ex profeso* sobre el tema, por lo que la consecuencia natural es encuadrar el hecho legal al marco jurídico ya existente, en observancia siempre de los derechos humanos al acceso a la justicia, debido proceso y respeto a los principios de legalidad, taxatividad y certeza.

En México, actualmente no existe ordenamiento jurídico en rango de ley, que se avoque al tema de la IA, no obstante, en el seno de la Cámara de Diputados del Congreso de la Unión, se presentó sin prosperar, una propuesta de *Ley para la Regulación Ética de la Inteligencia Artificial y la Robótica*², donde se busca establecer: i) Lineamientos para las políticas públicas en materia de IA; ii) La creación de Normas Oficiales Mexicanas en la materia; iii) Regular el uso de la IA y la robótica en su uso público gubernamental, financiero, comercial, administrativo y económico; y iv) Crear un “Consejo Mexicano de Ética para la IA”.

Llama la atención la definición que esta propuesta de ley establece sobre la IA, al señalar que se trata de “La disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico” (art. 2), también define a la robótica como “La técnica que aplica la informática al diseño y empleo de aparatos, que, en sustitución de personas, realizan operaciones o trabajos específicos” (art. 2). Insistimos que se trata simplemente de una propuesta de decreto de ley que no ha prosperado y evidencia las grandes áreas de oportunidad que existen en la regulación de la materia en México.

Sobre los sesgos en la IA, la exposición de motivos de esta propuesta de ley señala que:

“Uno de los elementos más problemáticos relativos al creciente uso de la ML, y su vínculo con principios y garantías constitucionales, lo encontramos en los sesgos que pueden derivar en discriminación, en contraste con el contenido del párrafo quinto del artículo 1º de nuestra Carta Magna. Las aplicaciones de la ML tienden a ser ofertadas y aceptadas como neutras ignorando que su operación depende de la visceralidad humana y de la hegemonía normativa ejercida por los países líderes en el desarrollo de estas tecnologías” (Loyola Vera 2023).

² Loyola Vera 2023

Existen sesgos en las bases de datos y arquitecturas de procesamiento, y es ahí donde se encuentran los orígenes de las asimetrías actuales. El factor humano es el que genera los sesgos, además existen *espacios vedados*, que no permiten el acceso a la información sobre la creación, operación y funcionamiento de los programas de IA. El llamado *velo corporativo* y los *derechos de propiedad industrial e intelectual* son los frenos que encubren los sesgos y generan inequidades a la sociedad.

Mucha de la literatura actual³ que analiza el tema de los sesgos presentes en la elaboración y uso de programas de IA, se orientan a cuestiones financieras, calificación de créditos bancarios, programas de vigilancia e identificación en materia de seguridad pública, así como el auxilio en la función judicial para gradar las sanciones en procesos criminales⁴.

Otro esfuerzo por regular la IA en nuestro país, se advierte en la “Iniciativa con proyecto de decreto por el que se expide Ley Federal que Regula la Inteligencia Artificial”, (Iniciativa IA)⁵ propuesta por el Senador Ricardo Monreal, el pasado 27 de febrero de 2024. En un punto relevante de dicha iniciativa, establece un enfoque regulatorio de los sistemas de inteligencias artificial (SIA), y siguiendo la metodología de la Unión Europea (UE), atendiendo al nivel de *Riesgo* que se pueda generar según la tecnología y el uso que se haga de ella, haciendo una clasificación de estos niveles en riesgo *inaceptable, alto y bajo*.⁶

La Iniciativa IA, considera como *riesgo inaceptable* a los efectos que son consecuencia del uso de “aquellos sistemas que mediante el uso de la IA, puedan provocar o provoquen graves perjuicios físicos o psicológicos a las personas, así como el uso de los mismos para la identificación biométrica”. Considera como de *riesgo alto*,

³ Op. Cit. Minsky, Simon y McCarty.

⁴ Véase el caso *State vs. Loomis en el Estado de Wisconsin, E.U.A.(2013)*, sobre el uso de la inteligencia artificial aplicado a las funciones judiciales.

⁵ Monreal Ávila 2024.

⁶ Iniciativa IA, artículos 7 a 10.

a aquellos sistemas de IA que puedan causar un perjuicio a la salud o la seguridad de las personas, o que produzcan un daño o menoscabo de los Derechos Humanos. Finalmente, considera como *riesgo bajo* a los SIA que durante su utilización no implican un peligro a la salud o la seguridad de las personas, o un daño o menoscabo de los Derechos Humanos.

Desprendido del contenido de la iniciativa de IA, podemos advertir que se hace un énfasis en la figura del *riesgo*, que bajo un aspecto jurídico normativo, una vez desarrollados a detalles tales conceptos con tintes tecnológicos podrán dotar de contenido legal específico al *deber de cuidado* que deberán observar los destinatarios de la regulación.

Por último, esta iniciativa de IA, define a los sistemas de inteligencia artificial (SIA), como “aquellos que implican el uso y aprovechamiento de las tecnologías de la información para crear programas informáticos capaces de ejecutar cálculos, operaciones, investigaciones, o razonamientos comparables con los que realiza la mente humana”.⁷

El desarrollador de IA, y productor de una de las líneas líderes en el mercado mundial de vehículos de conducción automatizada (VCA), Tesla, conceptualiza a sus *algoritmos de tecnología*, como:

“los algoritmos básicos que hacen funcionar el vehículo mediante la creación de una representación del mundo de alta fidelidad y planificación de trayectorias en ese espacio. Para entrenar las redes neuronales en la predicción de dichas representaciones, creamos de forma algorítmica, datos de verdad en tierra firme precisos y de gran escala a través de la combinación de información de los sensores del auto en el espacio y en el tiempo. Utilizamos técnicas de última generación para crear un sistema robusto de planificación y toma de decisiones que funcione en situaciones

⁷ Iniciativa IA, artículo 4, fracción V.

reales complicadas en situaciones de incertidumbre. Evaluamos los algoritmos a la escala de toda la flotilla”⁸

Algunas probables fallas en la operación de programas de IA se evidencian en accidentes automovilísticos donde la conducción se llevó por medio de funciones de manejo automatizado o autónomo, llamadas comercialmente *Autopilot*, los VCA.

Usuarios de internet⁹ llevan registros públicos de accidentes con desenlaces graves o fatales a los involucrados, acontecidos mayoritariamente en Estados Unidos de Norteamérica, el número de eventos documentados es de aproximadamente trescientos, sin considerar la cifra negra. Tales eventos se configuran bajo los siguientes rubros:

- 1) Colisión de dos automóviles, donde uno de ellos es un VCA.
- 2) Vehículos que golpean al VCA.
- 3) VCA mata a motociclista.
- 4) VCA conduce a través de una intersección al bosque.
- 5) VCA conduce a un edificio de oficinas y choca.
- 6) VCA mata a un peatón.
- 7) VCA golpea una banqueta o poste.
- 8) VCA se estrella contra un árbol.
- 9) VCA conduce fuera de la carretera.
- 10) VCA conduce hacia un estanque.

⁸ En el sitio de internet del fabricante Tesla, se publicita una serie de productos y especificaciones de funcionamiento de sus productos de IA y robótica. Disponible en: «https://www.tesla.com/es_mx/AI». Consultado el 1 de junio de 2024.

⁹ Para mejor referencia de la estadística citada, véase «<https://www.tesladeaths.com/es/>», Consultado el 30 de mayo de 2024.

11) VCA choca contra la barrera de la carretera

12) VCA mata a ciclista.

IV. NORMATIVIDAD DE DERECHO POSITIVO MEXICANO QUE DA RESPUESTA AL CASO HIPOTÉTICO PLANTEADO

1. El concepto jurídico de persona

El primer elemento que debemos destacar para abordar nuestro caso es el de la persona.

El derecho mexicano hace una distinción entre las personas físicas y las personas morales o jurídicas. No existe duda sobre el concepto de persona física en su dimensión humana, por lo que nos avocaremos a reflexionar sobre el concepto de las personas jurídicas y la incidencia de la IA.

Desde la doctrina del derecho civil, entendemos a la persona como “un sujeto de derechos y obligaciones”, se trata de un ente que detenta el vínculo legal con sus dos extremos: las obligaciones que tiene (se encuentra compelido a realizar conductas en términos de la ley) y la parte correlativa de derechos (reflejado en su capacidad legal de exigir a otros, ciertas conductas a su favor).

La persona es el eje de la construcción jurídica social, por lo que partimos del principio que *las cosas* no responden ante la ley, son siempre *las personas* quienes lo hacen.

2. Las personas jurídicas

Como elemento de la tónica jurídica mexicana, encontramos que las personas jurídicas o cualquier centro de imputación normativa deviene del reconocimiento que hace la ley al mismo, es decir, existe una fuente normativa para su creación o reconocimiento. Por consiguiente, son personas jurídicas aquellas a las que la ley les reconoce ese carácter.

El Código Civil aplicable en la Ciudad de México¹⁰ establece un catálogo que reconoce como personas jurídicas o morales a la nación, los estados y los municipios; las corporaciones públicas reconocidas como tales por la ley; las sociedades civiles o mercantiles; los sindicatos, asociaciones de profesionistas y confederaciones laborales; y básicamente cualquier asociación con fines políticos, científicos, artísticos, de recreo o cualquiera que tenga un fin lícito y que no sean desconocidas por la ley. Se establece que las personas morales pueden ejercer todos los derechos que sean necesarios para realizar el objeto de su institución.

La IA, al tratarse jurídicamente de *una cosa o un bien*, forma parte del patrimonio de alguna persona (pública o privada), por lo que existe jurídicamente siempre un titular; y debemos recordar que la condición de propietario o poseedor de un bien le genera al titular una serie de obligaciones por el simple hecho de serlo, es decir, la calidad de propietario o poseedor engendra obligaciones legales.

En el caso de la IA, bajo las reglas del derecho civil debemos identificar al titular responsable de ella, y esto lo podemos hacer mediante las reglas de propiedad común, propiedad industrial y propiedad intelectual o atendiendo a las causas generadoras de esa calidad, como puede ser la ley o los contratos.

3. *Temperamentos legales*

En derecho mexicano existen algunas *figuras o situaciones*, que formalmente no tienen el reconocimiento de persona jurídica, bajo la perspectiva de las doctrinas de derecho civil de la persona y sus atributos, pero encuentran un trato equiparado, en razón de los efectos que producen en la sociedad.

En nuestro país, el Código Civil de la Ciudad de México y la mayoría de los códigos de las entidades federativas, así como el Có-

¹⁰ Código Civil para el Distrito Federal. Artículos 25, 26, 27, 1798, 1910, 1913, 1916 y 1917.

digo Federal, abrevan las doctrinas fundantes de las teorías clásicas francesas y las italo-alemanas¹¹ sobre el tema de las *personas* y *bienes*, y del *hecho y acto jurídico*, pero diversos ordenamientos en materia tributaria o administrativa, reconocen algunas consecuencias de actos jurídicos, y en la práctica por ficción legal, se les da un tratamiento similar a las de la persona jurídica.

Dicho trato legal se enfoca en *los efectos o consecuencias* de carácter económico o tributario con fines recaudatorios, y no a sus causas. Se debe destacar que tales regímenes excepcionales rompen con la dogmática en materia civil y las teorías de la persona y personalidad tradicionales.

Tal es el caso de las *masas hereditarias*, algunas figuras tributarias, los fideicomisos, y todos aquellos que la doctrina llama *patrimonios de afectación*; otro ejemplo lo encontramos en el caso de las empresas o negocios, que sin contar con el formalismo legal de constituirse como persona jurídica ante las autoridades correspondientes, llevan a cabo actos materialmente propios de una persona jurídica o empresa, que aparejan efectos y consecuencias legales, se trata de las llamadas *sociedades irregulares*.

Doctrinalmente estas situaciones excepcionales, formalmente no constituyen una ampliación o distorsión en la institución jurídica clásica de la persona, bajo la dogmática del derecho civil mexicano.

4. Repensar el concepto de persona moral o jurídica

Si continuamos con esta lógica normativista que reconoce o crea situaciones jurídicas por efecto de la ley, llegamos al punto donde se evidencia la necesidad de reflexionar la incidencia de los adelantos tecnológicos, para reconocer y dar certeza jurídica a la nueva realidad presente en la sociedad como consecuencia de la interac-

¹¹ A mayor abundamiento véase la exposición de motivos del código Civil para el Distrito Federal. Publicado en el *Diario Oficial de la Federación* del 14 de enero de 1928.

ción de robots y vehículos que cuentan con conducción autónoma por medio de la inteligencia artificial.

Debemos destacar que la ley, es una de las llamadas fuentes formales del derecho, pero los adelantos de la ciencia y los hechos sociales constituyen parte de las llamadas fuentes reales del derecho.

Así las cosas, podríamos considerar algunos motivos para redefinir el concepto de persona moral o jurídica, en derecho mexicano:

- 1) Bajo una perspectiva pragmática y de política pública, el contar con una regulación expresa que confiera certeza jurídica sobre el concepto, alcance, funcionamiento y consecuencias, a las partes involucradas en una interacción con IA. Es decir, definir la masa de derechos y obligaciones de forma expresa y sencilla.
- 2) Desde el punto de vista de la teoría general de las obligaciones, el derecho y la obligación son contrapartes del vínculo jurídico de los entes sociales, lo que se traduce en la respuesta elemental de ¿qué puedo exigir? y ¿a qué estoy obligado a cumplir cuando tengo una contraparte que involucre a la IA? De nuevo, se lograría hacer explícitos los alcances del derecho y la obligación de las partes involucradas.
- 3) Bajo la perspectiva *ius filosófica* y política, que reconozca que las fuentes reales del derecho como lo es el adelanto de la tecnología y las interacciones humanas que emplean estas herramientas de inteligencia artificial, son el motivo suficiente para replantearnos los conceptos de persona y reconocer nuestra nueva realidad social, en razón de interacciones cada vez más complejas.

5. Un punto medio

La postura intermedia entre la rigidez de la actual dogmática civil mexicana, y la postura que se decante por la reformulación del concepto de persona jurídica o moral, puede echar mano de la actual tónica jurídica mexicana, por medio de los contratos de garantía que ya contempla la legislación (como una fianza, prenda,

hipoteca, obligación solidaria, etc.), que sirvan como respaldo y certeza ante las interacciones de la persona física con las personas jurídicas que involucre la IA.

La ruta por seguir sería constituir un *seguro de daños o riesgo*, que sea siempre concomitante a la interacción con funciones de IA, donde el titular patrimonial de la IA o un tercero, respondan por cualquier daño o perjuicio causado, el cual debería ser contemplado en la ley de manera específica.

Existe mucho margen de maniobra legal que puede activarse por medio de las cláusulas convenidas *inter partes*, o por medio de la disposición legal que pueda promulgarse al efecto.

6. El aparato legal mexicano

Hasta ahora, hemos establecido: i) Que la IA es una cosa, a la luz del derecho mexicano; ii) Que las cosas forman parte del patrimonio de las personas (públicas o privadas). En adelante analizaremos la responsabilidad que tienen las personas, por la afectación que puedan generar las omisiones del cuidado sobre sus cosas, específicamente la IA (con su carácter de cosa). Actualmente existen un par de materias legales que pueden dar respuesta a un caso hipotético, donde un vehículo automotor conducido de forma automatizada o autónoma genere lesiones, muerte o daño a la propiedad, se trata de la materia civil y penal.

Partiremos del axioma legal de que “las personas responden por sus actos (acciones u omisiones)”.

Cuando hablamos de actos, nos referimos a conductas, que pueden configurarse por acción (hacer algo) u omisión (no hacer, cuando se debería), en particular abordaremos aspectos de la llamada *teoría de los daños*, que de forma transversal se encuentra presente en la dogmática penal y civil.

Sobre la base orgánica donde reposan los procedimientos judiciales en la Ciudad de México, encontramos competencia en térmi-

nos de la Ley Orgánica del Poder Judicial de la Ciudad de México¹², donde se establece que el Poder Judicial cuenta con juzgados competentes en materia civil y penal.

7. La perspectiva del derecho civil mexicano

El Código Civil para el Distrito Federal regula la institución de la responsabilidad civil extracontractual de la siguiente manera:

“Cuando una persona hace uso de mecanismos, instrumentos, aparatos, vehículos automotores o sustancias peligrosas por sí mismos, por la velocidad que desarrollen, por su naturaleza explosiva o inflamable, por la energía de la corriente eléctrica que conduzcan o por otras causas análogas, está obligada a responder del daño que cause, aunque no obre ilícitamente, a no ser que demuestre que ese daño se produjo por culpa o negligencia inexcusable de la víctima. En todos los casos, el propietario de los mecanismos, instrumentos, aparatos, vehículos automotores o sustancias peligrosas, será responsable solidario de los daños causados” (art. 1913).

Y también dispone que las personas que causan un daño común deben responder solidariamente por la reparación a la víctima.

Otra institución jurídica que se contempla en la legislación civil, que puede dar pie a una acción judicial es la del daño moral causado. El Código Civil para el Distrito Federal lo concibe así: “Por daño moral se entiende la afectación que una persona sufre en sus sentimientos, afectos, creencias, decoro, honor, reputación, vida privada, configuración y aspectos físicos, o bien en la consideración que de sí misma tienen los demás. Se presumirá que hubo daño moral cuando se vulnere o menoscabe ilegítimamente la libertad o la integridad física o psíquica de las personas” (art. 1916).

¹² Véase la Ley Orgánica del Poder Judicial de la Ciudad de México, artículos 1, 32, 58, 59 y 60.

8. *La perspectiva del derecho penal mexicano*

La legislación penal de la Ciudad de México¹³, establece una serie de principios dogmáticos que resulta indispensable considerar para nuestro análisis:

- 1) Principio de prohibición de la responsabilidad objetiva. Es decir, siempre tiene que haber una acción a sancionar por aparte de la persona.
- 2) Principio del bien jurídico y de la antijuridicidad material. Implica que se debe lesionar o poner en peligro un bien jurídico, sin causa justa.
- 3) Principio de culpabilidad. Siempre se debe acreditar la culpabilidad del sujeto para sancionarlo por un hecho, y su pena será proporcional a su culpa.
- 4) Principio de acto. Es decir, que los delitos solo se pueden realizar por acción u omisión.
- 5) Omisión impropia o comisión por omisión. Se atribuye el resultado típico que se produzca, a quien omita impedirlo cuando se tenga la calidad de garante. Es procedente en los casos de delitos de resultado material.
- 6) La forma de ejecución del delito (instantáneo, continuo o continuado).
- 7) Dolo y culpa.
- 8) Formas de autoría y participación. Se prevén las siguientes hipótesis: i) Lo realicen por sí; ii) Lo realicen conjuntamente con otros; iii) Instrumentalicen a otro; iv) Determinen al autor a cometerlo; v) Presten auxilio para la comisión; o vi) Auxilien al autor con posterioridad al hecho.

¹³ Código Penal para el Distrito Federal, artículos 3 a 5, 15 a 18 y 22.

La base de nuestro sistema penal, gira en torno al concepto de *bien jurídico*.¹⁴

9. La responsabilidad penal de la persona moral o jurídica

El Código Penal para el Distrito Federal sanciona la responsabilidad de la persona jurídica, es decir, en nuestro caso hipotético puede derivar en responsabilidad de una persona física o de una persona moral, de la siguiente manera:

“Artículo 27 Bis.- (Responsabilidad Penal de una Persona Moral o Jurídica)

I.- Las personas morales o jurídicas serán responsables penalmente de los delitos dolosos o culposos, y en su caso, de la tentativa de los primeros, de todos los delitos previstos en el Código Penal, cuando:

a). - Sean cometidos en su nombre, por su cuenta, en su provecho o exclusivo beneficio, por sus representantes legales y/o administradores de hecho o de derecho; o

b). - Las personas sometidas a la autoridad de las personas físicas, realicen un hecho que la ley señale como delito por no haberse ejercido sobre ellas el debido control que corresponda al ámbito organizacional que deba atenderse según las circunstancias del caso, y la conducta se realice con motivo de actividades sociales, por cuenta, provecho o exclusivo beneficio de la persona moral o jurídica”.

Cuando la empresa, organización, grupo o cualquier otra clase de entidad o agrupación de personas no queden incluidas en los incisos a) y b), por carecer de personalidad jurídica y hubiesen cometido un delito en el seno, con la colaboración, a través o por medio de la persona moral o jurídica, el Juez o Tribunal podrá aplicarles las sanciones de i) Suspensión; ii) Prohibición de realizar actividades o negocios; iii) Intervención de la persona moral; iv)

¹⁴ Véase Constitución Política de los Estados Unidos Mexicanos, art. 22.

Clausura; v) Retiro de mobiliario urbano; y vi) Inhabilitación para obtener beneficios fiscales o sociales.

La legislación penal también tipifica los delitos de homicidio, lesiones y daño a la propiedad¹⁵.

V. LA TEORÍA DE LA IMPUTACIÓN OBJETIVA

En términos llanos, la imputación objetiva es un mecanismo para atribuir un resultado a una acción, cuando se ha creado un riesgo no permitido como consecuencia de esa acción, y se establece con claridad el nexo de causalidad.

Esta herramienta es clave para el adecuado procesamiento judicial en materia penal y civil, en búsqueda de la respuesta al caso hipotético que analizamos. Desde el enfoque civil, se extiende bajo las *teorías de los daños* y permite acreditar los extremos de la acción judicial civil.

Desde el enfoque penal, permite exponer los elementos objetivos del tipo penal (conducta), daño y nexo causal.

Sobre la teoría de la imputación objetiva, Carlos Daza refiere que la teoría actual se centra entre el finalismo y el funcionalismo y considera que no hay contradicción entre tipo penal de lo injusto e imputación objetiva; sino que la imputación objetiva es un regulador del tipo objetivo. Piensa que es necesario la teoría de la imputación objetiva, para fundamentar las diversas hipótesis que se presentan en la teoría causal ya que, sin ellos, no se puede dar solución a determinados casos (Daza Gómez: 1998).

La imputación objetiva, es la herramienta metodológica o mecanismo para atribuir responsabilidad a una persona o empresa. Tiene una base general, que, dependiendo la materia penal o civil, y atendiendo a los principios de legalidad, taxatividad y cer-

¹⁵ Código Penal para el Distrito Federal, artículos 123, 130 y 239.

teza, son contenidos con sus particularidades en las respectivas legislaciones.

Si bien cada legislación especifica los elementos de la imputación objetiva, siguiendo a Yuri López Casal (López Casal: 2016) podemos decir que los siguientes son los rasgos comunes a la figura:

- 1) *El riesgo general de la vida.* Se traduce en que no se puede imputar un daño al procesado, si el daño fue provocado dentro de los márgenes de acción de la propia víctima en su cotidianidad, es decir, si sufre el daño en razón de los riesgos naturales que cualquier persona pudiese sufrir (López Casal: 2016).
- 2) *La prohibición de regreso.* Aquí “se impide retroceder en la cadena causal desde que se verificó una intervención dolosa o negligente de un tercero, estamos ante la irrupción de un nuevo curso causal (conducta humana), en el ya iniciado por la conducta del procesado, que da lugar a un resultado que con aquella conducta o esta actividad no se hubiera alcanzado, o bien, hubiera sido diferente al finalmente acaecido” (López Casal: 2016). Por ejemplo, cuando se trata de atribuir una conducta al chofer de un vehículo, pero se determina que el mecánico que tuvo que poner el líquido de los frenos fue negligente y no lo hizo.
- 3) *Criterio de incremento del riesgo.*

“Esta figura descansa sobre dos ideas:

a) El resultado (daño) causado por el agente solo le es imputable (objetivamente) a la conducta del autor, si ha creado un peligro no cubierto por un riesgo no permitido, y ese peligro se ha realizado en el resultado concreto;

b) Si el resultado (daño) se presenta como realización de un peligro creado por el agente, por regla general es imputable (objetivamente). Pero excepcionalmente puede desaparecer la imputación si el alcance de la norma no abarca la evitación de tales peligros y repercusiones. Se plantea así una conexión entre este criterio de ‘incremento del riesgo’ y el siguiente del ‘ámbito de protección

de la norma fundamentadora de la responsabilidad” (López Casal: 2016: 130).

Para efecto de nuestro estudio, nos interesa retomar dos aspectos sobre *el incremento del riesgo*, a saber:

“La imputación en los casos de riesgo permitido. “Existen ciertas actividades realizadas por el humano, que, a pesar de que entrañan un peligro en si mismas, sin embargo, su realización es permitida por el legislador. En el ámbito de derecho de daños, estas actividades pueden encuadrarse, con carácter general, dentro del sector que ha venido a denominarse ‘responsabilidad objetiva’. Se trata de actividades que crean un riesgo jurídicamente relevante, pero que de modo general están permitidas” (López Casal 2016: 131). Un ejemplo es precisamente la conducción de vehículos en vías públicas.

“De este modo, en la responsabilidad objetiva, la imputación objetiva se establece por el simple desarrollo de la actividad que crea un riesgo jurídicamente relevante, aunque permitido” (López Casal 2016: 131).

La imputación en los casos de realización del riesgo no permitido. El incremento del riesgo. “La imputación objetiva presupone que en el resultado (daño) se haya realizado precisamente el riesgo no permitido creado por el posible responsable” (López Casal 2016: 131).

El procedimiento por seguir es: “Examínese qué conducta no se le hubiera podido imputar al autor según el principio del riesgo permitido como infracción del deber, compárese con ella la forma de actuar del procesado, y compruébese entonces si en la configuración de los hechos a enjuiciar la conducta incorrecta del autor ha incrementado la probabilidad de producción del resultado en comparación con el riesgo permitido” (López Casal 2016).

“De ello resulta que si la conducta del dañante ha incrementado la probabilidad de producción del resultado habrá imputación objetiva del resultado. Por el contrario, si no hay aumento del riesgo, habrá que negarse la imputación objetiva del mismo” (López Casal 2016).

VI. POSIBLES CAUSES QUE PUEDE TENER EL CASO BAJO EL ENFOQUE PENAL POR LA INTEGRACIÓN DEL ELEMENTO IA EN LA RESPUESTA LEGAL

El hecho genérico que analizamos se trata de un percance donde está involucrado un vehículo de conducción automatizada (CVA) por medio de IA, y genera como resultado muerte, lesiones o daño material.

Sobre los posibles causes y resultados legales en México, debemos considerar los siguientes aspectos:

- 1) *Casuística*. Como todo caso sometido a análisis jurídico y escrutinio de la autoridad judicial, se debe atender a los hechos en específico.
- 2) *Contradictorio*. La respuesta jurídica dependerá de la teoría del caso de la fiscalía o parte acusadora y del planteamiento de la teoría del caso de la defensa, para generar el resultado dialéctico.
- 3) *Los componentes no jurídicos*. En la práctica, el marco jurídico global, compuesto por la normatividad aplicable al caso (las normas de naturaleza sustantiva, orgánica y procesal), es un elemento sustancial que se mezcla con un par de componentes adicionales que deben ser tomados en consideración para proyectar un resultado jurídico, se trata de:
 - a) *La política pública aplicable al fenómeno social*. Si encuadramos el debate alrededor de la figura de la IA, advertimos que actualmente México no ha posicionado una agenda política clara, para generar la regulación especializada sobre el tema de la IA.

Parece que la reticencia política sobre la regulación de la IA, se entiende desde el pragmatismo político, pues especulamos que la ausencia regulatoria del tema, bajo el *principio de legalidad*, en su vertiente para el gobernado establece que “los particulares pueden hacer todo aquello que no les esté expresamente prohibido o restringido”, da margen a los particulares que desarrollan y operan

la IA para actuar sin restricción legal alguna. Si analizamos este vacío, bajo la óptica garantista de Ferrajoli se prende un foco rojo sobre el actuar de los *poderes salvajes* (Ferrajoli: 2011).

b) *Los aspectos fácticos o materiales.* En este punto nos referimos a los componentes de la realidad que permean las causas judiciales. Por ejemplo, el grado de capacidad y preparación de los operadores jurídicos es determinante en el resultado legal, pues es necesario contar con un perfil que incorpore conocimientos para el entendimiento de las funciones de la IA, y así plantear eficazmente los aspectos probatorios.

También es necesario un entendimiento básico sobre la IA, para hacer de manera adecuada una calificación jurídica de los hechos e invocar los preceptos de carácter técnico y normativo aplicables.

Para encontrar la respuesta legal a nuestro caso, los operadores de la procuración de justicia y la judicatura, juegan un papel fundamental, por ello la manera argumentativa en la calificación jurídica del hecho y el cúmulo probatorio, serán núcleo de la causa penal y consecuentemente de la sentencia.

A manera de ejemplo, encontramos que si inicialmente se plantea una teoría del caso centrada en la responsabilidad directa del conductor (bajo un enfoque dogmático del finalismo penal), la investigación seguramente se estacionará en la violación del deber de cuidado atribuible al conductor bajo los lineamientos del Reglamento de Tránsito aplicable, que impone deberes a los conductores de vehículos automotores, y actualmente no consideran la intervención y particularidades de los vehículos de conducción automatizada autónoma.

Sobre la definición y construcción del elemento *deber de cuidado* que debe observarse durante la operación de la conducción automatizada o autónoma, es importante considerar en el caso mexicano lo dispuesto en las diversas Normas Oficiales Mexicanas (NOM), Normas Mexicanas de Referencia (NMX) y las disposiciones del fa-

bricante sobre el vehículo de conducción automatizada (VCA) plasmadas en los manuales del usuario, avisos del fabricante y cláusulas contenidas en los contratos de adquisición o uso; de igual forma, también debe atenderse lo dispuesto en la normatividad administrativa como las leyes de movilidad y los reglamentos de tránsito.

Al respecto, el Reglamento de Tránsito de la Ciudad de México (RTránsito) dispone la forma en que el conductor debe operar durante la marcha de su vehículo, además resaltamos el hecho que ese cuerpo normativo actualmente no considera como opción, la conducción automatizada o autónoma de vehículos. Sobre el particular, establece:

“Artículo 37. Los conductores y ocupantes de los vehículos deben de cumplir con las disposiciones de seguridad indicadas en el presente artículo de acuerdo a la naturaleza propia de cada vehículo.

(...) Los conductores de vehículos motorizados deben:

a) Sujetar firmemente con ambas manos, el control de dirección y no permitir que otro pasajero lo tome parcial o totalmente;

b) Asegurarse que todos los pasajeros utilicen correctamente el cinturón de seguridad, además de colocarse el propio.

c) Circular con las portezuelas cerradas y, antes de abrirlas, verificar que no se interfiera en el flujo de peatones u otros vehículos; en su caso, no la mantendrán abierta por mayor tiempo que el estrictamente necesario para su ascenso o descenso;

d) Encender las luces cuando disminuya sensiblemente la visibilidad por cualquier factor ambiental o por las características de la infraestructura vial, evitando deslumbrar a quienes transitan en sentido opuesto; y

e) Colocar dispositivos de advertencia cuando por caso fortuito o de fuerza mayor se detenga en vías primarias. Si la vía es de doble sentido, los dispositivos de advertencia se colocarán 20 metros atrás del vehículo y 20 metros adelante en el carril opuesto (...).”

Sobre el tema del *conocimiento del riesgo*, hacemos notar que la página de internet de la marca Tesla¹⁶, adelanta a los usuarios una publicación sobre las condiciones de uso de su VCA, en la modalidad de *autopilot*, con la siguiente leyenda:

(...) “Piloto automático: La seguridad vanguardista del piloto automático y sus funciones de conveniencia están diseñadas para ayudarlo con las partes más tediosas de la conducción. El piloto automático presenta nuevas funciones y mejora las existentes para que, con el tiempo, su Tesla sea más seguro y tenga mayores capacidades. El piloto automático permite que su auto maniobre, acelere y frene automáticamente dentro de su carril. Las funciones actuales de piloto automático requieren una supervisión activa del conductor y que no deje que el vehículo sea autónomo”.

VII. UN BREVE EJERCICIO EN MATERIA PENAL

Si hacemos un somero ejercicio de posibles escenarios legales en materia penal, podemos trabajar sobre dos hipótesis básicas: i) La primera, considera que no existe falla alguna que involucre a la IA del VCA; ii) La segunda, parte de la existencia de fallas en el *software* o *hardware* de la IA en el VCA.

Primera hipótesis: El VCA tiene un funcionamiento de *software* y *hardware* óptimo, por lo que el conductor asume el riesgo de la conducción y puesta en operación de la modalidad de conducción automatizada o autónoma.

En esta hipótesis el Código Penal para el Distrito Federal prevé expresamente sobre las formas de comisión del delito:

¹⁶ Para mejor referencia de las condiciones de uso del VCA del fabricante Tesla, véase «https://www.tesla.com/es_MX/autopilot»

“Artículo 16 (Omisión impropia o comisión por omisión). En los delitos de resultado material será atribuible el resultado típico producido a quien omita impedirlo, si éste tenía el deber jurídico de evitarlo, si:

I. Es garante del bien jurídico;

II. De acuerdo con las circunstancias podía evitarlo; y

III. Su inactividad es, en su eficacia, equivalente a la actividad prohibida en el tipo.

Es garante del bien jurídico el que:

a). Aceptó efectivamente su custodia;

b). Voluntariamente formaba parte de una comunidad que afronta peligros de la naturaleza;

c). Con una actividad precedente, culposa o fortuita, generó el peligro para el bien jurídico; o

d). Se halla en una efectiva y concreta posición de custodia de la vida, la salud o integridad corporal de algún miembro de su familia o de su pupilo”.

Segunda hipótesis: El VCA tiene un funcionamiento de software y hardware deficiente, que es la causa de que el conductor no pueda controlar la conducción, aunque lo intente hacer.

En este caso, el dispositivo que establece las formas de comisión del delito mediante la omisión impropia o comisión por omisión debe ser complementado por la figura de la imputación objetiva, haciendo especial énfasis en el concepto del riesgo.

En términos llanos, básicamente se analiza la conducta y el rol que tuvo cada persona involucrada en la creación y operación del *software* y *hardware* que no funcionó adecuadamente y generó un riesgo, que posteriormente causó un daño.

Por lo que hace a las personas que pudiesen tener responsabilidad en la cadena de actos, en el sentido inverso, desde el evento dañoso hasta el momento de la conducta que generó el riesgo, aquí es donde tiene cabida la cadena causal: conductor-personas responsables del mantenimiento del VCA-personas responsables de la creación del VCA, etcétera, incluyendo en esta cadena a las personas que jurídicamente tengan una calidad de garante por disposición de ley o contrato.

Los ingenieros creadores y operadores, así como los responsables del buen funcionamiento del *software* y *hardware* de la IA, serán casuísticamente sujetos al escrutinio, hasta identificar el foco generador del riesgo, en virtud de no haber observado un deber objetivo de cuidado al que estaban obligados.

Respecto la *prohibición de regreso*: Por metodología, una vez analizada la conducta de cada persona, en razón del rol o papel que cumplen en el hecho, se va descartando la responsabilidad de cada uno, hasta encontrar al sujeto responsable de no cumplir su rol (creando el riesgo que genero daño).

VIII. UN POSIBLE CASO DE ATIPICIDAD POR AUSENCIA DE UNA ADECUADA REGULACIÓN ADMINISTRATIVA SOBRE LA IA

En el caso de una eventual falla en el *software* que emplea la IA, si no existe una regulación legal administrativa adecuada sobre su creación, funcionamiento o supervisión (por medio de la ley, reglamento, Norma Oficial Mexicana sobre IA), de la cual se pueda establecer con claridad un *deber objetivo*, podría invocarse la atipicidad de la conducta. Si esto ocurre, la conducta carecería de relevancia penal y sería competencia de la vía judicial civil¹⁷.

¹⁷ Véase la tesis del Poder Judicial de la Federación de México, visible en el Semanario Judicial de la Federación con número de registro digital 2022967 cuyo rubro es: Atipicidad del delito de lesiones culposas. Si no se acredita el elemento

A manera de corolario de la casuística que podemos encontrar, para judicializar las conductas relacionadas con los daños causados cuando se involucra un VCA que emplea IA, hay un par de causas judiciales que se han ventilado en los tribunales de los Estados Unidos de Norteamérica, que han cobrado relevancia mediática y de los cuales se observan los siguientes puntos jurídicos de interés:

1) *Primer caso*¹⁸ La defensa de parte, sostiene en su teoría del caso que el fabricante Tesla, sabía de fallas del *Autopilot* (IA), antes de ocurrir un accidente fatal en 2019.

“La demanda civil alega que el sistema de piloto automático provocó que el modelo 3 del propietario, se desviara de una carretera al este de los Ángeles a 105 kilómetros por hora, y chocara contra una palmera y estallara en llamas todo en cuestión de segundos (...) La demanda, presentada contra Tesla por los pasajeros, acusa a la empresa de haber estado informada de que el piloto automático y otros sistemas de seguridad estaban defectuosos cuando se vendió el vehículo” (Forbes 2023: s/p).

Por su parte, la empresa demandada *Tesla* afirma que el conductor consumió alcohol antes de ponerse al volante y, además, no está claro si el accidente ocurrió durante el funcionamiento del *Autopilot* (Forbes 2023).

Así quedaron establecidas las teorías del caso de la parte acusadora y la defensa, respectivamente

objetivo, consistente en el nexo de riesgo, mejor conocido como imputación normativa del resultado a la conducta, se actualiza la causa relativa prevista en el artículo 29, apartado A, fracción II, del Código Penal para el Distrito Federal, aplicable para la Ciudad de México, En relación con el diverso 405, fracción I, del Código Nacional de Procedimientos Penales; así como el artículo 29, apartado A, fracción II del Código Penal de la Ciudad de México.

¹⁸ Forbes Staff 2023.

2) *Segundo caso*¹⁹ Los involucrados del caso son el sistema *Autopilot* (IA), un *modelo S* de *Tesla*, y una conductora que sufrió un accidente mientras manejaba con la función en modo semiautónomo.

Según la demanda presentada, el sistema *Autopilot*, no funcionó adecuadamente y provocó la desviación del VCA hacia carriles centrales, de forma tan rápida, que no le permitió reaccionar a la conductora.

La empresa demandada señala la responsabilidad de la conductora, por el uso incorrecto del software: “la conductora uso el ‘*Autopilot*’ en las calles de la ciudad, a pesar de que el manual del usuario advierte que no debe hacerse”, además de que “los conductores deben estar listos para tomar el control en cualquier momento” (Prego 2023: s/p).

El jurado del caso se convenció de que “el software de conducción parcialmente automatizada no era un sistema autopilotado, por lo que se achaca lo ocurrido a una distracción de la conductora” (Prego 2023).

De igual manera, así se establecieron las teorías del caso de la parte acusadora y la defensa, respectivamente.

IX. CONCLUSIONES Y REFLEXIONES

- 1) Actualmente el adelanto de la tecnología en materia de IA, es una realidad que no debe soslayarse por el legislador.
- 2) El marco penal y civil en México, es suficiente para abordar las contingencias del caso que involucra a la IA en un hecho dañoso, pero el área de oportunidad para el legislador, se encuentra en la creación de regulación administrativa especializada.

¹⁹ Para un mayor abundamiento sobre el hecho, diríjase a Prego 2023.

- 3) Es necesario contar con los elementos legales normativos detallados, que den cuerpo a los elementos de responsabilidad penal específica, sobre los deberes objetivos de cuidado exigidos a cada interviniente.
- 4) Las reglas de trato ético sobre la IA son un buen marco referencial de inicio, pero urge detallar la aplicación práctica de la IA en la sociedad mexicana.
- 5) En el caso de México, la ley penal prevé la posibilidad de imponer condena penal a las personas jurídicas, materializándose la reparación integral del daño impuesto a la empresa responsable. Se trata de una figura poco empleada en México que es perfectible, pero la posibilidad legal existe.
- 6) En el caso mexicano, es un mito que, un vehículo de conducción automatizada o autónoma, mediante la inteligencia artificial, sea sujeto de alguna responsabilidad legal.
- 7) La persona (física y la jurídica eventualmente) que viole su deber objetivo de cuidado, respecto la operación de un vehículo de conducción automatizada o autónoma, mediante la inteligencia artificial que cause muerte, lesiones o daños en México será responsable legalmente.

BIBLIOGRAFÍA

- Beltrán Román, Víctor *et al.* (2020): “Inteligencia Artificial en el sistema de justicia criminal: Algunas reflexiones sobre su aplicación en el derecho chileno”, en *Revista Derecho Aplicado LLM UC*. Disponible en: «https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3660095» [Consultado el 10 de febrero de 2024]
- Comisión Europea sobre IA (2019): *Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico*

y Social Europeo y al Comité de las Regiones. Generar confianza en la inteligencia artificial centrada en el ser humano. Bruselas, Bélgica. Disponible en: «<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52019DC0168&from=PT>» [Consultado el 1 de febrero de 2024]

Daza Gómez Carlos (1998): “Teoría de la imputación objetiva”, en *Revista de la Facultad de Derecho. UNAM* N° 219. Disponible en: «<http://historico.juridicas.unam.mx/publica/librev/rev/facdermx/cont/219/dtr/dtr3.pdf>» [Consultado el 1 de febrero de 2024]

Ferrajoli Luigi (2011): “*Los poderes salvajes: La crisis de la democracia constitucional*”, Ed. Minima Trotta. Madrid, España. Disponible en: «<https://www.derechopenalened.com/libros/ferrajoli%20-poderes-salvajes-derecho-penal-en-la-red.pdf>» [Consultado el 7 de febrero de 2024]

Forbes Staff (2023): Tesla sabía de falla del autopilot antes de accidente fatal en EU en 2019: defensa, en *Forbes*, 24 octubre. Disponible en: «<https://www.forbes.com.mx/tesla-sabia-de-falla-del-autopilot-antes-de-accidente-fatal-en-eu-en-2019-defensa/>» [Consultado el 7 de febrero de 2024]

Gómez Liguierre Carlos et.al. (2020): “Responsabilidad por daños causados por la Inteligencia Artificial y otras tecnologías emergentes (Responsabilidad por Inteligencia Artificial y Otras Tecnologías Emergentes)”, en *Revista Crítica de Jurisprudencia Penal InDret, volumen 1*. Disponible en: «https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3535519» [Consultado el 6 de febrero de 2024]

Kuhn Thomas, S. (1962): *La estructura de las Revoluciones Científicas*, en página web de Fondo de Cultura Económica, México. Disponible en: «<https://materiainvestigacion.files.wordpress.com/2016/05/kuhn1971.pdf>» [Consultado el 15 de febrero de 2024]

López Casal Yuri (2016): “La imputación objetiva y sus criterios en el derecho de daños costarricense”, en *Revista Judicial de Costa Rica*, núm. 119, 119-152. Disponible en: «<https://www.corteidh.or.cr/tablas/r35174.pdf>» [Consultado el 1 de febrero de 2024]

Loyola Vera, Ignacio (2023): *Iniciativa con Proyecto de Decreto por el que se expide la Ley para la Regulación Ética de la Inteligencia Artificial y la Robótica (23/05/2023)*, suscrita por el diputado Ignacio Loyola Vera y legisladores integrantes del Grupo Parlamentario del Partido Acción Nacional. Disponible en: «https://infosen.senado.gob.mx/sgsp/gaceta/65/2/2023-05-24-1/assets/documentos/Inic_PAN_Dip_Ignacio_Loyola_inteligencia_artificial.pdf» [Consultado el 1 de febrero de 2024.]

McCarty, John (1979): *Inteligencia artificial. Un estudio general*

Minsky Marvin (2006): *La máquina de las emociones: pensamiento de sentido común, inteligencia artificial y el futuro de la mente humana*, ed. Debate, Argentina. Disponible en: «<https://es.slideshare.net/slideshow/minsky-m-la-mquina-de-las-emociones-ocr-2010pdf/265357760>» [Consultado el 1 de febrero de 2024]

Monreal Ávila, Ricardo (2024): *Iniciativa con proyecto de decreto por el que se expide Ley Federal que Regula la Inteligencia Artificial*. Disponible en: «https://ricardomonrealavila.com/wp-content/uploads/2024/02/Inic_Morena_inteligencia_artificial.pdf» [Consultado el 1 de junio de 2024]

Prego, Carlos (2023): “Quién es el responsable de un accidente si se utiliza el sistema Autopilot: Tesla ya tiene una sentencia pionera”, en página web oficial de Xataka, 23 abril. Disponible en: «<https://www.xataka.com/movilidad/esta-mujer-sufrio-grave-accidente-al-usar-su-tesla-autopi->

lot-ahora-sabemos-quien-cae-responsabilidad» [Consultado el 1 de febrero de 2024]

Rouhiainein, Lasse Petteri (2018): *Inteligencia Artificial: 101 cosas que debes saber hoy sobre nuestro futuro*, ed. Alienta, Barcelona. Disponible en: «https://proassetspdlcom.cdnstatics2.com/usuaris/libros_contenido/arxius/40/39307_Inteligencia_artificial.pdf» [Consultado el 10 de febrero de 2024]

Simon, Herbert (1969): *Las ciencias de lo artificial*

UNESCO (2021): *La ética de la inteligencia artificial*. París. Disponible en: «https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa» [Consultado el 1 de febrero de 2024]

UNICEF (2021): *Policy guidance on AI for children*. New York. Disponible en: «<https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf>» [Consultado el 1 de febrero de 2024]

ENSAYOS



En esta sección se contienen textos en los cuales las y los autores analizan, reflexionan, interpretan y evalúan distintas temáticas inmersas en los estudios sobre el género y los derechos humanos desde diversas perspectivas a través del género ensayístico.

Educación para la paz en tiempos de algoritmos y gadgets

MÓNICA ADRIANA MENDOZA GONZÁLEZ
Universidad Nacional Autónoma de México

SUMARIO: I.-Las violencias y discriminación automatizada. II.-Multi-pobrezas... más, pobreza tecnológica. III. Una alternativa, la educación para la paz y el silencio virtual.

Una de las asignaturas pendientes –si no la más importante–, en un mundo actual en crisis de muchos tipos, es la regulación de las relaciones humanas. Este tema, problema siempre presente, ha sido abordado desde la filosofía, la antropología, la sociología y prácticamente todas las ciencias humanas. En la era de la hiperinformación y de los condicionamientos en todos los ámbitos de la vida, resulta importante estudiar el uso de los sistemas complejos de procesamiento de datos, o lo que se ha llamado en la actualidad como *Deep Learning*.

En la travesía que he recorrido para escribir estas líneas, me he encontrado que todos los caminos en el conocimiento y análisis de lo que se ha llamado *inteligencia artificial*, nos llevan necesariamente a la ética, la educación, el derecho y por ende, a poner sobre la mesa el tema de la convivencia entre seres humanos. Ahora no solo necesitamos instrumentos que regulen la convivencia humana, sino instrumentos legales que regulen nuestra relación con las máquinas, ¡vaya lío! La transformación de la subjetividad e intersubjetividad a partir de los condicionamientos en la interrelación con los dispositivos, aplicaciones y software de todo tipo y sus vertiginosos –y por momentos abrumadores– avances, nos impelen a tomar de manera seria, las repercusiones que está teniendo en la vida cotidiana, personal, escolar, laboral y en todos

los ámbitos de nuestra vida. En este contexto, la bioética es uno de los campos que nos pueden ayudar y guiar en la intersección y multidimensionalidad de las relaciones que se tejen y que son un imperativo a atender.

Resulta relevante, y acompaño, lo que nos plantean el filósofo Alfredo Marcos (2024)¹ y el físico José Luis Mateos (2024)² en el sentido de que el título de *inteligencia artificial* no sólo es desacertado y desafortunado, sino que además mueve a una gran confusión de términos. En el campo de la filosofía, delimitar los términos empleados es muy importante, de ahí la importancia de analizar, por un lado, lo que entendemos por *inteligencia*, y por otro, lo que significa *artificial*. Resolver estas interrogantes conduce necesariamente a una reflexión profunda ontológica, ética, epistémica y hermenéutica, pero que también va de la semántica a la pragmática. Es decir, sobre el uso que se hace de estos términos en todos los campos de conocimiento y en las relaciones humanas que tenemos con dichos conceptos.

En este contexto, emplearé el término inteligencia artificial (IA) como sinónimo de cibernética, con la motivación de no alimentar la confusión y expectativas que genera la idea de que las máquinas piensan y toman decisiones. También en el sentido de caracterizar la IA-cibernética, como sistemas tecnológicos de procesamiento de datos, que, por más parámetros que se añadan en las llamadas *redes neuronales profundas*—por billones y trillones en el famoso

¹ Alfredo Marcos en su conferencia “Retos filosóficos de la IA” realizada en la Facultad de Filosofía y Letras de la UNAM, el 1º de febrero de 2024, nos dice que en 1956 en la Conferencia de Darmouth varios investigadores nombraron como inteligencia artificial (IA) a lo que se conocía como cibernética. Disponible en: «https://www.youtube.com/watch?v=n55nv1W_yQc».

² José Luis Mateos, conferencia “La revolución de la inteligencia artificial: de Alan Turing al ChatGPT y la IA generativa” realizada en el Centro de la Complejidad de la UNAM (C3), realizada el 5 de marzo de 2024. Disponible en: «https://www.youtube.com/embed/KV01-aIv7QY?si=na0UCEofbUDMSpR_».

ahora ChatGPT¹–, son finalmente datos que se relacionan dinámicamente para brindar resultados, que aunque sorprendentes, no dejan de ser automatizaciones que atienden parámetros específicos. No profundizaré en el tema en este espacio, las referencias citadas de Marcos y Mateos lo abordan con suficiencia teórica, solo enfatizaré en que no se puede predicar de las máquinas, la capacidad de la inteligencia, pensar y tomar decisiones, tal como se concibe para los seres humanos. Lo que ocupa en el presente escrito, es delinear algunos elementos de reflexión entre la necesidad de educar y formar para la paz en relación con la emergencia de la IA-cibernética.

I.-Las violencias y discriminación automatizada

Ante los nuevos escenarios que la educación formal está viviendo, se encuentran, entre otros, la brecha existente entre un profesorado no versado en el uso de herramientas digitales, interactuando con estudiantes –ya en su mayoría– nativos digitales, sin obviar que aún un sector estudiantil no tiene acceso tecnológico. La temprana exposición de niñas, niños y adolescentes a las tecnologías, les brinda amplio acceso a información de todo tipo, en todo momento y de cualquier lugar del planeta; ante sí, el mundo se abre con todos sus contrastes, desde lo sublime, hasta lo más cruento e ignominioso de la conducta humana. Esta brecha que se vive en la conflictividad entre educadores y educandos por los términos de relación y códigos de comportamiento que juegan en contraste, representa un reto no solo para docentes, sino para la política educativa que necesita atender los problemas del aquí y el ahora y colocar a la educación como posibilidad de acción desde su *thélos*, ante las demandas cotidianas.

Con la pandemia por COVID-19, los gobiernos realizaron en sus sistemas escolares –y México no fue la excepción–, esfuerzos

¹ El 30 de noviembre de 2022 sale a la luz ChatGPT (*Generative Pretrained Transformer*), tecnología de la empresa OpenAI.

importantes para migrar las clases presenciales a los ambientes virtuales, desplegando múltiples recursos para el acceso a las tecnologías y capacitación docente. De este modo, una gran cantidad de profesores/as conocieron y/o ampliaron sus conocimientos y habilidades tecnológicas en el uso de diversas aplicaciones, *gadgets* y plataformas educativas. Sin duda, el proceso de adaptación trajo consigo oportunidades no exploradas en las formas de enseñanza, pero como sabemos, las herramientas pueden ser usadas en un amplio espectro que va de lo más noble a lo indeseable, y, aunque se *bombardeó* a las y los docentes por doquier, de las bondades del uso de herramientas virtuales para la enseñanza y aprendizaje, la realidad fue develando sus verdades.

Otra de las preocupaciones de las y los educadores con la emergencia de las redes sociales y las TIC's, es la transformación de la subjetividad que observan en las y los estudiantes a partir del uso de las tecnologías, sin saber realmente cómo actuar; pues de manera general no se conciben preparados/as para entender las relaciones que se dan en las redes sociales y el uso de diversas herramientas tecnológicas, que viven como si fuera un universo paralelo a la realidad áulica del intercambio personal, y de las formas de aprendizaje de las que fueron herederos/as. En este contexto, se suma a la lista de dificultades de la interacción enseñanza-aprendizaje, una más: la lucha contra los dispositivos electrónicos como distractores en el salón de clase, como medios de reproducción de las violencias, pues los consideran una influencia con efectos perniciosos en el desarrollo de la personalidad de las y los educandos y por ende, en la sana convivencia. En los últimos años, las formas de violencia en donde están involucrados niñas, niños, adolescentes y jóvenes en las que son víctimas pero también reproductores de violencia, son alarmantes. Videos que se hacen virales en un sinfín de redes diversas, en donde se muestran peleas, tortura de animales, infancias armadas, retos virales², ciberacoso, asesinatos, ataques

² Existen retos virales de todo tipo, desde el #A4Waist que conduce a las y los adolescentes a mantenerse en extrema delgadez, hasta los que ponen en riesgo la vida como #comercanela o #hotwaterchallenge hasta el “desaparecer 48 ho-

a compañeros y docentes, entre otras acciones, son una realidad en nuestra sociedad, que deja inermes a madres, padres, responsables de crianza, docentes, directivos y autoridades educativas.

Como botón de muestra de las violencias digitales en un contexto escolar en nuestro país, está el caso de Diego “N”, el primer acusado por violencia digital en términos de la *Ley Olimpia*³. A partir del empleo de la IA-cibernética, la Fiscalía arrestó a un estudiante del Instituto Politécnico Nacional (IPN) acusado de alterar con IA miles de fotos de alumnas y venderlas como contenido íntimo (Aristegui Noticias 2023). En distintos medios de comunicación se señaló que este estudiante podría haber comercializado imágenes de hasta 1,000 mujeres y que le fueron encontradas aproximadamente 50,000 imágenes alteradas.

Desde el 2017 ya se habían denunciado mediáticamente las *Deepfakes* –imágenes y videos alterados con IA-cibernética–, en donde prácticamente se puede *robar* el rostro, la voz y el cuerpo de cualquier persona, para alterarlas al antojo de la persona victimaria. En la red hay un sinnúmero de ejemplos en donde figuras públicas como políticos, personas del espectáculo, youtubers y cualquiera, puede ser objeto de estas acciones.

La brecha digital hace de las suyas y coloca a las personas adultas ante un escenario que por momentos parece desolador. Los dispositivos electrónicos, extensión cuasi corporal, se han convertido en una herramienta de difusión-espectáculo de las múltiples violencias que afectan la vida social. Cada vez más, podemos ser espectadores en tiempo real del sufrimiento humano que daña la vida, no solo de las víctimas directas, sino de todo el entorno comunitario.

ras” o el famoso #ballenaazul que consiste en realizar 50 acciones y la última es el suicidio.

³ Se le nombra Ley Olimpia a un conjunto de reformas realizadas a la Ley General de Acceso de las Mujeres a una Vida Libre de Violencia.

Las violencias y formas de discriminación se han eficientado con su representación algorítmica, expresión de lo humano disfrazado de ciencia y tecnología neutral que se consume en imágenes por miles de millones en todo el mundo. Importante documental en este contexto, es el titulado *Coded Bias*, que pone al descubierto lo que llama el *sesgo codificado* o *prejuicio cifrado*. Algoritmos diseñados para exacerbar formas de exclusión, discriminación y/o violaciones a derechos humanos; como ejemplos presentan los casos de sistemas de reconocimiento facial que identifican más a hombres blancos que a mujeres o a personas afrodescendientes, mismas que son más identificadas como potenciales delincuentes; sistemas que deciden qué docente es apto o no para la enseñanza, sin importar la trayectoria que hayan observado a lo largo de su vida profesional. Estos algoritmos son empleados por los gigantes tecnológicos como Google, Facebook o Amazon que, con su poderío y capacidad de penetración, promueven tendencias de opinión, pero también de acción, a partir de la distorsión de la realidad o de presentar realidades útiles a sus negocios.

Sabemos que niñas, niños, y adolescentes cuentan por la experiencia de vida, con menos herramientas de comprensión de la realidad, siendo así susceptibles de manipulación conductual, que afecta gravemente su realidad, pues su proyecto de vida se modela a partir de la seducción que ejercen las miles de imágenes que reciben día a día.

A partir de las diversas necesidades a atender que surgen del uso de la IA-cibernética, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) tomó el liderazgo para la adopción –en 2021– y publicación –en 2022–, de la *Recomendación sobre la inteligencia artificial*⁴, en donde se parte de la

⁴ “La presente Recomendación trata de las cuestiones éticas relacionadas con el ámbito de la inteligencia artificial en la medida en que competen al mandato de la UNESCO. Aborda la ética de la IA como una reflexión normativa sistemática, basada en un marco integral, global, multicultural y evolutivo de valores, principios y acciones interdependientes, que puede guiar a las sociedades a la hora de afrontar de manera responsable los efectos conocidos y desconocidos de las

existencia de los riesgos que generan los sesgos, ya que pueden impactar la esfera social de maneras graves. En el Preámbulo, encontramos las siguientes consideraciones:

“Reconociendo las repercusiones positivas y negativas profundas y dinámicas de la inteligencia artificial (IA) en las sociedades... en particular en la mente humana, debido en parte a las nuevas formas que en su utilización influye en el pensamiento, las interacciones y la adopción de decisiones de los seres humanos y afecta a la educación, las ciencias sociales y humanas, las ciencias exactas y naturales, la cultura y la comunicación y la información...”

Considerando que las tecnologías de la IA pueden ser de gran utilidad para la humanidad... pero que también suscitan preocupaciones éticas fundamentales... en relación con los sesgos que puedan incorporar y exacerbar, lo que puede llegar a provocar discriminación, desigualdad, brechas digitales y exclusión y suponer una amenaza para la diversidad cultural, social y biológica, así como generar divisiones sociales o económicas; la necesidad de transparencia e inteligibilidad del funcionamiento de los algoritmos y los datos con los que han sido entrenados; y su posible impacto en, entre otros, la dignidad humana, los derechos humanos y las libertades fundamentales, la igualdad de género, la democracia, los procesos sociales, económicos, políticos y culturales, las prácticas científicas y de ingeniería, el bienestar animal y el medio ambiente y los ecosistemas...”

Como vemos, nuestras relaciones intersubjetivas están cada vez más condicionadas por la influencia de la IA-cibernética, desafortunadamente también para dañar la dignidad humana y para enfatizar los racismos, clasismos, sexismos y discriminaciones de todo tipo. Si el mundo adulto se encuentra afectado con estos

tecnologías de la IA en los seres humanos, las sociedades y el medio ambiente y los ecosistemas, y les ofrece una base para aceptar o rechazar las tecnologías de la IA. Considera la ética como una base dinámica para la evaluación y la orientación normativas de las tecnologías de la IA, tomando como referencia la dignidad humana, el bienestar y la prevención de daños y apoyándose en la ética de la ciencia y la tecnología”.

impactos, es más apremiante atenderlos en el caso de niñas, niños y adolescentes, pues estas tecnologías organizan y pueden modelar su conducta y el *deber ser* de sus mentes en formación. Por ello, se requiere de acompañamiento para un uso responsable que les proteja en su integridad física, mental y emocional. Resulta urgente la educación en materia de datos, la alfabetización mediática e informacional con un enfoque de derechos humanos, combatiendo el acceso asimétrico a los conocimientos y las tecnologías. Solo que lo anterior resulta complicado en comunidades en donde el acceso a servicios básicos no es una realidad, por momentos parece que solo la lista de pendientes aumenta con los vertiginosos avances tecnológicos y nos lanza a la indefensión aprendida.

II.-Multipobrezas... más, pobreza tecnológica

La investigadora Ma. Teresa Sirvent⁵, ya nos advertía hace poco más de 25 años, que la pobreza tiene múltiples dimensiones. Plantea que la carencia de recursos económicos nunca viene sola, la acompañan la pobreza de protección, la pobreza de entendimiento y la pobreza política o de participación. Considero que ahora podemos hablar de una pobreza más: la pobreza tecnológica, que implica la exclusión que de suyo conlleva la falta de acceso a dichas tecnologías, pues la brecha de desigualdad aumenta entre quienes cuentan con las herramientas, los dispositivos y conocimientos y quienes carecen de ellas. Alfabetizar y *educar* en la IA-cibernética no solo democratizando su acceso o en su dimensión cognitiva, sino principalmente en su dimensión ética-moral; es decir, formar en el para qué y por qué de su uso, y en las afectaciones e implicaciones individuales y sociales. Esta educación debe promover la responsabilidad, la consciencia social y ambiental, el respeto a la dignidad

⁵ Versión ampliada para Sirvent, Multipobrezas, Violencias y educación, Segundas jornadas de sociología de la UBA de la ponencia sobre Múltiples pobrezas presentado en el Panel de Apertura del Congreso Internacional de Educación celebrado en Buenos Aires el 24, 25 y 26 de julio de 1996. Facultad de Filosofía y Letras UBA.

de las y los demás, la privacidad de datos propios y ajenos. En muchos lugares de nuestro país, la escuela vive entre las multipobrezas y las transformaciones tecnológicas, entre el fango y el holograma.

Pero la primera pregunta que salta a la vista, es cómo hacemos para combatir o contrarrestar, o por lo menos prevenir principalmente en el ámbito educativo, todas las afectaciones que viven los actores de este entorno tan importante y que aún podemos considerar un factor-espacio de protección. Las llamadas *funas* que se dan en redes sociales y se viralizan en comunidades escolares específicas –a veces salen a la luz pública–, toman como rehén a estudiantes, docentes y personas trabajadoras o directivas, convirtiéndolas en centro de burlas, desprecio y exclusión que conduce al aislamiento, depresión, angustia e inclusive al suicidio. La pobreza tecnológica se acompaña de pobreza moral, de pobreza social, de alimentar la cultura del enemigo, del castigo social que impele a todas y todos a reproducir lo que hace la mayoría, al silencio por temor, al rechazo o al aislamiento, entre otras consecuencias.

No se puede predicar de los algoritmos, la moralidad ni la justicia, son automatizaciones que en el mundo actual reflejan algunas prácticas de la condición humana con sus insociables formas de exclusión y que ha facilitado también formas de destrucción. La escuela tiene el reto de promover cambios sustantivos por las mutaciones en las formas de aprender, enseñar, trabajar, evaluar y/o seleccionar al personal docente que estará en contacto directo con el estudiantado. Es importante traer a la memoria que en 1974 la Conferencia General de la UNESCO emitió la *Recomendación sobre la educación para la comprensión, la cooperación y la paz internacionales y la educación relativa a los derechos humanos y las libertades fundamentales* que nos dice en el apartado VII-33 relativo a la Preparación de los educadores:

“(Los estados miembro)... deberían, con ese fin: a) inculcar a los educadores las motivaciones de su acción ulterior: adhesión a la ética de los derechos humanos y al objetivo de cambiar la sociedad a fin de lograr la aplicación práctica de los derechos hu-

manos; comprensión de la unidad fundamental de la humanidad; capacidad para inculcar el aprecio de las riquezas que la diversidad de las culturas puede brindar a cada persona, grupo o nación; b) ofrecer un conocimiento interdisciplinario básico de los problemas mundiales y de los problemas de la cooperación internacional gracias, entre otros medios, a un trabajo dedicado a la solución de esos problemas; c) preparar a los educadores para que participen activamente en la elaboración de programas de educación para la comprensión internacional y de material y equipo de enseñanza, teniendo en cuenta las aspiraciones de los educandos y en estrecha colaboración con ellos...”.

Las escuelas son espacios que representan una gran ventana de oportunidad para sensibilizar en los temas de la conflictividad humana y deben velar por el bienestar de todas las personas que confluyen en ella. Ya la Declaración Universal de Derechos Humanos, particularmente en conformidad con el párrafo 2 del artículo 26, declara:

“La educación tendrá por objeto el pleno desarrollo de la personalidad humana y el fortalecimiento del respeto a los derechos humanos y a las libertades fundamentales; favorecerá la comprensión, la tolerancia y la amistad entre todas las naciones y todos los grupos étnicos o religiosos; y promoverá el desarrollo de las actividades de las Naciones Unidas para el mantenimiento de la paz”.

Es así que los espacios educativos no pueden ni deben aislarse a un tema tan importante en la actualidad, en donde todos los actores educativos se están viendo afectados por el uso de la IA-cibernética, pues en ella se viven, como caja de resonancia, las conflictividades que vienen del exterior y que emergen dificultades de nuevo sello. Es también pertinente retomar la famosa frase de la Constitución de la UNESCO en 1946, que dice: “Puesto que las guerras nacen en la mente de los hombres, es en la mente de los hombres donde deben erigirse los baluartes de la paz”. Las violencias cibernéticas en todas sus manifestaciones añaden

una tarea más a la escuela, encontrar caminos que mitiguen, prevengan o reparen los daños materiales, físicos o psicológicos que se gestan con la virulencia del uso de las tecnologías que emplean niñas, niños, adolescentes y jóvenes. Se puede considerar que la pobreza tecnológica puede alimentar y promover las conductas propias de la estupidez humana, pero a una escala viral.

III. Una alternativa, la educación para la paz y el silencio virtual

Las guías éticas y morales que nos ofrecen los instrumentos nacionales e internacionales que abordan el buen uso y aprovechamiento de la IA-cibernética, así como los principios rectores que arrojan luz sobre los métodos y enfoques de escuelas democráticas y libres de violencias, parece que no son suficientes y/o sólo se quedan en el ámbito declarativo de las buenas intenciones. Las *Alexas*⁶ no pueden ni deben sustituir a las y los docentes. Los desafíos van desde la revisión y actualización de los planes de estudio, a la mejora en la eficiencia operativa para sensibilizar en el uso de las tecnologías, que requiere ir acompañada del pensamiento crítico y creativo, y de una praxis significativa para estudiantes y docentes. La escuela debe de implementar de manera creativa alternativas atractivas para el estudiantado inmerso en la seducción de las pantallas y el espectáculo instantáneo pero permanente, que encuentran en las redes sociales. Abordar los problemas sociales y éticos que les hagan sentido, y que les permita discriminar la información y reconocer las repercusiones individuales y colectivas de sus acciones; huella digital que puede marcar de manera contundente sus vidas presentes y futuras.

En el mundo de la hiperinformación, nos encontramos en una encrucijada digital que por un lado nos obliga a destinar tiempo y recursos de todo tipo a la actualización en el uso de las tecnologías de la IA-cibernética, y por otro, a intentar preservar valo-

⁶ Asistente virtual desarrollado por Amazon.

res y prácticas de convivencia solidaria que vemos poco a poco desdibujarse. Se hace urgente priorizar los por qué y para qué del uso de estas herramientas que también se han integrado a una loca carrera de eficientismo y respuestas rápidas en el rendimiento académico tanto de docentes como de estudiantes; promoviendo la impaciencia y el cúmulo de información sin sentido, pero en la que ya es un imperativo responder en la inmediatez y en el corto plazo. Aprendizaje a la carta. Las nuevas formas de intersubjetividad, conducen a preguntarnos también sobre los fines de la educación y la información, aprendizaje automático pero efímero para la obtención de puntajes y credenciales en el menor tiempo posible y con el menor esfuerzo invertido, *a un clic de distancia*. ChatGPT y las llamadas IA generativas salen al rescate de niñas, niños, jóvenes y adultos para ganar la competencia de tareas escolares, informes, reportes y entregables en tiempo y forma. Si bien, el primer paso para cerrar la brecha digital es la accesibilidad de las tecnologías, el conocimiento y habilidades en el uso y manejo de hardware, software y aplicaciones diversas, es fundamental recorrer este camino con mínimos éticos que hagan de la alfabetización digital, una oportunidad de mejora de la convivencia social y no un camino más de las violencias.

La educación para la paz se erige como alternativa ante lo ya expuesto, pues concibe la conflictividad humana como una de sus tareas a analizar y a formar en el abordaje creativo pero pacífico y/o *no violento* de los conflictos. Las dificultades presentes en la convivencia humana se manifiestan en el aula, en el patio de juegos, en las calles, en los barrios, en las comunidades y entre países. Estamos en el aquí y ahora con conflictos armados internacionales que podemos *seguir* en nuestros dispositivos. En octubre de 1999 se aprobó la Declaración y Programa de Acción sobre una Cultura de Paz en el que se define la cultura de paz y se establece un programa de acción para los principales agentes en el plano nacional, regional e internacional. La cultura de paz es definida como:

“El conjunto de valores, actitudes, tradiciones, comportamientos y estilos de vida basados en el respeto a la vida, el fin de la violencia y la promoción y práctica de la no violencia por medio de la educación, el diálogo y la cooperación; el respeto pleno de los principios de soberanía, integridad territorial e independencia política de los Estados y de no injerencia en los asuntos internos; el respeto pleno y la promoción de los derechos humanos y las libertades fundamentales que incluye la igualdad de derechos y oportunidades de mujeres y hombres, el derecho a la libertad de expresión, opinión e información; el compromiso con el arreglo pacífico de los conflictos; los esfuerzos para satisfacer las necesidades de desarrollo y protección del medio ambiente de las generaciones presentes y futuras; la adhesión a los principios de libertad, justicia, democracia, tolerancia, solidaridad, cooperación, pluralismos, diversidad cultural, diálogo y entendimiento a todos los niveles de la sociedad y entre naciones”⁷.

La cultura de paz y su programa educativo está dirigida, entre otros elementos, a ser una educación ética para el conflicto que puede promover un uso responsable y prudente de la IA-cibernetica en las escuelas. Nos coloca como agentes de transformación y mira en cada persona la posibilidad de transformar las sociedades, concibiéndolas como ícono social que resiste las tentaciones de la seducción lumínica mediática. La educación para la paz es un proyecto que va de adoptar un estilo personal basado en la *noviolencia*, hasta las resistencias civiles pacíficas que buscan las transformaciones profundas sociales. Proyecto personal y colectivo que no es feliz con el estado de las cosas y que busca a partir de dinámicas lúdicas pero vivenciales, movilizar las emociones y la razón colectiva, para generar ambientes de pensamiento crítico, creativo, empático y solidario, sin dejar de reconocer que fluye en el interior humano los claroscuros y la contrariedad, pero también enfatiza que somos posibilidad y proyecto en cada decisión que tomamos. La educación para la paz se erige también anticolonialista de los territorios pero también del pensamiento; y el empleo irracional de las tecnologías que nos dejan en la alienación y minoría de edad, en donde

⁷ Resolución de Naciones Unidas A/RES/53/243.

somos proyecto de *otros* y no de nuestras más profundas decisiones y deseos. Lucha contra todas las formas de discriminación, nombre clasismo, apartheid, racismos, odios étnicos, misoginias, entre muchas otras. Nos ayudaría a mirar que la inteligencia artificial ni es inteligencia, ni es artificial, que pueden ser más sistemas de repetición y reproductores de la doxa masificada.

Sería ingenuo plantear un rechazo *per se* al uso de las tecnologías actuales, se trata más bien de entender y actuar, ya el adagio del Movimiento por la paz nos invitaba a “Pensar globalmente y actuar localmente”. Quizá en nuestros contextos, una tarea que podríamos ejercitar es el llamado silencio o desconexión virtual, nuestros niños –digo *nuestros* por corresponsabilidad y no como posesión– podrían disfrutar sin el Gran Hermano vigilando sus movimientos e incitándolos a realizar actividades riesgosas; podrían descubrir un espacio para encontrarse con el *otro* y por qué no, con toda su humana conflictividad.

Bibliografía

Marcos, Alfredo (2024): “Retos filosóficos de la IA”, Conferencia dictada en la Facultad de Filosofía y Letras, 1º de febrero, UNAM. Disponible en: «https://www.youtube.com/watch?v=n55nv1W_yQc».

Mateos, José Luis (2024): “La revolución de la inteligencia artificial: de Alan Turing al CHATGPT y la IA generativa”, Conferencia dictada en el Centro de la Complejidad (C3), 5 de marzo, UNAM. Disponible: «https://www.youtube.com/embed/KV01-aIv7QY?si=na0UCEofbUDMSpR_».

NOTAS



Esta sección contiene reflexiones y análisis sobre temas contemporáneos relevantes relacionados con los derechos humanos en perspectiva internacional y comparada.

La Inteligencia artificial como objeto y recurso en la educación jurídica

MORAMAY GUERRA GARCÍA

*Centro de Estudios sobre la Enseñanza y el
Aprendizaje del Derecho (CEEAD)*

SUMARIO: I. De los autómatas al ChatGPT-4. II. La inteligencia artificial como objeto de la educación jurídica. III. La inteligencia artificial como recurso en la educación jurídica. IV. Consideraciones finales.

I. De los autómatas al ChatGPT-4

Cuando era niña, encontré en un recoveco de la biblioteca pública un cuento sobre una nave futurista en la que había una máquina que respondía a todas las preguntas que los niños y las niñas le planteaban. Como yo estaba llena de preguntas, la idea me pareció fabulosa. Años después me enteré de que el interés por las máquinas con conocimientos y capacidad de razonar era una constante en la historia de la humanidad. Ejemplo de ello son las doncellas doradas descritas en la *Ilíada*, creadas por Hefesto, las cuales poseían entendimiento y habla. Con esta misma inquietud, a lo largo de los siglos se diseñaron autómatas y máquinas con sistemas lógicos, como el *Ars Generalis Ultima*, o la máquina de Turing. Es hasta mediados del siglo XX que dos hitos brindan bases sólidas al desarrollo de los sistemas de inteligencia artificial (IA): el primero, la construcción de la red neuronal computacional de Minsky, en la Universidad de Princeton y el segundo, la Convención de Dartmouth, en la cual John McCarthy acuña el término de inteligencia artificial (Oliver 2020).

Actualmente, la IA se define como sistemas de software –ocasionalmente también de hardware– diseñados por humanos que, dado

un objetivo complejo, actúan en la dimensión física o digital recopilando datos del entorno, interpretándolos, procesando la información derivada de dichos datos, y decidiendo las acciones que deben realizar para lograr el objetivo establecido (Zorrilla 2023).

Los sistemas de IA se han desarrollado bajo dos enfoques: el simbólico-lógico o *top down*, en el cual el ordenador debe programarse con el conocimiento y con reglas predeterminadas, y a partir de ello, el ordenador deriva nuevo conocimiento. Un ejemplo de este enfoque es Cyc, una IA que se apoya una base de datos con 100.000 conceptos y 1.000.000 aseveraciones, reglas o ideas del sentido común. Otro enfoque es el *bottom-up*, el cual postula que los ordenadores deben programarse para que *aprendan*; es decir, que los algoritmos se entrenen a partir de numerosas experiencias u observaciones de lo que se quiere que hagan. A este enfoque pertenecen el aprendizaje automático estadístico, los métodos de búsqueda, los sistemas de recomendación y personalización, entre otros. También se encuentran los modelos de redes profundas que tienen como base el algoritmo *backpropagation*, los cuales hacen predicciones, las comparan con la realidad, identifican el grado de error, y a partir de ello, ajustan las siguientes predicciones. Con el proceso iterativo, el error disminuye paulatinamente.

La capacidad de la IA se ha incrementado en los últimos años gracias al avance en las técnicas de aprendizaje estadístico por ordenador —*statistical machine learning*—, al desarrollo de procesadores más potentes, así como de las redes neuronales profundas y complejas. La IA que empleamos actualmente se clasifica como específica o *soft*, pues se dirige a realizar tareas concretas, pero se aspira al desarrollo de una IA general o *strong*, la cual se parecería a la inteligencia humana, siendo más amplia, flexible e incremental (Oliver 2020; Zorrilla 2023).

Recientemente, la IA ha ganado visibilidad gracias al ChatGPT-4 de la familia de *Generative Pre-trained Transformer* (GPT) de OpenAI. El ChatGPT-4 es una IA multimodal, es decir, que encuentra patrones en texto, audio e imágenes, lo que la hace más poderosa

que sus predecesoras —GPT (2018), GPT-2 (2019), GPT-3 (2020) y GPT-3.5 (2022)— que solo trabajaban con texto (Sargeant 2023).

En la actualidad convivimos cotidianamente con diferentes formas de IA, por ejemplo, con los sistemas de procesamiento natural del lenguaje, como *Alexa* o *Siri*, y otros asistentes personales de los dispositivos móviles u ordenadores; con los sistemas de búsqueda de información, contenido, productos o personas —Netflix, Facebook, Amazon—; con los sistemas incluidos en el software de las aplicaciones o de los dispositivos. Asimismo, hay IA implicadas en procesos como la autorización de créditos, de visas, o en los sistemas de vigilancia, entre otros. No siempre las personas usuarias saben que están interactuando con una IA, o bien, que su información está siendo recopilada, archivada y empleada por una de ellas (Oliver 2020).

Dadas las implicaciones éticas, jurídicas y sociales de la IA, constituye un tema relevante en la formación de las y los profesionales del derecho, y en consecuencia, para el campo de la educación jurídica (Ej) y las instituciones que la imparten.

II. La IA como objeto de la educación jurídica

Dada la creciente presencia de la IA en diferentes ámbitos, organismos internacionales han emitido documentos dirigidos a que su uso sea conforme al derecho internacional y en especial, a los derechos humanos (D.D.HH). Uno de los esfuerzos regulatorios más recientes y significativos es la Ley de Inteligencia Artificial, aprobada el Parlamento de la Unión Europea en marzo del 2024 (Expansión 2024), y cuya implementación se prevé para mayo o junio. La Ley tienen sus antecedentes en el proyecto de legislación surgido en 2021, y su finalidad es brindar un marco jurídico para la regulación de la IA, y prohibir aquellos sistemas que representen una amenaza para los derechos fundamentales, estableciendo, entre

otras cosas, obligaciones para desarrolladores e implementadores (Consejo de la Unión Europea 2024).

Por su parte, la UNESCO (2022) enfatiza el papel sustantivo de las instituciones de Educación Superior, y entre ellas las que imparten EJ, para promover “la participación cívica, las competencias digitales y la capacitación en materia de ética de la IA, la alfabetización mediática e informacional” (UNESCO 2022: 23) a fin de que todas las personas puedan adoptar decisiones informadas respecto al uso de las IA, y a la par, puedan protegerse de sus aspectos o efectos negativos. Particularmente, las instituciones de EJ pueden sumar a la formación de sus estudiantes la alfabetización en IA, incorporando en el currículo objetivos de aprendizaje relativos al funcionamiento y uso de estos sistemas, así como a su análisis ético, sociológico y jurídico –incluyendo el análisis de la normatividad específica para la IA–, teniendo como referente la legislación internacional y los DD.HH (Flores y García 2023). La previa capacitación del cuerpo docente sería una condición indispensable para incorporar dichos contenidos.

Algunos temas que pueden abordarse son:

- 1) La protección de datos personales y el derecho a la privacidad. Es necesario revisar y regular que los sistemas de IA recopilen, archiven y usen información respetando el derecho a la privacidad, y favoreciendo que las personas sepan qué información se recopila, cómo se usa y puedan tomar decisiones sobre sus datos (UNESCO 2022; Andrés y Ramón 2023).
- 2) Información errónea, sesgada o discriminatoria emitida por la IA. Se requiere explicitar que los modelos de IA son entrenados con información derivada de grupos sociales o acervos culturales específicos, por lo que refleja parcialmente la realidad y es posible que emita información falsa, sesgada o con base en estereotipos. Es necesario considerar los efectos de tal información, sobre todo si se usa para decisiones relevantes en la vida de las personas (Oliver 2020; UNESCO 2022).

- 3) Los contenidos ficticios o sintéticos. También conocidos como *Deepfake*, pueden ser empleados para manipular la opinión pública y la percepción social, promoviendo emociones o actitudes hacia personas o situaciones, además de ser utilizados para la comisión de delitos (Oliver 2020).
- 4) Decisiones automatizadas. Dada la capacidad de procesamiento de información de las IA, en algunos ámbitos se les ha cedido la toma de decisiones, lo cual debe regularse sobre todo cuando dichas decisiones inciden sustantivamente en la vida, salud o libertad de las personas (UNESCO 2022).
- 5) Los derechos de autor. Ya que los sistemas de IA crean contenidos a partir de diversas fuentes publicadas, lo cual pueden entrañar riesgos para la propiedad intelectual (Andrés y Ramón 2023).
- 6) La legislación específica sobre IA. Es necesario analizar en qué medida la legislación específica sobre IA coadyuva a la protección de los DD.HH, y aborda los riesgos potenciales, así como los desafíos e implicaciones de su implementación.
- 7) IA que realizan actividades de la profesión jurídica. Se requiere analizar los efectos del uso de sistemas que realizan prácticas como la revisión de documentos para determinar su utilidad jurídica; la predicción de fallos o decisiones de los tribunales; la redacción de documentos como contratos o actas, y la generación de respuestas a planteamientos o consultas legales, como DoNotPay (Zorrilla 2023).

III. La inteligencia artificial como recurso en la educación jurídica

La IA puede favorecer el acceso de las personas a la EJ, su inclusión y aprendizaje (Andrés y Ramón 2023), además de facilitar tareas administrativas rutinarias. Para ello, se requiere capacitar al claustro docente y crear condiciones institucionales para aprovechar el potencial de la IA. Su incorporación debe realizarse de for-

ma sistemática, accesible, crítica, ética y reflexiva sobre sus alcances y limitaciones.

Retomando a Flores y García (2023), la IA puede desempeñar distintos roles en la EJ:

- 1) Acceso universal. La IA puede ayudar a realizar adaptaciones que favorezcan la accesibilidad del conocimiento a todas las personas, por ejemplo, realizando traducciones simultáneas, o ajustes para personas con discapacidades visuales o auditivas.
- 2) Sistemas de tutoría inteligentes. La IA puede contribuir a crear ambientes de aprendizaje individual sobre un área de conocimiento específico, en el cual el sistema interactúe directamente con la o el estudiante y le brinde retroalimentación útil para su aprendizaje.
- 3) Facilitadores virtuales. La IA puede crear personajes virtuales realistas que reconozcan gestos e interactúen con el estudiantado, ayudándole en temas específicos. La animación por computadora también puede generar escenarios de realidad aumentada o experiencias de realidad inmersiva.
- 4) Contenido inteligente. La IA puede coadyuvar a la creación de *Smart content*, es decir, que considera las características de la persona usuaria, por ejemplo, interfaces de aprendizaje personalizadas, a través de medios como audio, vídeo y asistente en línea.
- 5) Espacios de colaboración entre profesores. La IA puede generar espacios y condiciones que faciliten que el claustro docente trabaje colegiadamente en actividades y materiales que favorezcan el aprendizaje.

- 6) **Análisis de contenido.** La IA puede identificar temas dentro de una gran cantidad de información y clasificarla, agrupar textos por su utilidad o pertinencia, analizar opiniones, así como generar recomendaciones personalizadas de materiales o información específica.
- 7) **Apoyo de tutorías fuera del aula.** La IA puede generar programas que ayuden al estudiantado en la preparación de tareas o exámenes.
- 8) **Automatizar las tareas de administración.** La IA puede facilitar actividades rutinarias, por ejemplo, calificar exámenes de opción múltiple o tareas, generar promedios o cumplir requerimientos de control escolar.

IV. Consideraciones Finales

Ante la creciente incorporación de la IA en la vida cotidiana, la Ej puede y debe contribuir a la formación profesionistas que coadyuven a que el funcionamiento de estos sistemas se apegue a los marcos normativos internacionales, y a que se priorice la dignidad humana y las prerrogativas que en ella se sustentan durante los procesos de recopilación de información, interpretación, procesamiento y toma de decisiones. A la par, las instituciones que imparten Ej pueden aprovechar los diferentes roles de la IA en los entornos educativos para enriquecer las experiencias de aprendizaje que ofrecen a sus estudiantes, además de capacitarles para el uso ético y responsable de los sistemas que realizan actividades del área jurídica, y cuya presencia, factiblemente, será cada vez más constante dentro de la profesión.

Bibliografía

- Andrés, Eva y Ramón, Francisca (2023): “La inteligencia artificial ‘ChatGPT’ versus la Ley y el Derecho. Jaque al derecho de la propiedad intelectual” en *Revista de Educación y Derecho*, núm. 28. Universidad de Barcelona, disponible en: «*Inteligencia Artificial: “chat GPT” versus la Ley y el Derecho. Jaque al derecho de la propiedad intelectual | Revista de Educación y Derecho (ub.edu)*» [Consultado el 5 de marzo de 2024].
- Consejo de la Unión Europea (2024): “Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Analysis of the final compromise text with a view to agreement”, disponible en: «<https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf>».
- Expansión (2024): “La UE aprueba la primera Ley para regular la IA” en *Expansión*. Disponible en: «https://expansion.mx/tecnologia/2024/03/13/union-europea-aprueba-prime-ra-ley-para-regular-ia?_amp=true».
- Flores, Jesús Miguel y García, Francisco José (2023): “Reflexiones sobre la ética, potencialidades y retos de la Inteligencia Artificial en el marco de la Educación de Calidad (ODS4)” en *Revista Científica de Educomunicación*, vol. XXXI, núm. 74. Disponible en: «[10.3916_C74-2023-03.pdf \(grial.eu\)](https://grial.eu/10.3916_C74-2023-03.pdf)» [Consultado el 3 de marzo de 2024].
- Oliver, Nuria (2020): *Inteligencia artificial, naturalmente. Un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos*, Ministerio de asuntos económicos y transformación digital del Gobierno de España, España.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) (2022): *Recomendación sobre la ética de la inteligencia artificial*, UNESCO, Francia.

Sargeant, Paul (2023): “Una guía sencilla para entender la IA” en *BBC News*. Disponible en: «<https://www.bbc.com/mundo/resources/idt-74697280-e684-43c5-a782-29e9d11fecf3>» [Consultado el día 10 de marzo de 2024].

Zorrilla Alonso, Paulina (2023): *Inteligencia artificial y derecho*, Centro de estudios sobre la Enseñanza y el Aprendizaje del Derecho (CEEAD), México.

RESEÑAS BIBLIOGRÁFICAS

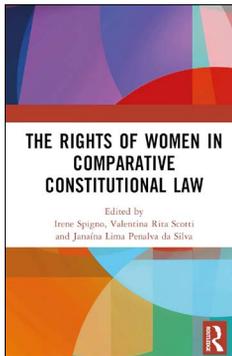


Esta sección está conformada por síntesis que informan sobre el contenido, aporte y significado de obras jurídicas académicamente relevantes en materia de derechos humanos; ya sea publicadas durante los últimos años, o bien editadas con anterioridad y consideradas esenciales en los temas abordados por esta publicación.

Reseña. Spigno, Irene; Scotti, Valentina Rita y Penalva da Silva, Janaína Lima, *The Rights of Women in Comparative Constitutional Law*, Nueva York, Routledge, 2023.

SANTIAGO DANIEL SÁNCHEZ JUÁREZ

*Academia Interamericana de Derechos Humanos
Universidad Autónoma de Coahuila*



La situación de las mujeres en el mundo sigue siendo analizada desde la desigualdad. Ya no se trata solo de las diferencias respecto a los hombres, sino también entre ellas por una cuestión geográfica. La historia, cultura, forma de gobierno o la tradición jurídica son elementos contundentes en una posible evaluación sobre el ejercicio efectivo de la igualdad de género.

Irene Spigno, Valentina Rita Scotti y Janaína Lima Penalva da Silva, conscientes de esta situación, desarrollaron una obra académica de gran valor para estudiar trece experiencias constitucionales respecto a los derechos de las mujeres. Dividida en tres partes, la primera reúne textos enfocados en el viejo continente.

Anita Blagojević abre este apartado con un estudio sobre la evolución y retroceso que existe en la República de Croacia. Le sigue Francesca Brunetta y la brecha legislativa de género en Alemania. Tímea Drinóczi y Lúdia Balogh hacen lo propio con la interpretación constitucional en Hungría. Por su parte, Valentina Rita Scotti analiza sublimemente el *diálogo* entre los legisladores y los jueces de la Corte Constitucional en Italia. Agnieszka Bień-Kacała, Julia Kapelańska-Pręgowska y Anna Tarnowska detallan el ascenso y caída de la igualdad de género en Polonia. Finalmente, Mar Aguilera Vaqués cierra el capítulo tratando el caso español.

En esta sección se destacan cuestiones que aún resultan problemáticas para los países seleccionados. Por ejemplo, el nivel de participación de las mujeres en los espacios públicos nacionales de Europa sigue estando por debajo del 45%, lo que significa una ausencia de paridad real. A nivel económico, los índices de igualdad de género también señalan un rezago generalizado tanto en la educación tecnológica, como en los ingresos. Incluso los procesos de transición democrática en Hungría y España muestran algunas mejoras, pero también retrocesos. En Italia, la fuerte cultura católica sigue representando un enorme obstáculo a superar.

Respecto a Polonia, se aporta un análisis interesante sobre la tradición liberal e iliberal en la política como factor de regresión en la protección de los derechos de las mujeres. Esto refleja que los éxitos obtenidos a través de grandes luchas sociales e históricas, pueden ser derrotados por agendas políticas conservadoras; y cómo el Tribunal Constitucional puede ser el último eslabón de defensa.

Una segunda parte de la obra está enfocada al estudio de los derechos de las mujeres en distintas constituciones del continente americano. Ana Micaela Alterio y Paulina Macías estudian el desarrollo legal argentino en materia de igualdad de género. A su vez, Janaína Lima Penalva da Silva y Vera Karam de Chueiri aportan una visión interesante en materia de inclusión del feminismo en el futuro del constitucionalismo brasileño. El siempre emblemático caso de México fue desarrollado y estudiado por Irene Spigno. Por último, la historia y el futuro del constitucionalismo estadounidense es puesto a prueba por Graziella Romeo.

Los tres casos latinoamericanos expuestos en este segundo bloque plantean de manera muy sólida una serie de problemáticas en la *efectividad* de normas y políticas públicas. El avance en las agendas públicas es palpable; algunos casos como el argentino han sido impulsados por movimientos sociales, mientras que en México, por su parte, es consecuencia de un trabajo legislativo vanguardista. Sin embargo, las autoras de estos capítulos concuer-

dan en que el siguiente paso es asegurar el cumplimiento de los estándares de protección de los derechos de las mujeres en temas tan relevantes como los derechos reproductivos o el acceso a una vida libre de violencia.

Aunque Estados Unidos se encuentra en América –igual que los países anteriores– su realidad social y jurídica es considerablemente distinta. Graziella Romeo identifica esos momentos clave en la historia estadounidense tanto en el aspecto social, como en su legislación y jurisprudencia. Especialmente en un país con una constitución poco extensa –a comparación de las constituciones iberofonas–, la protección del principio de igualdad entre hombres y mujeres ha sido impulsada por la Corte Suprema de los Estados Unidos a lo largo de su historia.

No obstante, y bien lo señala la autora, en el caso *Dobbs vs. Jackson Women's Health Organization*, este máximo tribunal ha reculado en su propia doctrina. Lo construido en *Roe vs. Wade* quedó sin efecto con el argumento que señala como verdadero responsable de la protección de derechos, en primera instancia, al Congreso. Esta decisión no solo modificó la posibilidad de interrumpir voluntariamente un embarazo, sino que sentó una nueva visión judicial de actuación, dejando un futuro incierto para las mujeres en este país.

El último apartado de la obra reúne 3 capítulos. El primero de ellos a cargo de Rosella Bottoni quien estudia la interacción secular, religiosa y consuetudinaria en Etiopía como entorno para el desarrollo de una agenda de igualdad de género. Elisa Bertolini, por su parte, señala la cuesta que tienen que superar las mujeres en Japón para encontrar una igualdad real. Y, finalmente, Asylai Akisheva estudia la situación constitucional en Kirguistán. De esta manera, aunque de manera breve, no se pierde de vista a los continentes asiático y africano.

En el primer caso, se refleja cómo las normas pueden no ser explícitamente patriarcales, pero tampoco ser promotoras de la

igualdad de género. Y ante esa supuesta neutralidad, son los contextos culturales y las tradiciones sociales las que inclinan la balanza en un sentido u otro; y en Etiopía se han reforzado los roles tradicionales de género.

Lo que se encontrará en el capítulo de Bertolini, es un maravilloso recorrido por la historia nipona, los actores políticos principales y la evolución legislativa en materia de género para abrir espacios que culturalmente se habían mantenido cerrados. La conformación de la Corte de Justicia es un claro ejemplo, por un lado, de la importancia en la representación femenina y, por el otro, de la apertura jurídica que se tuvo respecto a tratados internacionales como la Convención para la Eliminación de Todas las Formas de Discriminación Contra la Mujer (CEDAW).

Respecto al caso de Kirguistán, no puedo dejar pasar la oportunidad para mencionar lo llamativo que resulta la revisión de este país, toda vez que no se encuentra en los países foco para estudios tradicionales. De ahí el más grande reconocimiento a Asylai Aki-sheva por narrar la historia social y jurídica de las mujeres kirguís. Me parece particularmente interesante el tercer subtema del capítulo “When Law Meets Reality” acerca del funcionamiento judicial y la aplicación de garantías públicas en este país para asegurar una efectiva igualdad de género.

A manera de síntesis, en la obra se ofrecen reflexiones muy interesantes sobre los contextos históricos, sociales, políticos, económicos y religiosos que determinan la estrategia jurídica en distintos sentidos. Solo a partir de una visión multidisciplinaria se pueden lograr cambios verdaderos en las sociedades actuales. Ya no son suficiente las reformas legales; se necesita mucho más.

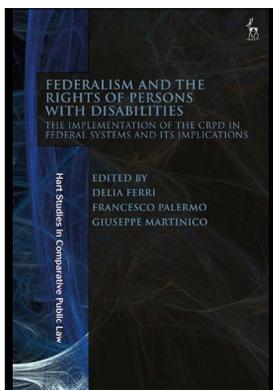
Cierto resulta que trece experiencias son pocas comparado con lo vasto que es el mundo. Sin embargo, lejos de ser una crítica a la obra, se vuelve una invitación a las editoras para seguir generando otros tomos que estudien más casos igualmente interesantes.

Los estudios africanos y asiáticos en la materia son escasos, pero no por ello dejan de ser relevantes.

Definitivamente este libro es un aporte doctrinal, académico y científico sumamente útil para mujeres y hombres, estudiantes y profesionistas, e incluso, para críticos del tema. Este ejemplar ha llegado para colocarse como un texto referencia sobre cómo se debe hacer el derecho comparado, y también sobre cuáles son los avances y retos en materia de protección constitucional de los derechos de las mujeres.

Reseña. Ferri, Delia; Palermo, Francesco y Martinico, Guiseppe, *Federalism and the rights of persons with disabilities. The implementation of the CRPD in federal systems and its implications*, Oxford, Hart Publishing, 2023.

DANIELA MONSERRAT HERNÁNDEZ ÁVILA
Academia Interamericana de Derechos Humanos
Universidad Autónoma de Coahuila



La discapacidad es un tema que ha lido cobrando relevancia en los últimos años. Con el paso del tiempo, los sistemas jurídicos han protegido cada vez más los derechos de las personas con discapacidad. El libro *Federalism and the rights of persons with disabilities* de Delia Ferri, Francesco Palermo y Giuseppe Martinico, precisamente aborda de manera muy atinada lo que implica la garantía de dichos derechos a través de la aplicación

de la Convención sobre Derechos de la Personas con Discapacidad (CDPCD) en países con sistemas federados.

A través de un estudio muy completo, las y los autores exponen los desafíos a los que se enfrenta la implementación de la CDPCD. Aún y cuando este tratado internacional es un instrumento que ha marcado un nuevo paradigma mundial en la promoción e inclusión de las personas con discapacidad a nivel global, los sistemas federados han planteado algunos retos, como lo son la coordinación de los niveles de gobierno en el cumplimiento de las obligaciones contraídas por los Estados.

Lo anterior se debe a la división de poderes que existe y las distintas responsabilidades que conlleva cada uno de ellos. Esta obra permite que el lector conozca semejanzas y diferencias entre los sistemas federados y la aplicación de la CDPCD en los siguientes

países: Alemania, Austria, Italia, España, Bélgica, Reino Unido, Canadá, India, Sudáfrica, Etiopía, Argentina, Brasil y México; además, también se aborda la Unión Europea (UE).

Cada país –y la UE– tiene destinado su propio capítulo, a los que se suma uno más de conclusiones. En estos apartados se explora cómo la CDPD se ha integrado a los distintos sistemas internos; así como las particularidades que implica su puesta en práctica a nivel nacional.

Un factor muy importante que destacar es la metodología utilizada en la obra. Las y los autores aclaran de manera muy acertada que la elección y orden de países corresponde, en primer lugar, a la geografía; y en segundo lugar al tipo de sistema jurídico-constitucional que defienden: monista, dualista y *mixto* o *en el medio*.

En ese sentido, los países objeto de análisis se dividen en tres grupos, facilitando al lector el estudio de las diferencias que hay en cada modelo de integración normativo. Para esto conviene recordar que un sistema monista es aquel que permite la aplicación directa de los tratados internacionales; mientras que los dualistas requieren de más acciones para convertirlos o ajustarlos a las disposiciones nacionales.

Por otro lado, se encuentra el grupo en el que es posible observar una dinámica que contiene un poco de ambos sistemas. Es decir, se consideran monistas, pero las acciones que llevan a cabo para la implementación de la CDPD tienen matices más dualistas o viceversa.

En el primer capítulo, Delia Ferri y Neža Šubic analizan la adhesión de la UE a la CDPD. A través de un estudio, se cuestionan si esta acción impacta en la división de competencias de los Estados Miembros y la UE.

Felix Welti, por su parte, se aproxima al tema central desde la perspectiva alemana. Nos brinda un breve contexto sobre la forma de gobierno del país y sobre los principales retos que ha enfrentado en la implementación de la Convención a nivel federal y local.

En el tercer capítulo, Peter Bubjäger ilustra el panorama del desempeño de la Convención y el gobierno en Austria. El autor destaca las dificultades de la implementación debido a la distribución de competencias en Austria y crítica que no exista un desarrollo pleno de las obligaciones internacionales derivadas de la CDPCD, evidenciando la ausencia de armonización legislativa.

Italia también presenta una serie de deficiencias legislativas a nivel central. Paolo Addis y Matteo Monti analizan la garantía y protección de la Convención en este país, destacando como principal obstáculo la reticencia local; algo que resulta evidente ante la pérdida de autoridad con la que cuentan al tener que adecuarse a una norma supraestatal.

Posteriormente, en el siguiente capítulo encontramos el caso de España. Maribel González Pascual subraya el impulso que trae consigo la Convención, al formar parte del ordenamiento interno de manera automática. Así, se revisa su cumplimiento impulsado por el gobierno central, quien ha creado resoluciones y políticas públicas que se implementan en todo su territorio.

En los siguientes capítulos (6, 7 y 8) desarrollados por Naomi Ghys, Freek Louckx, Daniel Dumont, Kasey McCall-Smith y Jonás-Sebastián Beaudry se exponen los escenarios de Bélgica, Reino Unido y Canadá, respectivamente. Tanto en Bélgica como Reino Unido, la CDPCD no representa un impacto directamente central. En el caso belga, la ausencia de jerarquías entre gobierno federal y local ha entorpecido la efectividad de la Convención. En el segundo caso, se identifica una apatía respecto al derecho internacional, por lo que tampoco existe un compromiso gubernamental.

Caso contrario es el de Canadá, donde encontramos una gran evolución normativa en materia de discapacidad. El autor señala acertadamente el papel que desempeñan los distintos niveles de gobierno para lograr una colaboración entre sí que les permite tener legislaciones armónicas.

Los apartados 9, 10 y 11 analizan a India (Amita Dahanda), Sudáfrica (Tinashe C. Chigwata y Robert D. Namina) y Etiopía (Yonatan T. Fessha y Beza Dessalegn). En el noveno capítulo encontramos un interesante recorrido procedimental sobre la adopción de documentos internacionales, la forma en la que se crean leyes y cómo son implementadas por cada nivel de gobierno indio. Esta estructura normativa es lo que, en palabras de la autora, determina un cumplimiento de la convención por parte de India, aunque con aspectos mejorables a futuro.

Por su parte, Sudáfrica a pesar de haber ratificado la CDPCD y su protocolo adicional, todavía no cuenta con una legislación específica sobre discapacidad que permita cumplir con sus obligaciones. Aun así, las tendencias expuestas en este capítulo son optimistas, por lo que se espera una evolución progresiva para consolidar la protección institucional de las personas con discapacidad.

Otro caso que sobresale por su ausencia es el destacado por Yonatan T. Fessha y Beza Dessalegn. Etiopía no ha tenido especial atención en el tema de protección a derechos de personas con discapacidad incluso cuando su constitución reconoce la igualdad como un derecho humano. Resulta interesante de este país la existencia de una doble facultad de promulgación internacional. Tanto el gobierno federal como el estatal tienen dicha posibilidad pero ninguno de los dos lo ha realizado en materia de discapacidad. Sin duda este mecanismo de coordinación federal es distinto al resto y merece la pena su profunda revisión.

En el capítulo 12, Franciso J. Bariffi destaca la complejidad que representa para Argentina la adopción de la Convención. También señala que a pesar de los retos a los que se enfrenta, se han tomado medidas que permiten la mejora de su implementación, por ejemplo, el convertir en leyes federales a los tratados internacionales en materia de derechos humanos considerados relevantes. Esta estrategia también resultará atractiva para quién lea la obra porque representa una forma interesante de armonizar lo interno y lo internacional.

Giblerto Ma. Rodrigues y Paloma Breit nos ofrecen un estudio enfocado en Brasil. En él, señalan el impacto que ha tenido la Convención desde su ratificación. Tanto Estados como municipios se han hecho de medios propios para garantizar los derechos de las personas con discapacidad como una forma de contrarrestar la ineficiencia de la federación en este tema.

El penúltimo capítulo es abordado por Irene Spigno y en él describe cómo es que México ha desarrollado una implementación centralizada de la Convención. Destaca la estrategia de réplica local a las acciones federales como una estrategia de armonización. Sin embargo, subraya que uno de los principales retos a los que se ha enfrentado el país es la erradicación de la brecha de protección creada a partir de la aplicación de leyes que se contraponen con la creación de ajustes razonables para la no discriminación.

Finalmente, el apartado 15 corresponde a los editores de la obra. En estas letras finales se encontrarán aportes y comentarios conclusivos sobre los casos estudiados en una obra que permite concentrar el tratamiento nacional de la CDPCD. Las experiencias federales son tan variadas que solo puedo recomendar enormemente este último apartado para identificar esos puntos de encuentro incluso en la diversidad jurídica.

En suma, el libro reseñado es un excelente producto académico que aborda, desde el derecho comparado, una visión enriquecedora de un tema sumamente relevante. Las personas con discapacidad han afrontado una gran cantidad de retos y desafíos. Desde lo internacional, la CDPCD ha impulsado una agenda de transformación. Aunque los ideales internacionales suelen enfrentarse con las realidades nacionales, son obras como esta las que permite generar propuestas y avances efectivos. Cualquier persona que lo lea disfrutará del rigor científico, sin perder de vista lo importante que resulta reflexionar en nuevas herramientas para garantizar los derechos de todas las personas.

AUTORES COLABORADORES

ARTÍCULOS DOCTRINALES

Ma. Elizabeth De Los Ríos Uriarte
*Profesora e Investigadora en la Facultad de Bioética
Universidad Anáhuac México*

Marcial Orlando Cabrera Cantarero
*Médico Pasante de Servicio Social en la Facultad de Bioética
Universidad Anáhuac México*

Inés Marquina Gurrea
*Médico Pasante de Servicio Social en la Facultad de Bioética
Universidad Anáhuac México*

Ana Martín Beringola
Universidad Internacional de Valencia

Alejandro Vilchis Robles
Poder Judicial de la Federación

José Rafael Grijalva Eternod
*Academia Interamericana de Derechos Humanos
Universidad Autónoma de Coahuila*

Juan Carlos Gámez Sagrero
Escuela Libre de Derecho

ENSAYOS

Mónica Adriana Mendoza González
Universidad Nacional Autónoma de México

NOTAS

Moramay Guerra García
*Centro de Estudios sobre la Enseñanza y el Aprendizaje del Derecho
(CEEAD)*

RESEÑAS BIBLIOGRÁFICAS

Santiago Daniel Sánchez Juárez
*Academia Interamericana de Derechos Humanos
Universidad Autónoma de Coahuila*

Daniela Monserrat Hernández Ávila
*Academia Interamericana de Derechos Humanos
Universidad Autónoma de Coahuila*