RIESGOS DEL USO DE ALGORITMOS PREDICTIVOS EN LA JUSTICIA PENAL[*]

Odone Sanguiné[**]

RESUMEN: 1.- Debilidades y riesgos del uso de algoritmos predictivos en las decisiones judiciales penales; 2.- Los déficits de objetividad de los datos de los algoritmos predictivos; 3.- El sesgo algorítmico y sus efectos discriminatorios; 4.- La opacidad algorítmica y el derecho al debido proceso; 5.- IA y violación del derecho a la presunción de inocencia; 6.- La falibilidad algorítmica y el derecho de defensa y al recurso; 7.- Deber de motivación, transparencia e IA explicable.

1.- **Debilidades y riesgos del uso de algoritmos predictivos en las decisiones judiciales penales.**

La introducción de la inteligencia artificial (en adelante IA) es un fenómeno inevitable y, en muchos aspectos, positivo, por lo que ha venido a quedarse en el ámbito de la Administración de Justicia, que también es permeable a las nuevas tecnologías y está abocado a convivir[1].

La implementación de la IA en el ámbito jurisdiccional penal se materializa, entre otros aspectos, en el uso de la llamada "justicia predictiva", la cual permite, a través de algoritmos predictivos, que el juez pueda adoptar decisiones judiciales dentro del proceso penal, dirigidas a obtener rapidez, eficacia y seguridad jurídica en la aplicación de la ley, encaminadas hacia la predicción del riesgo de reiteración delictiva por parte del imputado, sea en el ámbito de las medidas cautelares personales o bien, del riesgo de reincidencia en la etapa de ejecución de la pena o, en definitiva, como herramienta para evaluar el riesgo de violencia doméstica[2].

Para los defensores de este tipo de herramientas de evaluación de riesgos mediante algoritmos, su uso no se basa en la impresión subjetiva del juez sobre el grado de peligrosidad que presenta el sujeto, sino en un conocimiento riguroso motivado empiricamente sobre la existencia de dicho riesgo[3].

No obstante, aunque es cierto que las estimaciones de riesgo realizadas con herramientas automatizadas de IA tienen algunas ventajas respecto de los métodos tradicionales, por sus características: complejidad, opacidad, aprendizaje automático, continuo y autónomo, los algoritmos predictivos presentan una serie de déficits, riesgos, paradojas o contradicciones como herramienta jurisdiccional en los procesos penales que merecen ser tomadas en consideración por los juristas[4].

En primer lugar, la complejidad del propio funcionamiento de los sistemas de IA. Los algoritmos se basan en herramientas estadísticas muy complejas, pues funcionan a través de redes neuronales ("deep learning*"*) que, cuanto más complejas son, menos interpretables resultan, lo que deviene en

---

[*] Texto modificado y actualizado de la ponencia presentada en la XVIII Edición del "Corso Internazionale di Formazione in Diritto Penale, XVIII Edizione: AI, Metaverse and Criminal Law", 6 y 7/10/2023, realizado por el Dipartimento di Scienze Giuridiche dell'Università degli Studi di Salerno, Italia.

[**] Profesor Catedrático de Derecho Penal y Procesal Penal de la Facultad de Derecho – UFRGS – Brasil; Doctor por la Universitat Autònoma de Barcelona – España; Magistrado (jubilado) del Tribunal de Justicia del Estado RS (TJRS).

[1] M.D. García Sánchez, *El necesario balance entre la heurística algorítmica y judicial como garantia de los derechos procesales del justiciable; hacia una inteligência artificial explicable*, in P. Martín Ríos, C. Villegas Delgado (dirr.) *La tecnologia y la inteligencia artificial al servicio del processo*, Madrid 2023, 334; J.L. Gómez Colomer, *Derechos fundamentales, proceso e Inteligencia Artificial: una reflexión*, in Martín Ríos, Villegas Delgado (dirr.), *La tecnologia* cit. 259.

[2] R. Castillejo Manzanares, *Cuáles son las razones que obstaculizan la introducción de la IA en el processo judicial. Especial referencia al processo penal*, in Martín Ríos, Villegas Delgado, *La tecnologia* cit. 91.

[3] M. Llorente Sánchez-Arjona, *Inteligencia*, 378.

[4] L. Martínez Garay, A. García Ortiz, *Paradojas de los algoritmos predictivos utilizados en el sistema de justicia penal,* in *Inteligencia artificial y derecho. El cronista del Estado Social y Democrático de Derecho* 100 (2022) 162; Castillejo Manzanares, *Cuáles son las razones* cit. 85.

que cuanta más capacidad de comprensión, menor capacidad de explicar los resultados. Por eso existe un gran desconocimiento sobre lo que sucede en la "caja negra" ("black box") para la toma de decisiones jurisdiccionales penales[5].

En segundo lugar, la introducción de los datos parte de graves problemas. Por un lado, no representa todos los escenarios, no se ingresa mucha información y, la IA y el aprendizaje automático se alimentan de datos y estándares relacionados; por ejemplo, a veces la máquina no sabe por qué hay un escenario que no se introdujo y, en este caso, generaliza. Además, es inherente al propio sistema algorítmico, matemático o relacional, en que puedan existir sesgos o predisposiciones a la hora de aplicar la propia metodología de cálculo e interpretación. Si se encuentran sesgos en los datos de aprendizaje, la capacidad cognitiva dependerá de ellos[6].

Recientemente, el Reglamento de Inteligencia Artificial (UE) 2024/1689 (en lo sucesivo, RIA), aprobado por el Parlamento Europeo para regular y limitar los riesgos de los usos de la IA y garantizar la seguridad y el respeto a los derechos fundamentales, ha puesto de relieve que el sistema de IA es un conjunto de tecnologías disruptivas de rápida evolución que puede generar un amplio abanico de beneficios económicos y sociales en todos los sectores y actividades de la sociedad. Pero, a la vez, debido a su reciente y rápida evolución y dependiendo de las circunstancias de su aplicación y utilización concretas, la IA puede generar riesgos sobre los derechos fundamentales de la ciudadanía. Valga decir que el RIA califica de alto riesgo los sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para llevar a cabo evaluaciones de riesgos individuales de personas físicas, con el objetivo de determinar el riesgo de que cometan infracciones penales o reincidan en su comisión.

2.- **Los déficits de objetividad de los datos de los algoritmos predictivos.**

Dos cuestiones esenciales involucran la cuestión de la objetividad: a) la calidad de la información que entra y sale del sistema y, b) el problema de los sesgos algorítmicos ("algorithmic bias*"*), que se produce cuando un determinado componente de la IA produce resultados diferentes en relación con los sujetos, dependiendo de si la persona pertenece a un grupo específico, evidenciando un prejuicio subyacente hacia ese grupo. Cuando los algoritmos incluyen sesgos, su aplicación puede conllevar una posible discriminación social, ya que las decisiones replican dichas desviaciones y pueden reproducir o amplificar estándares de discriminación presentes en la sociedad[7].

A su turno, la herramienta depende de datos oficiales sobre nuevas detenciones, condenas o ingresos a prisión, por lo que reaparece el problema de la discrepancia entre la delincuencia real y los registros oficiales ("dark number*"*). Por ello, lo que estiman las herramientas automatizadas de evaluación de riesgos, no es tanto el riesgo de comisión de un nuevo delito, sino la actividad policial o judicial futura, es decir, el riesgo de ser detenido o condenado de nuevo. Además, el hecho de que no existan registros de arrestos o condenas de una persona en particular no significa que no haya cometido ningún delito anteriormente. Por otro lado, la delincuencia puede sobreestimarse, v.gr., si la herramienta tiene en cuenta arrestos en lugar de condenas o, por la incidencia de acuerdos de conciliación o procesamientos no penales, en los que el acusado opta por la justicia consensual para evitar un juicio con un resultado incierto, aunque no sea responsable de los cargos[8].

[5] Castillejo Manzanares, *Cuáles son las razones* cit. 85-89.
[6] Id., *Cuáles son las razones* cit. 86.
[7] Id., *Cuáles son las razones* cit. 88.
[8] Martínez Garay, García Ortiz, *Paradojas* cit. 166-167.

Por lo tanto, aunque el proyecto para su creación y validación sea impecable, si los "datos" con los que se construyó el modelo y/o con los que luego se trabaja en la realidad cotidiana presentan todos estos déficits, es inevitable que la calidad de las estimaciones se vea comprometida[9].

Así, es innegable que no existe neutralidad en los datos históricos introducidos, pues, aun cuando el algoritmo lo ejecuta un procesador, es un programador quien introduce los datos de entrada y, como todo ser humano, no está exento de predisposiciones o tendencias, y dichos datos pueden estar fundadamente parcializados, perpetuando – o incluso incrementando – los llamados sesgos o distorsiones cognitivas[10].

3.- **El sesgo algorítmico y sus efectos discriminatorios.**

Las decisiones algorítmicas revelan limitaciones, debido a los riesgos de discriminación por el uso de algoritmos instruidos con datos sesgados[11].

El famoso precedente judicial State v. Loomis, juzgado en 2016 por la Corte Suprema de Wisconsin, es un ejemplo paradigmático de la aplicación práctica de IA utilizada para predecir el comportamiento futuro, así como los riesgos que su uso puede implicar si no se observan garantías suficientes[12]. Debido a que el algoritmo utilizado por la herramienta llamada "COMPAS" consideró que el acusado presentaba un alto riesgo de reincidencia, en la fase de ejecución de la pena se le negó la libertad condicional y se le dictó una condena de seis años en la cárcel.

A pesar de declarar la constitucionalidad de COMPAS, el Tribunal de Wisconsin impuso numerosas restricciones a su uso: el algoritmo no se podía utilizar para determinar si un delincuente sería encarcelado, ni para calcular la duración de su sentencia; su utilización tenía que ir acompañada de una justificación independiente de la sentencia, y, además, cualquier informe de investigación de asistencia que tuviera la puntuación, tenía que contener una elaborada advertencia de cinco partes sobre la utilidad limitada del algoritmo[13].

Sin embargo, el análisis del caso LOOMIS revela los principales problemas a los que se enfrentan las decisiones basadas en algoritmos predictivos: a) en primer lugar, la imposibilidad de conocer los datos con los que se alimenta el sistema, si están sesgados o no, debido a que es método patentado por una empresa privada – es decir, secreto –, exento del requisito de transparencia que siempre debe acompañar el uso de la IA en el contexto jurisdiccional; b) en segundo lugar, el carácter opaco de estos sistemas para los operadores jurídicos, una especie de "caja negra" ("black box") que, en función de un resultado numérico, ofrece una puntuación que marca el decreto de una medida cautelar, un permiso penitenciario o el contenido de una sentencia, entre otros; c) en relación específicamente con el programa COMPAS, los riesgos de discriminación racial por el uso de indicadores de riesgo basados en datos sesgados marcadamente racistas, al multiplicar notablemente la puntuación del algoritmo simplemente por el hecho de que la persona investigada o acusada sea de origen afro-americano[14]; d) el algoritmo COMPAS tiene en cuenta 137 criterios, pero nadie sabe realmente cómo

[9] Ead., *Paradojas* cit.167.
[10] M.D. García Sánchez, *El necesario balance*, 318s.
[11] Martínez Garay, García Ortiz, *Paradojas* cit 168s.
[12] Llorente Sánchez-Arjona, *Inteligencia* cit. 384.
[13] E. Israni, *Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis*, recuperado 09/06/2024 de *https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1*.
[14] Gómez Colomer, *Derechos* cit. 287; P.P. Paulesu, *Inteligencia artificial y processo penal italiano: uma panorâmica*, in Martín Ríos, Villegas Delgado (dirr.) *La tecnologia* cit. 269; Castillejo Manzanares, *Cuáles son las razones* cit. 88.

se lleva a cabo la evaluación de riesgos; por esto, se dice en varios estudios, que la herramienta no es más correcta que una decisión humana y, además, es indudablemente racista[15].
El gran peligro, entonces, reside en que, el uso de la IA para la predicción judicial en materia penal, diga al juez lo que debe decidir en el caso, sustituye el razonamiento humano por uno artificial y, la decisión, en últimas, la decisión la toma una máquina, no un humano, surgiendo el tema del juez-robot[16].
Por ello, resulta prudente plantear algunos límites: a) exigir que su uso se limite a la ejecución de la condena, una vez establecida la culpabilidad del imputado; b) que el pronóstico haya sido probado; c) que se hagan públicos los algoritmos utilizados para calcular la peligrosidad, con el fin de poder defenderse de excesos o sesgos de la máquina[17] y, en definitiva, d) resulte indispensable la intervención y control del juez sobre la decisión final[18].

4.- **La opacidad algorítmica y el derecho al debido proceso.**

Otro gran riesgo de la implementación de la IA radica en la opacidad algorítmica, es decir, en la dificultad de explicar los resultados alcanzados por los algoritmos utilizados, la cual dificulta o, incluso, imposibilita el desarrollo de argumentos, lo que puede derivar en una violación del derecho de defensa, de rango, pues se ignora la heurística algorítmica seguida por el sistema predictivo para lograr sus resultados, por lo que al acusado le resultaría imposible refutar sus predicciones[19].
La falta de transparencia inevitablemente debilita el pregonado rigor científico de los algoritmos predictivos. Además de la violación del derecho a una defensa amplia y contradictoria, la falta de transparencia implica otro problema relacionado con la calidad científica de las valoraciones de riesgo que se realizan con sistemas automatizados[20]. De hecho, si uno de los principales argumentos para introducir sistemas basados en algoritmos automatizados para valorar el riesgo de reincidencia o violencia es el mayor rigor científico de sus resultados, es paradójico que no se cumpla con uno de los supuestos esenciales del método científico: facilitar toda la información necesaria para permitir el examen crítico de terceros[21].
La opacidad evita el análisis crítico por parte de terceros y la contestación de resultados, que lanza una importante sombra de duda sobre el rigor científico de las herramientas[22].
Los sistemas automatizados pretenden ser más transparentes que los juicios clínicos de peligrosidad, pero muchos de ellos protegen celosamente los detalles de los algoritmos con los que están construidos, lo que ensombrece sus pretensiones de cientificidad y choca frontalmente con el derecho constitucional a una defensa plena. Y esta oscuridad probablemente aumentará cuando la inteligencia artificial entre en este campo[23].
Los sistemas de IA suelen ser diseñados por empresas privadas que, amparadas por el secreto corporativo, no revelan su funcionamiento interno, lo que dificulta o, incluso, imposibilita la interpretación de los resultados obtenidos. En los sistemas predictivos dotados de aprendizaje

[15] J. Nieva-Fenoll, *Inteligencia artificial y processo judicial: perspectivas ante un alto tecnológico en el camino*, in *Inteligencia artificial legal y Administración de justicia*, Aranzadi 2022, 431s.
[16] Gómez Colomer, *Derechos* cit. 284s.
[17] Castillejo Manzanares, *Cuáles son las razones* cit.104.
[18] Llorente Sánchez-Arjona, *Inteligencia* cit. 385.
[19] García Sánchez, *El necesario balance* cit. 320s.
[20] Martínez Garay, García Ortiz, *Paradojas* cit.163.
[21] Ead., *Paradojas* cit.163.
[22] Ead., *Paradojas* cit.163.
[23] Ead., *Paradojas* cit. 172.

profundo ("deep learning*"*), la opacidad algorítmica empeora y, dada la infinidad de factores, combinaciones y posibilidades que resultan del funcionamiento de estas redes neuronales artificiales, los propios diseñadores pueden perder la capacidad de controlar la causalidad entre los datos y comprender sus decisiones. La paradoja es la siguiente: cuanto más datos se introduzcan en un sistema predictivo equipado con aprendizaje profundo, más precisos serán sus resultados, pero al mismo tiempo son más autónomos y difíciles de explicar. El algoritmo puede volverse no transparente, resultando el proceso de decisión, en ciertos casos, opaco para sus creadores. Este proceso de cálculo, llamado "heurística algorítmica", es imposible o muy difícil de rastrear incluso para los científicos o ingenieros de datos que crean el algoritmo, de manera que el proceso de decisión del algoritmo se ubica en la llamada caja negra ("black box")[24].

En este contexto, aunque el producto de esta interacción algorítmica muestra las probabilidades de éxito y los códigos introducidos por los programadores, no revela las múltiples variables, premisas y opciones que conducen a la solución, ni el peso relativo que se atribuye a cada una de ellas, dificultando su trazabilidad[25].

Por ello, tanto el uso de programas de predicción mediante algoritmos como la atribución de decisiones a jueces-robots plantean riesgos de violación de derechos constitucionales fundamentales, a saber: a) el principio de igualdad y prohibición de discriminación por los sesgos algorítmicos; b) el derecho a la defensa plena y a la contradicción, considerando que no se sabe cómo "razona" la máquina; c) el derecho a recurrir, ya que la máquina no motiva y, por tanto, el perjudicado desconoce el motivo por el que fue condenado[26].

Por otro lado, la opacidad y la consecuente falta de transparencia derivada de la falta de conocimiento técnico sobre el funcionamiento de un algoritmo o la interpretación de un código informático, puede afectar el derecho de confrontación y el principio de igualdad de armas entre las partes.

Los algoritmos, generalmente, están protegidos por el derecho de propiedad intelectual, lo que imposibilita el acceso al "código fuente", haciendo prácticamente imposible cuestionar sus resultados, lo que pone en grave peligro el derecho al debido proceso legal[27]. Ello produce una "asimetría o desequilibrio cognitivo", al permitir que el sector público y las grandes corporaciones tengan acceso a la tecnología más moderna en razón de la disposición de mayores medios económicos, que los que tienen los particulares[28].

La explicabilidad y trazabilidad del algoritmo se consolida como única forma de garantizar el derecho al recurso. Incluso si las herramientas modernas de estimación del riesgo de reincidencia tienen una base científica y, desde el punto de vista de su calidad científica, no son inferiores a los pronósticos de peligrosidad clásicos, su falta de transparencia debilita inevitablemente su pretensión de rigor científico. Considerando el carácter fundamental de los derechos constitucionales, aun cuando se materializara dicha explicabilidad total, sus predicciones no podrían, por sí solas, servir de base – algorítmica – para una decisión judicial, pues violaría el deber de motivación y la reserva de jurisdicción, por lo que dichos sistemas automáticos en todo caso solo podrían implicar apoyo a la labor judicial. El uso de IA genera el riesgo de automatizar las decisiones judiciales, implicando una reducción de su transparencia y justificación. Por tanto, la motivación de las sentencias es

[24] S. Muñoz Machado, *Prólogo*, in *Inteligencia artificial y derecho. El cronista del Estado Social y Democrático de Derecho* 100 (2022) 11; García Sánchez, *El necesario balance* cit. 320.
[25] García Sánchez, *El necesario balance* cit. 320.
[26] Gómez Colomer, *Derechos* cit. 263.
[27] Castillejo Manzanares, *Cuáles son las razones* cit. 90; Llorente Sánchez-Arjona, *Inteligencia* cit. 393.
[28] Id., *Inteligencia* cit. 393.

fundamental, pues permite a las partes conocer cómo el juez llegó a la conclusión que se refleja en sus decisiones[29].
Por lo tanto, en principio, se debería rechazar el uso de la IA como reemplazo total de la toma de decisiones humanas. La creación de un juez-robot vulnera los principios constitucionales de independencia e imparcialidad judicial[30].

5.- **IA y violación del derecho a la presunción de inocencia.**
Las herramientas tecnológicas predictivas, realizadas por los sistemas de IA basadas en multitud de datos, pueden dar lugar al establecimiento de estándares delictivos que ponen en grave riesgo el derecho fundamental a la presunción de inocencia, debido al hecho de que las características de un sospechoso coincidan con un perfil de riesgo sin haber realizado una mínima actividad probatoria[31], lo que implicaría la inversión de la carga de la prueba, obligando a los acusados a tener que demostrar su inocencia ante las predicciones lanzadas por un algoritmo[32].
Además, se afecta la máxima *in dubio pro reo*, al considerar que no puede haber control sobre la duda, ya que la máquina carece de sensibilidad para dudar[33].
Ahora bien, el RIA europeo de 2024 prohíbe expresamente, por su nivel de riesgo inaceptable, entre otras prácticas, la evaluación o predicción del riesgo criminal individual, teniendo en cuenta que se trata de un sistema de alto riesgo para los derechos fundamentales. Así queda vedado por el artículo 5.1, d) «el uso de un sistema de IA para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de que una persona física cometa un delito basándose únicamente en la elaboración del perfil de una persona física o en la evaluación de los rasgos y características de su personalidad; esta prohibición no se aplicará a los sistemas de IA utilizados para apoyar la valoración humana de la implicación de una persona en una actividad delictiva que ya se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva». Asimismo, el art. 10, II, de la reciente Resolución n. 615, del CNJ (Consejo Nacional de Justicia) de Brasil, de 11 de marzo de 2025, que prohíbe al Poder Judicial desarrollar valorar rasgos de personalidad, características o comportamientos de personas físicas o grupos de personas físicas, con el fin de evaluar o predecir la comisión de delitos o la probabilidad de reincidencia en la base de decisiones judiciales, así como con fines predictivos o estadísticos con el fin de fundamentar decisiones en materia laboral basadas en la formulación de perfiles personales.

6.- **La falibilidad algorítmica y el derecho de defensa y al recurso.**
La opacidad algorítmica imposibilita el desarrollo de argumentos lo que puede derivar en una afectación o aniquilación del derecho constitucional a una defensa adecuada, puesto que se desconoce la heurística algorítmica seguida por el sistema predictivo para lograr sus decisiones o, incluso, cómo llegó la máquina a ese resultado en particular, impidiendo al acusado refutar sus predicciones[34]. Ello genera indefensión, pues impide conocer los motivos que fundamentan una decisión y, por tanto, no permite interponer un recurso adecuado ante posibles sesgos o errores que el afectado podrá sospechar, pero no demostrar, precisamente porque no se le permite conocer el funcionamiento del

[29] García Sánchez, *El necesario balance* cit. 325-330.
[30] Gómez Colomer, *Derechos* cit. 285.
[31] Llorente Sánchez-Arjona, *Inteligencia* cit. 393.
[32] García Sánchez, *El necesario balance* cit. 322.
[33] Gómez Colomer, *Derechos* cit. 279.
[34] García Sánchez, *El necesario balance* cit. 321; Llorente Sánchez-Arjona, *Inteligencia* cit. 390-392.

algoritmo. La mera presentación de un nivel de riesgo como resultado de aplicar un algoritmo, cuando no se sabe cómo funciona ese algoritmo ni a qué factores da más peso o cómo juegan estos entre sí, no es ofrecer una motivación judicial transparente de la decisión, sino dar por descontada la motivación encubriéndose en el algoritmo[35].

Además, recurrir una decisión basada predominantemente en un algoritmo, es algo difícil y solo deja espacio a cuestionar el funcionamiento del algoritmo, explicando por qué no decidió correctamente en el caso concreto. En la argumentación del recurso de casación sería necesario analizar los datos estadísticos y las líneas jurisprudenciales o legales seguidas por el algoritmo, determinando su corrección, y en la parte de hechos, se podría discutir el funcionamiento de la máquina en términos de sus predicciones y evaluación de los antecedentes de sus bases de datos, los únicos a través de los cuales puede decidir, argumentando que la situación evaluada en el caso específico es diferente a aquella que sucedió en el pasado[36].

Así, el resultado algorítmico debe pasar por un filtro de control humano previo y posterior, que garantice la transparencia tanto en su funcionamiento como en relación con las personas que lo programan y supervisan[37]. Para garantizar un equilibrio justo entre las dimensiones digital y humana, es esencial que se haga la supervisión del juez, que nunca podrá ser sustituida por componentes artificiales[38].

Es fundamental contar con poderes públicos, autoridades y organismos públicos encargados de supervisar y certificar ante la ciudadanía que un sistema de IA es transparente, responsable y equitativo[39]. Así, el reciente RIA europeo de 2024, ha establecido: a) una serie de obligaciones de los proveedores antes de introducir un sistema de IA de alto riesgo en el mercado de la UE, y, además, b) el cumplimiento y supervisión del RIA por los Estados miembros de la EU, los cuales deberán designar una o varias autoridades nacionales competentes para supervisar la aplicación y ejecución, así como para llevar a cabo actividades de vigilancia del mercado[40].

En definitiva, es fundamental que todos los algoritmos utilizados en el sistema policial y penal sean transparentes para garantizar derechos constitucionales fundamentales de amplia defensa y contradictorios, y para poder detectar, discutir y, en su caso, corregir posibles predisposiciones o efectos discriminatorios. Asimismo, su funcionamiento podrá ser verificado por terceros[41].

7.- **Deber de motivación, transparencia e IA explicable.**

Así, el uso de estos sistemas tecnológicos no debe comprometer la garantía de una motivación dialéctica de las decisiones judiciales. Los algoritmos predictivos no puedan eliminar o reemplazar las heurísticas del juez debido al principio constitucional de reserva de jurisdicción. Por tanto, junto al proceso de determinación judicial de los resultados obtenidos con algoritmos predictivos, es necesario sumar la heurística del juez a la motivación de la decisión. Los resultados de la IA solo

[35] Martínez Garay, García Ortiz, *Paradojas* cit.163.
[36] J. Nieva-Fenoll, *Tecnología y derechos fundamentales en el proceso judicial. La tecnologia y la inteligência artificial al servicio del processo*, 2023, 224.
[37] Llorente Sánchez-Arjona*, Inteligencia* cit. 390-392.
[38] Paulesu, *Inteligencia* cit. 265-267.
[39] Muñoz Machado, *Prólogo* cit. 13.
[40] A.M. Paniagua Alario, *Análisis de la primera ley integral sobre inteligencia artificial en el mundo*, in *https://www.legaltoday.com/opinion/blogs/nuevas-tecnologias-blogs/blog-prodat/analisis-de-la-primera-ley-integral sobre-inteligencia-artificial-en-el-mundo-2024-02-05/accesso en 17/7/2024*. Asimismo, véase la reciente Resolución n. 615, del CNJ (Consejo Nacional de Justicia) de Brasil, de 11/03/2025.
[41] Martínez Garay, García Ortiz, *Paradojas* cit. 172.

ayudarían a reforzar, confirmar o refutar las heurísticas o decisiones del juez en el proceso penal[42], pero sin olvidar la capacidad humana crítica del juez, que deberá tener en cuenta, además del resultado de la estimación automatizada, otros factores que le parezcan relevantes para proporcionar una respuesta individualizada a cada caso concreto. Al menos en el sistema continental Europeo, no se puede utilizar un "juez-robot" en sustitución del juez, pues ello violaría directamente el principio constitucional de jurisdiccionalidad como actividad exclusiva de los jueces. De hecho, jueces robot podrían acelerar la justicia, pero no necesariamente harían que el sistema fuera justo. Además, la decisión automática choca con la noción constitucional de independencia judicial, ya que el juez pasa a depender del ingeniero o técnico de diseño[43].

A respecto, el RIA señala, en su considerando 61) que, habida cuenta de los efectos potencialmente importantes para la democracia, el Estado de derecho, las libertades individuales, el derecho a la tutela judicial efectiva y, a un juez imparcial, a fin de hacer frente al riesgo de posibles sesgos, errores y opacidades, «la utilización de herramientas de IA puede apoyar el poder de decisión de los jueces o la independencia judicial, pero no debe substituirlas: la toma de decisiones finales debe seguir siendo una actividad humana».

En definitiva, una justicia robótica desconectada de la función jurisdiccional nos situaría en escenarios distópicos difíciles de controlar[44]. Así, por su nivel de riesgo inaceptable, debe quedar prohibida la evaluación o predicción del riesgo criminal individual a través de algoritmos, sin la supervisión, control y motivación judicial.

Abstract.- Il presente contributo esamina in chiave critica l'impiego di algoritmi predittivi e dell'intelligenza artificiale (IA) nell'ambito delle decisioni di giustizia penale. Se, da un lato, tali strumenti possono accrescere l'efficienza e la coerenza applicativa, dall'altro la loro introduzione nel sistema giudiziario – in particolare nel campo della "predictive justice" – comporta rischi significativi per i diritti fondamentali. Le principali preoccupazioni riguardano l'opacità e la complessità dei sistemi algoritmici, la mancanza di obiettività e neutralità nei dati di "input" e la prevalenza di pregiudizi algoritmici che possono portare a risultati discriminatori. Il contributo, dunque, intende evidenziare le sfide che tali strumenti comportano per il principio al giusto processo, comprese le violazioni del diritto ad una difesa equa, della presunzione di innocenza e della possibilità di appellarsi alle decisioni. Viene, inoltre, evidenziata l'incompatibilità di strumenti di IA opachi con i principi costituzionali di indipendenza del giudice, trasparenza dell'azione giurisdizionale e obbligo di motivazione delle decisioni. In questo contesto, assumono rilievo le più recenti disposizioni normative: si pensi, da un lato, all'AI Act dell'Unione europea (2024) e, dall'altro, alla Risoluzione n. 615/2025 del Brasile, le quali – qualificando come rischio inaccettabile la valutazione o previsione del rischio criminale individuale – ne sanciscono espressamente il divieto, riconoscendo l'elevata pericolosità di tali pratiche per i diritti fondamentali. Lo studio sostiene, pertanto, la necessità di adottare sistemi di IA spiegabili, trasparenti e responsabili, concepiti per supportare, e non per sostituire, l'attività decisionale del giudice. In conclusione, pur potendo l'IA rappresentare un ausilio al lavoro giurisdizionale, le decisioni algoritmiche non possono in alcun caso surrogare il ragionamento giuridico umano e devono restare sottoposte a rigoroso controllo. Tutti gli strumenti

[42] García Sánchez, *El necesario balance* cit. 330-333.

[43] Castillejo Manzanares, *Cuáles son las razones* cit. 99-104; Gómez Colomer, *Derechos* cit. 285; Martínez Garay, García Ortiz, *Paradojas* cit.172; García Sánchez, *El necesario balance* cit. 101-104; Llorente Sánchez-Arjona, *Inteligencia* cit. 385; Paulesu, *Inteligencia* cit. 265-267.

[44] Id., *Cuáles son las razones* cit. 103.

algoritmici utilizzati in ambito penale devono, in ultima analisi, essere trasparenti, spiegabili e assoggettati a vigilanza giurisdizionale, al fine di mitigare i rischi sopra evidenziati.

This article critically examines the use of predictive algorithms and artificial intelligence (AI) in criminal justice decision-making. While AI can improve efficiency and consistency, its application in the judiciary – particularly in predictive justice – poses significant risks to fundamental rights. Key concerns include the opacity and complexity of algorithmic systems, the lack of objectivity and neutrality in input data, and the prevalence of algorithmic bias that can lead to discriminatory outcomes. This article highlights the challenges to due process, including violations of the right to a fair defense, the presumption of innocence, and the ability to appeal decisions. It also underscores the incompatibility of opaque AI tools with constitutional principles such as judicial independence, transparency, and the duty to provide reasoned decisions. It highlights recent regulations, such as the EU AI Act (2024) and Brazil's Resolution No. 615 (2025), expressly prohibits, due to its unacceptable high level of risk, among other practices, the evaluation or prediction of individual criminal risk, considering that it is a high-risk system for fundamental rights. The study advocates for explainable, transparent, and accountable AI systems to support – but not replace – judicial decision-making and, finally, conclude that while AI may support judicial work, algorithmic decisions cannot replace human judicial reasoning and must remain subject to strict oversight and all algorithmic tools used in criminal justice must be transparent, explainable, and subject to judicial oversight to mitigate these risks.