

Responsabilidad penal en decisiones automatizadas: ¿Quién responde cuando la inteligencia artificial se equivoca?

*Criminal Liability in Automated Decisions:
Who Is Responsible When Artificial
Intelligence Makes a Mistake?*

Gerardo Monroy Rosas

Egresado de la maestría en Ciencia Jurídico Penal
del INACIPE | Servidor público en la FGJ CDMX
ORCID: <https://orcid.org/0009-0007-6741-934X>
Correo electrónico: gerardomonroyrosas@gmail.com

Melany Jocelyn Monroy Santoyo

Estudiante de la maestría en Ciencia de Datos
en la Universidad Panamericana
ORCID: <https://orcid.org/0009-0001-8997-2978>
Correo electrónico: melany.monroy.santoyo@gmail.com

Responsabilidad penal en decisiones automatizadas:

¿Quién responde cuando la inteligencia artificial se equivoca?

Criminal Liability in Automated Decisions: Who Is Responsible

When Artificial Intelligence Makes a Mistake?

Gerardo Monroy Rosas

Melany Jocelyn Monroy Santoyo



Recepción: 6/03/2026



Aceptación: 14/04/2026



DOI: <https://doi.org/10.57042/rmcp.v9i29.1078>

Resumen

En los últimos años, la inteligencia artificial se ha incorporado en distintos ámbitos sociales, lo cual ha ocasionado que surjan nuevas interrogantes en el ámbito de la teoría del delito. En particular, por el uso de algoritmos capaces de tomar decisiones autónomas, que nos plantean desafíos relevantes en materia de atribución de responsabilidad penal cuando esas decisiones producen consecuencias jurídicas o daños a terceros.

Este trabajo analiza los vacíos normativos en contextos de decisión y examina si la teoría del delito ofrece herramientas suficientes para determinar la responsabilidad penal derivada de sistemas automatizados, con el fin de establecer un modelo de imputación coherente y eficaz.

Palabras clave

Inteligencia artificial, teoría del delito, responsabilidad penal, algoritmos, imputación objetiva, acción.

Abstract

In recent years, artificial intelligence has been incorporated into various social spheres, giving rise to new questions within the field of Criminal Theory. In particular, the use of algorithms capable of making autonomous decisions poses significant challenges regarding the attribution of criminal liability when such decisions produce legal consequences or harm to third parties. This paper analyzes the existing normative gaps in decision-making contexts and examines whether Criminal Theory provides sufficient tools to determine criminal liability arising from automated systems, with the aim of establishing a coherent and effective models of imputation.

Keywords

Artificial intelligence, criminal law, criminal liability, algorithms, automated decision-making.

Sumario

I. Enfoque y estado de la cuestión. II. Concepto y definición normativa de la IA en México. III. La inteligencia artificial frente al concepto de acción. IV. Vacíos de responsabilidad penal frente a decisiones automatizadas. V. Conclusiones. VI. Referencias.

I. Enfoque y estado de la cuestión

En los últimos años, el uso de la inteligencia artificial ha comenzado a ser muy relevante en distintos ámbitos de la vida: social, económica y tecnológicamente. La IA se encarga de analizar grandes volúmenes de

información, para así poder reconocer patrones y generar respuestas, todo esto de forma automatizada. Estas herramientas no sólo han impactado en los ámbitos previamente mencionados, sino que también han comenzado a impactar en el ámbito jurídico, empleándose para el análisis de información legal o como apoyo para la toma de decisiones.

En los últimos años, el uso de herramientas basadas en inteligencia artificial se ha expandido en distintos ámbitos institucionales, y claro, el ámbito jurídico no se queda atrás, pues ciertas tecnologías se utilizan para analizar información legal o como apoyo en la toma de decisiones.

Un ejemplo conocido del uso de sistemas algorítmicos en el ámbito judicial es el sistema COMPAS, utilizado en algunos estados de Estados Unidos, principalmente para estimar el riesgo de reincidencia de una persona acusada o condenada. Este sistema genera una puntuación del 1 al 10, basada en distintos factores como son los antecedentes penales y las condiciones sociales, y analiza respuestas a cuestionarios, todo esto con el objetivo de ayudar a jueces y autoridades a la hora de la toma de decisiones (Angwin, Larson, Mattu y Kirchner, 2016).

De manera similar, en algunos contextos policiales se han utilizado herramientas de análisis predictivo que emplean datos históricos de criminalidad para identificar zonas donde es más probable que ocurran determinados delitos (Perry, McInnis, Price, Smith y Hollywood, 2013).

Estas herramientas existen desde hace años, sin embargo, actualmente han alcanzado mayor auge. El crecimiento de la IA ha generado expectativas respecto a su potencial y, de igual manera, preocupaciones sobre qué pasará cuando las decisiones o recomendaciones generadas por estos sistemas automatizados puedan influir en

situaciones que afectan directamente los derechos de las personas, o a quién se le adjudicará la responsabilidad de sus actos, en el caso de que se conviertan en sistemas autónomos en su totalidad, ¿se podría culpar al creador, incluso si la inteligencia artificial llega a pensar por sí sola en algún momento?

Desde la perspectiva del derecho penal, estos cambios tecnológicos plantean desafíos significativos. Tradicionalmente, el derecho penal se ha construido sobre la idea de que las conductas reprochables son realizadas por individuos capaces de actuar con voluntad y conciencia. Conceptos fundamentales como acción, dolo, culpabilidad, imputabilidad, etcétera, parten del supuesto de que existe una persona que realiza una conducta y que puede ser considerada responsable de sus actos.

En este contexto, la incorporación de sistemas de inteligencia artificial en procesos de toma de decisiones genera nuevas preguntas jurídicas, ya que cuando una decisión automatizada por la inteligencia artificial produce consecuencias perjudiciales, resulta necesario preguntarse a quién corresponde atribuir la responsabilidad por estos resultados: a quienes desarrollaron el sistema, a las entidades que lo implementaron o a quienes decidieron utilizarlo.

Dado todo esto, el uso de sistemas de inteligencia artificial en procesos de decisión plantea nuevos desafíos para el derecho penal, especialmente en lo que respecta a la atribución de responsabilidad cuando estas decisiones generan consecuencias perjudiciales.

El uso e incorporación de estas herramientas han generado grandes expectativas respecto a su capacidad para mejorar la eficiencia de algunos procesos institucionales; sin embargo, su uso también nos plantea distintas

interrogantes desde la perspectiva del derecho penal. En particular, surge la cuestión de cómo debe atribuirse responsabilidad cuando una decisión generada o influida por un sistema automatizado produce consecuencias jurídicas negativas o afecta los derechos de terceros.

Debemos tomar en cuenta que, a diferencia de los sujetos tradicionales del derecho penal, los sistemas de inteligencia artificial no poseen voluntad, conciencia ni personalidad jurídica. Esto dificulta la visión y el encuadre dentro de las categorías clásicas mediante las cuales el derecho penal determina la responsabilidad de una conducta. En consecuencia, cuando una decisión automatizada provoca un daño o genera efectos relevantes, resulta necesario preguntarse quién debería responder por estas consecuencias.

En este contexto, el presente trabajo tiene como objetivo analizar los desafíos que plantea la atribución de responsabilidad penal frente al uso de sistemas de inteligencia artificial en procesos de toma de decisiones. Para ello, se examinará cómo el avance de la IA introduce nuevos cuestionamientos para el derecho penal y se analizarán posibles criterios que permitan fincar responsabilidad si llegara a haber consecuencias perjudiciales por su uso. La IA hace que el mundo avance bastante rápido y el derecho penal no debe ser una excepción.

Puede hablarse de responsabilidad penal derivada del uso de un sistema de inteligencia artificial en los siguientes casos:

- a. Una persona se sirvió de la IA, mediante una programación específica, para el hackeo de un sistema, o situaciones similares.
- b. El creador del sistema o el usuario presentaron

inobservancia de los deberes de cuidado, tras lo cual se valorará la responsabilidad.

- c. Los sistemas de IA más avanzados se consideran sujetos susceptibles de reproche penal, debido a que se ha abierto la posibilidad de que puedan adoptar decisiones al margen de la voluntad humana de su creador o usuario.

Pues bien, estos casos, así como sus posibles soluciones, se analizarán detenidamente a lo largo del presente artículo.

II. Concepto y definición normativa de la IA en México

El Grupo de Expertos de Alto Nivel en Inteligencia Artificial de la Unión Europea (HLEG AI) define la inteligencia artificial como aquellos sistemas de *software* (y posiblemente también de *hardware*) diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno y la adquisición de datos, interpretándolos de manera estructurada o no estructurada, razonando el conocimiento y procesando la información derivada de ellos, a fin de decidir las mejores acciones a realizar para alcanzar el objetivo dado (Comisión Europea, 2019).

Definición normativa en México

Por su parte, en la legislación mexicana ya se han establecido definiciones precisas que describen a la IA como un sistema informático o modelo algorítmico, caracterizado por su capacidad de procesar datos y generar contenidos (textos, imágenes, audios o videos) de manera autónoma o asistida,

mediante la imitación o sustitución de las funciones cognitivas e identidades humanas con apariencia de veracidad. Así, el Código Penal del estado de Durango establece:

Para efectos de esta disposición, se entenderá por inteligencia artificial (IA) cualquier sistema informático o modelo algorítmico, entrenado o programado, capaz de procesar datos y generar textos, imágenes, sonidos, videos o representaciones visuales, de manera autónoma o asistida, que imiten, reproduzcan o sustituyan conductas, rasgos o identidades humanas con apariencia de veracidad. (CPD, 2025: artículo 182 Quater)

Por su parte, el Código Penal para el estado de San Luis Potosí establece:

Para los efectos de este artículo, se entenderá por inteligencia artificial cualquier sistema, modelo computacional o algoritmo, incluidas las redes neuronales y las técnicas de aprendizaje automático o equivalentes, capaces de ejecutar, de manera autónoma o asistida, tareas que impliquen procesos de razonamiento, aprendizaje, percepción, interpretación o generación de información, imitando o sustituyendo funciones cognitivas humanas, y que permita producir, modificar, alterar o difundir contenidos digitales, tales como textos, imágenes, sonidos o videos, con apariencia verosímil o similar a la realidad. (CPSLP, 2025: artículo 187 Ter, cuarto párrafo)

Mientras que el Código Penal para el estado de Zacatecas establece la siguiente definición:

Para efectos de este artículo, se entenderá por inteligencia artificial, a los sistemas, modelos o algoritmos capaces de generar, procesar información visual, textos, imágenes, audios, o vídeos de manera automatizada o semiautomatizada, funciones de aprendizaje, análisis, predicción, clasificación, generación de contenidos, orientación y toma de decisiones, cuya operación pueda producir efectos jurídicos que dañen derechos de alguna persona. (CPZ, 2026: artículo 58 Quáter)

Se puede advertir que estas definiciones son un esfuerzo del legislador para respetar el principio de taxatividad que debe imperar en la creación de tipos penales; principio valioso de un estado de derecho para hacer posible la igualdad y la seguridad jurídica del gobernado.

El esfuerzo legislativo por establecer una definición de inteligencia artificial entra en el primer caso señalado al principio de este estudio, cuando se ejecutan delitos por una persona que se sirve de la IA para alcanzar un resultado típico; así están estructurados los tipos penales de algunos estados de la República, como se verá posteriormente.

III. La inteligencia artificial frente al concepto de acción

El concepto de acción es la categoría fundamental y el presupuesto básico de la teoría del delito, es evidente que la aparición de la inteligencia artificial pone en tela de juicio si aún la Teoría de la Acción Penal nos brinda el rendimiento necesario para sustentar un modelo de imputación que sea congruente con la sistemática penal y que sea capaz de soportar un modelo de imputación penal que resulte eficaz para el control social en las acciones delictivas.

La acción se ha definido como la conducta humana dominada o dominable por la voluntad. Para fines de este artículo, hay que tener presente que en el desarrollo del concepto de acción en la dogmática jurídico penal, ésta siempre ha estado provista del elemento "voluntad", como se advierte al echar un vistazo a las corrientes dogmáticas que la han desarrollado.

El Causalismo de Franz von Liszt y Ernst von Beling define a la acción como un movimiento corporal voluntario que causa una modificación en el mundo exterior, sin importar el contenido de esa voluntad (el dolo o la culpa).

Para el Finalismo de Hans Welzel, la acción se concibe como el ejercicio de la actividad final. El ser humano, gracias a su saber causal, puede prever las consecuencias de su actividad y dirigirla hacia una meta. Welzel sostiene que:

[...] la finalidad se basa en la capacidad de la voluntad de prever, dentro de ciertos límites, las consecuencias del engranaje de la intervención causal, y merced a ello dirigirla de acuerdo con un plan a la consecución del fin, es la voluntad consciente del fin, que rige el acontecer causal, la columna vertebral de la acción final. Ella es el factor de dirección que sobredetermina el acontecer causal exterior y en virtud de esto lo convierte en una acción dirigida finalmente. Sin ella, la acción quedaría destruida en su estructura material y rebajada a un proceso causal ciego. La voluntad final pertenece, por ello, a la acción como factor integrante, ya que, y en la medida en que, configura objetivamente el acontecer exterior. (2015, p. 66)

Para el Funcionalismo de Claus Roxin, la acción representa una manifestación de la personalidad, es decir, "todo lo

que puede atribuirse a un ser humano como centro psíquico-espiritual". De ahí que no pueda hablarse de acción cuando el individuo no participa en el acontecer causal como persona, sino únicamente como un cuerpo (2016, p. 79).

Desde el punto de vista de la conducta, la dogmática penal sostiene el principio de acto humano, lo que, en sentido estricto, excluye a las máquinas del concepto de acción. La IA no posee voluntad ni capacidad de autodeterminación; por lo tanto, sus procesos se consideran herramientas o medios comisivos y no acciones autónomas del sistema. La responsabilidad penal permanece anclada en el ser humano que programa, entrena o utiliza el algoritmo para vulnerar un bien jurídico.

Para el Normativismo de Günther Jakobs, la acción es el no reconocimiento de la vigencia de la norma; es una expresión de sentido que infringe una expectativa jurídica, tal y como se aprecia en el siguiente fragmento:

Un concepto jurídico-penal de acción debe combinar sociedad y Derecho Penal, So pena de degenerar hasta quedar convertido en un concepto de mera utilidad didáctica que describa un escalón inicial del delito, debe contener una teoría lo más completa posible del comportamiento jurídico-penalmente relevante. (1997, p. 102)

Por ende, como se observa desde la perspectiva de la acción, la dogmática penal sostiene el principio de acto humano, que se generó en la manifestación de la voluntad y se logra gracias a que el ser humano tiene autonomía de toma de decisiones, las cuales son relevantes para el derecho penal.

Cabe preguntarnos lo siguiente: ¿un sistema de inteligencia artificial posee autonomía plena? A esta interrogante podemos responder que, hasta este momento, aún

no está en posibilidad de tomar decisiones autónomas, entendiendo por tales aquellas en las que el sistema tenga la posibilidad de valorar qué conducta o comportamiento es el más adecuado para alcanzar un fin. En palabras de Welzel, que tenga un conocimiento causal, y "gracias a su saber causal, puede prever, dentro de ciertos límites, las consecuencias posibles de su actividad, ponerse, por tanto, fines diversos y dirigir su actividad, conforme a su plan, a la consecución de estos fines" (2015, p. 65).

Por lo tanto, considero que, a un sistema de inteligencia artificial, hasta este momento, no se le podría imputar que haya realizado una acción y aún menos una dolosa, pues el ser humano en el mundo fáctico no puede, hoy por hoy, desentenderse de las acciones del sistema de inteligencia artificial. En consecuencia, está excluido del concepto de "acción" en sentido estricto. Pues reiteramos que la inteligencia artificial no posee voluntad ni capacidad de autodeterminación.

Hasta este momento, la legislación considera a los sistemas de inteligencia artificial como herramientas o instrumentos para que un sujeto activo del delito realice la comisión del delito (contemplado en la ley penal); por lo tanto, los sistemas de inteligencia artificial no realizan acciones autónomas, ya que éstos no hacen más que obedecer/interactuar sobre los parámetros asignados por su creador y obedecer las instrucciones asignadas.

Existirá responsabilidad penal para el ser humano o persona jurídica que programa o utiliza el algoritmo para vulnerar un bien jurídico.

IV. Vacíos de responsabilidad penal frente a decisiones automatizadas

Ante la interrogante de quién debe responder penalmente por decisiones automatizadas que lesionan bienes jurídicos, recordemos que al principio de este estudio se estableció que la responsabilidad penal derivada del uso de un sistema de inteligencia artificial se puede presentar en los siguientes tres escenarios posibles:

- a. cuando un sujeto comete un hecho típico sirviéndose de la IA,
- b. cuando hay inobservancia de los deberes de cuidado de parte del creador del sistema de IA o el usuario;
- c. los sistemas de IA como sujetos susceptibles de reproche penal.

Cuando un sujeto comete un hecho típico sirviéndose de la IA

Respecto a este escenario, ya fue señalado los estados de la República que han incorporado tipos penales en los que ya se prevé la inteligencia artificial son pocos: San Luis Potosí, Zacatecas, Durango y Colima. En estos ordenamientos jurídicos de carácter penal se ha incorporado la inteligencia artificial de las siguientes formas:

La inteligencia artificial como instrumento para cometer el delito

Algunos códigos penales tipifican delitos como “violencia a la intimidad sexual” el uso de inteligencia artificial para generar, manipular o alterar material erótico de personas rea-

les sin su consentimiento, así por ejemplo tenemos el Código Penal del estado de Durango que establece como medio comisivo el uso de la inteligencia artificial, al establecer:

Comete el delito de violencia a la intimidad sexual mediante Inteligencia Artificial, quien genere, manipule, altere o simule material pudiendo ser estas imágenes videos o audios o cualquier otro, con contenido sexual mediante inteligencia artificial o herramientas digitales, que represente a una persona real sin su consentimiento, con la finalidad de distribuir, exhibir, comercializar o amenazar con difundir dicho contenido, se sancionará con pena de cinco a nueve años de prisión y multa de trescientas sesenta a seiscientos cuarenta y ocho Unidades de Medida de Actualización.

Esta pena aumentará una tercera parte si la víctima es menor de dieciocho años de edad o de personas que no tienen capacidad para comprender el significado del hecho o de personas que no tienen capacidad para resistirlo.

La pena se incrementará en una mitad cuando las imágenes, videos o audios hayan sido obtenidos sin autorización del titular desde redes sociales, dispositivos móviles, correos electrónicos u otras plataformas digitales, y posteriormente se modifiquen o transformen mediante inteligencia artificial para generar contenido sexual.

Para efectos de esta disposición, se entenderá por inteligencia artificial (IA) cualquier sistema informático o modelo algorítmico, entrenado o programado, capaz de procesar datos y generar textos, imágenes, sonidos, videos o representaciones visuales, de manera autónoma o asistida, que imiten, reproduzcan o sustituyan conductas, rasgos o identidades humanas con apariencia de veracidad. (CPD, 2025: artículo 182, Quater)

Así, también tenemos el Código Penal del estado de Zacatecas que establece:

Comete el delito de violación a la intimidad sexual, quien por cualquier medio incluyendo el uso de la inteligencia artificial u otras tecnologías emergentes, produzca, genere, altere, reproduzca, divulgue, comparta, distribuya, compile, comercialice, solicite, publique o amenace con publicar, imágenes, audios o videos de una persona parcial o totalmente desnuda; de contenido erótico o sexual, ya sea impreso, grabado o digital, sin el consentimiento, aprobación o autorización de la víctima. (CPZ, 2026: artículo 232, Ter)

Por su parte el Código Penal para el estado de San Luis Potosí tipifica el delito de uso indebido de inteligencia artificial para provocar alarma social, al establecer:

Manipulación institucional mediante inteligencia artificial. Comete el delito quien, a sabiendas de su falsedad y con la finalidad directa de alterar la confianza pública en las instituciones o poner en riesgo verificable la seguridad del Estado, genere, modifique o difunda, mediante sistemas de inteligencia artificial, contenidos digitales que simulen de manera verosímil declaraciones, comunicados o actuaciones de autoridades, instituciones públicas o cuerpos de seguridad. (CPSLP, 2025: artículo 272 TER (sic))

Así, este ordenamiento establece el delito de uso no autorizado de imagen o voz generada mediante inteligencia artificial, como se expone en el siguiente fragmento:

Comete el delito quien, sin consentimiento previo, expreso, específico e informado de la persona a la que corresponda

la imagen, la voz o la identidad digital, utilice sistemas de inteligencia artificial o programas automatizados para crear, reproducir, modificar, manipular o difundir contenidos que simulen la apariencia, la voz, los gestos o la identidad de una persona real. (CPSLP, 2025: artículo 187 TER.)

El uso de la inteligencia artificial como agravante de la pena

Algunos estados de la República han incorporado el uso de la inteligencia artificial como agravante genérica de la pena; así lo hace el Código Penal del estado de Zacatecas que contempla a la inteligencia artificial como un medio comisivo que agrava la pena en los delitos cuando se emplea en su realización el uso de la inteligencia artificial, al establecer lo siguiente:

A quien haga uso de la inteligencia artificial u otras tecnologías emergentes, *deepfakes*, clonación de voz o el uso de imagen, destinadas a crear apariencia de veracidad para la comisión de alguno o algunos de los delitos contemplados en este Código Penal, se le aumentarán las penas previstas en el delito ejecutado, hasta en una quinta parte más en su mínimo y máximo. (CPZ, 2026: artículo 58, Quáter)

Como también lo concibe el estado de Quintana Roo, al incluir en su Código Penal a la inteligencia artificial como medio comisivo que agrava la pena, al establecer lo siguiente:

Quien haga uso de la inteligencia artificial como herramienta o medio para la realización de hechos constitutivos de delitos contemplados así en este Código Penal, se le au-

mentarán las penas previstas en el delito que fuese realizado hasta en una mitad más.

Para efectos del párrafo anterior, se entenderá por inteligencia artificial a la capacidad de los sistemas tecnológicos, informáticos, *softwares* o aplicaciones de una máquina para simular las capacidades humanas como el razonamiento, el aprendizaje, la creatividad, la capacidad de planear y procesar datos para la realización de tareas específicas y autónomas. (CPOROO, 2026: artículo 20, Bis.)

De esta manera, podemos sostener que en pocos ordenamientos penales de nuestro país se ha contemplado la inteligencia artificial sólo como un medio comisivo o como una agravante genérica. Asimismo, se advierte que esas descripciones típicas citadas sólo se pueden cometer a título de acción dolosa.

Por otra parte, en los tipos penales que se han puntualizado, previstos por los códigos penales mencionados, consideramos que la forma de intervención en el delito, al emplear en su comisión un sistema de inteligencia artificial, se puede materializar bajo la figura de la *autoría mediata*, donde el sujeto humano (el "hombre detrás") instrumentaliza el *software* para alcanzar un resultado típico; es decir el autor ejerce, *un dominio del hecho*, ya sea cuando configura los parámetros del sistema de inteligencia artificial o da instrucciones, de acuerdo con su saber causal o cuando pone en marcha el curso de los hechos para que el algoritmo lo lleve a la consumación del delito.

La responsabilidad del creador del sistema de IA o del usuario con base en la inobservancia de los deberes de cuidado

El problema se complica. Por tal razón, considero que en este apartado surge el primer vacío que existe y que dificulta sostener un modelo de imputación penal al creador de un sistema de inteligencia artificial, o bien, contra el usuario por la inobservancia de los deberes de cuidado, cuando se genera el delito sin ser ésta la voluntad inicial del individuo, lo que en estos casos abre la puerta a la imputación a título culposo; pero en este caso como pregunta obligada: ¿Cuál sería el fundamento de imputación penal si nuestros ordenamientos jurídicos, a mi juicio, aún no regulan esta situación fáctica? Nuevamente surge la exigencia por parte del control penal, respetar el principio de taxatividad para poder imputar responsabilidad en un hecho de tal naturaleza.

Ante este problema, sostenemos que la teoría de la imputación objetiva aún nos puede proporcionar rendimiento para poder lograr una imputación ante ese hecho fáctico y poder resolver quién debería responder penalmente por el acto ilícito producido por la elevación del riesgo y, de esta manera, atribuirle al autor el resultado típico producido por un sistema de inteligencia artificial.

Consideramos que la teoría de la imputación objetiva sería la que nos daría el andamiaje para poder sostener que el resultado causalmente le es atribuible al sujeto que inobservó los deberes de cuidado en la utilización de un sistema de inteligencia artificial, como pudiera ser la utilización de un dron que está programado para realizar determinadas funciones, pero antes de realizarlas se precipita y causa la muerte y lesiones de varias personas.

Como se sabe, la teoría de la imputación objetiva fue desarrollada por Roxin al principio de 1970. Esta teoría, de forma esquematizada, establece que un resultado le es objetivamente imputable a una persona siempre que se demuestren los siguientes presupuestos:

- Creación de un riesgo no permitido: el sujeto que utiliza un sistema de inteligencia artificial debe haber realizado una conducta que sobrepase los niveles de riesgo que la sociedad considera aceptables o necesarios para su desarrollo, es decir, el sujeto al inobservar un deber jurídico penal de cuidado elevó el riesgo autorizado.

Se vuelve necesario determinar cuáles son las normas de cuidado que se tienen que respetar y que deban ser exigibles a los ciudadanos en el uso de sistemas inteligentes para poder delimitar quién ha cometido el delito. Por lo que se requiere la regulación de protocolos de actuación, para que no existan estos vacíos a la hora de realizar un juicio de imputación penal.

Cuando, al utilizar un sistema de inteligencia artificial, se produzca un resultado lesivo y se demuestre que éste deriva de una conducta desarrollada dentro del ámbito del riesgo permitido, conforme a la normativa administrativa existente, no se podrá imputar ese resultado acaecido, en razón de que faltaría el desvalor de acción. El sujeto no respondería ni siquiera por tentativa.

- Concreción del riesgo en el resultado: este riesgo jurídicamente desaprobado al utilizar un sistema de inteligencia artificial, inobservancia de los de-

beres de cuidado, ya sea por su creador o por su operador del sistema, debe ser precisamente el que se materializa en el resultado típico.

Para que un resultado sea imputable al autor, es preciso que además de la relación de causalidad exista una "relación de riesgo", es decir, que como consecuencia del riesgo creado por la conducta se produzca el resultado. Si no existe esa relación de riesgo, no se puede imputar el resultado (De la Cuesta Aguado, 1988, p. 148).

- **Ámbito protector de la norma:** el resultado producido por la inobservancia del deber jurídico penal de cuidado ya sea por el fabricante, creador o programador del sistema de inteligencia artificial, debe ser un resultado en los que la norma penal tenía la intención y finalidad de prevenir.

Los sistemas de IA como sujetos susceptibles de reproche penal autónomo

En lo referente a la cuestión de si se puede generar un título de imputación, así como un juicio de reproche penal a una IA, consideramos que no, debido a que actualmente los sistemas de inteligencia artificial no poseen autonomía e independencia de su creador o desarrollador, por lo tanto, el título de imputación, hasta este momento, será para el ser humano que utilice los sistemas como medio para cometer un ilícito o realice acciones culposas por inobservancia a los deberes jurídicos penales de cuidado.

V. Conclusiones

Con el desarrollo de los sistemas de inteligencia artificial, se ha llegado a aseverar que la teoría del delito ya no proporciona el rendimiento suficiente para fundamentar y sostener la imputación a las acciones realizadas por los sistemas de inteligencia artificial; a este respecto, me parece que el *quid* de esta discusión se desvanece porque aún no se puede atribuir autonomía a los sistemas con inteligencia artificial.

La teoría del delito desarrollada hasta este momento es totalmente válida para resolver los diferentes problemas causales y de imputación penal, ya sea en un hecho doloso, cuando se emplea la inteligencia artificial como medio comisivo, o al tratarse de un hecho culposo, cuando en el manejo de los sistemas de inteligencia artificial ya sea como creador o programador del sistema o como usuario, no se cumplió con la observancia de los deberes de cuidado. En este sentido, se considera que la teoría del delito que se ha desarrollado hasta nuestros días puede quedar intacta ante la existencia de los sistemas de inteligencia artificial.

No se debe perder de vista que los sistemas de inteligencia artificial han sido elaborados y desarrollados por el ser humano, y la inteligencia artificial aún no goza de absoluta autonomía, por lo tanto, no se puede considerar que los sistemas de inteligencia artificial puedan realizar acciones independientes y que sean jurídicamente relevantes para el derecho penal.

Si consideramos el concepto jurídico penal de acción, los sistemas de inteligencia artificial aún no tienen voluntad propia, por lo que son incapaces de actuar con dolo o alcanzar una independencia total, al menos en este momento histórico, lo cual resulta fundamental para delimitar el ámbito de la responsabilidad penal derivada de actuaciones

ilícitas en las que se ve implicado un sistema de inteligencia artificial. Por ello, el sujeto de imputación será siempre el ser humano, el único ente que goza de libre albedrío y es capaz de determinarse conforme a la norma.

El modelo de imputación que se encuentra regulado hasta este momento en los códigos penales de la República Mexicana que consideran a la inteligencia artificial consiste en prever la utilización de estos sistemas como medio o como instrumento para cometer los tipos penales establecidos y en regular el uso de la IA como agravante genérica de la pena cuando se utiliza en la comisión de delitos.

Puede sostenerse que el modelo de imputación penal que ofrece mayores herramientas para abordar los casos en los que tanto el creador de un sistema de inteligencia artificial como su usuario incumplen los deberes de cuidado y cometan conductas culposas, es la teoría de la imputación objetiva. Ello se debe a que esta construcción permite fundamentar normativamente la conexión del nexo causal con el resultado provocado por la acción culposa del autor que desarrolló o utilizó un sistema de inteligencia artificial.

VI. Referencias

- Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner (2016, mayo 23). "Machine bias. There's software used across the country to predict future criminals. And it's biased against blacks". *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- CE: Comisión Europea (2019). "A definition of ai: Main capabilities and disciplines"/ Definición de Inteligencia Artificial: principales capacidades y disciplinas científicas. Bruselas:

- CE Unión Europea. <https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>
- CPD: Código Penal de Durango (2009, 14 de junio). *Código Penal del estado Libre y Soberano de Durango*, última reforma 25 de diciembre de 2025. [https://congresodurango.gob.mx/Archivos/legislacion/CODIGO%20PENAL%20\(NUEVO\).pdf](https://congresodurango.gob.mx/Archivos/legislacion/CODIGO%20PENAL%20(NUEVO).pdf)
- CPSLP: Código Penal del estado de San Luis Potosí (2014, 28 de septiembre). *Instituto de Investigación y Evaluación Legislativa*, última reforma 8 de mayo de 2026. <https://congresosanluis.gob.mx/legislacion/codigos>
- CPQROO: Código Penal del estado de Quintana Roo (2016, 23 de junio). *Código Penal para el Estado Libre y Soberano de Quintana Roo*, última reforma 10 de abril de 2026. <https://documentos.congresoqroo.gob.mx/codigos/C6-XVIII-10042026-20260413T103001-C1820260410217-C%C3%B3digo-Penal.pdf>
- CPZ: Código Penal para el estado de Zacatecas (1986, 17 de mayo). *Suplemento al Periódico Oficial del estado de Zacatecas*, última reforma 04 de marzo de 2026. <https://www.congresozac.gob.mx/65/ley&cual=103>
- De la Cuesta Aguado, Paz (1998). *Tipicidad e imputación objetiva*. Argentina: Ediciones Jurídicas Cuyo.
- Jakobs, Günther (1997). *Estudios de Derecho Penal*. España: Civitas.
- Perry, Walter, Brian McInnis, Carter Price, Susan Smith y John Hollywood (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. RAND Corporation.
- Roxin, Claus (2016). *La teoría del delito en la discusión actual*, tomo II. Grijley.
- Welzel, Hans (2015). *Derecho penal alemán: parte general*. Chile: Editorial Jurídica de Chile.

