

Monográfico: «Digitalización y algoritmización de la justicia» (coord.: F. Miró)

# Algoritmos predictivos y perspectiva de género en el proceso penal

Ana Montesinos García

Universitat de València

Fecha de presentación: mayo 2022

Fecha de aceptación: julio 2023

Fecha de publicación: noviembre 2023

## Resumen

La introducción en el proceso judicial de herramientas algorítmicas de evaluación del riesgo para auxiliar al juez debe necesariamente venir precedida de un análisis exhaustivo con perspectiva de género acerca de su impacto sobre los derechos fundamentales de las partes que pueden verse seriamente afectados, principalmente el de igualdad.

Se reflexiona en este trabajo acerca de la posible incorporación de herramientas predictivas en diferentes etapas de un proceso penal para pasar a continuación a examinar el problema de los sesgos de género que pueden contener estas y sus consecuencias en el marco de un proceso.

## Palabras clave

algoritmos; inteligencia artificial; proceso penal; perspectiva de género; sesgos

## *Predictive algorithms and gender perspective in criminal proceedings*

## Abstract

*The introduction in the judicial process of algorithmic risk assessment tools to assist the judge must necessarily be preceded by a comprehensive analysis with a gender perspective about their impact on the fundamental rights of the parties that may be seriously affected, mainly that of equality.*

*This work reflects on the possible incorporation of predictive tools at different stages of a criminal process to then examine the problem of gender biases that they may contain and their consequences in the framework of a process.*

## Keywords

*algorithms; artificial intelligence; criminal process; gender perspective; biases*

## Introducción

El nuevo paradigma de justicia evoluciona hacia un modelo más tecnológico y digital, en el que la inteligencia artificial (en adelante, IA) comienza a cobrar cierto protagonismo. Paralelamente, la perspectiva de género se está incorporando a la toma de conciencia de quienes imparten justicia (Llorente Sánchez-Arjona, 2022, pág. 259) en aras de eliminar los obstáculos que dificultan la igualdad efectiva entre hombres y mujeres. De manera que nos encaminamos hacia una justicia digital y algorítmica que debe regularse y administrarse en clave de género.

La intersección entre la perspectiva de género, la inteligencia artificial y el derecho y, en particular, el derecho procesal, todavía está en ciernes.<sup>1</sup> De ahí que consideremos que la introducción en el proceso judicial de nuevas herramientas algorítmicas que ayuden al juez venga necesariamente precedida de un análisis exhaustivo acerca de su impacto sobre los derechos fundamentales de las partes, que pueden verse seriamente afectados si se parte de datos sesgados (especialmente, los derechos a la no discriminación e igualdad).

Dicho esto, entendemos que la perspectiva de género en la justicia no se agota en el modo en el que los jueces aplican el derecho, sino que debemos ir un paso más allá y extenderla también a las herramientas digitales y algorítmicas que se utilicen en el seno del proceso. Si se augura que en un futuro próximo los jueces van a auxiliarse de sistemas algorítmicos, deviene imprescindible garantizar que no

incorporen sesgos de género que acaben reproduciendo, o incluso acrecentando, determinados estereotipos que perjudican o discriminan a las mujeres.<sup>2</sup>

Es cierto que los sistemas algorítmicos pueden proporcionar resultados más precisos, con una exactitud y celeridad infinitamente superior a la de una persona, dado que analizan una cantidad ingente de datos, difícilmente abordable por un ser humano. Pero también lo es que pueden contener sesgos (entre otros, de género)<sup>3</sup> y, como consecuencia, la respuesta que facilite el algoritmo puede generar decisiones injustas y discriminatorias (Belloso Martín, 2022, pág. 49). Por todo ello, la perspectiva de género debe estar presente en el diseño, desarrollo y empleo de estas herramientas; especialmente si van a utilizarse en un proceso judicial.

En definitiva, la transformación digital de la justicia que estamos presenciando, y que va a intensificarse en los próximos años, debe llevarse a cabo desde una perspectiva de género, que permita alcanzar una justicia libre de estereotipos, así como garantizar un adecuado respeto del derecho constitucional a la tutela judicial efectiva en igualdad de condiciones.

## 1. Incorporación de algoritmos evaluadores del riesgo en el proceso penal

Sin perjuicio del empleo de algoritmos predictivos por la policía en tareas de prevención o en la fase investigación

1. En sentido similar, pero con referencia al derecho privado, se pronuncia Navas Navarro (2021, pág. 1).
2. En este sentido, la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, en su artículo 23 dispone que «las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio». En la misma línea, la Carta de Derechos Digitales adoptada el 14 de julio de 2021 prevé el derecho a la igualdad y a la no discriminación en el entorno digital. En concreto, fomenta que los procesos de transformación digital apliquen la perspectiva de género y adopten, en su caso, medidas específicas para garantizar la ausencia de sesgos de género en los datos y algoritmos usados (VIII). En esta materia, destáquese la labor de la Unesco y su «Recomendación sobre la ética de la inteligencia artificial», adoptada el 23 de noviembre de 2021, cuyo ámbito de actuación n.º 6 versa sobre el género. Véase, además, sus informes «Artificial intelligence and gender equality» (2020), donde propone que la igualdad de género se constituya en un principio autónomo dentro del elenco de sus principios éticos y «I'd Blush if I Could: closing gender divides in digital skills through education» (2019), en el que aborda el potencial de los sesgos algorítmicos de propagar y reforzar estereotipos de género.
3. Por eso, O'Neil califica los algoritmos como armas de destrucción matemática que, bajo la apariencia de neutralidad, pueden arrojar decisiones que perpetúen las desigualdades (O'Neil, 2017).

de un proceso,<sup>4</sup> vamos a centrarnos, por razones de espacio, en la utilización de estos sistemas en el estricto marco de un proceso penal. Concretamente analizamos aquellos que tratan de predecir la reincidencia o, mejor dicho, de realizar una estimación probabilística del riesgo que un sujeto puede sostener a la hora de cometer un nuevo delito.<sup>5</sup> Estas herramientas han ido evolucionando hasta alcanzar su completa automatización y las más sofisticadas se sirven de algoritmos para emitir el pronóstico.<sup>6</sup>

El uso de herramientas algorítmicas de evaluación del riesgo de reincidencia en el proceso penal está implementado en muchos países (Martínez Garay, 2022, pág. 160), fundamentalmente en Estados Unidos.<sup>7</sup> Sin embargo, por el momento, son escasas en nuestro ordenamiento, limitándose prácticamente a los sistemas VioGén y Riscanvi. Pero dado que ya existen (y pueden aportar algunas ventajas frente a los tradicionales sistemas de evaluación del riesgo), nos cuestionamos si conviene emplearlas en nuestro sistema procesal y, en su caso, en qué fase del proceso.

### 1.1. Fase de ejecución

En primer lugar, barajamos<sup>8</sup> su empleo en la fase de ejecución de un proceso penal, puesto que resulta menos proble-

mática al contar ya con una condena.<sup>9</sup> Podrían servir, como un elemento informador más,<sup>10</sup> para ayudar a establecer el riesgo de violencia que es determinante en la adopción de decisiones tales como: la concreción del régimen penitenciario, incluida la concesión de la libertad condicional, la clasificación de los reclusos, la concesión de permisos penitenciarios, la suspensión de la condena e incluso la revisión de la prisión permanente revisable. De hecho, en las prisiones catalanas se cuenta desde el año 2009 con la herramienta Riscanvi, desarrollada por el Grupo de Estudios Avanzados en Violencia de la Universidad de Barcelona y los Servicios Penitenciarios de la Generalitat de Cataluña, para evaluar la conducta de los internos (riesgo de violencia)<sup>11</sup> y, en función del resultado, asistir en la toma de decisiones acerca de su situación penitenciaria, como puede ser la obtención de permisos, el régimen de clasificación o la adopción de determinadas medidas de supervisión.

### 1.2. Medidas cautelares

En segundo lugar, consideramos la utilización de herramientas predictivas para calcular el riesgo de reincidencia que debe valorar el juez cuando decide determinadas medidas cautelares, como la prisión provisional, u otras medidas de seguridad cuyo presupuesto de adopción re-

4. Los sistemas predictivos de riesgos pueden convertirse en herramientas esenciales en la prevención y en la investigación delictiva. Repárese, como señala Barona Vilar (2022, pág. 161), que la coyuntura ideológica y política de control y la seguridad han abierto los brazos, con fascinación, a estos sistemas como los mejores posibles para alcanzar sus objetivos.
5. Como detalla Solar Cayón (2020, pág. 130), están basados en modelos estadísticos generados automáticamente mediante técnicas de *machine learning* que, a partir del análisis de grandes volúmenes de datos correspondientes a casos pretéritos, son capaces de identificar una serie de correlaciones entre determinados factores personales sociales y el riesgo de comisión de futuros delitos. Normalmente toman en cuenta indicadores relativos a circunstancias personales, sociodemográficas y al historial judicial del individuo, como factores predictores del riesgo de reincidencia, asignando a cada uno de ellos un peso relativo en función del análisis de los casos pretéritos.
6. Incluso recurriendo a la IA. Desmarais y Zottola (2020, págs. 813-816) mencionan los beneficios que la IA puede aportar a los instrumentos de evaluación del riesgo, a la vez que ponen de manifiesto los peligros que conlleva (aunque sin descartar que puedan ir resolviéndose en un futuro).
7. Siendo, sin lugar a duda, COMPAS el algoritmo que mayor repercusión ha tenido, a raíz del caso *State vs. Loomis* que lo ha dado a conocer mundialmente. Un análisis sobre este caso puede verse en Freeman (2016).
8. Barajamos, es decir, nos planteamos su empleo. Lo que no quiere decir que las defendamos rotundamente, ni mucho menos. Su utilización vendría condicionada por múltiples factores, garantías y requisitos, en los que, por cuestiones de espacio, no podemos detenernos.
9. Como señala Nieva Fenoll (2018, pág. 66), una cosa es establecer la posibilidad de reiteración delictiva de un reo condenado, es decir, de aquel con respecto al que ya se comprobó su responsabilidad, y otra muy distinta hacer idéntico juicio con alguien que aún no ha sido juzgado.
10. Tráigase a colación el artículo 6.1 del Reglamento Penitenciario titulado «Limitación del uso de la informática penitenciaria», que establece que ninguna decisión de la Administración penitenciaria que implique la apreciación de comportamiento humano de los reclusos podrá fundamentarse, exclusivamente, en un tratamiento automatizado de datos o informaciones que ofrezcan una definición del perfil o de la personalidad del interno. Adviértase al respecto que no se impide el empleo de estas herramientas, simplemente se prohíbe que sean el único elemento de valoración.
11. En concreto, evalúa cuatro comportamientos violentos por parte de los reclusos: el quebrantamiento de la condena, la violencia dentro de la institución, la reincidencia violenta y la violencia autodirigida.

side en la peligrosidad del sujeto, tal y como ocurre en el ámbito de la violencia de género.<sup>12</sup>

Si estas herramientas coadyuvan a objetivar el *periculum* - que hoy se decide con base en heurísticos y otros factores contextuales-, podrían ser bienvenidas (Simón Castellano, 2021, pág. 94) como «colaboradoras», cuyo resultado no pueda ser, en ningún caso, decisivo.<sup>13</sup> De manera que, en lugar de evaluar el riesgo de reincidencia de manera intuitiva a partir de la experiencia, como vienen haciendo los jueces, se objetive este riesgo. De hecho, los algoritmos predictivos hacen algo similar, salvando las distancias, a lo que hacen los jueces. Como refiere Simón Castellano, con las decisiones cautelares por parte del juez se produce una suerte de automatismo en relación con casos parecidos que han sido resueltos con anterioridad, dado que, ante la falta de actividad probatoria, difícilmente se pueden apoyar en algo más (2021, pág. 122). Precisamente esto es lo que hacen los sistemas algorítmicos, que analizan resoluciones pasadas (referidas a experiencias anteriores, de las que se conoce los resultados que produjeron) y extraen correlaciones, es decir, identifican qué situaciones han llevado a determinados resultados y toman esos criterios como base para las decisiones (Huergo Lora, 2020, pág. 35).

Ciertamente, el riesgo de reiteración delictiva, que actúa como presupuesto para la adopción de algunas medidas cautelares, podría objetivarse a través de sistemas algorítmicos capaces de arrojar un pronóstico (o estimación probabilística) respecto a la posibilidad de que un sujeto vaya a reincidir. El resultado ofrecido podría ser tenido en consideración por el juez, al valorar los datos obrantes en la causa que justifican la idoneidad de la medida. Para ello,

devendría fundamental una regulación que contemplara los presupuestos y garantías procesales que necesariamente deben preservarse ante los evidentes riesgos que plantea su uso en esta fase del proceso (Planchadell Gargallo, 2021, pág. 414).<sup>14</sup>

### 1.3. Sentencia

Hasta aquí, no sin cierta preocupación, podría llegar a aceptarse su empleo, si se respetan los principios y garantías del proceso. Lo verdaderamente inquietante de estos programas informáticos es que se propone trasladarlos al momento de dictar sentencia. De modo que la decisión final del proceso dependa de un juicio predictivo, en el sentido de que, a mayor riesgo de reincidencia, se imponga una pena de mayor intensidad, mientras que, a menor riesgo, la pena será inferior e incluso podrá imponerse una pena alternativa a la prisión. En suma, se propone que estos programas asesoren al juez en el dictado de la sentencia.<sup>15</sup>

La cuestión clave reside, por tanto, en reflexionar en torno a si la función jurisdiccional debe abrazar la predicción, pues son indudablemente graves las consecuencias para el proceso penal.<sup>16</sup> Baste al respecto con realizar tres advertencias de enorme calado.

**a)** En primer lugar, como bien sabemos, el proceso penal consiste en enjuiciar hechos pasados, reconstruir la verdad de lo sucedido y demostrarla más allá de toda duda razonable. Las herramientas predictivas de riesgos, por el contrario, lo que hacen es predecir comportamientos o hechos futuros que pueden acontecer o no. Al respecto, manifestamos con rotundidad que no puede juzgarse a una

12. Tal vez sea este el ámbito donde podrían enraizar los instrumentos valoradores del riesgo, que serían presentados con un halo de perfección como un auxilio judicial. En este sentido, Julià Pijoan (2020, pág. 230). En España contamos desde el año 2007 con el sistema VioGén (Sistema de Seguimiento Integral en los casos de Violencia de Género). Esta herramienta informática permite el seguimiento y protección de las víctimas de violencia de género y de sus hijos e hijas. Entre sus objetivos, destaca la realización de valoraciones policiales del riesgo de las víctimas denunciantes de sufrir una nueva agresión, y en función del resultado, poder protegerlas. Para dicha tarea se sigue un protocolo en el que se emplean dos instrumentos complementarios: la Valoración Policial del Riesgo (VPR4.0) para realizar una estimación inicial y la Valoración Policial de la Evolución del Riesgo (VPER4.0) para el seguimiento del caso. Si bien VioGén se diseñó inicialmente para evaluar el riesgo de reincidencia de una agresión, se ha creado un nuevo algoritmo para valorar la posibilidad de sufrir violencia letal (formulario VPR5.0).

13. Desde luego, como señala De Hoyos Sancho (2021, pág. 146), es complicado admitir juicios predictivos de peligrosidad de un sujeto cuando todavía tiene la presunción de inocencia a su favor. No obstante, puesto que las medidas cautelares son necesarias y se acepta que uno de sus fundamentos es evitar el riesgo de reiteración delictiva, no tiene sentido limitar los instrumentos cognitivos del juez.

14. Por cuestiones de espacio no podemos detenernos en las garantías procesales que pueden verse afectadas. Así que nos remitimos a nuestro trabajo en Montesinos García (2023).

15. Este planteamiento se ha venido a denominar: *evidence based sentencing*.

16. Julià Pijoan. Conferencia impartida en el «Debate sobre el Derecho Procesal y la Justicia Digital: presente y futuro», Instituto Iberoamericano de Derecho Procesal, 30 de abril de 2021. Reflexionan al respecto, entre otros, Chanenson y Hyatt, (2016); Garrett y Monahan, (2019); Monahan y Skeem (2016) y Christin (2017).

persona por lo que potencialmente podría hacer. Se corre, además, con ello el peligro de que el juez antes incluso de oír al sujeto en el juicio, se haga una idea predeterminada, lo que incide claramente en la presunción de inocencia, porque ya lo considera culpable. La imparcialidad del juez podría, por ende, verse seriamente afectada. Por todo ello, entendemos que no debe admitirse bajo ningún concepto una determinación de la pena basada en pronósticos.

**b)** En segundo lugar, la predicción, como advierte Julià Pijoan (2020, pág. 223), conlleva la responsabilidad por «el ser», es decir, por las características personales, familiares, sociales, etc., del acusado que, además, son ponderadas a partir de las actuaciones de otras personas. De modo que, dependiendo de los resultados de dicha ponderación, se le puede llegar a intensificar la condena. Al respecto, debemos señalar que no podemos responsabilizar a un sujeto por sus características personales, familiares, sociales, etc. Aunque los sistemas algorítmicos sean capaces de crear perfiles muy precisos de determinadas personas, no obviemos que son confeccionados con base en su pertenencia a un grupo, sin tener en cuenta sus especificidades, lo que puede conducir a decisiones erróneas «ya que, lo que es verdad para el grupo, no tiene por qué serlo necesariamente para una persona en concreto» (Zarsky, 2013, pág. 1560).<sup>17</sup>

**c)** En tercer y último lugar, debemos tener en cuenta que los sistemas predictivos no son infalibles. Como señala Martínez Garay (2018, pág.4), diversos estudios demuestran que, si bien la mayor parte de los modernos instrumentos de predicción ofrecen valores elevados de sensibilidad y especificidad, así como unos niveles relativamente bajos de falsos negativos, su principal debilidad radica en el bajo valor predictivo positivo que alcanzan, es decir, en que sobrestiman la peligrosidad y generan, por tanto, muchos falsos positivos. Además, a pesar de que se afirma que la información proporcionada por estos programas informáticos es objetiva y, por tanto, preferible a la valoración personal del juez, no debemos olvidar que puede estar sesgada. A los posibles sesgos algorítmicos nos referimos a continuación.

## 2. Herramientas predictivas con sesgos de género

Uno de los principales problemas que suscita el empleo de herramientas predictivas en la justicia penal reside en la posible existencia de sesgos en sus resultados, lo que puede conducir a una situación de discriminación y, con ello, a una afectación del principio de igualdad (Angwin, Larson, Mattu y Kirchner, 2016; Harcourt, 2015; Starr, 2014).<sup>18</sup> Veamos con detalle de qué estamos hablando.

A pesar de que las decisiones adoptadas por los sistemas algorítmicos suelen ser más precisas que las tomadas por seres humanos en casos similares, debemos tener presente que pueden arrojar resultados erróneos, al haberse introducido sesgos durante la configuración de su base de datos, en su entrenamiento e incluso tras su puesta en funcionamiento. Por ello, esa «mayor objetividad» de los algoritmos debe ponerse en duda (Soriano Arnanz, 2021, pág. 90; Soriano Arnanz y Simó Soler, 2021, pág. 193).

Ahora bien, como recuerda Miró Llinares, tampoco olvidemos que estas herramientas predictivas no vienen más que a hacer lo que se hace hoy en día de manera tradicional y probablemente con los mismos sesgos o más, añadiendo, en algunos casos, una metodología más sistemática o científica. Las decisiones sobre la adopción de una medida cautelar o una medida de protección, o sobre la concesión de un permiso de salida de prisión, entre otras, se tienen que seguir tomando, y «no cabe duda alguna que todo aquello que aporte información científica y supere los prejuicios subjetivos mejorando la toma de decisiones debería ser visto como algo positivo. Pero esto debe tomarse con cautela. En primer lugar, porque frente a la utopía de una IA futura imparcial [...], lo que sabemos hasta el momento nos dice que los algoritmos, que reflejan con precisión nuestro mundo, parecen reflejar también nuestros prejuicios» (Miró Llinares, 2018, pág. 126).

17. Soriano Arnanz (2021, pág. 93).

18. Cuestión distinta, y que excede del ámbito de este trabajo, es la discusión acerca de si las herramientas algorítmicas, especialmente las de valoración del riesgo, pueden trabajar con variables tales como el género, o por el contrario deberían ser retiradas de los algoritmos. Reflexiona al respecto Miró Llinares (2018, pág. 124).

## 2.1. Los sesgos de género en las herramientas algorítmicas empleadas en el marco de un proceso

Nos encontramos con sesgos algorítmicos de género cuando un sistema informático propone o toma decisiones erradas que replican estereotipos de género,<sup>19</sup> que pueden derivar en una discriminación algorítmica basada en el género (Corvalán, Macchiavelli, Cinalli, Danesi, Dávila, Tolosa, Sánchez Caparrós, Mafud, 2021).

Como señalan Xenidis y Senden (2020, págs. 4-8), los algoritmos refuerzan y propagan patrones de desigualdad que ya existen en el tejido social. Un refrán muy conocido entre los informáticos, «Garbage in, garbage out» (si entra basura, sale basura), transmite la idea de que cualquier resultado algorítmico discriminatorio procede de prejuicios inyectados en los algoritmos por seres humanos. En otras palabras, y tal y como lo replantea Mayson, «Bias in, bias out» (Mayson, 2018).

Son diversas las razones por las que un sistema algorítmico puede generar resultados discriminatorios. Aunque puede suceder que los sesgos sean aprendidos por la herramienta una vez puesta en marcha (a medida que se alimenta de nueva información), el origen de estos proviene, principalmente, de los datos de entrada con los que se nutre el algoritmo antes de ponerse en funcionamiento. Esto engloba aquellos supuestos en los que los datos reflejan:

- a)** prejuicios de quienes diseñan la base de datos (que de manera voluntaria o no introducen estereotipos que tienen sobre determinados grupos),
- b)** prejuicios ya existentes en la realidad social (cuando los datos son por sí mismos discriminatorios)<sup>20</sup> o,
- c)** cuando la muestra de datos no es representativa de la población sobre la cual el sistema adopta decisiones, sobre o infrarrepresentando a determinados colectivos (Belloso Martín, 2022, págs. 57 y ss.).

En todo caso, no siempre resulta sencillo identificar la procedencia de la discriminación algorítmica.

Trasladado al marco de un proceso judicial, imaginemos una herramienta predictiva que valora una agresión sexual y que se nutre de un enorme repertorio de sentencias, muchas de las cuales contienen estereotipos (sesgos) de quienes han dictado tales pronunciamientos. Por ejemplo, presentan prejuicios sobre el comportamiento que debe adoptar la víctima (esto es, debe haberse resistido, denunciado inmediatamente, olvidado su vida social, etc.). En tal supuesto, la información que proporcione la herramienta al juez estará sesgada, por lo que probablemente arrojará un resultado que poco o nada tenga que ver con la realidad de la mujer que ha sufrido este tipo de violencia en el caso concreto que está siendo enjuiciado.

Otro ejemplo claro de las consecuencias nocivas que pueden tener en la justicia penal lo representan los *softwares* de reconocimiento facial, sobre los que se alerta acerca de su potencial discriminatorio (si parten de bases de datos de imágenes sesgadas). Precisamente, se advierte que su precisión decrece cuando se utiliza con colectivos que no son hombres caucásicos. Por lo tanto, fallan más con mujeres y personas de piel oscura (European Union Agency for Fundamental Rights, 2020).<sup>21</sup>

Por último, los sesgos también pueden provenir del modo en que se entrena al algoritmo (a la hora de seleccionar y procesar los atributos que se le proporcionan -Corvalán y Macchiavelli, 2021-). Pensemos en un ejemplo ficticio, planteado por Martínez, Borges y Simó, con el que podríamos encontrarnos en el seno de un proceso penal: se crea un *software* para ayudar al juez a valorar la declaración de la víctima y ese algoritmo ha sido entrenado bajo el criterio de que la tardanza en denunciar es relevante para valorar su credibilidad, fruto de un estereotipo humano vigente en los tribunales. Ante un supuesto en el que la víctima denunciara los hechos inmediatamente después de que ocurrieran, su nivel de credibilidad sería alto. Sin embargo, si la víctima tardara meses en denunciar, el

19. Entendido este como una opinión o un prejuicio generalizado acerca de los atributos o características que hombres y mujeres poseen o deberían poseer o de las funciones sociales que ambos desempeñan o deberían desempeñar (Danesi, 2021, pág. 161).

20. Puesto que estos sistemas procesan la realidad social, y esta está construida sobre estructuras que discriminan a las personas pertenecientes a grupos con desventajas, los algoritmos reproducen y perpetúan estas estructuras de discriminación (Soriano Arnanz, 2022, pág. 139).

21. Véase European Union Agency for Fundamental Rights. *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, 2020, pág. 4.

algoritmo le reportaría al juez un grado de credibilidad mínimo. Estaríamos rotundamente ante un supuesto de discriminación algorítmica (2021, pág. 6).

## 2.2. Medidas para minimizar los sesgos algorítmicos de género

Lo descrito en las líneas anteriores evidencia la necesidad de adoptar medidas para evitar que el empleo de herramientas algorítmicas sesgadas suponga la vulneración de los derechos que principian el proceso, concretamente, el de igualdad (que es uno de los que puede verse afectado de manera más inmediata).

Una primera medida que se propone (aunque no sea propiamente ni tecnológica ni jurídica) implica actuar *ex ante*, y se refiere a la necesidad de que en el diseño de este tipo de herramientas intervengan no solo equipos interdisciplinarios (integrados por técnicos en IA y expertos en ciencias sociales y jurídicas),<sup>22</sup> sino también una participación equilibrada de hombres y mujeres (Navas Navarro, 2021, pág. 10).<sup>23</sup> Sin embargo, la realidad con la que nos encontramos es la siguiente: existe un radical desequilibrio entre hombres y mujeres en las denominadas carreras STEM (ciencias, tecnología, ingeniería y matemáticas), que propicia equipos de desarrolladores muy masculinizados (Fernández Benítez y García del Blanco, 2022).<sup>24</sup>

Convendría, además, que se tratara de equipos formados (en género e igualdad), capaces de introducir la perspec-

tiva de género en el diseño del propio sistema. Veamos de nuevo un ejemplo aplicado al proceso penal que refleja esta necesidad. Imaginemos que se crea una herramienta algorítmica para detectar denuncias falsas y una de las variables que evaluar fuera la retirada de la denuncia. Variable introducida debido al completo desconocimiento por parte del equipo que entrena la herramienta acerca del proceso por el que atraviesan las mujeres que interponen una denuncia por violencia de género y las múltiples causas o móviles que originan tal desistimiento. Esta herramienta, con toda seguridad, generaría resultados erróneos (Martínez García, 2021, págs. 222-223).

En segundo lugar, los algoritmos deben ser auditados y controlados (tanto de manera previa a su puesta en marcha como de forma periódica). El riesgo de que estas herramientas puedan discriminar resulta inaceptable cuando nos encontramos en el ámbito de la administración de justicia. Tengamos en cuenta que estamos ante sistemas de alto riesgo, dado que se destinan a auxiliar a los jueces, por lo que pueden tener efectos importantes, entre otros, para el derecho a la tutela judicial efectiva y a un juez imparcial.<sup>25</sup> Por ello, entendemos, deben ser sometidos a supervisión por parte de organismos o agencias independientes, como podría ser en nuestro caso la Agencia Española de Supervisión de Inteligencia Artificial.<sup>26</sup>

Para lograr un efectivo control, deviene necesario que los programas que se utilicen sean, en la medida de lo posible, transparentes y explicables.<sup>27</sup> La cuestión es que, de

22. En un sentido similar se pronuncia la «Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno» de 2018, que fomenta análisis científicos multidisciplinares para combatir la discriminación.

23. Como señalan Ortiz de Zárate y Guevara Gómez, la diversidad en los equipos de diseño de la IA puede ofrecer nuevas perspectivas y traer a colación experiencias que permitan reconocer los sesgos y trabajar para corregirlos (2021, pág. 24).

24. La falta de diversidad de género en el campo de la IA dificulta la capacidad para detectar los prejuicios y estereotipos de género durante el diseño de las bases de datos y el aprendizaje automático de la IA (AHN y Costigan, 2019, pág. 3). Según recoge el informe «Assessing Gender Gaps in Artificial Intelligence», *The Global Gender Gap Report 2018*, elaborado por el World Economic Forum, solo el 22 % de profesionales de la IA a nivel mundial son mujeres, frente al 78 % que son hombres (datos de 2018).

25. Así lo considera la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial, de 21 de abril de 2021.

26. Adscrita a la Secretaría de Estado de Digitalización e IA, dentro del Ministerio de Asuntos Económicos y Transformación Digital.

27. Advirtase en todo caso que, si bien los algoritmos aumentan los problemas de discriminación en algunos casos, también ofrecen la oportunidad de reducir la arbitrariedad mediante una mayor explicabilidad de los procedimientos de toma de decisiones. Mientras que las decisiones humanas podrían asimismo calificarse de «caja negra» por su naturaleza opaca y no reproducible, los algoritmos de aprendizaje automático ofrecen la posibilidad de una toma de decisiones más responsable, siempre que se cumplan ciertos requisitos de transparencia. Mientras que las decisiones humanas no pueden reproducirse cambiando un factor para comprobar de dónde procede la discriminación, sí pueden hacerlo las decisiones algorítmicas. Por lo tanto, ciertos principios como la transparencia, la explicabilidad y la rendición de cuentas son fundamentales para desarrollar aplicaciones de IA si el objetivo es convertir los riesgos existentes de discriminación en una oportunidad para aumentar la igualdad. Xenidis y Senden (2020, pág. 30).

ser trazable o explicable, esto mitigaría el problema de los sesgos, en tanto en cuanto «se podría permitir, *a priori*, su corrección y perfeccionamiento» (Macchiavelli, 2022, pág. 64). A lo que debemos añadir que la falta de transparencia provoca una indefensión que no puede tolerarse en un proceso penal porque impide que se puedan conocer las razones que fundamentan una decisión y, por ende, la posibilidad de defenderse y rebatir las decisiones judiciales adoptadas con el apoyo de tales herramientas.

## Conclusiones

Los sistemas algorítmicos de evaluación del riesgo podrían auxiliar al juez en determinadas (y contadas) decisiones que deben adoptarse a lo largo del proceso y aportar beneficios en términos no solo de eficacia y eficiencia, sino también de mejora en la calidad de la justicia. Pero no olvidemos que, a su vez, son numerosos los problemas que entrañan y los riesgos que amenazan tanto su eficacia predictiva o rigor científico como su objetividad.

Si queremos incorporarlos en nuestro ordenamiento, deben previamente adoptarse toda una serie de medidas legales, tecnológicas y éticas que garanticen el respeto de los derechos fundamentales, entre los que se incluye la igualdad. Deben, por tanto, regularse y establecerse unos límites. Precisamente uno de esos límites consiste en que del empleo de estas herramientas en el marco de un proceso penal no puedan derivarse resultados discriminatorios. En particular, ha de procurarse que no reproduzcan sesgos discriminatorios de género.

Las reflexiones vertidas en este trabajo no pueden llevarnos sino a concluir que una futura regulación de nuevos sistemas algorítmicos predictivos debe necesariamente contar con una perspectiva de género.

## Reconocimientos

Estudio redactado en el marco del proyecto *Claves para una justicia digital y algorítmica con perspectiva de género*, PID2021-1231700B-I00 financiado por MCIN/AEI/10.13039/501100011033.

## Referencias bibliográficas

- AHN, S.; COSTIGAN, A. (2019, diciembre). «Trend Brief: How AI Reinforces Gender Stereotypes». *Catalyst*, n.º 5 [en línea]. Disponible en: <https://www.catalyst.org/research/ai-gender-stereotypes/>
- ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L. (2016, mayo). «Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks». *ProPublica* [en línea]. Disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- BARONA VILAR, S. (2022). «Sistemas algorítmicos en el proceso penal. Su incidencia en The Judge Craft». En: PÉREZ JUAN, J.A. y SANJUAN ANDRES, F.J. (dirs.). *La transformación algorítmica del sistema de Justicia penal*. Cizur Menor: Aranzadi, págs. 159-174.
- BELLOSO MARTIN, N. (2022). «La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos?». En: LLANO ALONSO, F. (dir.). *Inteligencia artificial y filosofía del derecho*, Murcia: Laborum, págs. 45-78.
- CASTRO LIÑARES, D. (2019) *Los instrumentos de valoración y gestión de riesgos en el modelo de penalidad español*. Editorial Reus. DOI: <https://doi.org/10.30462/9788429021882>
- CHANENSON, S.; HYATT, J. (2016). «The Use of Risk Assessment at Sentencing: Implications for Research and Policy». *Working Paper Series*, Villanova University Charles Widger School of Law [en línea]. Disponible en: <https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps>
- CHRISTIN, A. (2017, diciembre). «The Mistrials of Algorithmic Sentencing». *Logic*, n.º 3, Justice [en línea]. Disponible en: <https://logicmag.io/justice/the-mistrials-of-algorithmic-sentencing/>
- CORVALAN, J.; MACCHIAVELLI, N.; CINALLI, C.; DANESI, C.; DÁVILA, L.; TOLOSA, P.; SÁNCHEZ CAPARRÓS, M.; MAFUD, M. (2021). «Sesgos algorítmicos de género». IALAB/Universidad de Buenos Aires [en línea]. Disponible en: [https://www.researchgate.net/publication/359115499\\_Sesgos\\_algoritmicos\\_de\\_genero](https://www.researchgate.net/publication/359115499_Sesgos_algoritmicos_de_genero)
- DANESI, C. (2021). «Sesgos algorítmicos de género con identidad iberoamericana: las técnicas de reconocimiento facial en la mira». *Revista Derecho de Familia*. n.º 100, págs. 159-168 [en línea]. Disponible en: [https://www.ceciliadanesi.com/\\_files/ugd/e2918c\\_c96111f9656461986249acd26b49c2e.pdf](https://www.ceciliadanesi.com/_files/ugd/e2918c_c96111f9656461986249acd26b49c2e.pdf)
- DE HOYOS SANCHOS, M. (2021). «Premisas y finalidades del Libro Blanco sobre Inteligencia Artificial de la Comisión Europea: perspectiva procesal del nuevo marco regulador». En: BARONA VILAR, S. (ed.). *Justicia algorítmica y neuroderecho, Una mirada multidisciplinar*. Valencia: Tirant Lo Blanch, págs. 129-155.
- DESMARAIS, S.L.; ZOTTOLA, S.A. (2020). «Violence risk assessment: current status and contemporary issues». *Marquette Law review*, vol. 103, n.º 3, págs. 793- 817.
- EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS (2020). *Facial recognition technology: fundamental rights considerations in the context of law enforcement*. European Union Agency for Fundamental Rights [en línea]. Disponible en: <http://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-law>
- FERNÁNDEZ, A. (2019). «Inteligencia artificial en los servicios financieros». *Boletín económico - Banco de España*, n.º 2.
- FERNÁNDEZ BENÉITEZ, A.; GARCÍA DEL BLANC, I. (2022, junio). «Inteligencia artificial, igualdad real». *El país* [en línea]. Disponible en: <https://elpais.com/tecnologia/2022-06-01/inteligencia-artificial-igualdad-real.html>
- FREEMAN. K. (2016). «Algorithmic injustice: How the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis». *North Carolina Journal of Law & Technology*, vol. 18, n.º 5 [en línea]. Disponible en: <https://scholarship.law.unc.edu/cgi/viewcontent.cgi?article=1332&context=ncjolt>

- GARRETT, B.; MONAHAN, J. (2019). «Assessing Risk: The Use of Risk Assessment in Sentencing». *Judicature, Bolch Judicial Institute Duke Law School*, vol. 103, n.º 2 [en línea]. Disponible en: <https://judicature.duke.edu/articles/assessing-risk-the-use-of-risk-assessment-in-sentencing/>
- HARCOURT, B.E. (2015). «Risk as a Proxy for Race. The Dangers of risk assessment». *Federal Sentencing reporter*, vol. 27, n.º 4, págs. 237- 243. DOI: <https://doi.org/10.1525/fsr.2015.27.4.237>
- HUERGO LORA, A. (2020). «Una aproximación a los algoritmos desde el Derecho administrativo». En: *La regulación de los algoritmos*. Cizur Menor: Aranzadi, págs. 23-87.
- JULIÀN PIJOAN, M. (2020). *Proceso penal y (Neuro) ciencia: una interacción desorientada. Una reflexión acerca de la neuropredicción*. Madrid: Marcial Pons.
- LLORENTE SÁNCHEZ-ARJONA, M. (2022). La inteligencia artificial como nueva estrategia de prevención en los delitos de violencia sexual. En: COLOMER HERNÁNDEZ, I. (dir.). *Uso de la información y de los datos personales en los procesos: los cambios en la era digital*. Cizur Menor: Thomson Reuters Aranzadi, págs. 259-282.
- MACCHIAVELLI, N. (2022). «La violencia de género y el uso de algoritmos como herramienta efectiva para la protección de los derechos fundamentales». *Anuario de filosofía del derecho, XXXVIII*, págs. 59-74. DOI: <https://doi.org/10.53054/afd.vi38.9738>
- MAYSON, S. G. (2018). «Bias In, Bias Out». *Yale Law Journal*, vol. 128, n.º 8, págs. 2218- 2300 [en línea]. Disponible en: <https://www.yalelawjournal.org/article/bias-in-bias-out>
- MARTINEZ GARAY, L. (2022). «Paradojas de los algoritmos predictivos utilizados en el sistema de Justicia Penal». *El Cronista*, n.º 100, págs. 160-173.
- MARTÍNEZ GARAY, L. (2018, julio). «Peligrosidad, algoritmos y Due Process: el Caso State V Loomis». *Revista de Derecho Penal y Criminología*, n.º 20, época. 3, págs 485-502. DOI: <https://doi.org/10.5944/rdpc.20.2018.26484>
- MARTÍNEZ GARCÍA, E.; BORGES BLÁZQUEZ, R.; SIMÓ SOLER, E. (2021, enero). «Inteligencia artificial y perspectiva de género en la justicia penal». *Diario La Ley*, n.º 47, Sección Ciberderecho [en línea]. Disponible en: <https://diariolaley.laleynext.es/diI/2021/01/20/inteligencia-artificial-y-perspectiva-de-genero-en-la-justicia-penal>
- MARTINEZ GARCÍA, E. (2021). «Justicia e Inteligencia Artificial sin género». En: BARONA VILAR, S. (ed.). *Justicia algorítmica y neuroderecho*. Valencia: Tirant Lo Blanch, págs. 209-227.
- MIRÓ LLINARES, F. (2018). «Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots». *Revista de Derecho Penal y Criminología*, n.º 20, época. 3, págs. 87-130. DOI: <https://doi.org/10.5944/rdpc.20.2018.26446>
- MONAHAN, J.; SKEEM, J.L. (2016). «Risk assessment in criminal sentencing». *Annual Review of Clinical Psychology*, vol. 12, págs. 489- 513. DOI: <https://doi.org/10.1146/annurev-clinpsy-021815-092945>
- MONTESINOS GARCÍA, A. (2023). «Afectación de los derechos y garantías procesales por el empleo de algoritmos predictivos». En: ASECIO MELLADO y FUENTES SORIANO (dirs.). *El proceso como garantía*. Barcelona: Atelier, págs. 703-713.
- NAVAS NAVARRO, S. (2021). «La perspectiva de género en la inteligencia artificial». *La Ley digital. Ciberderecho*, vol.1, n.º 1, págs. 1-7.
- O'NEIL, C. (2017). *Weapons of math destruction*. Nueva York: Penguin.
- ORTIZ DE ZÁRATE ALCARAZO, L.; GUEVARA GÓMEZ, A. (2021). «Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España». *Fundación alternativas*, n.º 101 [en línea]. Disponible en: <https://fundacionalternativas.org/publicaciones/inteligencia-artificial-e-igualdad-de-genero-un-analisis-comparado-entre-la-ue-suecia-y-espana/>

- JULIÁN PIJOAN, M. (2020) *Proceso penal y (Neuro) ciencia: una interacción desorientada. Una reflexión acerca de la neuropredicción*. Madrid: Marcial Pons.
- NIEVA FENOLL, J. (2018). *Inteligencia artificial y proceso judicial*, Madrid: Marcial Pons.
- PLANCHADELL-GARGALLO, A. (2021). «Inteligencia Artificial y medidas cautelares». En: BARONA VILAR, S. (ed.). *Justicia algorítmica y neuroderecho, Una mirada multidisciplinar*. Valencia: Tirant Lo Blanch, págs. 389- 419.
- SENDEN XENIDIS, R.; SENDEN, L. (2020). «EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination». En: BERNITZ, U. et al (eds.). *General Principles of EU law and the EU Digital Order*. Países Bajos: Kluwer Law International.
- SIMÓN CASTELLANO, P. (2021). *Justicia Cautelar e inteligencia artificial*. Barcelona: Bosch. DOI: <https://doi.org/10.2307/j.ctvtqcxbh>
- SOLAR CAYÓN, J. I. (2020). «Inteligencia artificial en la justicia penal: los sistemas algorítmicos». En: *Dimensiones éticas y jurídicas de la Inteligencia artificial en el marco del Estado de Derecho*. Cuadernos de la Cátedra de Democracia y Derechos Humanos, 16. Universidad de Alcalá. Defensor del Pueblo, págs. 125-172.
- SORIANO ARNANZ, A. (2021). «Decisiones automatizadas: problemas y soluciones jurídicas. más allá de la protección de datos». *Revista de Derecho público: teoría y método*, vol. 3, págs. 85-127. DOI: [https://doi.org/10.37417/RPD/vol\\_3\\_2021\\_535](https://doi.org/10.37417/RPD/vol_3_2021_535)
- SORIANO ARNANZ, A.; SIMÓ SOLER, E. (2021). «Machine learning y Derecho: aprendiendo la (des) igualdad». En: BARONA VILAR, S. (ed.). *Justicia algorítmica y neuroderecho*. Valencia: Tirant Lo Blanch, págs.183-207.
- SORIANO ARNANZ, A. (2022). «Discriminación algorítmica: garantías y protección jurídica». En: COTINO HUESO, L. (dir.). *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*. Cizur Menor: Aranzadi, págs.139-169.
- STARR, S.B. (2014). «Evidence-based sentencing and the scientific rationalization of discrimination». *Stanford Law review*, n.º 66, pág. 803-872 [en línea]. Disponible en: [http://www.stanfordlawreview.org/wp-content/uploads/sites/3/2014/04/66\\_Stan\\_L\\_Rev\\_803-Starr.pdf](http://www.stanfordlawreview.org/wp-content/uploads/sites/3/2014/04/66_Stan_L_Rev_803-Starr.pdf)
- WARREN, R. (2010). «Evidence-Based Sentencing: Are We Up to the Task?». *Federal Sentencing Reporter*, vol. 23, n.º 2, págs. 153 - 158. DOI: <https://doi.org/10.1525/fsr.2010.23.2.153>
- WORLD ECONOMIC FORUM. «Assessing Gender Gaps in Artificial Intelligence». *The Global Gender Gap Report 2018* [en línea]. Disponible en: <https://www.weforum.org/reports/reader-global-gender-gap-report-2018/in-full/assessing-gender-gaps-in-artificial-intelligence/>
- ZARSKY, T. (2013). «Transparent predictions». *University of Illinois Law Review*, n.º 4, págs. 1503-1570 [en línea]. Disponible en: <https://www.illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf>.

### Cita recomendada

MONTESINOS GARCÍA, Ana (2023). «Algoritmos predictivos y perspectiva de género en el proceso penal». En: Miró, F. (coord.). «Digitalización y algoritmización de la justicia». *IDP. Revista de Internet, Derecho y Política*, núm. 39. UOC [Fecha de consulta: dd/mm/aa]

<http://dx.doi.org/10.7238/idp.v0i39.416439>



Los textos publicados en esta revista están –si no se indica lo contrario– bajo una licencia Reconocimiento-Sin obras derivadas 3.0 España de Creative Commons. Puede copiarlos, distribuirlos y comunicarlos públicamente siempre que cite su autor y la revista y la institución que los publica (*IDP. Revista de Internet, Derecho y Política*; UOC); no haga con ellos obras derivadas. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-nd/3.0/es/deed.es>.

### Sobre la autoría

Ana Montesinos García  
 Universitat de València  
[ana.montesinos@uv.es](mailto:ana.montesinos@uv.es)

Profesora titular de Derecho Procesal de la Universitat de València y codirectora del máster en Derecho y Violencia de Género de dicha universidad. Investigadora principal del proyecto de investigación de generación del conocimiento *Claves para una justicia digital y algorítmica con perspectiva de género*, PID2021-1231700B-I00 financiado por MCIN/ AEI/10.13039/501100011033. Ha formado parte del equipo investigador de siete proyectos de investigación financiados por los Ministerios de Educación y Ciencia, Economía y Competitividad, todos ellos calificados como muy satisfactorios.

Desde los inicios de su carrera investigadora pertenece al grupo de investigación de excelencia Me-dArb de la Universitat de València. Sus líneas de investigación más destacadas son: el arbitraje, las acciones colectivas, las reclamaciones de escasa cuantía y la digitalización de la justicia. Los estudios y trabajos que ha llevado a cabo se han visto reflejados en diversas publicaciones tanto de ámbito nacional como internacional, concretamente en cuatro monografías como autora única y numerosos capítulos de libro y artículos de revista.

