

Capítulo 7

La transparencia y explicabilidad algorítmica de la Inteligencia Artificial aplicada en la investigación penal y el derecho a la prueba

Sidney Ernestina Marcos Escobar

Marcos Escobar, S. E. (2026). La transparencia y explicabilidad algorítmica de la Inteligencia Artificial aplicada en la investigación penal y el derecho a la prueba. En A. B. Benalcázar (Coord). *Ciencias sociales y humanidades en América Latina. Investigaciones disciplinares e interdisciplinarias desde la región (Vol. III)*. (pp. 168-195). Religación Press. <http://doi.org/10.46652/religacionpress.469.c961>



La transparencia y explicabilidad algorítmica de la Inteligencia Artificial aplicada en la investigación penal y el derecho a la prueba

Resumen

Esta investigación tiene como objetivo reflexionar acerca del empleo incipiente de la Inteligencia Artificial en los sistemas de justicia penal acusatorio de países del Sur Global. Para ello se parte del análisis de reglas y principios comunes en la etapa de investigación penal, en los Sistemas de Justicia Penal de nuestra Cultura Jurídica. Posteriormente, se problematiza presentando reflexiones acerca de la transparencia y la explicabilidad de los sistemas de Inteligencia Artificial que se han empleado en algunas experiencias internacionales, para advertir que aspectos tales como el secreto industrial, el tipo de datos con que se abastece la Inteligencia Artificial y la automatización de los procedimientos para obtener resultados, generan un escenario de poca transparencia y explicabilidad de los resultados obtenidos mediante el uso de Inteligencia Artificial para los efectos del proceso penal. En este sentido, se analizan las implicaciones del uso de Inteligencia Artificial específicamente en la etapa de investigación penal, señalando críticamente las debilidades y fortalezas de los elementos probatorios generados con esta tecnología, así como su incidencia en el respeto al derecho a la prueba y al principio de contradicción que rige el proceso penal acusatorio. Finalmente, se plantean algunas cuestiones que podrían ser claves para comenzar a pensar en la implementación de alguna Inteligencia Artificial en los sistemas de justicia penal del Sur Global y que, a su vez fuera fiable, respetuosa del derecho a la prueba y de los principios rectores que rigen en los sistemas acusatorios.

Palabras clave: Inteligencia Artificial; investigación penal; transparencia; explicabilidad; prueba: sistema penal.

Introducción

Actualmente, las sociedades transitan un contexto de cambio tecnológico. El uso de la inteligencia Artificial se ha hecho presente en diversos sectores y esto implica que en el futuro inmediato se presentarán nuevas formas de realizar tareas que anteriormente solo estaban destinadas a ser ejecutadas por seres humanos. El sector jurídico no ha sido la excepción y a la fecha es posible afirmar que, en sistemas de justicia en materia penal, de diversas experiencias internacionales la Inteligencia Artificial se ha incorporado incipientemente como herramienta para facilitar tareas que originariamente eran realizadas por las personas operadoras del sistema penal.

El empleo de este tipo de tecnología ha tenido un impacto relevante, sobre todo, en sede de órganos de procuración e impartición de justicia penal. Sin embargo, pese a que en diversas experiencias internacionales ya es posible advertir el uso de la Inteligencia Artificial aplicada en la persecución e investigación penal, y en menor medida, en órganos jurisdiccionales, una de las preocupaciones que han surgido en estos contextos se centra en la transparencia y explicabilidad de los procesos empleados para la generación de los resultados a través de Inteligencia Artificial y su uso como elementos probatorios en el proceso penal.

En efecto, una de las cuestiones más problemáticas acerca del empleo de la Inteligencia Artificial en los sistemas de justicia penal es la poca o nula transparencia y explicabilidad de los procedimientos que subyacen a los resultados que, posteriormente, en algunas experiencias, se emplean como elementos probatorios a considerar y valorar para la toma de decisiones judiciales. Esa falta de claridad y explicabilidad se advierte sobre todo en Inteligencias Artificiales que emplean procesos automatizados para la emisión de sus resultados, pues aunado a esta automatización, lo que, en la mayoría de los casos ocurre, es que estos desarrollos tecnológicos están protegidos por el secreto industrial de los fabricantes, lo que impide, por un lado, adentrarse a cuestionar su metodología de automatización, pero también por la propia

complejidad operativa de la Inteligencia Artificial esa explicabilidad se vuelve casi imposible.

En este contexto, se hará alusión a una de las Inteligencias Artificiales pioneras no especializada en el sector jurídico, pero sí empleada en procesos de índole criminal en Estados Unidos de América: el Sistema Compas. Compas fue desarrollado en el sector privado y ofrece servicios de análisis e innovación geoespacial, esto es, mapeo de datos para la toma de decisiones, a partir de la identificación de riesgos y monitoreo de cambios con mayor rapidez que si se realizara por la actividad meramente humana. Incluso ha servido de apoyo para la predicción de riesgos procesales en procesos de índole criminal. Sin embargo, hay experiencias litigiosas en cortes estadounidenses que han culminado en desestimar las pruebas surgidas por esta Inteligencia Artificial, basadas en evidencia que permite señalar que el algoritmo a partir del cual se ha producido el resultado en casos concretos ha considerado un conjunto de datos sesgados e incluso discriminatorios a partir de la identificación de las personas por su origen nacional, étnico, económico, laboral e incluso religioso, esto es, este sistema algorítmico predictivo ha empleado variables subjetivas ponderadas, lo que implicaría en un sistema de justicia que asuma el compromiso del respeto a los derechos humanos una violación flagrante a los mismos y principalmente, al derecho de igualdad y no discriminación de las personas.

Sin adentrarse de forma amplia en el caso de Compas, para los efectos de este análisis se toma como referencia para poner de manifiesto algunos de los debates en materia de transparencia, explicabilidad y ética que impactan en el debido proceso al hacer uso de inteligencias artificiales abastecidas con información sesgada, las cuales reproducen estereotipos y prejuicios, aunado a la poca o nula claridad sobre la explicabilidad de los procedimientos automatizados con los que elaboran sus resultados.

Asimismo, se analizarán desde una perspectiva crítica los aspectos de transparencia y explicabilidad, de los cuales se ocupan las Directrices éticas para una IA fiable de la Unión Europea con el pro-

pósito visibilizar las preocupaciones que grupos de personas expertas han planteado con el objetivo de generar contextos más claros para la implementación del uso de Inteligencia Artificial en la vida, enfocándolos en esta ocasión en los sistemas de justicia penal. A partir del análisis de los debates y propuestas generadas en experiencias internacionales, se intentará dilucidar desde la perspectiva del Razonamiento Probatorio, de qué forma podrían garantizarse el derecho a la prueba y la observancia del principio de contradicción al emplearse inteligencia artificial en la actividad probatoria que se realiza en la etapa de investigación, en un sistema de justicia penal. Las anteriores reflexiones se presentan tomando en cuenta que uno de los derechos humanos que deben garantizarse en un Estado Constitucional y Democrático de Derecho es la seguridad jurídica de las personas.

Por las razones anteriores es que se plantea como problemático el empleo de inteligencias artificiales en la investigación penal sin marcos jurídicos robustos que establezcan reglas mínimas para la producción probatoria en sede de investigación penal con uso de esta tecnología, que permitan a las partes procesales tener meridiana claridad sobre la información con la cual han sido abastecidas y por ende, se comprometa la garantía de dotar a las personas que intervienen en el proceso penal de algún grado de explicabilidad y transparencia del proceso automatizado con el que se ha obtenido un resultado probatorio o prueba con la Inteligencia Artificial generativa, en su caso.

Aunado a lo anterior, se parte de la idea de que también resulta necesario que se establezcan a nivel normativo criterios orientadores que permitan a las personas operadoras del sistema de justicia penal realizar el análisis acerca de la admisión o exclusión de elementos probatorios generados a partir del uso de la Inteligencia Artificial, sin que esto signifique la tasación de la prueba obtenida mediante inteligencia artificial, garantizando la libre valoración racional de la prueba.

Desde luego, que la propuesta de construcción de los marcos normativos que se advierten como necesarios, parte de la idea de que el empleo de la Inteligencia Artificial en el sector jurídico debe ser espe-

cializada en este ámbito, teniendo en cuenta un enfoque ético, de respeto a derechos humanos, así como a los principios propios que rigen el proceso penal acusatorio.

Finalmente, se presentan algunas reflexiones a modo de conclusión, sin que estas sean acabadas en virtud de ser un tema incipiente en muchos de los sistemas de justicia de Iberoamérica.

Método

La investigación que se presenta ha partido de una metodología cualitativa a partir del empleo de una técnica documental al consultar directrices o proyectos para la construcción de normas marco en materia de inteligencia artificial, informes de grupos de trabajo especializados en la materia, así como el análisis de disposiciones legales de carácter procesal penal a fin de confrontar reglas de debido proceso y principios sobre ética y responsabilidad estatal en el uso de la inteligencia artificial, en sistemas de justicia pioneros en uso de esta tecnología.

Asimismo, se han consultado fuentes primarias como datos estadísticos originales del uso del sistema Compas, pionero en materia de inteligencia artificial empleada en el sector jurídico, en la experiencia estadounidense y otros registros que han sido productos de investigaciones encaminadas al análisis de sistemas de inteligencia artificial aplicados a la investigación penal en experiencias de países que hoy en día son los más avanzados del sector.

A partir de ello, con un enfoque interpretativo hermenéutico, así como crítico se presentará el problema de la transparencia y explicabilidad acerca de los procesos automatizados que caracterizan a las Inteligencias Artificiales para el sector jurídico, visibilizando la tensión que se configura con el principio de contradicción y el derecho a la prueba de las partes procesales que rige el sistema penal acusatorio.

En efecto, esta investigación presenta conceptos base para la discusión acerca de la tensión generada desde la perspectiva ética y de derechos humanos en el uso de la Inteligencia Artificial con los derechos y principios que caracterizan a un sistema penal acusatorio, tales como el derecho a la prueba y el principio de contradicción, poniendo de manifiesto que estas tensiones comprometen la seguridad jurídica de las personas justiciables en un Estado Constitucional y Democrático de Derecho.

En suma, las técnicas de investigación empleadas en esta investigación consistentes en el análisis exegético de textos jurídicos, así como el procesamiento y presentación de información a partir de categorías conceptuales propias tanto del sistema penal acusatorio, las tecnologías de la información y particularmente, de la Inteligencia Artificial permitirá plantear un panorama actual de los aspectos más problemáticos que un sistema penal acusatorio debería tener en cuenta para incluir herramientas de inteligencia artificial en la investigación penal y al mismo tiempo ser respetuoso de los derechos humanos, así como de los principios de contradicción, seguridad jurídica y derecho a la prueba en el proceso penal.

Para la obtención de datos se ha adoptado el método jurídico normativo, recopilando información acerca del tema en fuentes primarias especializadas, tales como normas jurídicas en sentido estricto (Código Nacional de Procedimientos Penales de México), directrices y literatura científica sobre Inteligencia Artificial. De igual forma, se ha acudido a las notas de clase del curso formativo impartido por el Centro de Estudios de Justicia de las Américas en materia de Inteligencia Artificial y Sistemas de Justicia, segunda edición 2026 que se cursó durante los meses de marzo a mayo del presente año. También se ha acudido a fuentes complementarias tales como literatura especializada sobre la investigación penal, el sistema penal acusatorio, democracia y ética en sistemas de justicia, consultando bases de datos de libre acceso y revistas especializadas.

Resultados

Como parte de los resultados obtenidos en esta investigación se advirtió que es común la preocupación por parte de la comunidad tecnológica, así como del sector jurídico, respecto a la ausencia de marcos jurídicos robustos y éticos que fijen ciertos criterios para la utilización de la Inteligencia Artificial en los sistemas de justicia penal y, específicamente, en la investigación penal en la que se origina la conformación de elementos probatorios.

Para ello, se ha consultado la base de datos sobre la Regulación sobre IA en América Latina y el Caribe, en específico, el Repositorio Regulación sobre la IA en ALC, de la Universidad de los Andes y el sitio web oficial de la Cámara de Diputados de México lo que ha permitido advertir la incipiente actividad legislativa para regular el uso de la Inteligencia Artificial en el sector público, en países de nuestra región.

Sin embargo, ante ese escenario analizado se advirtió que ninguna de las iniciativas de reforma legislativa está encaminada a regular de forma puntual la actividad probatoria en sede de órganos de investigación penal con uso de la Inteligencia Artificial.

Este escenario de nula actividad legislativa no ha frenado los debates acerca de la tensión que se configura debido a la poca transparencia y explicabilidad de los procesos automatizados que caracterizan a la Inteligencia Artificial por sí misma y que, se replicaría también en las Inteligencias Artificiales que se podrían emplear para obtener resultados en la investigación penal, con la finalidad de ser empleados posteriormente como elementos probatorios en el proceso penal. En efecto, la poca transparencia y explicabilidad de los procesos automatizados se debe en gran medida a que los desarrollos tecnológicos de Inteligencia Artificial están protegidos por los derechos de propiedad intelectual y esto imposibilita transparentarlos de forma pública. Asimismo, se suman a este escenario los riesgos que el uso de la Inteligencia Artificial trae aparejada por sí misma. Para ello es pertinente, en primer lugar, la identificación de riesgos y clasificación de los mismos con la finalidad de que los Estados sean conocedores de estos y pue-

dan tomar decisiones respecto de cuáles asumir y cuáles no, responsabilizándose acerca de aquellos que se traduzcan en la producción de restricciones de derechos o daños a las personas. Esta cuestión es una toma de decisión política y de la Gobernanza exigida en materia de Inteligencia Artificial (Gutiérrez, 2026), pero, en ese sentido, los Estados a fin de mantener la ética y fomentar la confianza en el uso de la Inteligencia Artificial en el proceso penal debieran asumir tales costos de forma expresa, en un marco jurídico claro y a partir de la propia capacidad institucional que conlleva la adopción de esta tecnología.

De lo anterior, también se puede señalar como resultado del análisis realizado que sin la comprensión profunda del funcionamiento de la inteligencia Artificial y los riesgos propios de su utilización no es posible construir un marco jurídico claro y robusto en materia probatoria penal. Aunado a ello, en todo caso para la creación de una legislación que regule la utilización de la Inteligencia Artificial en el proceso penal debe contemplar la supervisión y responsabilidad indelegable humana especializada, esto permitirá hacer posible la protección a las personas usuarias del sistema de justicia penal que, eventualmente, resientan los estragos de los resultados obtenidos por medio de Inteligencia Artificial dentro del proceso penal.

Asimismo, garantizar la transparencia y explicabilidad de los procedimientos automatizados con que opere una inteligencia artificial empleada para la generación de elementos probatorios en el proceso penal permitiría hacer posible el derecho a acceder a un medio de impugnación y la tutela judicial efectiva de las partes en el proceso penal, así como el respeto a los derechos humanos de cualquier persona que intervenga en este y que resienta alguna afectación por el uso de la inteligencia artificial.

Finalmente, se concluye presentando algunos lineamientos para la construcción de un modelo de marco jurídico que regule el uso de la Inteligencia Artificial en la conformación de elementos probatorios en la etapa de investigación penal y su uso en el proceso penal acusatorio. En este sentido, se considera necesario observar las reglas probatorias

previstas en el marco jurídico procesal penal, así como del principio de contradicción para asegurar el derecho a la prueba de las partes en el proceso penal acusatorio.

El contexto de la conformación de elementos probatorios en la investigación penal

La investigación penal en el proceso penal acusatorio es la etapa procesal en la cual las partes deben realizar su actividad probatoria con relación a su postura acusadora o defensiva. En países de Iberoamérica que han configurado sistemas penales acusatorios la observancia del principio de contradicción y el derecho a la prueba permiten a las partes, desde etapas preliminares, como en la de investigación, ofrecer elementos probatorios para demostrar sus posturas fácticas, siempre y cuando estos se hayan practicado de conformidad con los procedimientos previstos en las legislaciones procesales, sean lícitos y se hayan aportado oportunamente a la carpeta de investigación para posteriormente ser desahogados en las audiencias tanto preliminares como en la de juicio oral y de esta forma tener la posibilidad de debatir los elementos probatorios ofrecidos por la contraparte en el proceso penal.

En ese sentido, en los sistemas penales acusatorios de nuestra Cultura Jurídica se ha asumido que es la parte acusadora quien tiene la carga de la prueba con relación a los hechos delictivos que le imputa a la persona investigada. Aunque esta postura ha sido analizada doctrinariamente y en la actualidad hay un sector de la academia que postula que se trata de un principio que debiera replantearse en el proceso penal (Nieva, 2019), la mayoría de los sistemas de justicia penal de Iberoamérica aun lo comparten y lo han asumido en sus legislaciones procesales. No es momento aquí de profundizar en esta cuestión, pero sí se dirá que al asumir este principio en el proceso penal acusatorio la institución de la Fiscalía, así como la víctima y su asesoría jurídica tienen la obligación de probar aquellas proposiciones fácticas que afirmen en el proceso penal con elementos probatorios que sean adecuados, perti-

nentes, lícitos, que se rijan por la libertad probatoria y que permitan, en sede judicial, la observancia del principio de contradicción.

Por ello, la investigación penal exige no solo la exhaustividad de prácticas probatorias para poder generar líneas de investigación acerca de la ocurrencia de los hechos delictivos y así arribar al esclarecimiento de la verdad, sino que además estas prácticas probatorias deben ser respetuosas de las reglas previstas en la legislación procesal de la materia y de los derechos procesales y humanos de las partes. En ese sentido, de forma incipiente, algunos sistemas de justicia de nuestra Cultura Jurídica han apostado a la incorporación de la Inteligencia Artificial para la conformación de elementos probatorios o de juicio, como lo llama la literatura especializada (Ferrer, 2007). Sin embargo, de la revisión de los marcos jurídicos y trabajo legislativo de veintisiete países de América Latina y el Caribe, es posible afirmar que el material documental hallado se trata en su mayoría de proyectos de ley o iniciativas (Gutiérrez & Hurtado, 2026) y que tales propuestas o iniciativas no necesariamente están encaminadas a establecer disposiciones legislativas que regulen el uso de inteligencia artificial en el proceso penal o, en específico, en la etapa de investigación penal (Gutiérrez & Hurtado, 2026).

Sin embargo, este mismo escenario de incipiente normatividad es el área de oportunidad para considerar que cualquier propuesta de regulación o construcción de marcos jurídicos sobre el uso de la Inteligencia Artificial aplicada a la investigación penal debiera partir de la idea de garantizar los derechos que son propios a las partes que participan del proceso penal: derecho a la prueba, a la seguridad jurídica y a la observancia del principio de contradicción y a su vez, dotar de claridad sobre los tipos de actos de investigación en los cuales estaría permitido el uso de la Inteligencia Artificial, para ello se deben establecer criterios mínimos indispensables para poder garantizar el uso de la inteligencia artificial sin sacrificar los derechos ya garantizados a las partes en el proceso penal acusatorio y con ello fortalecer la confianza ciudadana en el sistema de justicia penal y no generar lo contrario.

Lo anterior es un verdadero reto frente a la protección legal que está detrás de los desarrollos de Inteligencia Artificial que hasta el momento se han desarrollado, preponderantemente, en el sector privado. Es por ello, que, para proponer la incorporación de la Inteligencia Artificial en la investigación penal tendría que partirse de que el desarrollo tecnológico se realice en el sector público, por ejemplo, en sede de las propias Fiscalías para los efectos de determinados actos de investigación.

Aunado a lo anterior, se debe partir de la idea que en los sistemas de justicia penal las Fiscalías desarrollan una función de investigación de los hechos delictivos para lograr su esclarecimiento. Que en esa función deben observar los principios de legalidad, objetividad, eficiencia, profesionalismo y respeto a los derechos humanos, realizando una investigación exhaustiva, inmediata, eficiente e imparcial, libre de estereotipos y discriminación, orientada a explorar todas las líneas de investigación posibles que permitan esclarecer el hecho que la ley señala como delito (Cámara de Diputados, 2026). De ahí que, en principio su actividad investigativa no tenga como única finalidad acusar a una persona específica, sino indagar con completitud y exhaustividad, generando lo que la doctrina ha denominado líneas de investigación, las cuales permitan explicar el hecho y aclararlo, para poder llevar a proceso penal a las personas genuinamente responsables y evitar la acusación de inocentes o falsos positivos como lo analiza ampliamente la literatura especializada (Laudan, 2013).

Si eso es así, una Inteligencia Artificial desarrollada por el sector público, específicamente, por la institución del Estado que tiene a su cargo la investigación de los delitos debería garantizar que las personas investigadas tengan oportunidad de estar no solo informadas de los actos de investigación en los que se aplique inteligencia artificial, sino poder confrontarlos o debatir tanto sus fuentes o datos de abastecimiento, algoritmos empleados y su resultado. Sin embargo, esto no es tan sencillo por la propia complejidad con la que opera la inteligencia artificial como se verá a continuación.

El derecho a la prueba de las partes en el proceso penal

Hasta el momento la doctrina ha puesto de relieve que el contexto de la investigación penal es propiamente un contexto de descubrimiento, en el sentido de que a partir de la práctica probatoria es posible comenzar a dilucidar la verdad de los hechos (Guillén, 2023). En los sistemas de justicia penal acusatorio el derecho a la prueba se trata de un derecho que debe ser garantizado a las partes, pues éste permite lograr las exigencias epistemológicas del propio proceso penal (Ferrer, 2007).

La doctrina señala que, en general, se considera que el derecho a la prueba “como una especificación, un derivado, del derecho a la defensa” (Gimeno, 1988). Asimismo, este derecho a la prueba ha sido reconocido a nivel constitucional en Estados constitucionales de nuestra Cultura Jurídica (Ferrer, 2020) e incluso, incorporado a instrumentos internacionales en materia de derechos humanos, como es el caso del artículo 6.3.d) del Convenio Europeo para la Protección de los Derechos Humanos y las Libertades Fundamentales (Consejo de Europa, 1950); así como en el caso del artículo 8.2.c) de la Convención Americana de Derechos Humanos en el cual se presenta como un derecho incorporado a su vez en el derecho de defensa, el cual ha generado línea jurisprudencial en sede de la Corte Interamericana de Derechos Humanos para hacer patente el derecho a la prueba como derecho humano, pero además destacar que en las decisiones judiciales los tribunales deben expresar los argumentos que razonadamente les permitieron llegar a las conclusiones que señalan en las decisiones judiciales (CoIDH, 2008).

De lo anterior es posible sostener que no basta con que un ordenamiento jurídico prevea a favor de las partes el derecho a la prueba, ya sea contemplándolo como derecho procesal o incluso como derecho constitucional, sino que esto obliga a su vez, sobre todo a los órganos de impartición de justicia (juzgados y tribunales) a señalar de forma expresa las razones por las cuales se les ha de conferir o no valor pro-

batorio y realizar una argumentación racional de esos elementos considerados como pruebas, tanto en lo individual, como en su conjunto, que les han llevado a concluir tener o no por probados ciertos hechos de un caso específico.

La idea anterior no es novedosa, sino que data de la doctrina procesalista más aceptada y que ha sido adoptada en los sistemas penales acusatorios de Iberoamérica (Comoglio, 1970; Cappelletti- Vigoriti, 1971).

Algunas experiencias de inteligencia artificial aplicadas a la investigación penal y sus problemáticas los sistemas de justicia penal que a la fecha han incorporado la inteligencia artificial a la investigación penal lo han realizado para facilitar actuaciones específicas que se realizan en sede de los órganos de investigación, tales como: reconstrucción de lugar de los hechos, reconocimientos faciales u otro tipo de reconocimientos biométricos o bien, en la realización de la evaluación de riesgos para los efectos de medidas cautelares, entre otras tareas. Sin embargo, aunque a primera vista resulte innovadora la inclusión de esta tecnología en la realización de actos de investigación complejos, facilitando y acelerando su práctica en comparación al trabajo meramente humano, uno de los puntos más débiles y cuestionados en algunas experiencias se centra en la poca transparencia de los procesos automatizados con los que operan estas inteligencias artificiales e incluso en los sesgos que están detrás del algoritmo con el que arrojan sus resultados. Tal es el caso en el sistema de justicia estadounidense con el uso del sistema COMPAS (*Correctional Offender Management Profiling Alternative Sanctions*).

En efecto, para ejemplificar una de las experiencias más polémicas y sesgadas acerca del uso de la inteligencia artificial en el proceso penal se encuentra el sistema COMPAS, invención de la empresa Northpoint (Serventich, 2022). Esta inteligencia artificial opera a través de un algoritmo protegido por los derechos de propiedad intelectual, esto implica que no se trata de un algoritmo público, pues ni siquiera puede ser desvelado a favor de las partes que actúan en el proceso penal en

el cual se emplea. La función de esta inteligencia artificial es predecir el riesgo de incidencia delictiva de una persona a partir del análisis de variables con las que se ha abastecido, realizando una especie de *check list* acerca de las condiciones personales de quien tiene el carácter de persona investigada en el proceso penal para ofrecer un resultado que permite clasificarla como una persona peligrosa o no, y en todo caso, arrojando una clasificación de dicha peligrosidad.

El uso de la anterior inteligencia artificial se ha extendido en el sistema judicial estadounidense. Sin embargo, su legitimidad ha sido puesta en duda por diversos estudios doctrinarios (Danesi, 2020) e incluso a través de litigios se han logrado desestimar los resultados de esta inteligencia artificial para los efectos probatorios de procesos de índole criminal en los cuales se ofrecieron (CEJA: 2026).

Los estudios realizados al sistema COMPAS han advertido sobre los sesgos raciales y discriminatorios que están detrás del algoritmo a partir del cual se producen los resultados en casos concretos. En efecto, la investigación realizada por la ONG ProPublica demostró que “en varios casos a las personas afroamericanas se les asignaba un puntaje más alto, a pesar de que luego no reincidían, y, a la inversa, personas caucásicas a las que se les fijaba puntaje bajo, aunque después reincidían” (Requert, citado por Serventich, 2022). En ese sentido, abastecer una Inteligencia Artificial con información acerca del origen nacional, étnico, económico, laboral e incluso religioso de las personas sin mayores respaldos empíricos de riesgo o peligro procesal a los efectos del proceso penal, así como programar un algorítmico sesgado y discriminatorio, implica la obtención de un resultado con las mismas cualidades, lo que en un sistema de justicia penal que se diga comprometido con el respeto a los derechos humanos de las personas implica una violación flagrante a los mismos.

Otros ejemplos del uso de la inteligencia artificial en la etapa de investigación penal que también se advierten en la experiencia estadounidense son aquellos que se han dirigido a desarrollar técnicas forenses de segunda generación (Polanski, 2020). Estos desarrollos de

inteligencia artificial están enfocados hacia el análisis de ADN, el rastreo satelital, análisis biométricos de reconocimiento facial y de iris (Serventich, 2022). El funcionamiento de estos desarrollos tecnológicos se ve mejorado en la medida en que son abastecidos con más información. A partir del cúmulo de datos que posee, la inteligencia artificial identifica patrones complejos y con ello, predice comportamientos, detectando similitudes y anomalías, para así tomar decisiones afines a su codificación (Polanski, 2020).

No obstante, los anteriores ejemplos del uso de la inteligencia artificial en el sistema de justicia estadounidense, destinados a la investigación penal son solo algunas experiencias dirigidas a actos de investigación concretos que en ningún sentido acotan la diversidad de actuaciones de indagación de los hechos que conlleva un sistema de justicia penal. En efecto, hay que partir de la idea de la complejidad de algunos hechos delictivos, los diversos tipos de evidencia, así como los contextos sociales y de incidencia delictiva en los que ocurren los hechos relevantes para el Derecho Penal, mismos que pueden llegar a ser sumamente especializados y exigir la aplicación de ciencias forenses específicas.

El problema de la transparencia y explicabilidad de la inteligencia artificial

Para continuar con la discusión acerca del problema de la transparencia y explicabilidad del uso de la Inteligencia Artificial en la investigación penal conviene referirse a un concepto de Inteligencia Artificial. En realidad, intentar presentar una definición exacta de lo que es la Inteligencia Artificial es sumamente complejo. La doctrina especializada señala que parte de esa complejidad se encuentra en el propio concepto de inteligencia, respecto del cual aún no hay un consenso generalizado.

Sin embargo, cuando se ha intentado definir qué es la Inteligencia Artificial la literatura ha coincidido en al menos una idea central

por su ubicación en las disciplinas de la Ciencia y la Tecnología. En ese sentido se afirma que la Inteligencia Artificial:

es la subdisciplina del campo de la Informática que busca la creación de máquinas que puedan imitar comportamientos inteligentes. Estos comportamientos pueden ser muy diversos: conducir, analizar patrones, reconocer voces o ganar juegos. Son muchas las formas mediante la cual una máquina puede simular un comportamiento inteligente. (DotCSV: 2026)

En ese sentido, hay áreas en las cuales la Inteligencia Artificial tiene mejor desempeño que en otros, e incluso un desempeño mayor o mejor al que tendría un ser humano, pero esta no es la regla. Incluso dentro de las inteligencias artificiales que se han desarrollado a la fecha existen aquellas destinadas a realizar tareas muy específicas, que al intentar emplearlas para alguna diversa a la que fue programada pueden dar resultados equívocos o desastrosos. Incluso este es uno de los temas que se siguen estudiando en los centros de investigación que tienen como objetivo el desarrollo de Inteligencia Artificial, pues las multitareas que, por ejemplo, un ser humano realiza cuando habla, camina, ríe y se emociona al mismo tiempo, son un conjunto de actividades que las inteligencias artificiales en el mercado aún no logran (DotCSV, 2026).

Las características que presentan las Inteligencias Artificiales ha llevado a que se propongan ciertas clasificaciones. Una primera clasificación es aquella que distingue entre Inteligencias Artificiales débiles y fuertes. Se consideran débiles aquellos sistemas de Inteligencia Artificial que únicamente pueden cumplir con un conjunto muy limitado de tareas. Por el contrario, las fuertes hacen referencia a aquellas que son capaces de aplicarse a un conjunto de problemas y dominios diferentes. La mayoría de las Inteligencias Artificiales en el mercado actual pertenecen al primer grupo (CEJA, 2026).

Dentro del concepto de inteligencia artificial es importante el elemento consistente en “imitar comportamiento inteligente” pues es la clave para intentar clarificar los demás conceptos relacionados con esta tecnología. Imitar no es entendido necesariamente como un comportamiento cognitivo, sino que se refiere al resultado de una programación con que opere una Inteligencia Artificial, esto es, si está programada para arrojar un resultado a partir de una orden ejecutiva que se introduce y esa orden ejecutiva corresponde a las capacidades tecnológicas con las que cuenta la inteligencia artificial, arrojará un resultado acorde a ello. Visto así, ese ya es considerado un comportamiento inteligente dentro del campo de la Inteligencia Artificial (DotCSV, 2026).

También es preciso señalar que las bases de datos de estos sistemas de Inteligencia Artificial son algoritmos o también llamados códigos fuente. Estos, por lo general, son abundantes y a partir de ellos se producen los resultados, imitando la actividad humana de procesar información (Serventich, 2022). Aun cuando un sector de la doctrina especializada refiere que las inteligencias artificiales “realizan una actividad de pensar o imitar el pensamiento humano” (Serventich, 2022), aquí me aparto de esa postura, pues aun cuando se puede reconocer que la actividad algorítmica es de suma complejidad, parecida a la complejidad que representa la actividad neuronal humana, considero que el resultado algorítmico de una Inteligencia Artificial es producido de forma bruta y automatizada, esto es, en el sentido de que es carente de un enfoque afectivo o empático, mucho menos racional de tipo explicativo, como acontece en la actividad neuronal humana.

Siguiendo con los tipos de Inteligencia Artificial se advierten diversas subcategorías que se pueden identificar con diversos comportamientos considerados inteligentes: i) aquellas enfocadas en movimientos, desarrolladas por la Robótica; ii) las dirigidas a la capacidad de entender el lenguaje, del cual se ocupa el *Natural Language Processing*, contemplando la conversión de voz a texto o de texto a voz. Estas capacidades conforman campos de estudio propios dentro del campo de la Inteligencia Artificial, pero el campo que permite considerar a un

comportamiento como inteligente dentro de este sector es el *Machine Learning*. El *Machine Learning* o Aprendizaje Automático es la rama de la Inteligencia Artificial que se encarga del estudio de cómo dotar a las máquinas de la capacidad de aprendizaje, entendida ésta como la generación del conocimiento a partir de un conjunto de experiencias o procesamiento de datos previos. (Irazabal, 2025) Este tipo de aprendizaje se clasifica a su vez en tres grupos: a) aprendizaje supervisado; b) no supervisado y c) reforzado. De esta forma el *Machine Learning* es un componente nuclear dentro del campo de la Inteligencia Artificial, pues esto ha permitido dentro del sector identificar sistemas que realizan tareas porque los propios sistemas han aprendido a realizarlas.

Dentro del propio *Machine Learning* hay diversidad de técnicas que sirven para cubrir diferentes tipos de aplicación, tales como: árboles de decisión, modelos de regresión, modelos de clasificación, técnicas de caracterización, entre otras (DotCVS, 2017). Sin embargo, la técnica que ha dado mayor visibilidad al *Machine Learning* ha sido la de redes neuronales, las cuales son capaces de aprender de forma jerarquizada, esto es, por capas. Regularmente, en las primeras capas se encuentran conceptos muy concretos y en las capas posteriores se utilizan otros conceptos aprendidos previamente para aprender conceptos más abstractos. (Arellano, 2019; Cotino y Castellanos, 2022). Lo anterior implica que, a mayor número de capas, la información es más abstracta. La tendencia en este tipo de algoritmos es que agreguen más capas convirtiéndose en algoritmos más complejos. La complejidad de estos algoritmos es lo que hace que sean conocidos como *Deep Learning* (Irazabal, 2025).

De ahí que como lo han afirmado autores como Dupuy y Corvalán (2021), la actividad de una inteligencia artificial es disruptiva y engloba varios tipos. El tipo de inteligencia artificial que se ha empleado en el desarrollo de aquellas dirigidas a la investigación penal es del tipo de aprendizaje automatizado o mejor conocido como *machine learning*, el cual permite ejecutar una multitud de tareas, apelando a una serie de órdenes o bien llamados *prompts* y desde luego, los datos o información con los que se haya abastecido la inteligencia artificial.

Asimismo, dentro del tipo de inteligencia artificial de *machine learning* se encuentra el subtipo conocido como aprendizaje profundo o *deep learning*, el cual se trata de “un modelo inteligente que procesa información a través de interacciones complejas, algo así como redes artificiales que imitan el funcionamiento de las neuronas y mecanismos básicos del cerebro” (Serventich: 2022).

Dicho lo anterior, una vez revisados los documentos que a la fecha han propuesto ciertas directrices para el uso de la Inteligencia Artificial tanto en el sector público, como en el sector jurídico, se advierte que las referencias que realizan se dirigen, preferentemente, a la actividad jurisdiccional y no propiamente a la conformación de elementos probatorios en la etapa de investigación penal (UNESCO, 2026). A propósito de ello, las experiencias que se han observado a nivel global y que han permitido reflexionar para proponer una serie de directrices aplicables al sistema de justicia de los Estados han partido de documentar aquellos sistemas de Inteligencia Artificial que han sido utilizados por los Poderes Judiciales “para apoyar tareas de gestión (por ejemplo, toma de notas e informes), tareas documentales (por ejemplo, clasificación de documentos), funciones judiciales (por ejemplo, identificar la jurisprudencia pertinente) y corregir y sintetizar textos bajo la supervisión del usuario” (UNESCO, 2026, . p. 9).

De esta forma, las experiencias sobre la utilización de la Inteligencia Artificial en sistemas de justicia están encaminadas a tareas de procesamiento de cúmulos de información, pero no para la conformación o producción de elementos probatorios para efectos del proceso penal.

En este contexto de construcción o conformación de elementos probatorios en el proceso penal es de suma relevancia la transparencia y explicabilidad de los procedimientos con los que opera la Inteligencia Artificial, pues una de las finalidades del principio de contradicción, como se dijo, es que las partes procesales puedan cuestionar tanto el origen de los elementos probatorios aportados por la contraria, en el sentido amplio de poder someter a debate desde la información que

ha servido para su producción, como el procedimiento o técnicas que se han aplicado en su realización, lo que en el caso de las inteligencias artificiales es de suma complejidad.

La transparencia y la explicabilidad se consideran principios éticos en el contexto de los sistemas de Inteligencia Artificial (CE, 2018). Ambos principios guardan una relación estrecha, pues el principio de explicabilidad incluye necesariamente a la transparencia tanto de los datos, el sistema y el modelo de negocio u operación (CE, 2018). Por lo tanto, la transparencia de la Inteligencia Artificial implica observar la trazabilidad, explicabilidad y comunicación de ésta.

En la doctrina especializada la trazabilidad es “el conjunto de datos y los procesos que dan lugar a la decisión del sistema de IA, incluidos los relativos a la recopilación y etiquetado de los datos, así como a los algoritmos utilizados” (UNESCO, 2026). La trazabilidad en una Inteligencia Artificial permitirá identificar los motivos o razones de una decisión o resultado erróneo por parte del sistema, lo que, a su vez, ayudará a identificar, reconfigurar y prevenir errores en el futuro. Por lo tanto, la trazabilidad ayuda la auditabilidad y explicabilidad de la inteligencia artificial (UNESCO, 2026).

Por su parte, la explicabilidad consiste en “la capacidad de explicar los procesos técnicos de un sistema de IA como las decisiones humanas asociadas... La explicabilidad técnica requiere que las decisiones que adopte un sistema de IA sean comprensibles para los seres humanos y estos tengan la posibilidad de rastrearlas” (UNESCO, 2026). La exigencia de la explicabilidad de la IA en el sector jurídico y, sobre todo, aplicable a la generación de actos de investigación en el proceso penal es tal pues la producción de elementos probatorios tiene un impacto significativo en la vida de las personas. En ese sentido debería “ser posible reclamar una explicación adecuada del proceso de toma de decisiones del sistema de IA” (UNESCO, 2026).

En estos casos, no solo debería tenerse en cuenta que la IA deba ser explicada a las personas justiciables, sino que dicha explicación sea proporcionada de forma clara, de tal suerte que una persona no exper-

ta en el uso de IA pueda comprender la información que se le presenta. No hay que olvidar que una de las justificaciones para la implementación de los sistemas penales acusatorios en países de nuestra región fue, entre otras cosas, poder facilitar el arribo de información de las partes procesales a los órganos de investigación penal y de impartición de justicia; que uno de los argumentos de tipo sociológico jurídico que está detrás de la oralidad de los procesos penales es que la comunicación con el tribunal pueda realizarse en lenguaje comprensible y sencillo para las personas justiciables, y que a su vez, tanto los órganos de investigación como de impartición de justicia comuniquen de la forma más clara posible las implicaciones de las actuaciones que desarrollan a las personas que participan del proceso penal.

Ahora bien, todas las aplicaciones de inteligencia artificial hoy en día realizan su actividad a partir de redes neuronales profundas. Éstas intentan simular el aprendizaje humano y funcionan a partir de una gran cantidad de datos y a través de procesos de automatización difíciles de explicar o dilucidar, esto es, saber exactamente cómo una red neuronal va aprendiendo los grandes modelos de lenguaje y explicar con exactitud y secuencia qué hacen cada una de las conexiones de una Inteligencia Artificial con cada dato es imposible (Irazabal: 2026).

Conclusiones

El escenario de funcionamiento de la inteligencia artificial que se ha planteado es el que, en todo caso, tendría que considerarse en un sistema penal acusatorio que incorporara la inteligencia artificial en la etapa de investigación penal, manteniendo el respeto del derecho a la prueba y la garantía de observancia del principio de contradicción. Sin embargo, como se ha planteado, el escenario representa un reto ante las propias dificultades explicativas de los sistemas de Inteligencia Artificial automatizados, que, por los propios procesos de operación, explicar la serie de conexiones neuronales que llevan a cabo para obtener un resultado es de suma complejidad o imposibilidad.

En ese sentido, el sistema de justicia penal que incorporara la aplicación de la Inteligencia Artificial a la etapa de investigación pe-

nal, tendría que, en primer lugar, determinar en específico qué acto o actos de investigación podrían ser susceptibles de realizarse con apoyo de esta tecnología, previendo desde luego tal cuestión en la legislación procesal correspondiente, en la que además, se consideraran los supuestos en los cuales las partes procesales o en específico, la parte que esté inconforme con la operación o resultado obtenido por parte de la Inteligencia Artificial pudiera cuestionarlo en sede judicial, esto es, en audiencia mediante la interposición de un medio de impugnación.

Aunado a lo anterior, probablemente, tendría que considerarse que si la inconformidad de alguna de las partes se sostiene con relación a la operación y resultado de la Inteligencia Artificial aplicada en un determinado acto de investigación la audiencia en la que se sostenga tal debate ante el órgano jurisdiccional tendría que ser de carácter privado, es decir, este tipo de audiencia debería preverse dentro de las excepciones al principio de publicidad que también rige en la metodología de audiencias de los sistemas penales acusatorios.

Lo anterior, con la finalidad de que, ante la develación de cierta información a las partes sobre el funcionamiento de la inteligencia artificial, primero se garantizara la transparencia y explicabilidad de su operación, pero con la previsión de mantener privada la audiencia de debate, se protegieran los derechos de propiedad intelectual.

Desde luego, que, en este punto, como lo he sostenido con anterioridad, me decanto porque los sistemas de Inteligencia Artificial aplicados para la conformación de actos de investigación se traten de desarrollos del sector público, encabezados por las Fiscalías, que se tratan de los entes competentes en un Estado de Derecho para la investigación de los hechos delictivos. Asimismo, que para la protección de tales derechos de propiedad intelectual, se prevea que las partes procesales, en los casos donde se haya realizado el uso de cierta inteligencia artificial para la conformación de elementos probatorios, primero tengan garantizado su derecho a impugnar, así como a acceder a la información que sirvió tanto de abastecimiento de datos, como a aquella que se refiera a la operación de la propia inteligencia artificial

—pese a sus limitaciones explicativas propias por lo complejo de la tecnología en cuestión— y a su vez, se tuviera la previsión de su obligación de no revelar tal información o emprender alguna acción que pudiera comprometer la vulneración de derechos de propiedad intelectual.

El panorama para la incorporación de la Inteligencia Artificial en los sistemas de justicia penal y, específicamente, en la investigación penal, aún está en vías de configuración. Como se señaló con anterioridad, algunos países han comenzado a revisar sus marcos jurídicos para legislar en materia de Inteligencia Artificial, pero en ninguna de las experiencias de la región de América Latina y El Caribe podría sostenerse en la actualidad que se haya logrado lo anterior y mucho menos que los marcos jurídicos se traten de normas robustas, claras y que permitan la transparencia de los procesos de la Inteligencia Artificial.

Probablemente, lo anterior se explica porque aun con el trabajo especializado que se está realizando en el sector tecnológico, en el sector jurídico las instituciones responsables de la innovación y modernización del sistema de justicia aún no han mirado con relativo compromiso la incorporación de esta tecnología por todos los retos que conlleva, pues desde luego, entre otras cosas, amerita contar con un presupuesto para poder impulsar un proyecto de esta magnitud en sede de órganos operadores del sistema de justicia penal, que amerita, desde luego, contemplar un cierto tiempo tanto en la creación o desarrollo tecnológico, pilotaje o periodo de prueba, supervisión de resultados, identificación de errores, reconfiguración y seguimiento para su puesta en marcha. Sin embargo, pienso que mientras más pronto cada país comience a cuestionarse tan solo acerca del uso de la Inteligencia Artificial en sus sistemas de justicia penal se estará incorporando a uno de los escenarios indiscutibles por los que atravesará el sector jurídico a nivel global.

Referencias

- Arellano, W. (2019). El derecho a la transparencia algorítmica en big data e inteligencia artificial. *RGDA Iustel*, 50.
- Cámara de Diputados. (2026). *Código Nacional de Procedimientos Penales*.
- Cappelletti, M., y Vigoriti, V. (1971). Il diritto costituzionale delle parti nel processo civile italiano. *Rivista di diritto processuale*, (1).
- CEJA. (2026). *Curso Inteligencia Artificial y Sistemas de Justicia: Perspectivas éticas, experiencias y regulación jurídica*. Centro de Estudios de Justicia de las Américas.
- Comoglio, L. (1970). *La garanzia costituzionale dell' azione e il processo civile*, Padova. Cedam.
- Consejo de Europa. (1950). *Convenio Europeo para la Protección de los Derechos Humanos y las Libertades Fundamentales*. Council of Europe.
- CoIDH. (2008). *Caso Apitz Barbera y otros ("Corte Primera de lo Contencioso Administrativo") vs. Venezuela*. Corte Interamericana de Derechos Humanos.
- Cotino, L. y Castellanos J. (2022). *Transparencia y Explicabilidad de la Inteligencia Artificial y "Compañía" (Comunicación, Interpretabilidad, Inteligibilidad, Auditabilidad, Testabilidad, Comprobabilidad, Simulabilidad...)*. Para qué, para quién y cuánta. *Transparencia y explicabilidad de la Inteligencia Artificial*. Tirant Lo Blanch.
- Danesi, C. (2020). *Inteligencia artificial, tecnologías emergentes y derecho*. Hammurabi.
- DotCVS. (2017). ¿Qué es el Machine Learning? ¿Y el Deep Learning? Un mapa conceptual. [video]. Youtube. <https://www.youtube.com/watch?v=KytW151dpqU>
- Ferrer, J. (2007). *La valoración racional de la prueba*. Marcial Pons.
- Gimeno, J. (1988). *Constitución y proceso*. Tecnos.

- Guillén, G. (2023). *Técnicas, actuaciones y actuaciones de investigación criminal en el Código Nacional de Procedimientos Penales*. Universidad Nacional Autónoma de México, Instituto de Investigaciones Jurídicas.
- Gutiérrez, J. (2026). *Directrices para el uso sistemas de IA en cortes y tribunales*. UNESCO.
- Gutiérrez, J. D. & Hurtado, S. (2026). *Regulación sobre inteligencia artificial en América Latina y el Caribe (Versión V.3.0)*. Universidad de los Andes.
- Irazabal, E. (2025). Legaltech e Inteligencia Artificial: adaptarse o quedarse atrás. La Tertul-IA Podcast: Inteligencia Artificial. [video]. Youtube. <https://www.youtube.com/watch?v=koj8-WHewLU>
- Irazabal, E. (2026). *Tercera sesión de la Segunda Edición del Curso Inteligencia Artificial y sistemas de justicia*. CEJA.
- Laudan, L. (2013). *Verdad, error y proceso penal. Un ensayo sobre epistemología jurídica*. Marcial Pons.
- Nieva, J. (2019). *Contra la carga de la prueba*. Marcial Pons.
- Riquert, M. A. (2022). *Inteligencia artificial y derecho penal*. Ediar.
- Serventich, C. (2022). Inteligencia artificial en el proceso penal. ¿Más vale humano conocido o algoritmo por conocer? *Revista Jurídica Austral*, 3(2), 869-880.

Sidney Ernestina Marcos Escobar

Universidad Veracruzana | Xalapa | Veracruz | México

<https://orcid.org/0000-0001-8366-4037>

smarcos@uv.mx

sidneymarcosescobar@gmail.com

Doctora en Derecho Procesal, por el Centro Mexicano de Estudios de Posgrado, Maestra y Licenciada en Derecho por la Universidad Veracruzana, Investigadora del Sistema Nacional de Investigadoras e Investigadores (SNII) SECIHTI México, Investigadora adscrita al Instituto de Investigaciones Jurídicas de la Universidad Veracruzana

Transparency and Algorithmic Explainability of Artificial Intelligence Applied to Criminal Investigation and the Right to Evidence

Abstract

This research aims to reflect upon the incipient use of Artificial Intelligence in the adversarial criminal justice systems of Global South countries. To this end, it begins with an analysis of common rules and principles in the criminal investigation stage within the Criminal Justice Systems of our Legal Culture. Subsequently, it problematizes the issue by presenting reflections on the transparency and explainability of Artificial Intelligence systems that have been employed in some international experiences, in order to highlight that aspects such as trade secrets, the type of data used to supply Artificial Intelligence, and the automation of procedures to obtain results generate a scenario of limited transparency and explainability of the results obtained through the use of Artificial Intelligence for the purposes of criminal proceedings. In this sense, the implications of the use of Artificial Intelligence specifically in the criminal investigation stage are analyzed, critically pointing out the weaknesses and strengths of the evidentiary elements generated with this technology, as well as their impact on respect for the right to evidence and the adversarial principle that governs the accusatorial criminal process. Finally, some key issues are raised that could serve as a starting point for considering the implementation of Artificial Intelligence in the criminal justice systems of the Global South that would be, at the same time, reliable, respectful of the right to evidence, and of the guiding principles that govern adversarial systems.

Keywords: Artificial Intelligence; criminal investigation; transparency; explainability; evidence; criminal justice system.

Transparência e explicabilidade algorítmica da Inteligência Artificial aplicada à investigação penal e ao direito à prova**Resumo**

Esta pesquisa tem como objetivo refletir acerca do emprego incipiente da Inteligência Artificial nos sistemas de justiça penal acusatório de países do Sul Global. Para tanto, parte-se da análise de regras e princípios comuns na etapa de investigação penal, nos Sistemas de Justiça Penal de nossa Cultura Jurídica. Posteriormente, problematiza-se apresentando reflexões acerca da transparência e da explicabilidade dos sistemas de Inteligência Artificial que têm sido empregados em algumas experiências internacionais, para advertir que aspectos tais como o segredo industrial, o tipo de dados com que se abastece a Inteligência Artificial e a automatização dos procedimentos para obtenção de resultados geram um cenário de pouca transparência e explicabilidade dos resultados obtidos mediante o uso de Inteligência Artificial para os efeitos do processo penal. Nesse sentido, analisam-se as implicações do uso de Inteligência Artificial especificamente na etapa de investigação penal, apontando criticamente as fragilidades e fortalezas dos elementos probatórios gerados com essa tecnologia, bem como sua incidência no respeito ao direito à prova e ao princípio do contraditório que rege o processo penal acusatório. Finalmente, levantam-se algumas questões que poderiam ser chave para começar a pensar na implementação de alguma Inteligência Artificial nos sistemas de justiça penal do Sul Global que, ao mesmo tempo, fosse confiável, respeitosa do direito à prova e dos princípios norteadores que regem nos sistemas acusatórios.

Palavras-chave: Inteligência Artificial; investigação penal; transparência; explicabilidade; prova; sistema penal.