

CENTRO DE ESTUDIOS DE DERECHO E INVESTIGACIONES PARLAMENTARIAS



INTELIGENCIA ARTIFICIAL Y DELITO. NUEVOS RETOS EN EL ÁMBITO LEGISLATIVO

MARIO EDUARDO MALDONADO SMITH

DOCUMENTO DE TRABAJO

2024



PODER LEGISLATIVO FEDERAL
CÁMARA DE DIPUTADOS



CENTRO DE ESTUDIOS DE DERECHO
E INVESTIGACIONES PARLAMENTARIAS



INTELIGENCIA ARTIFICIAL Y DELITO. NUEVOS RETOS EN EL ÁMBITO LEGISLATIVO

ARTIFICIAL INTELLIGENCE AND CRIME. NEW CHALLENGES IN THE LEGISLATIVE REALM

Mario Eduardo MALDONADO SMITH¹

RESUMEN: En este documento se presentan consideraciones generales sobre las implicaciones de la inteligencia artificial (IA) en el ámbito criminal. Se desarrollan tres temas en los que la IA tiene relevancia: su empleo en conductas criminales, en la prevención del delito y en la individualización y determinación de la responsabilidad penal. El objetivo es problematizar cada caso para identificar vacíos normativos y posibilidades de intervención legislativa y se concluye con principios que pudieran ser objeto de una eventual regulación. Como métodos de trabajo, se han empleado el deductivo, documental e inferencial.

PALABRAS CLAVE: Inteligencia artificial, delito, responsabilidad, tecnología.

ABSTRACT: *This document presents general considerations on the implications of Artificial Intelligence (AI) in the criminal domain. Three topics are developed where AI is relevant: its use in criminal behavior, its use in crime prevention and its use in individualization and determination of criminal responsibility. The objective is to problematize each case to identify regulatory gaps and possibilities for legislative intervention, concluding with principles that could be subject to eventual regulation. Deductive, documentary, and inference methods have been employed as working methods.*

KEYWORDS: *Artificial Intelligence, crime, responsibility, technology.*

SUMARIO: *I. Introducción. II. Consideraciones generales sobre la inteligencia artificial. III. Empleo de la inteligencia artificial en el ámbito delictual. IV. Empleo de la inteligencia artificial como herramienta de prevención del delito. V. Empleo de la inteligencia artificial como herramienta de individualización y sanción del*

¹ Investigador A del Centro de Estudios de Derecho e Investigaciones Parlamentarias de la Cámara de Diputados, profesor de la Universidad Internacional de la Rioja y doctor en Derechos Humanos por la Università degli Studi di Napoli, Federico II.



delito. VI. Reflexiones de orden legislativo y criminal. VII. Conclusiones. VIII. Bibliografía.

I. INTRODUCCIÓN

En 1973 se exhibió la película *Soylent Green*, que se convirtió en una de las figuras más emblemáticas de culto. La obra, traducida como “Cuando el destino nos alcance”, es un ejemplo de ciencia ficción “distópica”, donde se presenta un futuro apocalíptico con escasez de recursos, desastre ambiental y un permanente control industrial maquinista.

Bajo ciertas salvedades –aunque no tantas, en realidad–, podemos decir que el destino nos ha alcanzado, tanto en lo que respecta a la escasez de recursos y a los enormes desastres ambientales, como a la permanente influencia de las máquinas, y no de cualquiera de ellas, sino de un tipo específico que hoy hace parte de nuestra vida cotidiana; nos referimos a las máquinas con inteligencia artificial (en adelante IA).

La IA no es algo nuevo. Se trata de una tecnología que, si bien sólo en los últimos años ha experimentado una nueva primavera o “bum”, en realidad fue pensada y desarrollada desde hace casi un siglo cuando en los albores de la Segunda Guerra Mundial se buscaban nuevas tecnologías para descifrar códigos y dar un giro a la guerra.² Desde aquel tiempo y hasta nuestros días, la IA ha experimentado un constante desarrollo; hoy asistimos a una nueva “primavera”³ mediante el empleo de modelos de lenguaje de gran tamaño (*Large Language Model* o LLM), los cuales permiten implementar estrategias predictivas de lenguaje y de respuesta, tales como: *ChatGPT*, *Google BERT*, *YouChat*, *Bloom* o *Perplexity AI*.

² Se trata del caso emblemático de Alan Turing y la decodificación de la máquina alemana “Enigma”. Véase Simón Castellano, *Justicia cautelar e inteligencia artificial. La alternativa a los atávicos heurísticos judiciales*, 52.

³ De hecho, el término *inteligencia artificial* fue acuñado en 1956 por John McCarthy, durante una conferencia en el *Dartmouth Summer Research Project on Artificial Intelligence*. Véase Abadías Selma, *Justicia juvenil e inteligencia artificial en la era de la cultura “Touch”*, 29; Valls Prieto, “Sobre la responsabilidad penal por la utilización de sistemas inteligentes”, 2-3; Soria Olivias, Rodríguez Belenguer *et. al.*, *Inteligencia Artificial. Casos prácticos con aprendizaje profundo*, 16.



Ahora bien, la IA se proyecta en una enorme variedad de manifestaciones que no sólo incluye a los LLM. Tras una ligera observación de nuestro entorno, podemos apreciar que la IA se encuentra por doquier: en nuestro teléfono digital, en los algoritmos predictivos utilizados por *Google* o cualquier otro buscador, al realizar una compra en línea, al acudir al banco, al ser observado por cientos de cámaras instaladas en la calle (incluyendo los satélites que desde cientos de kilómetros recopilan información), al utilizar un automóvil, al presionar un identificador biométrico, en los videojuegos, al interactuar con *Siri*, *Cortana*, *xAlexa* o, simplemente, al manipular artefactos como una lámpara o una aspiradora moderna.⁴ En estos casos –y muchos otros–, las máquinas recopilan información permanentemente acerca de nosotros; la estudian, analizan y, con base en complejos algoritmos, establecen predicciones.⁵

Esta tecnología ha contribuido enormemente a la simplificación de actividades que antes suponían el empleo de una gran de tiempo y de recursos; en cierta medida, nos hace más sencilla la vida. Sin embargo, el empleo de esta tecnología no ha estado exento de riesgos, retos y dilemas sobre los que vale la pena reflexionar.

En este documento procuramos acotar el enorme ámbito en el que la inteligencia artificial se expresa para concentrarnos en someras reflexiones de índole penal; asimismo, dilucidaremos ciertos rubros en los que valdría la pena un desarrollo de orden legislativo, sobre todo, atendiendo a los potenciales riesgos de la IA de cara a principios constitucionales, entre los que se incluyen el respeto a los derechos humanos. Este documento pretende ser una primera aproximación al fenómeno de la IA para expresar consideraciones generales que, en su oportunidad, podamos desarrollar con mayor amplitud en trabajos posteriores.

Nuestra investigación encuentra justificación en la relevancia que la IA ha tenido en el mundo, en general, y en México, en lo particular, tal como se refleja en su incorporación dentro del ámbito legislativo a través de diversas propuestas

⁴ Morán Espinosa, “Responsabilidad penal de la Inteligencia Artificial (IA). ¿La próxima frontera?”, 294.

⁵ O’Neil, *Weapons of Math Destruction. How big data increases inequality and threatens democracy*, 70.



normativas.⁶ Como hipótesis, se plantea que la IA es susceptible de ser empleada de modo tal que puede afectar bienes jurídicos fundamentales y, en consecuencia, resulta necesario emprender acciones para regularla.

Como objetivo, se pretende plantear algunas “luces” y directrices para que, en su caso, sean valoradas por las legisladoras y los legisladores. Para abordar el fenómeno se acudió a los métodos deductivo, documental e inferencial. En cuanto a la estructura del documento, en primer lugar, se brinda una descripción general sobre la inteligencia artificial y su impacto; se continúa con un análisis del empleo de esta tecnología con fines criminales, seguido de su utilización en la prevención delictiva y como técnica de individualización de responsabilidad penal. Finalmente, se presenta una serie de reflexiones legislativas y, como corolario, las conclusiones.

II. CONSIDERACIONES GENERALES SOBRE LA INTELIGENCIA ARTIFICIAL

De acuerdo con José Cabanelas Omil, la IA puede entenderse como:

la habilidad y capacidad de un ordenador, red de ordenadores o red de robots controlados por ordenadores para realizar las tareas comúnmente asociadas a seres humanos inteligentes. Es una rama de la informática-computación que se ocupa de la simulación del comportamiento inteligente.⁷

Asimismo, en su recientemente aprobado Reglamento de Inteligencia Artificial, el Parlamento Europeo afirma que un sistema de IA es:

Un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que

⁶ Como ejemplo, a febrero de 2024, las Comisiones Unidas de Seguridad Ciudadana y de Ciencia, Tecnología e Innovación de la Cámara de Diputados se encuentran en proceso de dictaminación de una eventual Ley de Ciberseguridad. Véase “Comisiones Unidas definen metodología para analizar iniciativa que expide la Ley Federal de Ciberseguridad”, acceso el 24 de marzo de 2024, <https://comunicacionsocial.diputados.gob.mx/index.php/boletines/comisiones-unidas-definen-metodologia-para-analizar-iniciativa-que-expide-la-ley-federal-de-ciberseguridad>.

⁷ Cabanelas Omil, “Inteligencia artificial, ¿Dr. Jekyll o Mr. Hyde?”, 5.



recibe la manera de generar información de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.⁸

De esta manera, el Parlamento Europeo afirma que por medio de esta tecnología, entre otras cuestiones, es posible derivar predicciones. Ello se realiza a través de un proceso en el que la máquina recibe una previa alimentación de datos, los cuales son filtrados aplicando diversas técnicas (como la de redes neuronales artificiales) mediante las cuales es posible identificar patrones que, a su vez, son complementados con información adicional del entorno.⁹ Con base en un algoritmo previamente elaborado, esa información se interpreta y permite derivar las predicciones o inferencias.

Este procedimiento, mucho más complejo en realidad, sorprende por su capacidad deductiva (atributo que hoy reservamos a la especie humana), de ahí que se le haya denominado *inteligencia artificial*. La acuñación de dicho término se dio en un momento en el que la IA se encontraba bajo reflexiones teóricas y sólo en tiempos relativamente recientes hemos podido trasladarla al ámbito empírico; ello, como consecuencia de dos grandes coincidencias: el desarrollo tecnológico computacional y la disponibilidad de enormes bases de datos, estas últimas, recopiladas durante varios años y que día con día continúan desarrollándose.¹⁰

La IA maravilla por sus potencialidades que podrían repercutir en beneficio de toda la humanidad; considérense sus aplicaciones para la detección del cáncer y otras enfermedades, en su empleo para la agricultura y la ganadería, en sus beneficios para la enseñanza y la educación, entre otras expresiones. La IA

⁸ Art. 3.º, numeral 1, Reglamento de Inteligencia Artificial, 13 de marzo de 2024, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf (aceso el 4 de abril de 2024).

⁹ La IA trabaja con un conjunto de diversas herramientas, técnicas y modelos matemáticos de procesamiento de datos; no es un conglomerado unitario sino una amplia pluralidad de estrategias; algunas de ellas son: el aprendizaje automático (*Machine Learning*), el aprendizaje profundo o por capas (*Deep Learning*), la inteligencia generativa (*GenAI*), técnicas de procesamiento de lenguaje natural (LLM), de ingeniería del conocimiento (*Knowledge Engineering*), redes neuronales artificiales (*Artificial Neural Networks*), sistemas multiagente y sistemas expertos (*Multi Agent Systems* y *Expert Systems*), redes bayesianas (*Bayesian Networks*), redes semánticas (*Semantic Networks*), técnicas de *Big Data* y de *Data Mining* (detección de patrones), entre muchas otras. Véase Miró Llinares, “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, 91; Martínez Bahena, “La inteligencia artificial y su aplicación al campo del Derecho”, 828-832; Valls Prieto, *op. cit.*, 9.

¹⁰ O’Neil, *op. cit.*, 75-76.



trabaja con diversas técnicas en cada una de sus aplicaciones y una de ellas es particularmente interesante por la manera en que funciona, aspecto que además de sorprender, también preocupa profundamente por sus posibles efectos secundarios; nos referimos al aprendizaje automático (*Machine Learning*).

Mediante el *Machine Learning*, la máquina no sólo es capaz de derivar predicciones, sino también de aprender y de mejorar constantemente; la clave de ello es el algoritmo mediante el cual está configurada. Un algoritmo es:

una secuencia finita de instrucciones repetibles y únicas que indican una combinación de acciones que realizar para resolver un problema, transformando los datos entrantes (*input*) en información de valor saliente (*output*). En consecuencia, las decisiones automáticas o robóticas son aquellas que involucran los denominados algoritmos de razonamiento que, dadas unas premisas, llegan a conclusiones mediante una programación lógica expresada con un lenguaje capaz de realizar operaciones automáticas a partir de dicho conocimiento preliminar.¹¹

En suma, un algoritmo es un conjunto de reglas que indican a la máquina cómo debe resolver un problema. El algoritmo no da por sí mismo la respuesta, sino que le indica a la máquina cómo debe proceder, cuáles son los pasos, secuencias o estrategias para dar con la respuesta. Mediante el mecanismo de aprendizaje automático, permite a la máquina aprender y mejorar constantemente para encontrar en cada caso la mejor de todas las posibles respuestas, incluso, una que ni siquiera el programador haya eventualmente pensado.¹²

Pues bien, mediante estos algoritmos, la máquina no sólo deduce una posible respuesta a partir de la información que le es suministrada en su programación originaria, sino que, al retroalimentarse de los datos de su entorno, presenta proyecciones que constantemente se reelaboran a partir del cúmulo de información con el que se entrena –y perfecciona–. Un eventual problema se plantea cuando la máquina ya no trabaja bajo los *inputs* originales con la que ha

¹¹ Caterini, “El sistema penal en la encrucijada ante el reto de la inteligencia artificial”, 4.

¹² Un ejemplo clásico que evidencia esta situación es el de la tecnología de *DeepMind*, la cual, en 2014 ideó una estrategia que nunca antes había sido pensada para su victoria en el juego de Atari. Así también, el caso de *AlphaGo* que logró vencer al mejor jugador de “Go” en aquel momento, lo curioso de todo fue que la tecnología logró identificar una estrategia que hasta aquel entonces jamás había sido pensada en el milenar juego de “Go”. Véase Tegmark, *Ser humano en la era de la inteligencia artificial*, 97-105.



sido creada, sino con los que son resultado de su proceso autónomo de aprendizaje que, incluso podrían ser desconocidos para el programador.

Esta cuestión ha conducido a diversas y profundas reflexiones, tanto por lo que hace a la ausencia de una regulación de estos supuestos como a la determinación de una eventual responsabilidad. Sobre todo, en este último aspecto se plantean innumerables complejidades al tratarse de una tecnología que, como se ve, podría haber sido pensada para una determinada función pero que, como consecuencia del aprendizaje automático, a la postre pudiera desempeñar tareas que no se corresponden con su original diseño de programación.

En todo caso, es importante tener presente que un primer elemento consubstancial a la IA es que ésta sólo puede funcionar a partir de los datos que se le suministran. Estos datos son información objetiva, fáctica, que se corresponde con la realidad, pero la realidad muchas veces es injusta y se construye a partir de subjetividades y prejuicios humanos a los que históricamente damos continuidad. Además, quien alimenta a la máquina puede decidir qué información se incluye en su programación y cuál no, por lo que nuevamente se presenta la subjetividad como un riesgo latente.

El derecho moderno aspira a ser garante de la tutela de los derechos fundamentales, pero, para hacerlo, se inserta en el mundo del “deber ser”, porque en los hechos el “ser” continúa siendo, en mayor o menor medida, injusto, discriminatorio, racista, sexista, cruel, inhumano y un largo etcétera de consideraciones negativas. Esta distinción entre el “ser” y el “deber ser” es incapaz de ser procesada por los sistemas informáticos; por ello, aún se requieren diseños tecnológicos que permitan incorporar estas valoraciones, donde, si bien se trabaja con datos reales, también son la consecuencia de condicionamientos y prejuicios que hoy buscamos superar.¹³

A la incapacidad para procesar la distinción entre lo fáctico y las expectativas en el derecho, pueden sumarse otras consideraciones que, si bien son jurídicas, antes, incluso son de tipo ético y moral, tales como la idea de justicia, el bien común, la buena fe, la solidaridad, el compromiso social, la compasión y el

¹³ Maldonado Smith, “Defensa de un modelo principialista en sociedades complejas y plurales”, 112-114.



propio sentido de humanidad; cuestiones que escapan a un análisis matemático, cuantitativo, utilitario o procesal.

Como se infiere, bajo esta situación no parece sorprendente que se pueda apreciar a la IA como un beneficio para nuestra sociedad, pero también como un peligro en múltiples campos; en esta ocasión nos referiremos a los riesgos delictuales.

III. EMPLEO DE LA INTELIGENCIA ARTIFICIAL EN EL ÁMBITO DELICTUAL

La tecnología no es ni buena ni mala, ni justa o injusta; estos atributos siempre son relativos y dependen del uso que el ser humano da a una determinada cuestión.¹⁴ De esta manera, así como la IA puede ser empleada para facilitarnos la realización de actividades cotidianas, a la par, constituye una tecnología susceptible de ser empleada por la delincuencia en la comisión de una inmensa cantidad de delitos.

El campo en donde pueden presentarse conductas criminales con el empleo de IA es, por excelencia, el contexto de la ciberdelincuencia; y esto incluye: estafas, fraudes, suplantaciones de identidad, *deepfakes* o contenido falso; evasión fiscal, blanqueo de capitales, pornografía infantil, delitos en materia de intimidad y libre desarrollo de la personalidad (incluyendo ciberacoso y *sexting*); violaciones al derecho de autor, propiedad intelectual e industrial; acceso no consentido a bases de datos (*hackeo*), *malware*, secuestro de información (*ransomware*), ataques a sistemas informáticos mediante el empleo de *bots*; entre otros.¹⁵

A tal efecto, caben algunas reflexiones. En el terreno del *deepfake*, mediante el empleo de IA es posible recrear de manera idéntica la voz e incluso imagen de políticos y altas personalidades comunicando cierta información o mostrando determinadas conductas. Se trata de contenido en donde las personas involucradas aducen que no es real; sin embargo, la tecnología de IA es tan

¹⁴ Muñoz Vela, “Avanzando en la regulación de la IA. Hacia un equilibrio entre ética, protección de los derechos fundamentales, seguridad, confianza, innovación, desarrollo y competitividad”, 3.

¹⁵ Merlo y Ulloa, “Inteligencia artificial. Su impacto en la justicia penal y en las ciencias forenses: desafíos y oportunidades”, 6.



avanzada que resulta sumamente complejo determinar la autenticidad de esa información.¹⁶ Esta misma situación puede presentarse para cualquier persona y comprometerla en escenarios que pudieran ser delictuales.

Otro ejemplo es el que concierne a la intimidad, la imagen personal, la reputación o el libre desarrollo de la personalidad. Aquí también hemos conocido casos en donde con tecnología de IA ha sido posible elaborar imágenes de contenido sexual de personas que, naturalmente, no han dado su consentimiento para el tratamiento de esa información.¹⁷

Al verse expuestas mediante esta tecnología, las víctimas resienten un profundo daño en múltiples bienes jurídicos (intimidad, imagen, reputación, libre desarrollo de su personalidad, dignidad, entre otros) y, sin embargo, persistirían dudas respecto de si tal conducta podría ser susceptible de sancionarse por alguna figura penal, toda vez que el marco jurídico aún no prevé el empleo de IA como una herramienta o medio comisivo para la realización de esas conductas. En consecuencia, atendiendo al principio de legalidad y más concretamente a su subprincipio de taxatividad, la conducta puede resultar atípica por no contemplarse específicamente dentro de las hipótesis normativas previstas.¹⁸

El caso del *deepfake* es de especial preocupación en un contexto moderno en donde la masificación de la información puede ser compartida en instantes y con un impacto a gran escala en todo el mundo, no sólo por lo que respecta al contenido falso, sino sobre todo por el empleo que de esa información puede hacerse en términos de manipulación y desinformación.¹⁹ De esta manera, bastaría un *clic* para dar inicio a eventuales escenarios de efervescencia social, exacerbación de los ánimos, florecimiento de discursos de intolerancia, odio, instigación y deseos de violencia. Naturalmente, esto constituye a su vez todo un riesgo para la estabilidad de los Estados, los sistemas democráticos, la seguridad pública, la paz y la tranquilidad.²⁰

¹⁶ *Ibid.*

¹⁷ Rodríguez, Andrés, “Acusado un universitario de alterar con inteligencia artificial miles de imágenes de alumnas para venderlas como pornografía”.

¹⁸ Sobre la ausencia de normas penales aplicadas a hipótesis específicas en las que la IA incide en bienes jurídicos fundamentales, véase Morán Espinosa, *op. cit.*, 309.

¹⁹ O’Neil, *op. cit.*, 42.

²⁰ Merlo y Ulloa, *op. cit.*, 6-9.



Además de estos ejemplos, no perdamos de vista que la IA puede ser incorporada a estrategias de guerra (de hecho, ya existen algoritmos de combate empleados en conflictos armados en curso),²¹ terrorismo, crimen organizado y delincuencia ordinaria. También podemos decir que hoy en día existen armas que incluso funcionan de manera autónoma con tecnología de IA;²² por ejemplo, el caso de los drones.

A propósito, otro de los grandes retos legislativos es el de reflexionar en torno a la responsabilidad penal que puede derivarse cuando la IA se encuentra de por medio. El caso de los drones es un estupendo ejemplo, aunque en realidad puede aplicarse a cualquier máquina que incorpore tecnología de IA.

Sobre este aspecto, hoy es claro que una entidad dotada de IA es susceptible de lesionar bienes jurídicos tutelados por el derecho penal y que se trata de entidades con una capacidad autónoma de decisión. Ahora bien, ¿puede hacerse responsable a una máquina de ello? El escenario no es para nada futurista. En este caso, la legislación tradicional puede plantear dos respuestas: 1) la responsabilidad para el programador y 2) para el fabricante.

Pasemos primero a la hipótesis más sencilla: la del fabricante. Aquí, la responsabilidad únicamente podría derivarse como resultado de un defecto de fabricación, que, como consecuencia de ello, derivase en una lesión o puesta en peligro de los bienes jurídicos tutelados por el ordenamiento (vida, salud, integridad personal, etc.).²³ Sin embargo, en este supuesto la IA no supondría un defecto de fabricación, antes bien, nos encontraríamos ante la existencia de un sistema operativo que es funcional y que no presenta defectos en su fabricación. En este escenario, lo que ocurre es que la máquina finalmente es capaz de aprender por sí misma y de manera autónoma dentro del entorno o ambiente en el que se sitúa, cuestión que hipotéticamente podría ser reconducida a una responsabilidad para el programador (y, como autoría mediata, a la persona jurídica o empresa que brinda dichas instrucciones).

²¹ Núñez Zorrilla, *Inteligencia artificial y responsabilidad civil. Régimen jurídico de los daños causados por robots autónomos con inteligencia artificial*, 15.

²² Santos González, “Regulación legal de la robótica y la inteligencia artificial: retos de futuro”, 34.

²³ De la Cuesta Aguado, “Inteligencia artificial y responsabilidad penal”, 55-58; Núñez Zorrilla, *op. cit.*, 18-19.



Sin embargo, la respuesta tampoco es clara en este segundo supuesto. En efecto, el programador habría elaborado un código destinado para una actividad específica que, con el paso del tiempo, la propia máquina (mediante su sistema de aprendizaje autónomo y de redes neuronales auto corregibles) podría hacerlo distinto del planteado originalmente, en estos casos, se observa que la máquina aprende de manera autónoma y desarrolla criterios de inferencia que incluso resultan desconocidos para los programadores. Sobre este punto, es de observar que incluso en los dispositivos que incorporan IA de última generación existe ya una cierta autonomía decisional.²⁴

De esta manera, el comportamiento de la máquina pudiera resultar imprevisible para el propio trabajador; así, parece poco adecuada extenderle la responsabilidad por una conducta que no tiene una relación causal y directa con lo que él ha realizado. Además, este escenario plantea un problema de constitucionalidad, pues atribuirle una responsabilidad sobre lo que él no ha hecho o en lo que no ha contribuido de manera efectiva, supondría un esquema de responsabilidad objetiva; esto es, se haría responsable a una persona que no ha cometido el delito y que estaría respondiendo por los actos de alguien más (en este caso, de una entidad no humana), proceder que contravendría los principios de culpabilidad y de responsabilidad penal individual (artículos 14, segundo párrafo, y 18, párrafo segundo de la Constitución Política de los Estados Unidos Mexicanos).

Como se muestra con este ejemplo, se requieren directrices de orden legislativo que permitan determinar cómo responsabilizar a alguien o a “algo” en estos casos. En el ámbito doctrinal no es raro el planteamiento de propuestas que permitirían una tercera categorización de responsabilidad jurídica,²⁵ en este caso, una forma de responsabilidad para el “sujeto virtual”.²⁶ Esta hipótesis, para

²⁴ Es el caso de la tecnología *Human-out-of-the-loop* (HOOL), donde la intervención humana es inexistente. Sobre este punto, véase Simón Castellano, *op. cit.*, 51; Caterini, *op. cit.*, 4; Basile, “Intelligenza artificiale e diritto penale quattro possibili percorsi di indagine”, 7.

²⁵ Hernández Giménez, “Inteligencia artificial y derecho penal”, 801; Cappellini, “Macchina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale”, 10-13.

²⁶ En caso de optarse por esta vía, habría que darse una respuesta convincente en torno a argumentos jurídico penales como ¿Una máquina puede tener dolo? ¿Puede fallar y, por tanto, obrar con culpa? ¿Podría derivarse una autoría mediata en la que, por ejemplo, la persona humana fuese un instrumento? ¿Puede fincarse una suerte de culpabilidad o de reprochabilidad? ¿Sobre qué base se fundaría, libre arbitrio? ¿Existe en las máquinas capacidad de querer y entender?



nada es descabellada pues, en realidad, la posibilidad de imputar responsabilidad penal a una entidad no humana es algo que ya existe, es el caso de la responsabilidad penal de las personas jurídicas que, tanto en México como en otras partes del mundo ya se regula y que resulta independiente de la responsabilidad que pueda darse para la persona física en lo individual. Si se reflexiona sobre este punto, tratándose de las personas morales, parece más dramático el hecho de imputar responsabilidad penal a una entidad que en el plano real no existe, sino que se trata de una entequeia jurídica. En tal supuesto, puede incluso resultar más plausible²⁷ atribuir responsabilidad a una entidad que en el plano de los hechos es real y que además puede tomar sus propias decisiones.

Otra vía defendida en la doctrina pasa por una anticipación a la eventual lesión de bienes jurídicos, hipótesis que nos evitaría tener que llegar a la aún compleja determinación de decidir si se atribuye responsabilidad penal a una máquina.

Esta vía de anticipación se refiere al código original de programación. En este caso, podría llegarse a un consenso para que toda forma de IA respete una serie de principios inviolables, dentro de ellos se incluiría el de no lesión y el de sujeción al control humano. Así, la omisión de estas pautas podría ser planteada como una hipótesis de responsabilidad para el fabricante o programador, en el sentido de que, si no se respetan estos parámetros mínimos derivados del consenso global, entonces tendrían que responder quienes no incorporaron dichos principios desde la configuración originaria de la máquina (el fabricante o programador).²⁸

Naturalmente, esta segunda vía podría suponer la existencia de sistemas paralelos que garantizaran la extrema dificultad para *hackear* las instrucciones originarias con las que la máquina ha sido fabricada o, incluso mejor, garantizar su invulnerabilidad, así como herramientas que permitan combatir el empleo de

²⁷ Morán Espinosa, *op. cit.*, 304.

²⁸ Esta segunda vía tiene antecedentes en la ciencia ficción, concretamente en los trabajos de Isaac Asimov, quien en obras como *Yo robot* expresaba las 3 leyes de la robótica: “1) Un robot no debe dañar a un ser humano o, por su inacción, dejar que un ser humano sufra daño; 2) 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) Un robot debe proteger su propia existencia, hasta donde esta protección no entre en conflicto con la primera o segunda Leyes”, Asimov, *Yo Robot*, 7.



la IA por parte de la delincuencia. El reto adicional sería el de lograr ese consenso global frente a principios inviolables que toda máquina con tecnología de IA habría de respetar.

Sea como sea y se elija la vía que se desee, la realidad es que a la fecha no se tiene certeza respecto de la responsabilidad que puede derivarse de máquinas que emplean IA, estamos ante un escenario de entera desregulación. Esta situación es altamente preocupante pues coloca en estado de indefensión a quienes son víctimas de conductas lesivas a sus bienes jurídicos fundamentales, un escenario que, a futuro, no hará más que incrementarse, por ello se requieren herramientas legislativas que permitan dar certeza a esta situación.²⁹

IV. EMPLEO DE LA INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA DE PREVENCIÓN DEL DELITO

Una de las principales preocupaciones de la ciudadanía tiene que ver con la seguridad pública, la cual supone prevención del delito y persecución de los delincuentes. A la par, el empleo de mejores herramientas de inteligencia que permitan la valoración del riesgo delictual para, a partir de ello, generar estrategias de combate a la criminalidad.

La IA puede en este rubro ser una estupenda herramienta al incorporarla en los sistemas de videovigilancia, biométricos, financieros, de geolocalización, inclusive, en *software* de identificación de zonas de riesgo o de alta incidencia delictiva (*Hotspots*).³⁰ La IA, a su vez, es capaz de elaborar mapas criminales a partir de diversas bases de información existentes que permitirían a las autoridades estar en el lugar y momento adecuado en el que una conducta delictiva podría estar cometiéndose.³¹ Esta tecnología resulta aplicable para la prevención y detección de cualquier delito, aunque es claro que en aquellos que se realizan mediante sistemas o plataformas digitales (la ciberdelincuencia) las oportunidades son inmensas.

En estos casos, se trabaja con información que permite realizar predicciones muy cercanas a la realidad que, mediante la tecnología de IA, puede darnos

²⁹ Salcedo Beltrán, “Inteligencia artificial y derechos sociales”, 44.

³⁰ Merlo y Ulloa, *op. cit.*, 3-5.

³¹ Hernández Giménez, *op. cit.*, 816.



respuestas inmediatas para atender situaciones de riesgo. Como hemos dicho, la tecnología es neutra y puede ser nuestra mejor aliada para la prevención de la delincuencia; o bien, puede ser un factor más de riesgo si se emplea inadecuadamente, en este caso, cuando no es utilizada exclusivamente con la finalidad de prevención criminal. Sobre este aspecto se presentan preocupaciones en cuanto a la privacidad y proporcionalidad de estas herramientas.

En el primer caso, al ser hoy la tecnología omnipresente, una irrupción de mecanismos de inteligencia y de espionaje público en nuestros teléfonos, computadoras, videocámaras y dispositivos de todo tipo, coloca en riesgo el ámbito de la intimidad y de la privacidad humana, recordándonos escenarios como el planteado por George Orwell en su novela *1984*. Así, bajo una pretendida alegación de combate a la delincuencia, esta tecnología puede socavar la libre expresión, el debate, el disenso, la crítica y demás elementos que conforman las bases de nuestro sistema democrático, sin dejar de lado la tutela a aquel espacio de la persona en la que solamente ella pueda decidir quién accede y quién no (a su intimidad).

Por lo que respecta a la proporcionalidad, hoy podemos advertir en otros contextos cómo la IA puede ser incorporada en herramientas de combate, como el empleo de drones, ya no bajo el control humano sino mediante la tecnología de IA. Estas herramientas también son empleadas en la prevención y combate a la delincuencia; en tal supuesto, surgen dudas respecto de la capacidad de una máquina para identificar y cuantificar una situación de peligro que permita tomar decisiones de manera proporcional al riesgo identificado. Una inadecuada valoración del riesgo puede dar lugar a excesos y, en el peor de los casos, a la pérdida de bienes jurídicos fundamentales como la vida.

Así, en la prevención del delito habría de incorporarse directrices como el control humano en todo momento y no relegarlo a la decisión autónoma de una máquina a fin de evitar situaciones desproporcionadas. A la par, la tecnología de IA tampoco debería de ser incorporada como herramienta en ámbitos que no son los estrictamente señalados para el combate a la criminalidad, pues tal supuesto podría dar lugar a incursiones no consentidas en la privacidad e intimidad de las personas.



V. EMPLEO DE LA INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA DE INDIVIDUALIZACIÓN Y SANCIÓN DEL DELITO

Ahora, reflexionemos acerca del empleo y de los peligros que la IA representa dentro de la fase de individualización de una probable responsabilidad criminal; es decir, en lo que concierne al ámbito jurisdiccional. En este punto, es sabida la histórica demanda ciudadana de justicia y su correlativa deuda por parte de los tribunales, sobre todo porque la justicia pocas veces se brinda de manera pronta y expedita. Por todos es sabido que los procesos jurisdiccionales demoran años y que no siempre se resuelven de una manera “justa”, pues en toda decisión judicial siempre existe un margen de discrecionalidad que puede influir de una manera u otra en la decisión final.

También, históricamente la figura del juez ha sido pensada como el vigilante o el custodio de la ley, expresión de la voluntad soberana y, en un modelo ideal, como un sujeto capaz de conocer a plenitud el ordenamiento jurídico para poder brindar una respuesta exacta a cada cuestión que se le plantea (modelo del juez “Hércules” de Ronald Dworkin).³² Claramente, dicho modelo ideal es incapaz de ser realizado por personas de carne y hueso, aunque sí por máquinas, las cuales tendrían la capacidad de conocer todas las leyes existentes, nacionales e internacionales, así como los precedentes judiciales que se han vertido a lo largo de los años. Con toda esa información, la “máquina-juez” podría darnos una respuesta concreta y exacta a cada problema jurídico que se le plantea, más aún, hacerlo de manera expedita, “imparcial” y eficiente.³³

Este escenario ya no es una ficción. En algunos países la impartición de justicia incorpora *software* que funciona con IA y que permite presentar escenarios predictivos antes, durante y con posterioridad a un proceso:³⁴ con antelación al proceso, mediante el empleo de herramientas preventivas; durante, con medidas cautelares o de seguridad a fin de evitar la continuación de la situación delictiva; y con posterioridad al proceso, en el momento de la

³² Dworkin, *Taking Rights Seriously*, 106-110. Asimismo, sobre el modelo de juez ideal: Esposito, “Note sparse sull’intelligenza artificiale”, 40.

³³ Morán Espinosa, *op. cit.*, 294.

³⁴ En Estados Unidos, además del programa COMPAS se emplean otras aplicaciones como el “Palantir Gotham”; en Inglaterra “Hart”; en Argentina “Prometea”; en Colombia “Pretoria”; entre otros. Véase Hernández Giménez, *op. cit.*, 833-834.



determinación de la responsabilidad penal. Estos escenarios incorporan una enorme pluralidad de datos que permiten, por ejemplo, proponer una determinada medida con base en la peligrosidad y el riesgo de reincidencia delictual (prisión en libertad, prisión cautelar, una pena alta ante el riesgo de reincidencia, negación de beneficios procesales, etcétera).³⁵

En el plano de los beneficios, pensemos cómo podría impactar esta situación en contextos como el de nuestro país donde existe un número abismal de expedientes en trámite en las jurisdicciones federales y locales. Así también, en escenarios donde continuamente subyace el velo de las subjetividades e incluso, de sospechas de corrupción.³⁶ Con herramientas de este tipo, podría pensarse en respuestas inmediatas y objetivas que finalmente concretizaran un auténtico derecho de acceso a la justicia para la ciudadanía.³⁷

Ahora bien, en los contextos comparados, la incorporación de esta tecnología no ha supuesto alcanzar ese escenario ideal; por tanto, tampoco ha existido una inmediata celeridad procesal ni una mayor imparcialidad o eficiencia. Antes bien, la incorporación de esta tecnología ha supuesto grandes problemas y críticas; el caso que por excelencia refleja esta situación es el del programa COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) empleado en Estados Unidos.³⁸

Se trata de un programa informático desarrollado por la empresa Northpoint (hoy llamada Equivant),³⁹ que incorpora tecnología de IA para elaborar predicciones de riesgo delictual. Con base en este programa, los jueces disponen de información que les indica la probabilidad de que una persona sea o no responsable de un delito y, con apoyo en esta herramienta, dictan sentencia. Aquí, la valoración se funda en el riesgo delictual; es decir, en la probabilidad de que esa persona pueda cometer otro delito en el futuro.

³⁵ Se trata de sistemas expertos de IA que, con base en información sobre el caso y el imputado, permiten arrojar predicciones sobre la responsabilidad. Algunos autores condenan abiertamente que el factor a considerar sea la reincidencia porque se está juzgando a una persona no por lo que específicamente ha hecho en un caso concreto, sino que se le juzga por su pasado, por lo que hizo antes y por lo que ya ha pagado en otro proceso penal. Véase Salcedo Beltrán, *op. cit.*, 49.

³⁶ Esposito, *op. cit.*, 42.

³⁷ Basile, *op. cit.*, 16; Merlo y Ulloa, *op. cit.*, 50.

³⁸ Caterini, *op. cit.*, 6; y Miró Llinares, *op. cit.*, 108-109.

³⁹ Basile, *op. cit.*, 17.



Un primer problema en la utilización de esta herramienta es que permite inferir una responsabilidad, no con base en evidencia concreta relacionada con el caso en cuestión, sino con la probabilidad de que una cierta persona pueda o no delinquir. En otras palabras, trabaja sobre la base del riesgo y no sobre la concreta responsabilidad, un proceder que es típico de modelos de anticipación delictual de los llamados “derecho penal de autor” y “derecho penal del enemigo”.⁴⁰ Tales modelos se apoyan, no en la conducta concreta desplegada y en la efectiva lesión de un bien jurídico, sino en la peligrosidad del agente; se trata de modelos que no son nuevos, pues nos remiten a escenarios como el de la escuela positiva del derecho penal y a uno de sus mayores exponentes, Cesare Lombroso, quien pretendía identificar a posibles criminales desde su nacimiento a través de rasgos somáticos (el criminal *nato*).⁴¹

La tecnología que se encuentra detrás del programa COMPAS no es muy distinta al pensamiento de Lombroso,⁴² pues las predicciones delictivas son alimentadas bajo algoritmos complejos que incorporan variables no dependientes de la conducta concreta desplegada (el delito), sino de cuestiones más bien correlacionadas con la vida y el contexto de la persona. Dentro de estas variables se encuentran las siguientes: educación, historia personal, barrio donde vive, edad, antecedentes criminales, nivel socioeconómico, estilo de vida, personalidades, actitudes, situación familiar y laboral, círculo de amistades, experiencias traumáticas, historial médico, origen étnico, entre otras.⁴³

Estos factores han levantado enormes críticas en tanto que, con tales insumos, el programa dispone de diversos sesgos, cuestión que ha sido confirmada en la mayor proporción de sentencias en las que se condena a personas de un cierto origen étnico y racial. Incluso, la cuestión ha sido llevada ante la Corte Suprema de Wisconsin, la cual ha reconocido el sesgo existente; sin embargo, también ha considerado que dicha herramienta puede seguir aplicándose en tanto que, para el poder judicial, dicho programa no es la única

⁴⁰ Sobre estas figuras, permítase el reenvío a Maldonado Smith, “Pena y reinserción social, reflexiones en torno al ergastolo encubierto”, 117.

⁴¹ Moccia, “I nipotini di Lombroso: neuroscienze e genetica nel diritto penale”, 681-684.

⁴² Basile, *op. cit.*, 10.

⁴³ Caterini, *op. cit.*, 6; Basile, *op. cit.*, 17.



herramienta de prueba a disposición del juez, sino tan sólo una de las múltiples opciones.⁴⁴

En efecto, el programa COMPAS no es racista por sí mismo. Ninguna forma de IA es capaz de crear algo de la nada, pues se alimenta de los datos con que ha sido programada y con lo que aprende de su entorno.⁴⁵ Así, la componente racial, sexista, de origen étnico, social, económica y cultural que COMPAS incorpora, en realidad ha estado presente a lo largo de la historia de la impartición de justicia en Estados Unidos, y no sólo ahí, sino en muchas otras latitudes.⁴⁶ De esta manera, el programa sólo da continuidad a los prejuicios, estereotipos y sesgos ya existentes desde hace mucho tiempo.⁴⁷ Como señala O'Neil, estos programas funcionan a partir de modelos matemáticos que, por su naturaleza, son programados con la información del pasado, con la expectativa de que lo que ha ocurrido vuelva a repetirse (probabilidad y estadística).⁴⁸

Aun así, en el hipotético caso de conceder algún valor a estos elementos, como hemos dicho, no estaríamos sino reproduciendo pensamientos deleznable que fundan los orígenes de la delincuencia en el color de la piel, el origen nacional o étnico, la religión, la educación, el poder adquisitivo u otros elementos.

Ahora bien, COMPAS y otros programas que emplean IA también son preocupantes por otro aspecto y es el que se refiere a los algoritmos empleados

⁴⁴ El caso emblemático es *State of Wisconsin vs. Eric L. Loomis*, 881 N.W.2d, 749, de 13 de julio de 2016. En este caso, pese a reconocer la existencia de sesgos, la Corte de Wisconsin determinó que el empleo de COMPAS resultaba lícito en tanto que constituía sólo una de las herramientas a disposición del juez, no siendo definitiva para la determinación de responsabilidad penal. La sentencia está disponible en <https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690>, acceso el 17 de febrero de 2023. Sobre este tema, véase Freeman, “Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in *State vs. Loomis*”, 76-106; De Miguel Beriain y Pérez Estrada, “La inteligencia artificial en el proceso penal español: un análisis de su admisibilidad sobre la base de los derechos fundamentales implicados”, 546-549.

⁴⁵ O'Neil, *op. cit.*, 20; Valls Prieto, *op. cit.*, 10.

⁴⁶ Hernández Giménez, *op. cit.*, 824-825.

⁴⁷ Miró Llinares, *op. cit.*, 123-124. Otro aspecto que eventualmente podría dar lugar a alguna forma de responsabilidad (vía la igualdad y no discriminación), pero de tipo laboral y administrativo, es la que se refiere al empleo de programas con IA usados en el ámbito privado para la contratación y el desempeño de las actividades. En estos casos, se ha acreditado la existencia de sesgos fundados en factores como la edad o el género. Véase Criado Pérez, *Invisible women*, 164-168.

⁴⁸ O'Neil, *op. cit.*, 38.



para su funcionamiento. El problema deriva de que, quien dispone de la tecnología necesaria para el desarrollo, innovación e incorporación de esta tecnología no es el Estado o la sociedad civil, sino grandes y poderosas empresas (como ocurre en el caso de COMPAS). Las disposiciones nacionales e internacionales disponen, a su vez, de estrategias normativas por las que se busca incentivar la inversión en el desarrollo tecnológico; una de ellas es el secreto empresarial.⁴⁹

Así, grandes empresas invierten desarrollando tecnología que, en principio, permitiría obtener mejores productos a un menor costo de producción y, por ende, ofrecerlo a un precio más accesible para el consumidor. Como contrapartida, los Estados disponen de herramientas jurídicas que garantizan beneficios empresariales por la vía de la propiedad intelectual, industrial y del secreto empresarial. En este esquema, la información relativa al funcionamiento de programas informáticos que emplean IA no es para nada pública, antes bien, se encuentra protegida por el secreto empresarial.

La consecuencia de esta protección es que resulta sumamente complejo poder conocer a detalle cuál es el mecanismo y cuáles son las variables de información que son empleadas para el funcionamiento de dichos programas y, por ende, se dificulta conocer el mecanismo o proceso por el que se llega a una determinada conclusión; es decir, funcionan bajo la opacidad. En este punto, si bien hemos dicho que el programa no es racista por sí mismo, sí puede serlo el programador que nutrió al sistema mediante la incorporación –consciente o no– de determinada información; en todo caso, no podemos saberlo porque dicha información es parte del secreto empresarial.⁵⁰

Así, nos encontramos con una situación en la cual, por una parte, se aplican estas herramientas para facilitar la toma de decisiones, pero, por otra, no puede conocerse cuál es el procedimiento bajo el cual toman sus decisiones.⁵¹ Se piense, por ejemplo, en el programa COMPAS que puede ser empleado por el juzgador para tomar su decisión; sin embargo, el acusado se encuentra

⁴⁹ Simón Castellano., *et. al.*, *Derecho Digital. Esquemas y casos prácticos para su estudio*, 108-109.

⁵⁰ Hernández Giménez, *op. cit.*, 824-826.

⁵¹ Sobre la secrecía de estos sistemas, la necesidad de medidas de transparencia y garantías para una defensa adecuada: Caterini, *op. cit.*, 6.



imposibilitado para organizar una defensa estratégica al no disponer de la mínima información respecto del porqué el programa informático tomó la decisión en cuestión. Como se advierte, es un área en donde no existe disposición regulatoria que permita compatibilizar los intereses económicos en juego (de las empresas), con la tutela de garantías sustantivas y procesales de orden penal.

De ahí que otra de las demandas en este rubro sea la de incorporar principios de transparencia que nos permitan conocer a detalle cuál es el proceso de la toma de decisiones, cómo se alimenta, cuáles son sus variables y cómo se estructura un determinado programa. Asimismo, también es importante la “traducción” de esta información no sólo al lenguaje matemático e informático, sino también al jurídico (del juez y la defensa), así como al de los ciudadanos comunes, que son quienes pueden verse afectados por el empleo de este tipo de tecnología.

VI. REFLEXIONES DE ORDEN LEGISLATIVO Y CRIMINAL

Como hemos expresado, la inteligencia artificial se encuentra presente prácticamente en todos los órdenes, incluido el del derecho penal. En este caso, hemos presentado reflexiones desde el análisis criminal, que, no obstante (y salvo sus respectivas adaptaciones), pueden extrapolarse a cualquier ámbito. En todo caso, lo cierto es que nos encontramos frente a una tecnología desregulada y que, por tanto, genera dudas y preocupaciones colocándonos en los terrenos de la inseguridad jurídica.

La preocupación que aquí se avanza no sólo es de México, sino de todos los países; por tal razón, en otras latitudes, como el contexto europeo, se han emprendido acciones comunitarias para regular la IA.⁵² El 13 de marzo de 2024

⁵² Algunos antecedentes son los siguientes: resolución del Parlamento Europeo, de 16 de febrero de 2017, sobre normas de derecho civil y robótica; resolución del Parlamento Europeo sobre aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, de 20 de octubre de 2020; Propuesta de Reglamento del Parlamento Europeo y del Consejo, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, de 21 de abril de 2021; Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital, aprobada por el Parlamento Europeo, el Consejo y la comisión el 26 de enero de 2022. El 9 de diciembre de 2023, fue aprobado en el seno del Parlamento Europeo y del Consejo, un acuerdo para la versión final del Reglamento de Inteligencia Artificial para la Unión Europea (UE), mismo que ha sido aprobado en el mes de marzo de 2024. Su entrada en vigor se prevé por etapas para que opere plenamente en 2026. En otros contextos, además del europeo, se destaca la Recomendación sobre Inteligencia Artificial adoptada por la Organización para la Cooperación y el Desarrollo Económico (OCDE) el 22 de mayo de 2019 y la Recomendación sobre la Ética



fue aprobada por el Parlamento Europeo la primera normativa internacional que regula el empleo de la tecnología de IA; se trata de un reglamento vinculante para todos los Estados miembros de la Unión Europea (UE) y que tiene por objetivo reconducir todo empleo de esta tecnología a principios y valores comunes de la UE, como lo son el respeto y la tutela de los derechos humanos.

Como se observa, en el contexto europeo se trabaja de manera conjunta sobre un fenómeno que no es consubstancial de un solo Estado, sino de toda la Unión, en donde, por tanto, se requieren de medidas conjuntas. En esta perspectiva de derecho comparado, para un fenómeno que afecta ya no sólo a la UE sino a todo el mundo, los lineamientos generales identificados pueden ser de utilidad.

Así, por ejemplo, en dicha normativa se distinguen diversos grados de desarrollo de la IA que determinan mayor o menor control en su uso; en algunos casos, el empleo de esta tecnología sería incluso prohibido ante la peligrosidad de esta tecnología. Esto es, todo proyecto legislativo habría de distinguir la proporción de IA que se requiere emplear para cada caso.⁵³

Asimismo, se identifican principios comunes a observar en toda aplicación de la IA y que guardan coherencia con los problemas que en los apartados anteriores hemos expresado; estos principios serían los siguientes: transparencia y explicabilidad (hipótesis en las que el secreto empresarial debería de ceder, así como la traducción de esa información al lenguaje jurídico y común); uso seguro y fiabilidad durante toda su vida útil (garantía de que el producto es seguro); trazabilidad (el recorrido que la tecnología ha tenido a fin de poder fincar eventuales responsabilidades); respeto a los derechos fundamentales (lo que supone también entrenar a las máquinas en este concepto); perfiles de no discriminación (no estigmatización); tutela de la privacidad y gobernanza de los propios datos (en el caso de particulares); consentimiento informado (por ejemplo, en el tratamiento de datos); rendición de cuentas (que pueda revisarse, que sea auditable); programación con perspectiva de género (nuevamente, entrenar a las máquinas en lo que supone este concepto y sus implicaciones); uso responsable y sostenido del medio ambiente; análisis y gestión de riesgos (antes

de la Inteligencia Artificial adoptada por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), en 2021. Véase Muñoz Vela, *op. cit.*, 5-17.

⁵³ En la doctrina se distingue entre una IA débil, fuerte o general y una súper IA. Asimismo, el Reglamento de Inteligencia Artificial de la UE distingue entre una IA de riesgo inexistente, bajo, alto e inaceptable. Véase Abadías Selma, *op. cit.*, 43-45.



de su lanzamiento y durante todo su funcionamiento); control y supervisión humana (que también permita fincar eventuales formas de responsabilidad y anticiparse a escenarios en donde la máquina sobrepasa al humano); mecanismos de verificabilidad de la información (para evitar *deepfakes*), entre otros.⁵⁴

Como se advierte, el destino nos ha alcanzado; las máquinas y la IA se encuentran por doquier simplificándonos la vida, pero también generando nuevos escenarios de riesgo. En este contexto, se observa como impostergable el dotarnos de herramientas jurídicas que nos permitan afrontar una realidad tan compleja en la que necesitamos instrumentos que nos den certeza y seguridad. En este trabajo, además, resulta iluso pretender afrontar la situación únicamente mediante el empleo de juristas; en esta labor debe contarse con la participación de programadores, informáticos, analistas de riesgos, con las empresas y con la ciudadanía.

VII. CONCLUSIONES

Como resultado de la investigación que aquí se ha planteado podemos derivar las siguientes conclusiones:

La IA no es una cuestión del futuro, es del ahora inmediato y está presente en una amplia pluralidad de situaciones de las que muchas veces no somos conscientes.

La IA nos proporciona una gran cantidad de beneficios que permiten hacer nuestra vida más fácil; sin embargo, también supone riesgos y peligros por el empleo que de esta tecnología puede hacer la delincuencia. El riesgo es aún mayor porque se trata de un ámbito desregulado, es decir, no existen normas jurídicas que permitan dar respuesta a complejidades que la IA ha supuesto y en la que potencialmente pueden resultar afectados bienes jurídicos de orden fundamental.

En este trabajo se han planteado tres escenarios en el ámbito criminal: el empleo de la IA como instrumento delictual; el empleo de la IA como

⁵⁴ Principios extraídos a partir del Reglamento de Inteligencia Artificial, *op. cit.* Asimismo: Muñoz Vela, *op. cit.*, 5-9; Santos González, *op. cit.*, 32-36; Criado Pérez, *op. cit.*, 165.



herramienta para la prevención del delito, y el empleo de la IA en la individualización y aplicación de responsabilidades penales.

Con relación al empleo de la IA como instrumento delictual, la utilización de esta tecnología para la realización de actividades delictuales es enorme, particularmente en el ámbito de la ciberdelincuencia. Resulta necesario adaptar nuestro ordenamiento jurídico a nuevas formas delictivas que ya operan y en donde, de no hacerlo, pueden crearse espacios para la impunidad; en particular, se ha destacado el caso del *deepfake* y la afectación a bienes jurídicos como la imagen, el honor, la intimidad, el libre desarrollo de la personalidad y la libertad sexual. A la par, una agenda pendiente es la determinación de la responsabilidad que puede derivarse por el empleo de la IA. Sobre este último punto se ha destacado la prevención mediante principios base en la programación, así como la eventual responsabilidad que pueda darse para entidades como las máquinas.

Acerca del empleo de la IA como herramienta de prevención delictual, se han destacado los enormes beneficios que pueden derivarse de emplear esta herramienta en el combate a la delincuencia; sin embargo, se ha advertido de los riesgos que pueden derivarse cuando el empleo de dicha tecnología no es proporcional y trasgrede los objetivos inicialmente planteados de prevención delictiva. En este punto, es necesario regular el empleo de dichas herramientas para garantizar la privacidad, la intimidad y la autodeterminación en el manejo de los propios datos, así como observar en todo momento el control humano de esta tecnología.

Con relación al empleo de la IA como herramienta para la individualización y determinación de la responsabilidad penal, se ha presentado cómo hoy en día existen herramientas usadas en el ámbito judicial que permiten generar escenarios predictivos en los que los jueces pueden basarse para tomar sus decisiones. Se ha expresado que en esos escenarios pueden estar presentes sesgos, estereotipos o prejuicios, derivados de los datos con los que ha sido alimentado el programa informático. En este punto, resulta importante saber bajo qué parámetros ha sido elaborado el programa, cómo se ha alimentado y con qué algoritmos toma sus decisiones. Para ello, es importante compatibilizar el secreto empresarial y las actividades de impartición de justicia. Asimismo, estas herramientas no pueden ser un criterio unívoco para la toma de decisiones judiciales, ni tanto menos ser empleadas cuando no valoran la conducta



efectivamente desplegada por el agente (derecho penal del acto), sino solamente su pasado o su historia (derecho penal de autor).

Acerca de las recomendaciones que eventualmente pudieran ser incorporadas en una ley que regule el empleo de la IA, se han tomado como ejemplo las directrices que al día de hoy se aplican en el contexto europeo y que pudieran servirnos de ejemplo. Como grandes principios a seguir en una regulación de IA se podrían incorporar las siguientes: distinción entre los diversos grados o implicaciones que la IA puede suponer (proporcionalidad); transparencia y explicabilidad; uso seguro y fiabilidad; trazabilidad; respeto a los derechos fundamentales; perfiles de no discriminación; tutela de la privacidad y gobernanza de datos; consentimiento informado; rendición de cuentas; programación con perspectiva de género; uso responsable y sostenido del medio ambiente; análisis y gestión de riesgos; control y supervisión humana; mecanismos de verificabilidad de la información; entre otros.

VIII. BIBLIOGRAFÍA

- Abadías Selma, Alfredo. *Justicia juvenil e inteligencia artificial en la era de la cultura "Touch"*. Valencia: Editorial Tirant Lo Blanch, 2022.
- Asimov, Isaac. *Yo Robot*. Traducido por Manuel Bosch Barrett. Buenos Aires: editorial Edhasa, 2021.
- Basile, Fabio. "Intelligenza artificiale e diritto penale quattro possibili percorsi di indagine", *Diritto penale e uomo* (Milán), (septiembre 2019): 1-33. <https://archiviodpc.dirittopenaleuomo.org/upload/3089-basile2019.pdf>.
- Cabanelas Omil, José. "Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde?", *Mercados y negocios* (Jalisco), (julio-diciembre 2019): 5-16. <https://www.redalyc.org/journal/5718/571860888002/571860888002.pdf>.
- Cappellini, Alberto. "Macchina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale", *Criminalia. Annuario di scienze penalistiche* (Pisa), (2019): 499-520. <https://discrimen.it/machina-delinquere-non-potest-brevi-appunti-su-intelligenza-artificiale-e-responsabilita-penale/>.
- Caterini, Mario. "El sistema penal en la encrucijada ante el reto de la inteligencia artificial", *Revista de Internet, Derecho y Política* (Barcelona) 35, (marzo 2022): 1-19. <https://dialnet.unirioja.es/descarga/articulo/8399208.pdf>.



- Criado Pérez, Caroline. *Invisible women*. Nueva York: Editorial Abrams Press, 2021.
- Parlamento Europeo. “Reglamento de Inteligencia Artificial, Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, COM(2021)0206-C9-0146/2021-2021/0106(COD)”. 1
https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_ES.pdf.
- De la Cuesta Aguado, Paz. “Inteligencia artificial y responsabilidad penal”, *Revista Penal Mexicana* (Ciudad de México) 16-17, (marzo 2019-febrero 2020): 51-62.
<https://revistaciencias.inacipe.gob.mx/index.php/01/article/view/336/266>.
- De Miguel Beriain, Iñigo y Miren Josune Pérez Estrada. “La inteligencia artificial en el proceso penal español: un análisis de su admisibilidad sobre la base de los derechos fundamentales implicados”, *Revista de Derecho UNED* (Madrid) 25, (2019): 531-561.
<https://revistas.uned.es/index.php/RDUNED/article/download/27013/21089>.
- Dworkin, Ronald. *Taking Rights Seriously*. Cambridge: Harvard University Press, 1978.
- Esposito, Andreana. “Note sparse sull’intelligenza artificiale”, en Giuliano, Balvi, Federica De Simone, Andreana Esposito y Stefano Manacorda (coordinadores). *Diritto penale e intelligenza artificiale, “Nuovi scenari”*. Materiali e Studi di Diritto Pubblico. Turín: Editorial Giappichelli, 2022: 37-49.
<https://www.giappichelli.it/media/catalog/product/openaccess/9788892177710.pdf>.
- Freeman, Katherine. “Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in *State v. Loomis*”, *North Carolina Journal of Law & Technology* (Chapel Hill) 18, n.º 5, 2016: 75-106.
<https://scholarship.law.unc.edu/cgi/viewcontent.cgi?article=1332&context=ncjolt>.



- Hernández Giménez, María. “Inteligencia artificial y derecho penal”, *Actualidad Jurídica Iberoamericana* (Valencia) 10 bis, (junio 2019): 792-843.
<https://www.revista-aji.com/wp-content/uploads/2019/06/792-843.pdf>.
- Maldonado Smith, Mario Eduardo. “Defensa de un modelo principialista en sociedades complejas y plurales”, *Iuris Tantum* (Ciudad de México) 37 (enero-junio 2023): 96-124.
- “Pena y reinserción social, reflexiones en torno al ergastolo encubierto”, *Revista Quórum Legislativo* (Ciudad de México) 136, (diciembre 2021): 93-125.
<https://portalhcd.diputados.gob.mx/PortalWeb/Micrositios/edc19728-d484-4cfd-ba42-e4501170f621/PUBLICACIONES/QUORUM%20LEGISLATIVO/Quorum-136.pdf>.
- Martínez Bahena, Goretty Carolina. “La inteligencia artificial y su aplicación al campo del Derecho”, *Alegatos* (Ciudad de México) 82, (septiembre-diciembre 2012): 827-846.
<https://alegatos.azc.uam.mx/index.php/ra/article/view/205>.
- Merlo, Guillermo y José Ricardo Ulloa. “Inteligencia artificial. Su impacto en la justicia penal y en las ciencias forenses: desafíos y oportunidades”, *Revista pensamiento penal* (Buenos Aires) 475, (julio 2023).
<https://www.pensamientopenal.com.ar/doctrina/90869-inteligencia-artificial-su-impacto-justicia-penal-y-ciencias-forenses-desafios-y>.
- Miró Llinares, Fernando. “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, *Revista de Derecho Penal y Criminología* (Madrid) 20, (julio 2018): 87-130.
doi: <https://doi.org/10.5944/rdpc.20.2018.26446>.
- Moccia, Sergio. “I nipotini di Lombroso: neuroscienze e genetica nel diritto penale”, *Diritto Penale e Processo* (Milán) 5, (2016). 681-684.
<https://www.rivisteweb.it/>.
- Morán Espinosa, Alejandra. “Responsabilidad penal de la Inteligencia Artificial (IA). ¿La próxima frontera?”, *Revista del Instituto de Ciencias Jurídicas de Puebla* (Puebla) 48, n.º 15 (julio-diciembre 2021): 289-323.
<https://www.scielo.org.mx/pdf/rius/v15n48/1870-2147-rius-15-48-289.pdf>.
- Muñoz Vela, José Manuel. “Avanzando en la regulación de la IA. Hacia un equilibrio entre ética, protección de los derechos fundamentales, seguridad, confianza, innovación, desarrollo y competitividad”, *Derecho*



- digital e innovación* (Madrid) 14, (octubre-diciembre 2022): 1-46.
<https://web.laley.es/revistas-laley/derecho-digital/>.
- Núñez Zorrilla, Ma. Carmen. *Inteligencia artificial y responsabilidad civil. Régimen jurídico de los daños causados por robots autónomos con inteligencia artificial*. Colección Scientia Iuridica. Madrid: Editorial Reus, 2019.
- O’Neil, Cathy. *Weapons of Math Destruction. How big data increases inequality and threatens democracy*. Nueva York: Editorial Crown, 2016.
- Rodríguez, Andrés. “Acusado un universitario de alterar con inteligencia artificial miles de imágenes de alumnas para venderlas como pornografía”, *Inteligencia Artificial*, *El País*, 14 de octubre de 2023, <https://elpais.com/mexico/2023-10-14/acusado-un-universitario-de-alterar-con-inteligencia-artificial-miles-de-imagenes-de-alumnas-para-venderlas-como-pornografia.html>.
- Salcedo Beltrán, Carmen. “Inteligencia artificial y derechos sociales”, en Abadías Selma, Alfredo y Guillermo García González (coordinadores). *Protección de los trabajadores e inteligencia artificial: la tutela de los derechos sociales en la cuarta revolución industrial*. Barcelona: Editorial Atelier, 2022: 43-82.
- Santos González, María José. “Regulación legal de la robótica y la Inteligencia Artificial: retos de futuro”, *Revista Jurídica de la Universidad de León* (Castilla y León) 4, (2017): 25-50.
doi: <https://doi.org/10.18002/rjule.v0i4.5285>.
- Simón Castellano, Pere. *Justicia cautelar e inteligencia artificial. La alternativa a los atávicos heurísticos judiciales*. Barcelona: Editorial J. M. Bosch, 2021.
- Simón Castellano, Pere, Alfredo Abadías Selma, Mario Eduardo Maldonado Smith, Javier Dorado Ferrer y Guillem Castro Izquierdo. *Derecho digital. Esquemas y casos prácticos para su estudio*. A Coruña: Editorial Colex, 2022.
- Soria Olivas, Emilio, Pablo Rodríguez Belenguer, Enrique García Vidal, Fran Vaquer Estalrich, Juan Vicent Camisón y Jorge Vila Tomás. *Inteligencia Artificial. Casos prácticos con Aprendizaje Profundo*. Madrid: Ra-Ma editorial, 2022.
- Tegmark, Max. *Vida 3.0. Ser humano en la era de la Inteligencia Artificial*. Traducido por Marcos Pérez Sánchez. Taurus: Madrid, 2018.



Valls Prieto, Javier. “Sobre la responsabilidad penal por la utilización de sistemas inteligentes”, *Revista Electrónica de Ciencia Penal y Criminología* (Granada) 24, (2022): 1-35.
<https://digibug.ugr.es/bitstream/handle/10481/87192/recpc24-27-1.pdf?sequence=1&isAllowed=y>.

