

# INTELIGENCIA ARTIFICIAL Y DERECHO

## REFLEXIONES JURÍDICAS PARA EL DEBATE SOBRE SU DESARROLLO Y APLICACIÓN

Martín A. Rocha Espíndola

Daniel Sansó-Rubert Pascual

Nuria Rodríguez Dos Santos



No está permitida la reproducción total o parcial de este libro, ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea éste electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito del editor. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (art. 270 y siguientes del Código Penal).

Diríjase a Cedro (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra. Puede contactar con Cedro a través de la web [www.conlicencia.com](http://www.conlicencia.com) o por teléfono en el 917021970/932720407

Este libro ha sido sometido a evaluación por parte de nuestro Consejo Editorial  
Para mayor información, véase [www.dykinson.com/quienes\\_somos](http://www.dykinson.com/quienes_somos)

©Copyright by los autores  
Madrid, 2023

Editorial DYKINSON, S.L.  
Meléndez Valdés, 61 - 28015 Madrid  
Teléfono (+34) 915442846 - (+34) 915442869  
e-mail: [info@dykinson.com](mailto:info@dykinson.com)  
<http://www.dykinson.es>  
<http://www.dykinson.com>

ISBN: 978-84-1122-949-4  
Depósito Legal: M-19169-2023  
DOI: 10.14679/2053

ISBN electrónico: 978-84-1170-411-3

New Garamond Diseño y Maquetación, S.L.

# EMPLEO DE SISTEMAS ALGORÍTMICOS DE EVALUACIÓN DE RIESGOS EN MATERIA PENAL. ESTÁNDARES MÍNIMOS PARA UN USO ACORDE A LAS EXIGENCIAS DEL DEBIDO PROCESO\*

Ximena Marcazzolo Awad

*Profesora Investigadora del Centro de Derecho Regulatorio y Empresa*

*Universidad del Desarrollo*

Nathalie Walker Silva

*Profesora de Derecho Civil*

*Universidad Andrés Bello*

**DOI: 10.14679/2057**

## 1. INTRODUCCIÓN

El gran interés que suscita la inteligencia artificial traspasa, desde hace algún tiempo, el umbral de la comunidad científica. El concepto se utiliza con propiedad en la prensa mundial y, casi a diario, es posible obtener información acerca de los múltiples y sorprendentes avances que experimenta en diversos ámbitos de la vida moderna. La inteligencia artificial está presente en nuestros hogares, se utiliza en la contratación de personas, en la medicina, en el ámbito bancario, en el transporte, en el denominado “criptoarte” y un largo etcétera.

Una de las razones por las cuales la inteligencia artificial atrae poderosamente nuestra atención es la impredecibilidad de sus usos. Los seres humanos ignoramos los límites a su potencial expansión y, es eso mismo lo que mantiene con vida la eterna promesa de que algún día existirán máquinas con consciencia de sí mismas o, incluso, con una inteligencia superior a la nuestra.

Predicciones aparte, las aplicaciones de la inteligencia artificial son cada vez más variadas, extendiéndose a ámbitos que hasta solo hace una década habríamos considerado imposibles de abarcar por la tecnología, como es el caso de la resolución de conflictos de índole jurídico. Este cambio ha sido paradigmático en el campo de los mecanismos alternativos de resolución de conflictos –también denominados con frecuencia por su sigla “MASC” o “ADR” (*Alternative Dispute Resolution*)–, los cuales han experimentado un alto

---

\* Investigación vinculada al proyecto Jorge Millas DI-03-JM/22, Universidad Andrés Bello, del cual Nathalie Walker es investigadora responsable.

grado de desarrollo debido, en parte, a la creciente automatización de ese tipo de procedimientos. En particular, el auge de los sistemas de resolución de disputas en línea (RDL) se explica, en gran medida, por la implementación de tecnologías de inteligencia artificial.

En este orden de cosas, es posible visualizar cómo los avances logrados –primero en el terreno de lo privado y, luego, en lo público– se han extendido de manera progresiva al proceso judicial, gracias a las ventajas de esos sistemas en términos de costos y de rapidez en la entrega de soluciones, lo que se ha traducido en una mayor posibilidad de acceso de los ciudadanos a la administración de justicia. Gracias a este panorama, muchos ven con esperanza la evolución de los RDL, en cuanto podría hacer realidad la idea de los “tribunales multipuertas” (*Multidoor Court-Houses*); vale decir, un sistema judicial en que las demandas puedan ser dirigidas a diferentes mecanismos de solución de conflictos, de modo que cada caso sea conducido y resuelto a través de la vía más idónea, teniendo en cuenta sus características particulares<sup>1</sup>.

Los usos judiciales de la inteligencia artificial tuvieron un comienzo tímido, que se produjo con cierto retraso, manifestándose a través del uso de variables estadísticas. Más tarde, la tecnología se incorporó en los procesadores de texto, para optimizar sus resultados y, en forma más reciente, en los buscadores de jurisprudencia y en la práctica de algunas pruebas científicas<sup>2</sup>. Hoy, el avance de la ciencia nos permite contar con herramientas de inteligencia artificial más sofisticadas, como ocurre en materia de evaluación y predicción de riesgos, las cuales serán objeto de especial estudio en este trabajo.

La primera parte de este estudio estará dedicada, en primer lugar, a la noción de sistemas algorítmicos de evaluación de riesgos y la función que éstos pueden cumplir en un proceso judicial de carácter penal. Luego, se efectuará una síntesis de los principales problemas que ha planteado el uso de sistemas algorítmicos de predicción de riesgos en aquellos ordenamientos en donde se han empleado.

La segunda parte de este estudio se destinará a la búsqueda de criterios para hacer compatible el uso de sistemas algorítmicos de predicción de riesgos con las exigencias del debido proceso en materia penal. Así, en base a la experiencia existente, se hará el intento de determinar qué condiciones deberían darse para que esos sistemas funcionen sin faltas al debido proceso y sin atropello de las garantías de los justiciables.

## 2. LOS SISTEMAS ALGORÍTMICOS DE EVALUACIÓN DE RIESGOS EN LA TOMA DE DECISIONES JUDICIALES

En la última década, el grado de avance en la automatización de tareas de índole jurídica ha sido sorprendente. En efecto, ha habido un notable desarrollo de programas y aplicaciones que optimizan la búsqueda y procesamiento de la información de rele-

---

<sup>1</sup> WALKER SILVA, N., “Resolución alternativa de conflictos en línea. Perspectivas de futuro”, *Revista Iberoamericana de Derecho Privado*, núm. 13, 2021. Accesible en: <https://latam.ijeditores.com/pop.php?option=publicacion&idpublicacion=76&idediccion=4883>

<sup>2</sup> NIEVA FENOLL, J., *Inteligencia artificial y proceso judicial*, Madrid: Marcial Pons, 2018, pp. 23 y 24.

vancia jurídica. En esa línea, algunas de esas aplicaciones son también utilizadas en el ámbito del proceso judicial, como es el caso de los sistemas de análisis predictivo, que permiten seleccionar la mejor estrategia de defensa para los clientes a través de la identificación de patrones de comportamiento de jueces y tribunales. Lo mismo ocurre con las herramientas de codificación predictiva, usadas por las partes en varios países del *common law* en la fase *pre-trial* para escoger el material que será considerado relevante en el proceso<sup>3</sup>.

Pero los avances no se circunscriben a lo antes señalado. También se emplea inteligencia artificial en el ejercicio mismo de la función jurisdiccional, esto es, más allá del uso de la tecnología como mero auxilio en la búsqueda de legislación y jurisprudencia. Esto ha ocurrido durante los últimos años en varios países que han introducido en sus sistemas de administración de justicia ciertos mecanismos de resolución de disputas en línea, basándose en casos exitosos en el ámbito del comercio electrónico, en los que la experiencia privada ha sido pionera en la solución automatizada de conflictos. Un ejemplo pionero lo encontramos en Canadá, en donde el *British Columbia Civil Resolution Tribunal* ha resuelto miles de disputas en línea desde el año 2016. O el caso del Reino Unido, en que, en el marco de una reforma a la justicia civil, incorpora la eventual creación de un tribunal en línea –la denominada *On line Solutions Court*– para reclamaciones de una cuantía de hasta 25.000 libras<sup>4</sup>. Iniciativas similares también se han proyectado en países como China, Francia y Estonia. En Latinoamérica, en el ámbito del análisis predictivo para la elaboración de propuestas de resolución judicial, encontramos el caso del software *Prometea*<sup>5</sup>, en Argentina; y, para auxiliar a los tribunales de justicia en la toma de decisiones, se ha implementado el programa *Pretoria*<sup>6</sup>, en Colombia<sup>7</sup>.

---

<sup>3</sup> SOLAR CAYÓN, J. I., “Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos”, en VVAA, Solar (Ed.), “Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho”, *Cuadernos Democracia y Derechos Humanos* núm. 16, Madrid: Universidad de Alcalá/Defensor del Pueblo, 2020, pp. 125 y 126.

<sup>4</sup> SOLAR CAYÓN, J. I., “Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos”, cit., p. 126.

<sup>5</sup> *Prometea* es un sistema de inteligencia artificial creado en Buenos Aires, en el ámbito del Ministerio Público Fiscal. Este sistema informático se utiliza para resolver casos en diversas materias, pero de fácil solución, tales como infracciones menores, accidentes de tráfico o políticas sociales. *Prometea* cuenta con habilidades que van desde la automatización hasta la predicción, aunque fue originalmente construida con la finalidad de optimizar la justicia. MUÑOZ RODRÍGUEZ, A. B., “El impacto de la inteligencia artificial en el proceso penal”, *Anuario de la Facultad de Derecho*, Universidad de Extremadura 36, 2020, p. 704.

<sup>6</sup> El sistema *Pretoria* se basa en la experiencia argentina de *Prometea*. Fue puesto en marcha en julio de 2020 por la Corte Constitucional de Colombia y es “un sistema predictivo basado en aprendizaje automático supervisado que selecciona, lee y elabora resúmenes sobre miles de sentencias en apenas unos segundos, de manera que puede realizar automáticamente la búsqueda de información de interés para la selección de los casos, la categorización de los mismos atendiendo a criterios de relevancia establecidos por la Corte y la elaboración de estadísticas que permitan visualizar de manera íntegra las tutelas presentadas en ese ámbito”. SOLAR CAYÓN, J. I., “Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos”, cit., p. 127.

<sup>7</sup> *Ibid.*, p. 127.

Unido a lo anterior, gracias al empleo del *big data* y de la inteligencia artificial, han proliferado diversas iniciativas que contemplan el uso de algoritmos de apoyo a las decisiones judiciales, dentro de los cuales se encuentran los sistemas de evaluación de riesgos. En atención a que éstos constituyen la materia objeto de este trabajo, en la primera parte de este capítulo nos dedicaremos al estudio de su naturaleza y utilidad.

### 3. ¿QUÉ SON Y CUÁL ES LA FUNCIÓN DE LOS SISTEMAS ALGORÍTMICOS DE EVALUACIÓN DE RIESGOS?

Los sistemas algorítmicos de evaluación de riesgos son herramientas especialmente diseñadas con la finalidad de predecir la probabilidad de reincidencia criminal de una persona que ha delinquido<sup>8</sup>. Se sustentan en el fenómeno de la automatización de las herramientas de valoración del riesgo, y son potencialmente aplicables a los procesos de toma de decisiones vinculadas con individuos que se ven envueltos en el proceso judicial, tanto en el desarrollo del mismo proceso –como, por ejemplo, en la aplicación de medidas cautelares– como tras la condena. Se trata de herramientas que auxilian en la toma de decisiones importantes, como la concreción del régimen penitenciario, la concesión de libertad provisional y la libertad condicional, entre varias otras<sup>9</sup>.

A partir de los datos disponibles, se busca el establecimiento de patrones y de relaciones entre diversas variables, a fin de establecer una predicción respecto del comportamiento futuro de los individuos. Así, a través de formulaciones especialmente adaptadas a las complejidades existentes, se persigue la disminución de las barreras relacionadas con el almacenamiento y procesamiento de datos, para diseñar una estrategia procesal eficiente<sup>10</sup>. De este modo, frente a un caso determinado los algoritmos cuantifican la probabilidad de que un individuo vuelva a cometer un delito y establecen, en función de esa probabilidad, el nivel de riesgo del sujeto<sup>11</sup> (que se cataloga de riesgo alto, medio o bajo).

Estas herramientas actuariales automatizadas se apoyan en datos personales relacionados con factores individuales, sociales y ambientales, y en una configuración propia de valoración del riesgo de violencia y reincidencia basada en metodologías cualitativas y cuantitativas diversas<sup>12</sup>. Al respecto, la literatura científica ha concluido que las variables de predicción más fuertes de la reincidencia radican en las necesidades criminógenas, la historia criminal y de comportamientos antisociales, los logros sociales, la edad, el género, la raza y los factores de índole familiar. Otros predictores de menor fuerza son el desarro-

---

<sup>8</sup> Ibid., p. 129.

<sup>9</sup> MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de resultados lesivos causados por robots”, *Revista de Derecho Penal y Criminología*, 3ª época, núm. 20, 2018, p. 108.

<sup>10</sup> MUÑOZ RODRÍGUEZ, A. B., “El impacto de la inteligencia artificial en el proceso penal”, cit., p. 701.

<sup>11</sup> SOLAR CAYÓN, J. I., “Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos”, cit., p. 129.

<sup>12</sup> MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de resultados lesivos causados por robots”, cit., p. 108.

llo intelectual, los factores de angustia personal y el nivel socioeconómico en la familia de origen<sup>13</sup>.

Si bien todos estos programas han tenido éxito en términos de eficiencia –posibilitan el procesamiento y análisis de ingentes cantidades de información, evitando el uso de recursos humanos en labores rutinarias y destinándolos a otras tareas en donde presenten respecto de las máquinas– las decisiones que se toman producto de la recomendación de la inteligencia artificial pueden resultar injustas o poco equitativas para individuos o grupos particulares. Esto puede deberse no sólo a situaciones de sesgo algorítmico –tema que trataremos más adelante–, sino también a otras causas, como la cantidad, calidad y representatividad de los datos empleados. Y, si bien ello puede producirse sin una intención premeditada, es habitual que los desarrolladores, usuarios y otros involucrados no prevean los efectos discriminatorios o atentatorios contra los derechos de ciertos grupos o individuos<sup>14</sup>. Esto es especialmente delicado porque las decisiones que se toman a partir de recomendaciones dadas por algoritmos, en muchos casos, tienen consecuencias graves en la vida de las personas, como, por ejemplo, dejar de recibir un crédito que se necesita, no conseguir un trabajo, ir a la cárcel, o sufrir el atropello de derechos fundamentales.

#### 4. PRINCIPALES PROBLEMAS QUE HA PLANTEADO EL USO DE SISTEMAS DE EVALUACIÓN DE RIESGOS EN EL PROCESO PENAL

A continuación, se hará una síntesis de los principales problemas que ha planteado el uso de sistemas algorítmicos de predicción de riesgos en aquellos ordenamientos en donde se han empleado. En consecuencia, se examinarán las dificultades introducidas a los procesos judiciales de carácter penal por el sesgo algorítmico, la falta de transparencia y de comprensión de los algoritmos, el uso del secreto empresarial como escudo para no dar a conocer a las partes el proceso de toma de decisiones y, finalmente, los atropellos al debido proceso que se han producido por el uso indebido de herramientas tecnológicas predictivas de riesgo.

##### 4.1. Sesgo algorítmico y falta de imparcialidad

Un sistema algorítmico con aprendizaje profundo funciona por medio de redes neuronales que se pueden organizar por capas. Una primera capa de inicio al procesamiento, la segunda toma el resultado de la primera y lo elabora un poco más, y así, la información pasa sucesivamente, de capa en capa, hasta llegar a las últimas neuronas que aportan el resultado final<sup>15</sup>. Así, mientras más compleja es la red, más provista de capas estará y sus funciones podrán ser más complejas. Las redes neuronales son construidas por los programadores, quienes determinan la arquitectura del modelo. Con posterioridad, esas redes se entrenan con datos y con los pesos que se les asignan a las diferentes conexiones entre

---

<sup>13</sup> Ibid., p. 107.

<sup>14</sup> COECKELBERGH, M., *Ética de la inteligencia artificial*, Madrid: Cátedra, 2021, p. 107.

<sup>15</sup> LATORRE, J. I., *Ética para máquinas*, Barcelona: Ariel, 1ª edición, 2019, p. 108.

las capas<sup>16</sup>, varias o muchas de las cuales pueden estar ocultas. Por lo mismo, para entrenarse, las redes requieren de muchos datos. De modo que el algoritmo tendrá la calidad que tengan los datos con los cuales se ha entrenado. Y ocurre que, en muchos casos, un algoritmo cometerá los mismos errores implícitos en los datos empleados para entrenarlo –si los datos contienen sesgos, el modelo los reproducirá, e, incluso, los amplificará–<sup>17</sup>. En eso consiste el sesgo. En términos estadísticos, es una inclinación sistemática presente en el proceso de recopilación de datos y su efecto principal es crear discriminaciones en contra o a favor de individuos o grupos concretos, dando lugar a resultados desviados y engañosos<sup>18</sup>.

Hay varios casos emblemáticos de sesgos. Por ejemplo, los programas de reconocimiento facial, como el de Facebook (el que adolece de un importante defecto: le cuesta reconocer caras de personas de color). O el caso del programa COMPAS<sup>19</sup>, empleado en varios Estados de EEUU para determinar la reincidencia criminal (y que, entre otros resultados reprochables, da más falsos positivos –acusados respecto de los cuales se predijo que volverían a delinquir y no lo hicieron– en personas de raza negra, y más falsos negativos –acusados que, se suponía, no volverían a delinquir, pero sí lo hicieron– en personas de raza blanca). También encontramos los algoritmos de policía predictiva que, basados en datos poco representativos, estigmatizan la pobreza y la criminalizan<sup>20</sup>. O el bullado caso de la selección de *curriculum vitae* por parte de un algoritmo implementado por la empresa Amazon, que discriminaba a las mujeres en el reclutamiento para puestos laborales<sup>21</sup>.

En el ámbito específico del proceso judicial, la imparcialidad de los jueces es un principio y un derecho esencial. El núcleo de la imparcialidad radica en que el juez no esté más próximo a una parte que a la otra, tanto en lo personal como en el objeto por el cual se litiga. En tal sentido, se espera que el juez no aplique todo el rigor de las leyes, sino que sea

<sup>16</sup> En efecto, se puede asignar un valor al peso de una conexión. Mientras más grande es el peso, más intensa es la conexión. Al contrario, a menor peso, la conexión se torna irrelevante. De esta forma, se puede experimentar con los pesos entre las neuronas para representar las sinapsis y así poder resolver problemas complejos. *Ibid.*, p. 108.

<sup>17</sup> AMUNÁTEGUI PERELLÓ, C., *Archana Technicae. El Derecho y la inteligencia artificial*, Valencia: Tirant lo Blanch, 2020, pp. 45 y 46.

<sup>18</sup> COECKELBERGH, M., *Ética de la inteligencia artificial*, cit., p. 167.

<sup>19</sup> El programa COMPAS (la sigla abrevia su nombre completo: *Correctional Offender Management Profiling for Alternative Sanctions*) tiene por objeto reducir la población carcelaria mediante la utilización de una serie de algoritmos que, según los antecedentes penales del acusado, predicen el posible nivel de reincidencia penal. MUÑOZ RODRÍGUEZ, A. B., “El impacto de la inteligencia artificial en el proceso penal”, cit., p. 703.

<sup>20</sup> Un ejemplo está dado por *PredPol*, herramienta de policía predictiva que se ha usado en Estados Unidos para predecir la probabilidad de comisión de delitos en zonas específicas de las ciudades y, así, recomendar la distribución de los policías. El principal problema que se ha diagnosticado en el funcionamiento de *PredPol* es que el sistema pareciera estar sesgado en contra de barrios pobres o personas pertenecientes a ciertas razas, y en que la vigilancia policial segmentada disminuiría la confianza de la gente en esas áreas, transformando esa predicción en una profecía autocumplida. COECKELBERGH, M., *Ética de la inteligencia artificial*, cit., p. 109.

<sup>21</sup> Véase, al respecto, la noticia en el sitio web: <https://www.bbc.com/mundo/noticias-45823470>



capaz de adaptarlas a lo que sea más justo en el caso concreto y, por esa misma razón, se acepta que tenga un amplio margen de apreciación en la determinación de las distintas figuras delictivas o de las penas concretas que imponer a un delito<sup>22</sup>.

En atención a que las decisiones judiciales varían, porque los criterios o nociones particulares de equidad, buena fe o prudencia de los jueces son cambiantes y dinámicos, con ayuda de la inteligencia artificial alguien podría intentar la incorporación en algoritmos de diversos “factores de corrección” a fin de aportar una visión más empática o “humana” de dichas nociones, para hacerlas aplicables de una forma más uniforme al caso concreto. Sin embargo, esa idea colisionaría con el hecho de que solo los humanos podemos actuar con empatía. Las máquinas actuarían con frialdad, de manera que serían incapaces *a priori* de considerar todos los factores que modulan esa empatía, los cuales se vinculan con los detalles concretos de la historia de cada pleito y que serían muy probablemente ignorados por la inteligencia artificial. Así, ocurriría que la máquina o el programa empleado multiplicaría los errores, en atención a que las circunstancias sociales y los acuerdos sobre ciertos valores son de suyo cambiantes. En consecuencia, no parece conveniente introducir en los algoritmos esas herramientas de imparcialidad, sino que deberían mantenerse en manos de los jueces; pese a que este camino –lo sabemos de sobra– tampoco está exento de dificultades<sup>23</sup>.

#### 4.2. Falta de transparencia y explicabilidad de los algoritmos

Uno de los aspectos más controversiales de los sistemas algorítmicos de evaluación de riesgos radica en su opacidad. Varios de los modelos más complejos de aprendizaje automático (*machine learning*), al funcionar a través de una estructura de aprendizaje profundo (*deep learning*), con muchas capas neuronales, operan como “cajas negras”. Esto quiere decir que no son totalmente inteligibles o escrutables, debido a que no presentan una correlación clara entre las características con las cuales funcionan y los resultados que obtienen<sup>24</sup>. Dicho en términos más simples, significa que se puede rastrear lo que el algoritmo hizo durante el proceso de aprendizaje, pero no por qué lo hizo. En este tipo de sistemas, la posibilidad de comprender por qué tomó cierta determinación, o que llevó a un resultado específico, es mínima.

Como se ha señalado con anterioridad, varios Estados pertenecientes a los Estados Unidos de América utilizan algoritmos de *deep learning* para predecir el riesgo<sup>25</sup>, pero no hay evidencia de que todos ellos hayan realizado estudios que permitan evaluar la fiabilidad de los sistemas. Esto es especialmente problemático, por cuanto se deja en manos

---

<sup>22</sup> NIEVA FENOLL, J., *Inteligencia artificial y proceso judicial*, cit., p. 136.

<sup>23</sup> Ibid., pp. 138 y 139.

<sup>24</sup> NARVÁEZ LÓPEZ, C., “La inteligencia artificial entre la culpa, la responsabilidad objetiva y la responsabilidad absoluta en los sistemas jurídicos del derecho continental y anglosajón”, en VVAA, *El derecho de daños en la era digital*, Bogotá: Ibáñez, 2021, p. 193.

<sup>25</sup> Entre ellos se encuentra, a modo ejemplar, el Estado de Montana, que emplea el denominado “*Montana Offender Reentry*” y el “*Risk Assessment Tool*”; o Arkansas, que utiliza el “*Parole Risk Assessment Tool*”.

de empresas privadas el diseño de programas que emplean algoritmos secretos, sin que conste que los poderes públicos los hayan sometido a procesos de evaluación o los hayan auditado. En consecuencia, la tensión entre intereses públicos y privados se torna evidente, primando estos últimos, en la medida en que los sistemas siguen operando sin mayores límites, pese a que existen estudios que demuestran que pueden resultar abiertamente discriminatorios y poco fiables<sup>26</sup>.

En ese sentido, uno de los principales problemas que presentan este tipo de aplicaciones de inteligencia artificial es que gran parte de ellas mantienen su funcionamiento bajo un estricto secreto industrial. Mientras más sofisticadas son, mayor es el celo de las empresas para que mantener su funcionamiento en reserva, a fin de desincentivar el plagio por parte de sus competidores o de otros sujetos que puedan aprovechar de forma gratuita esa tecnología por la cual ellos han trabajado y desembolsado cuantiosas sumas de dinero. En este escenario, la dificultad radica en que las empresas no están dispuestas, bajo ninguna circunstancia, a revelar el todo o parte de su secreto empresarial; ni siquiera en el ámbito específico de los procesos judiciales. En tal sentido, suelen aminorar las consecuencias del conflicto provocado por el desconocimiento en cuanto a su forma de operar, haciendo énfasis en las bondades de eficiencia en el manejo de la información y en su contribución en el ámbito de la administración de justicia<sup>27</sup>.

En el sistema judicial norteamericano, el programa COMPAS ha sido implementado en varios Estados como un instrumento de apoyo a los jueces. Este auxilio se traduce en que indica qué condena imponer a cada persona privada de libertad y valora la posibilidad de reincidencia de cada una de ellas<sup>28</sup>. Una de las causas judiciales más emblemáticas y controvertidas en las que se ha utilizado esta herramienta es el caso *State v. Loomis*, el cual, por su particular importancia será objeto de un análisis más detallado en el apartado II de este capítulo. Por ahora, baste decir que, en ese proceso, tras haberse sometido al cuestionario de COMPAS, un ciudadano estadounidense fue sentenciado a siete años de prisión, habida cuenta de que el programa arrojó un resultado de riesgo de reincidencia alto. El señor Eric LOOMIS recurrió la condena, basándose en que la empresa que fabricaba el algoritmo (Northpointe Inc.) impedía averiguar los detalles de su funcionamiento, lo que se transformaba en la “caja negra” del sistema<sup>29</sup>.

Por supuesto que COMPAS no es el único programa utilizado en este ámbito<sup>30</sup>. Con propósitos similares, pero con un alcance geográficamente más acotado, el equivalente

---

<sup>26</sup> HERNÁNDEZ GIMÉNEZ, M., “Inteligencia artificial y Derecho Penal”, *Actualidad Jurídica Iberoamericana*, núm. 10 bis, 2019, p. 826.

<sup>27</sup> A modo ejemplar, en el caso *State v. Loomis*, la empresa desarrolladora de COMPAS argumentó que lo importante no era cómo funcionaba el programa, sino la ayuda que con él brindaban a la justicia de Estados Unidos. HERNÁNDEZ GIMÉNEZ, M., “Inteligencia artificial y Derecho Penal”, cit., p. 824.

<sup>28</sup> Ibid., p. 824.

<sup>29</sup> Ibid., p. 824.

<sup>30</sup> Su especial cobertura es consecuencia del emblemático caso ya mencionado de *State v. Loomis*, por cuanto el imputado recurrió una sentencia judicial luego de que la empresa ProPublica hiciera conocido un reportaje que alertaba sobre las falencias del sistema. En efecto, pese a que el programa lleva varios años en funcionamiento, sólo en forma reciente se ha hecho más visible, debido a una resolución judicial y a un reportaje periodístico que se hizo sobre el mismo. El reportaje lo

europeo de COMPAS podría ser HART (*Harm Assessment Risk Tool*) programa utilizado en Durham para informar sobre el riesgo de reincidencia de personas que habían delinquido y, sobre la base de este pronóstico, decidir sobre la puesta en libertad o no de los sujetos<sup>31</sup>.

Un aspecto clave en la utilización de este tipo de algoritmos es que, al igual que acontece con los procesos de toma de decisiones por humanos, ninguno resulta ser infalible. Es cierto que, en los sistemas de aprendizaje profundo, el algoritmo, a medida que aprende, va minimizando los errores que comete. Pero esta minimización no llega al valor cero; al menos no en el estado actual de desarrollo de la inteligencia artificial. Por lo mismo, es imprescindible que el uso de los sistemas de predicción de riesgos esté supervigilado por humanos, desde el diseño de los mismos, hasta la aplicación de procesos de auditoría, que permitan evaluar su funcionamiento, no sólo en términos de eficiencia, sino también –y en esto radica el aspecto más relevante a considerar– en términos del respeto a los derechos fundamentales de las personas involucradas en los procesos judiciales. Sobre todo, en un ámbito de carácter tan garantista como el proceso penal.

En consecuencia, es posible avanzar en la transparencia y explicabilidad de los algoritmos utilizados en los sistemas predictivos de riesgo. En la actualidad existen varios esfuerzos orientados a tratar de que tales sistemas resulten inteligibles y escrutables para las personas ajenas a su diseño y entrenamiento<sup>32</sup>. Uno de los factores que contribuyen a lograr ese objetivo es que las empresas que los desarrollan den a conocer –al menos en el ámbito del proceso judicial penal– los valores que han tenido en cuenta para ponderar el riesgo de reincidencia de cada sujeto. Asimismo, que dichos valores no estén vinculados a elementos potencialmente discriminatorios, como la raza, el nivel educacional, el sexo o factores socioeconómicos que redunden en una estigmatización de la pobreza, más que en la predicción para la toma de decisiones racionales y justas.

Resulta, entonces, imperativo que los sistemas de inteligencia artificial estén orientados por los principios de transparencia y de explicabilidad, de modo que las personas puedan tener conciencia de que están interactuando con aquéllos y puedan, con conocimiento de causa, oponerse a los resultados de esa interacción. Por ello, no sólo es conveniente, sino

---

publicó la web ProPublica, con el sugerente título: “*El sesgo de la máquina: Existe un software que se está usando en todo el país para la predicción de futuros criminales. Y está sesgado contra los negros*”. En dicho reportaje se analizaba el nivel de fiabilidad de la predicción realizada por COMPAS y se ponía en duda la equidad del sistema, que parecía discriminar en forma significativa a ese grupo de la población estadounidense. MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de resultados lesivos causados por robots”, cit., p. 109.

<sup>31</sup> Ibid., p. 110.

<sup>32</sup> Ello se logra, por ejemplo, a través del mecanismo de la retropropagación (*back-propagation*). Debido a que un error en una capa oculta de una red neuronal no se conoce, porque los datos de entrenamiento del algoritmo no dicen cuál es el valor que han de tomar los nodos ocultos (los nodos son las unidades de una red neuronal), se puede “propagar hacia atrás” el error desde la capa de salida a las capas ocultas. Así, mediante ecuaciones matemáticas, es posible detectar dónde el error se produce (una suerte de rastreo hacia atrás) y corregirlo para que el algoritmo prosiga su entrenamiento disminuyendo el margen de error. RUSSELL, S. y NORVIG, P., *Artificial Intelligence. A modern approach*, New Jersey: Prentice Hall, third edition, 2010, p. 733.

imprescindible, sobre todo en el especial contexto del proceso penal, que el funcionamiento de la inteligencia artificial sea explicable en un lenguaje accesible para todas las personas y que sea posible analizar los resultados intermedios del sistema –no sólo los resultados finales–, para poder validarlos<sup>33</sup> en forma adecuada, con pleno respeto de las garantías procesales que asisten a los imputados.

#### 4.3. Vulneraciones al debido proceso en las resoluciones judiciales

Uno de los principales temores que surgen de la aplicación de la inteligencia artificial en un proceso judicial es que una máquina acabe sustituyendo la función del juez, de modo que los seres humanos terminen siendo juzgados por agentes artificiales, los cuales carecen de conciencia, de criterio jurídico y son ajenos a la noción de justicia. Una aplicación de inteligencia artificial no está capacitada para valorar los hechos como puede hacerlo una persona humana. Por lo mismo, al decidir, lo hará siempre de la misma forma, sin adaptarse a los cambios o situaciones particulares, con la consecuente tendencia a “fossilizar” las decisiones que adopte<sup>34</sup>.

En el caso particular –pero emblemático– de COMPAS, los tribunales de Wisconsin declararon la validez de la sentencia dictada en contra del señor LOOMIS, en atención a que se habían valorado más pruebas que la aportada por dicho programa. La Corte Suprema del Estado de Wisconsin ratificó la sentencia, desestimando las pretensiones de la defensa, argumentando –en idéntico sentido– que COMPAS era solo uno de los muchos factores que se habían considerado y ponderado en la sentencia. Con todo, ello no ha logrado despejar la interrogante en torno a si los jueces realmente fueron, en ese caso específico y lo serán en procesos futuros, capaces de apartarse de las conclusiones arrojadas por los algoritmos<sup>35</sup>. Al respecto, no se debe olvidar que este programa aprende a partir de sentencias que ya han sido ejecutadas, razón por la cual, si estas sentencias contienen un sesgo relacionado con un colectivo determinado, aplicará este sesgo a sus propias recomendaciones<sup>36</sup>.

En vista de casos como el anterior, parte de la doctrina ha podido concluir que la inclusión en un proceso de las técnicas de predicción judicial debe realizarse, en toda circunstancia, con el debido respeto a los derechos fundamentales de las partes o intervinientes, así como de sus garantías procesales. En tal sentido, se torna absolutamente necesaria y pertinente la adopción de una normativa en el ámbito procesal “que, asentada bajo los principios de la transparencia algorítmica y la imparcialidad del validador,

---

<sup>33</sup> CORVALÁN, J. G., “Inteligencia artificial: retos, desafíos y oportunidades. *Prometea*: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia”, *Revista de Investigações Constitucionais*, vol. 5, núm. 1, 2018, p. 311; SAN MIGUEL CASO, C., “Las técnicas de predicción judicial y su repercusión en el proceso”, en VVAA, Conde y Serrano (Eds.), *La justicia digital en España y la Unión Europea: situación actual y perspectivas de futuro*, Barcelona: Atelier, 2019, p. 46.

<sup>34</sup> MUÑOZ RODRÍGUEZ, A. B., “El impacto de la inteligencia artificial en el proceso penal”, cit., p. 708.

<sup>35</sup> HERNÁNDEZ GIMÉNEZ, M., “Inteligencia artificial y Derecho Penal”, *Actualidad Jurídica Iberoamericana*, núm. 10 bis, 2019, pp. 824 y 826.

<sup>36</sup> *Ibid.*, p. 825.

conjugue todos los elementos descritos anteriormente, con la finalidad de avalar, con las máximas garantías, la utilización de la inteligencia artificial en el proceso, evitando, de este modo, la vulneración sistemática del derecho de defensa. Del mismo modo, es necesario diseñar un sistema de corresponsabilidad frente a los daños que se pudieran derivar del uso de estas técnicas de predicción judicial de manera que, tanto los desarrolladores del algoritmo como la propia Administración de Justicia, respondan por los errores técnicos y jurídicos que pudieran resultar del empleo de las mismas<sup>37</sup>.

## 5. ¿CÓMO COMPATIBILIZAR EL USO DE ESTOS SISTEMAS ALGORÍTMICOS Y LAS EXIGENCIAS DEL DEBIDO PROCESO EN MATERIA PENAL?

En el apartado anterior se identificaron los principales problemas detectados por la doctrina y jurisprudencia en relación con el uso de sistemas algorítmicos de predicción y de evaluación de riesgos por parte de los tribunales de justicia. En el presente, se explicitarán cuáles son las condiciones que dichos sistemas deben cumplir para cautelar de manera adecuada los derechos fundamentales que integran el debido proceso de los justiciables.

Que los sistemas algorítmicos de predicción y evaluación de riesgos utilizados para auxiliar a los jueces no afecten de los derechos de las personas sometidas a un proceso penal es, a todas luces, esencial a la hora de evaluar su aplicación y viabilidad en el sistema de justicia criminal. Bajo la premisa anterior resulta evidente la importancia de dar respuesta a la siguiente pregunta de investigación: ¿cómo compatibilizar el uso de estos programas algorítmicos con las exigencias del debido proceso en materia penal?

Para dar respuesta a esta interrogante, esta sección se dividirá en dos subapartados. En el primero, se abordarán las respuestas que ha dado la doctrina y la jurisprudencia para superar las dificultades detectadas en el uso de estos sistemas. En concreto, se analizarán los problemas sobre sesgo algorítmico o discriminación, secreto empresarial, explicabilidad y transparencia. La segunda parte se destinará a la identificación de algunos lineamientos para hacer compatible el uso de sistemas algorítmicos de predicción de riesgos con las exigencias del debido proceso en materia penal. De esta manera, en base a la experiencia existente, se procurará identificar qué condiciones deberían darse para que esos programas funcionen sin vulnerar el debido proceso o atropellar las garantías de los justiciables.

### 5.1. ¿Qué ha señalado la jurisprudencia y la doctrina en relación con el uso de sistemas algorítmicos de predicción de evaluación de riesgos?

Las soluciones proporcionadas por la doctrina y jurisprudencia respecto de los problemas derivados del uso de sistemas de algorítmicos de predicción y evaluación de riesgos, en su gran mayoría, se han originado en pronunciamientos derivados de algunos casos que pueden calificarse de fundamentales. Estados Unidos de América es uno de los paí-

---

<sup>37</sup> SAN MIGUEL CASO, C., “Las técnicas de predicción judicial y su repercusión en el proceso”, p. 49.

ses<sup>38</sup> en los que se aplican extendidamente estos modelos predictivos, por lo que no es de extrañar que los casos destacados hubiesen tenido lugar en dicho país. En términos cuantitativos, más de veinte Estados<sup>39</sup> emplean regularmente dichos sistemas de inteligencia artificial para auxiliar a los jueces al momento de determinar la pena aplicable; uno de ellos es COMPAS<sup>40</sup>.

El primer caso que abrió el debate anglosajón fue *Malenchik vs. State of Indiana*, de 9 de junio de 2010. El pronunciamiento judicial corresponde a la Suprema Corte del Estado de Indiana. El imputado MALENCHIK fue condenado por un delito contra la propiedad. La sentencia fue de seis años, dos de los cuales se otorgaron con suspensión. La defensa presentó un recurso de apelación sobre la base de dos impugnaciones: 1) El tribunal de primera instancia, de manera equivocada, habría considerado como circunstancia agravante de responsabilidad penal el puntaje entregado por un sistema de predicción de riesgos (LSI-R, SASSI) empleado por el departamento de libertad condicional del condado de Tippecanoe y 2) esto se habría traducido en una sentencia errónea. El tribunal de la segunda instancia rechazó las dos reclamaciones. La decisión de fondo del tribunal aludió a que los modelos predictivos y de evaluación de riesgos, si bien no pueden reemplazar la decisión judicial, sí pueden ser utilizados para apoyarla y, por ende, ser legítimamente tomados en consideración por las autoridades encargadas de la decisión. El laudo agrega que el uso de la inteligencia artificial sirvió para complementar otras pruebas con que contaba el tribunal y, en modo alguno, fue la única que sirvió al órgano jurisdiccional para que sustentara su decisión<sup>41</sup>.

Un segundo caso es *State vs. Loomis*, el cual es mencionado de manera prácticamente unánime por la doctrina especializada. Esta sentencia fue pronunciada por el Tribunal Supremo de Wisconsin el día 13 de julio de 2016. Eric LOOMIS fue imputado por la Fiscalía por haber participado en cinco delitos, uno de los cuales consistía en haber intervenido en un incidente donde se usaron armas de fuego. El imputado alcanzó un acuerdo con la Fiscalía, aceptando dos de los cinco cargos que originalmente le habían imputado. El tribunal, al momento de dictar sentencia, tuvo en consideración un *presentence investigation report*. Este reporte recogió los resultados de COMPAS que, como sistema de predicción y evaluación de riesgo, valora la reincidencia y el perfil criminal de los imputados<sup>42</sup>. El medio

<sup>38</sup> Mucho más radical que estos sistemas, puesto que derechamente importa resolución de conflictos a través de sistemas automatizados, es el caso canadiense con el *British Columbia Civil Resolution Tribunal* que proporciona soluciones *on line* para diversos conflictos jurídicos. También en el reino Unido se puede observar un sistema *on line* denominado *Online Solution Court*, que resuelve reclamaciones que no sobrepasen cierto monto de libras. Véase, al respecto, la nota número 4.

<sup>39</sup> Arkansas (*Parole Risk Assessment Tool*), Louisiana (LARNA); Montana (*Montana Offender Reentry y Risk Assessment Tool*); Alabama (VPRAI/Jefferson County) Kentucky (PSA/Statewide). En *Algorithms in the Criminal Justice System. AI and Human Rights: Criminal Justice System* EPIC.ORG. Accesible en: <https://epic.org/ai/criminal-justice/index.html>.

<sup>40</sup> CASTELLANOS CALARMUNT, J. y MONTERO CLARO, D., "Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales", *Revista Ius et Ciencia*, vol. 6, núm. 2, 2020, p. 73.

<sup>41</sup> *Malenchik vs. State of Indiana*, 9 de junio de 2010.

<sup>42</sup> El sistema COMPAS emplea más de 100 factores como sexo, edad, etc. Con estos elementos clasifica a las personas según grupos de riesgos, la máxima puntuación es 10 y se estima que de

tecnológico reportó que el condenado tenía un alto riesgo de reincidencia, la que además fue catalogada como violenta.

En la condena, el tribunal rechazó someter al condenado a un sistema de libertad a prueba porque existía un riesgo importante de reincidencia. La sanción impuesta consistió en seis años de privación de libertad y cinco de supervisión después que diera cumplimiento a dicha pena. La defensa recurrió por diversas razones. Una de estas consistió en haber considerado dentro del cálculo de la pena los delitos que quedaron fuera del acuerdo. Para la presente investigación, lo más relevante es la alegación relativa a que la fundamentación de la sentencia se había basado en el sistema COMPAS, cuya utilización habría afectado el debido proceso durante el juicio seguido contra el mencionado imputado.

La defensa recurrió para ante el Tribunal Supremo de Wisconsin. El recurso se fundamentó en lo siguiente: 1) Tránsito del derecho a ser condenado en base a información fiable y precisa. Esta alegación se justificó en la circunstancia de que el secreto empresarial impide conocer el algoritmo que realiza la evaluación del riesgo, lo que redundaría en la imposibilidad de cuestionar la valoración del peligro realizada por el modelo predictivo, porque no se conoce cómo opera; 2) Vulneración de su derecho a contar con una sentencia individual y 3) La decisión del juez se basó en su género<sup>43</sup>.

La Corte de Wisconsin rechazó el recurso planteado por la defensa de LOOMIS. La argumentación en la que basó su decisión consistió en que, si bien existe un secreto de fábrica, en virtud del cual la empresa no está obligada a develar cómo funciona el algoritmo de COMPAS, esto no obsta a que entregue información general sobre el programa. En la especie, el condenado habría tenido acceso a la información que utilizó el tribunal para condenarlo, cuestión que se evidencia en que pudo revisar el nivel de riesgo asignado por el sistema y también conocía su propio historial criminal que fue considerado en el cálculo. Así las cosas, se consideró que no se vulneró su derecho a ser condenado sobre la base de información fiable y precisa. De hecho, fue sometido a un cuestionario, de manera tal que pudo refutar la información. En relación con la segunda alegación planteada por la defensa, relativa a la afectación de la individualización de la sentencia, tampoco se habría verificado esta infracción. El motivo que esgrimió el tribunal se basó en que la información general que propone COMPAS se contrasta con la que se encuentra en su poder respecto de la persona que está siendo juzgada, de forma tal que el juez pondera y realiza un juicio sobre la base de toda la información con que cuenta. Respecto del tercer cuestionamiento esgrimido, la Corte también descartó dicha alegación, indicando que en el juicio las partes aceptaron que uno de los elementos para valorar el riesgo es el género de los imputados y que los indicadores disponibles dan cuenta que los hombres arrojan mayores resultados en relación con el riesgo. De ello se deriva que éste es un elemento que contribuye a dar

---

1 a 4 puntos las personas se consideran de bajo riesgo; de 5 a 10 de riesgo medio o alto, tal como lo mencionan CASTELLANOS CALARMUNT y MONTERO CLARO. CASTELLANOS CALARMUNT, J. y MONTERO CLARO, D., "Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales", cit., p.77.

<sup>43</sup> MARTÍNEZ GARAY, L., "Peligrosidad, algoritmos y *due process*: el caso *State vs. Loomis*", *Revista de Derecho Penal y Criminología*, 3.ª Época, núm. 20, 2018, pp. 490-491.

resultados más precisos, por lo que no podría sostenerse que esto perjudique, sino que, al contrario, beneficia a la administración de justicia.

A la par que la jurisprudencia, la doctrina también ha sido crítica con COMPAS y, en general, con los modelos predictivos y de evaluación de riesgos. Uno de los cuestionamientos efectuados se relaciona con las discriminaciones que se producen en el trato que se despliega respecto de las personas blancas y de las de color, o más específicamente, la disparidad de puntajes otorgados a unos y otros. Las personas de color obtienen puntajes más altos al momento de cuantificar los riesgos y además aumenta la probabilidad de incurrir en delitos violentos. Todavía más decisivo que estos criterios es el porcentaje de error al momento de determinar los resultados en materia de reincidencia, toda vez que, en relación con la población blanca, se incurre en muchas más equivocaciones que respecto de la población de color<sup>44</sup>.

También los autores se han pronunciado sobre el complejo escenario que genera para el debido proceso el uso de este tipo de herramientas. Por ejemplo, en relación con el hecho de que las condenas se basen en información que sea fiable, precisa y que la defensa tenga la posibilidad de poder revisarla. Sobre el punto, la doctrina norteamericana ha manifestado la importancia de permitir que el algoritmo que se emplea en los sistemas de evaluación de riesgo sea abierto, para que pueda ser conocido por las defensas. Esto supone hacer primar el debido proceso por sobre la protección de los derechos empresariales. Otra opción es permitir que empresas auditoras externas hagan el proceso de validación para garantizar su precisión y que su uso pueda ser legitimado<sup>45</sup>. En suma, si pudiera lograrse que los códigos de los sistemas sean abiertos y los defensores puedan conocerlos, comprendiendo la forma en que generan sus reportes, estos podrían ser revisados por expertos convocados por esta parte del proceso penal. Esto, por ejemplo, ha ocurrido con la tecnología que se emplea para determinar los ADN, toda vez que el conocimiento de los procesos permite que, durante el contra examen de los peritos, se formulen preguntas que manifiesten posibles errores o defectos de dichos sistemas. También la existencia de códigos abiertos permite que estos sean objeto de investigación, lo que, a la postre, facilita evaluar externamente sus resultados. En todo caso, los sesgos en relación con la población de color no se superan con la existencia de códigos abiertos<sup>46</sup>.

En el mismo orden de ideas, un estudio realizado en Estados Unidos explica las fortalezas y debilidades del sistema COMPAS. Al respecto señala que lo destacable y positivo de esta herramienta es que resulta sencilla de usar para jueces y otras autoridades que puedan servirse de sus reportes. También, que permite que respecto de todas las personas se evalúen los mismos criterios o categorías, lo que proporciona generalidad e igualdad. Además, tendría la virtud de que permite comparar cuán elevada es la puntuación de un imputado respecto de otro en relación con dichos parámetros. Sobre la base de estos criterios se determina lo que “parece ser” un riesgo de reincidencia o la existencia de ciertos factores criminógenos. El

---

<sup>44</sup> FREEMAN, K., “Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in *State vs. Loomis*”, *North Carolina Journal of Law & Technology*, vol. 18, Issue 5, 2016, p. 84.

<sup>45</sup> FREEMAN, K., “Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in *State vs. Loomis*”, cit., p. 101.

<sup>46</sup> Ibid., pp. 102-104.



problema es que realmente existe muy poca evidencia de que esto sea lo que realmente dicho sistema evalúa. En este sentido, por ejemplo, cuando COMPAS señala que valora problemas de drogas y alcohol para predecir reincidencia, lo que no se sabe es si realmente “evalúa” dichos problemas de drogas y alcohol, lo que naturalmente conlleva incertidumbre. Esto, porque no se conoce si COMPAS utiliza parámetros validados para determinar un consumo problemático de drogas y alcohol, lo que conlleva que sus resultados sean inciertos. Lo explicitado trasunta que los programas de rehabilitación que se elaboran puedan no resultar útiles para sus destinatarios y lo mismo puede decirse de los informes o reportes entregados a los jueces al momento de valorar una sanción u otra. Adicionalmente, en el estudio se indica que no existe mucha evidencia acerca de que el sistema COMPAS realmente prediga la reincidencia. Tampoco se ha demostrado que la escala de esta herramienta prediga actuaciones violentas. La tercera conclusión de este trabajo es que no existe evidencia de que las escalas o baremos que utiliza el sistema se modifiquen durante el transcurso del tiempo y que, por lo tanto, se adapten a los factores criminógenos, de suyo cambiantes<sup>47</sup>. Lo mencionado anteriormente revela por qué es necesario que el sistema sea auditado y valorado externamente y no solo por el propio fabricante. No ha resultado suficiente que las defensas conozcan los cuestionarios y los elementos con los que se alimentan los mencionados programas.

En suma, sistematizando los principales problemas que los programas algorítmicos de predicción y evaluación de riesgos pueden conllevar para las garantías del debido proceso, es posible mencionar los siguientes:

- Afectación del derecho a defensa y a prepararla adecuadamente. Si bien las empresas que elaboran los algoritmos explican y entregan información respecto de éstos, no es posible conocer cómo funcionan cabalmente. Lo expresado obedece a que darlo a conocer sería sinónimo de develar un secreto empresarial, lo que perjudica el negocio. Con motivo de la existencia de un secreto de fábrica las instituciones no entregan información, impidiendo a las defensas conocer el funcionamiento preciso de la herramienta y, por ende, de los resultados que arroja.

El derecho a defensa comprende el conocimiento cabal de los antecedentes que serán presentados, especialmente la de cargo. Ello se explica en que el derecho a la defensa material requiere que el defensor conozca cómo se obtuvieron, de qué manera, dónde y cuál es su contenido, ello con la finalidad de poder controvertirlos; de otro modo la garantía se puede ver seriamente afectada. De esta forma, si los reportes que fijan puntajes de riesgos para los imputados pueden ser empleados por los tribunales como antecedentes para determinar la procedencia de una pena sustitutiva o para conceder un beneficio intrapenitenciario, naturalmente, esta información debe ser sometida al mismo estándar de contradicción que afecta a todo aquello que pueda fundar una decisión en su contra.

- Tutela Judicial efectiva. Esta alegación se basaría en que la condena no se fundamentaría –al menos exclusivamente– en el razonamiento judicial, sino más bien

---

<sup>47</sup> SKEEM J. y ENO LOUDEN, J., “*Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)*”, Center for Public Policy Research, University of California, 2007, p. 29.

en el resultado emanado de una herramienta de inteligencia artificial. Esta situación conllevaría indefensión, toda vez que la decisión se radicaría en un sistema informático. Lo expresado se explica en que no es posible conocer todos los antecedentes que llevaron al juez a resolver o razonar de una determinada manera. Tampoco sería claro cuál es la incidencia que el sistema algorítmico tuvo en la decisión del juez<sup>48</sup>. Esto resulta de la mayor relevancia para el debido proceso, que queda integrado por la garantía de fundamentación judicial, cuya finalidad es reproducir el razonamiento del órgano decisor justificando racionalmente cómo y por qué arribó a un determinado pronunciamiento.

La ausencia de explicabilidad de la decisión jurisdiccional se agrava si consideramos lo reseñado en el punto anterior respecto del secreto de fábrica y la falta de transparencia en relación con el programa de predicción del riesgo. La cuestión se torna circular: no existe un cabal conocimiento acerca de cuál es el algoritmo, ni cómo funciona al momento de evacuar sus reportes, por lo que los puntajes contenidos en éstos tampoco resultan explicables, tornando la decisión judicial, al menos, en parcialmente ciega desde la perspectiva del justiciable.

- Se afecta la imparcialidad del juzgador. Tal como se explicó en la primera parte de este trabajo, uno de los problemas de los sistemas predictores y de evaluación de riesgos es que pueden afectar la imparcialidad de los jueces que los utilizan. Lo anterior obedece a la posibilidad de que los sesgos de los modelos informáticos se traspasen al tribunal, influyendo en la imparcialidad que debe primar en la resolución del caso sometido a su conocimiento. Lo afirmado es consecuencia de que estos programas se alimentan, en una primera etapa, de la información que incorporan sus programadores, la que responde a sus creencias respecto de la edad, sexo, etnia, etc. Esto, a su vez, influye en la información que se agrega, generando una cierta concepción de estos parámetros, la que, a la postre, afectará los reportes que el sistema evacúe. Adicionalmente, cuando el sistema se encuentre en funcionamiento, recopilará los datos que emanan de casos anteriores, los que, pese a corresponder a una determinada localidad y un contexto acotado, provocará que este sesgo quede incorporado en la base del sistema de inteligencia artificial, lo que finalmente se traducirá en traspasos de sesgos que pueden influir en la decisión que adopte el tribunal<sup>49</sup>. El problema de los sesgos y la consecuente discriminación de los justiciables durante el proceso penal fue tratado anteriormente en este trabajo, por lo que nos remitimos a ello<sup>50</sup>.

De acuerdo con la posición que se asume en este trabajo, la utilización de instrumentos de inteligencia artificial al servicio del proceso penal necesariamente debe respetar los de-

---

<sup>48</sup> CASTELLANOS CALARMUNT, J. y MONTERO CLARO, D., "Perspectiva constitucional de las garantías de aplicación de la inteligencia artificial: la ineludible protección de los derechos fundamentales", cit., pp. 76-80.

<sup>49</sup> HERNÁNDEZ GIMÉNEZ, M., "Inteligencia artificial y Derecho Penal", cit., pp. 824-825.

<sup>50</sup> Parte I, Sección 2.1.

rechos esenciales de los justiciables. La versión contraria no puede tener cabida, toda vez que el respeto de los derechos fundamentales de los intervinientes es esencial para un Estado democrático de derecho. El carácter utilitario que pueda tener el uso de la inteligencia artificial debe ser bienvenido, pero no a costa del sacrificio de aquello que configura el estatuto de garantías de los justiciables frente al Estado. Por lo tanto, a la hora de analizar su uso, es relevante que logre superar el examen de garantías. Por este motivo, los programas de predicción y evaluación de riesgos, desde el debido proceso de los imputados, debería precaver y dar solución a estos aspectos críticos. En el acápite siguiente se consignarán algunos estándares que deberían considerarse en el evento de que se opte por usar estos modelos predictivos en el sistema chileno.

## 5.2. ¿En qué aspectos debemos poner especial atención al momento de pensar en la posibilidad de aplicar estos sistemas y en el ordenamiento interno?

La doctrina clasifica los instrumentos de inteligencia artificial en materia penal distinguiendo su uso al servicio del *policing* y su empleo para el *sentencing*. El primer caso se orienta a la prevención de los delitos y al análisis de información policial<sup>51</sup>. El segundo, a prestar auxilio al juez en la toma de algunas decisiones<sup>52</sup>. Ejemplos del primer grupo son: la georreferenciación o mapas de los delitos, programas de análisis o cruce de información, empleo de drones, entre otros. Dentro del segundo grupo se encuentran los programas de predicción y evaluación de riesgos que se emplean para determinar la peligrosidad y la posibilidad de reincidencia de los imputados.

Este trabajo se ha centrado en el uso de la inteligencia artificial para el apoyo de decisiones que se toman durante el proceso penal, en particular, en relación con el debido proceso de los justiciables. En ese contexto, se ha relatado cómo el uso de esa inteligencia artificial para apoyar la toma de decisiones de los jueces es utilizado en sistemas procesales penales de diversos países, aunque con una especial ascendencia en los de origen anglosajones, como el caso norteamericano, británico y el canadiense. Respecto de Europa, su implementación ha sido más tímida, no existiendo a la fecha un uso generalizado, en parte, debido a las reservas existentes en materia de derechos fundamentales. Pese a ello, queda claro que existe una cierta tendencia a incorporar la inteligencia artificial, lo que se manifiesta en un cúmulo de instrumentos europeos que dan cuenta de aquello<sup>53</sup>.

---

<sup>51</sup> Sobre el uso de inteligencia artificial por parte de las policías y, concretamente, por el Departamento de Policía de Chicago, se puede revisar el trabajo de BELTRÁN ROMÁN, V. y PREMINGER SAMET, D., “Inteligencia artificial en el sistema de justicia criminal: algunas reflexiones sobre su aplicación en el derecho chileno”, *Revista de Derecho Aplicado LLM UC* 5, 2020, pp. 6-8.

<sup>52</sup> MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de resultados lesivos causados por robots”, cit., p. 97.

<sup>53</sup> Entre otros, la Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 26 de abril de 2016, relativa al tratamiento de datos personales y también la Carta Ética Europea sobre el Uso de la Inteligencia Artificial en los Sistemas Judiciales y en su Entorno, de diciembre de 2018. También es posible mencionar el Libro Blanco de la Inteligencia Artificial.

El ordenamiento jurídico chileno<sup>54</sup> paulatinamente ha ido siguiendo la misma tendencia de ampliar el recurso a instrumentos de inteligencia artificial. Si bien todavía el poder judicial local no aplica modelos predictivos y de evaluación de riesgos para apoyar las decisiones de los jueces penales, se emplea, en la actualidad –en el contexto del sistema de justicia criminal interno– el sistema IGI, cuya implementación corresponde a Gendarmería de Chile<sup>55</sup>. Este mecanismo no puede ser calificado como una herramienta de inteligencia artificial, toda vez que se trata de un instrumento actuarial<sup>56</sup> que se organiza a través de un conjunto de elementos o factores usados para la valoración, puntuación y registro de información en las diversas secciones que lo configuran, lo que sumado a cierta discrecionalidad que se le otorga al evaluador, permite determinar el riesgo de reincidencia de una persona en específico y elaborar un plan de intervención, según sea el caso<sup>57</sup>.

Después de haber pasado revista a los problemas que conlleva para el debido proceso el uso de sistemas predictivos y de valoración de riesgos en apoyo de las decisiones judiciales, en adelante, se propondrán algunos estándares mínimos que deberían tenerse en consideración al momento de implementar este tipo de programas en los sistemas de justicia criminal, a fin de evitar que se afecten los derechos procesales esenciales de los justiciables. La propuesta que se formula a continuación no pretende agotar el tema, sino

---

<sup>54</sup> En este trabajo se deja fuera el uso de herramientas de inteligencia artificial por parte del Poder Judicial chileno en materia de gestión de casos, videoconferencias, etc. Dichas herramientas son tratadas por otros autores e, incluso, han sido recomendadas como una manera de facilitar el acceso a la justicia y la tutela judicial efectiva, especialmente respecto de ciertos miembros de la sociedad que pueden ver limitados sus derechos debido a su situación económica, la distancia, la ausencia de información, entre otros factores. En LILLO, R., et al., “El sistema judicial en crisis: la inteligencia artificial como solución al acceso a la justicia”, *Facultad de Derecho de la Universidad Diego Portales, Semillero de Derecho Procesal*, 2020, p. 7. Accesible en: <https://semillero derechoprocesal.udp.cl/wp-content/uploads/2021/01/2020-el-sistema-judicial-en-crisis-la-inteligencia-artificial-como-solucion-al-acceso-a-la-justicia.pdf>

<sup>55</sup> Gendarmería de Chile es un Servicio Público que depende del Ministerio de Justicia, cuya finalidad es atender, vigilar y contribuir a la reinserción social de las personas detenidas o privadas de libertad (artículo 1º de la Ley Orgánica Constitucional de Gendarmería núm. 2.859 de 12 de septiembre de 1979).

<sup>56</sup> Son instrumentos que se basan en la estadística. Son estructurados y desarrollados para estimar el riesgo de reincidencia, detectar las necesidades de intervención y, sobre la base de estos resultados, construir un plan de intervención para una persona determinada. Al utilizarse estos instrumentos actuariales, se logra construir el perfil de cada individuo. El evaluador logra identificar un determinado peligro de reincidencia, conjugando los factores generales de riesgo, que corresponden a aquellos que tendrían una mayor incidencia el comportamiento ilícito. En LETELIER G.; CERDA PEREIRA, A. y DÍAZ BÓRQUEZ, D. *Informe final evaluación programas gubernamentales (EPG)*, Programas de Rehabilitación y reinserción social, 2018-2019. Accesible en: [https://www.dipres.gob.cl/597/articles-189326\\_informe\\_final.pdf](https://www.dipres.gob.cl/597/articles-189326_informe_final.pdf)

<sup>57</sup> Este instrumento es utilizado por Gendarmería de Chile desde el año 2013. En la actualidad, se emplea al momento de sugerir a los tribunales la aplicación de alguna de las penas sustitutivas de libertad vigilada o libertad vigilada intensiva. También en los informes sobre libertad condicional debe incluirse una mención al nivel de riesgo y elaborar el plan de intervención. Asimismo, se utiliza el instrumento IGI respecto de algunos privados de libertad que se encuentran próximos terminar su condena (un año aproximadamente para salir) con el objetivo de determinar si merecen o no un plan de intervención para prepararlos al momento de reincorporarse a la sociedad. En Gendarmería de Chile. *Informe de avances en reinserción social. Informe de gestión*, 2014/2017, pp. 68 y ss. Accesible en: [https://html.gendarmeria.gob.cl/doc/ARS\\_informe\\_de\\_gestion.pdf](https://html.gendarmeria.gob.cl/doc/ARS_informe_de_gestion.pdf).

tan solo proponer unas orientaciones iniciales para evitar, o al menos minimizar, el riesgo de afectación de los mencionados derechos procesales. Estos lineamientos básicos son los siguientes:

- Someter a los sistemas de predicción y evaluación de riesgos a auditoría y certificaciones. Sobre este punto, se debe tener en consideración que la propia Comisión Europea para la Eficacia de la Justicia<sup>58</sup> ha aludido a la necesidad que estos programas sean certificados y sometidos a pruebas que acrediten su calidad y seguridad. En este sentido, la Carta ética aprobada el año 2018 reconoce expresamente el principio de calidad y seguridad, lo que deriva en la necesidad de que para el procesamiento de decisiones judiciales y de datos se usen “*fuentes certificadas y datos intangibles con modelos concebidos en un sistema multidisciplinario, en un entorno tecnológico seguro*”<sup>59</sup>. Precizando este tema, la Carta consigna que “*Los datos basados en decisiones judiciales que se ingresan en un software que implementa un algoritmo de aprendizaje automático deben provenir de fuentes certificadas y no deben modificarse hasta que realmente hayan sido utilizados por el mecanismo de aprendizaje. Por lo tanto, todo el proceso debe ser rastreable para garantizar que no se haya producido ninguna modificación para alterar el contenido o el significado de la decisión que se está procesando*”<sup>60</sup>.

Una manera de disminuir los problemas que produce la falta de fiabilidad de los datos, derivados del desconocimiento de su origen y la manera como son procesados, consiste en realizar auditorías y certificaciones que permitan conocer cómo se ha ejecutado el poblamiento de la información y su trazabilidad. Estas actuaciones permiten determinar que los antecedentes con los que se alimenta el modelo puedan ser confiables. Lo explicitado soluciona, al menos parcialmente, el problema que se produce en relación con la fiabilidad de los datos que sirven de base a un reporte, el que, a su vez, consigna una puntuación en relación con la posibilidad de reincidencia de una persona. Uno de los aspectos centrales de la auditoría y certificación debería centrarse en evitar o minimizar la existencia de sesgos vinculados a la información que se incorpora al sistema, lo que puede evitar la discriminación que contienen algunos reportes.

En cualquier caso, la auditoría y certificación interna es la primera parte y debería existir para efectos de garantizar la validez del programa de predicción y evaluación de riesgo que se fiscaliza. Adicionalmente, la información sobre auditorías y certificaciones debería ser puesta a disposición de los operadores del sistema de justicia criminal, como jueces y ministerio público; también ser entregada a los intervinientes. De esta forma no solo podrán estar informados de la validación de las herramientas, sino que de aquellos aspectos críticos que arroja el examen de estos.

---

<sup>58</sup> Esta Comisión fue creada el 18 de septiembre de 2002, a través de Resolución (2002)12 del Comité de Ministros del Consejo de Europa. Su rol es incrementar los niveles de eficacia en torno a los sistemas de justicia de los Estados miembros.

<sup>59</sup> Carta Ética Europea sobre el Uso de la Inteligencia Artificial en los Sistemas Judiciales y en su Entorno, de diciembre de 2018, pp. 8 y ss.

<sup>60</sup> Ibid.

- Los reportes emanados por los sistemas de predicción de riesgos no deberían ser elementos o medios de prueba que determinen la decisión del juicio penal. Ni siquiera deberían ser conocidos durante la etapa de juicio, esto es, antes que el órgano jurisdiccional arribe a alguna conclusión de absolución o de condena sobre el caso que conoce. Esto, porque el tribunal dicta la sentencia dando por acreditados ciertos hechos en relación con el delito y la participación de los acusados, de acuerdo con la prueba rendida durante el procedimiento y el reporte no debería ser considerado como uno de dichos medios.

En el mismo orden de ideas, los reportes de peligrosidad de reincidencia no deberían ser conocidos por el tribunal antes del veredicto condenatorio ni absolutorio porque, de otro modo, inducen al órgano jurisdiccional a inclinar sus decisiones en un sentido determinado, afectando su imparcialidad. Lo explicitado obedece a que, si llega información al juez sobre el comportamiento anterior del justiciable –por ejemplo, que señale que es peligroso porque ha cometido delitos anteriormente– dicho conocimiento puede influir en la decisión de la controversia jurídico penal. La sentencia no debería estar supeditada o influida por la forma de ser de un individuo, ni menos aún debido a la predicción de un comportamiento futuro posible, que pueda emanar de una herramienta de inteligencia artificial. Lo contrario equivale a la instauración de un sistema criminal que juzga a las personas por su forma de ser y no por sus actos, cuestión que es rechazada por el sistema de derecho penal liberal, donde el juicio de culpabilidad se vincula con un comportamiento típico y antijurídico, mas no por las características o juicios respecto de las personas sometidas a un sistema procesal penal.

Esto, además, estaría acorde con lo dispuesto en el artículo 343 inciso final<sup>61</sup> del Código Procesal Penal, que alude a la audiencia de determinación de pena. Al respecto, la disposición señala que solo una vez que el tribunal pronuncie el veredicto de condena, se discute y toma conocimiento de los antecedentes que son ajenos al hecho punible y que sirven para determinar la pena y su cumplimiento. Este sería el momento, no antes, en que el juez podría tener acceso a los reportes, ya que, de esta manera, se evita afectar su imparcialidad respecto del juicio de condena o de absolución.

Con posterioridad al juicio de culpabilidad, la utilización de estos reportes podría ser útil desde la perspectiva de la definición de los planes concretos de intervención que pueden elaborarse respecto de uno u otro imputado, en consideración a los fines de prevención especial y, en menor medida, de prevención general. Al respecto, el tribunal debería evidenciar qué otros elementos han tenido en consideración, además de este reporte, para evaluar el riesgo de reincidencia de cada justiciable, fundamentando su decisión. Resulta esencial que la resolución jurisdiccional se justifique más allá de lo recomendado por el programa de inteligencia artificial. Este punto será desarrollado en el siguiente lineamiento.

---

<sup>61</sup> Artículo 343 inciso final del Código Procesal Penal chileno: “En el caso de condena, el tribunal deberá resolver sobre las circunstancias modificatorias de responsabilidad penal en la misma oportunidad prevista en el inciso primero. No obstante, tratándose de circunstancias ajenas al hecho punible, y los demás factores relevantes para la determinación y cumplimiento de la pena, el tribunal abrirá debate sobre tales circunstancias y factores, inmediatamente después de pronunciada la decisión a que se refiere el inciso primero y en la misma audiencia. Para dichos efectos, el tribunal recibirá los antecedentes que hagan valer los intervinientes para fundamentar sus peticiones, dejando su resolución para la audiencia de lectura de sentencia”.

- Los intervinientes en el proceso penal y, particularmente, el imputado y su defensor, deberían tener la posibilidad de conocer, comprender y revisar los fundamentos de las decisiones jurisdiccionales, para lo cual resulta, también, fundamental que se encuentre en condiciones de entender el funcionamiento del sistema algorítmico y los datos que utiliza para dar lugar al reporte que podría sustentar dichas resoluciones judiciales.

Lo afirmado conlleva dos implicancias: en primer lugar, respecto de la fundamentación de la resolución judicial que se basa en la información contenida en el reporte. Debe explicitarse por qué y en qué sentido este antecedente puede ser relevante para la decisión que está adoptando dicho magistrado. Además, esta información no debería ser la única que justifique la decisión jurisdiccional sino, por el contrario, debería ir acompañada por otros antecedentes que la sostengan, a fin de evitar la mecanización en torno a este tipo resoluciones.

En segundo lugar, es relevante que pueda transparentarse la forma de funcionamiento del sistema algorítmico y el proceso para arribar a sus reportes. En el mismo sentido, la Carta Blanca Ética consagra el principio de transparencia, imparcialidad y equidad, cuyo postulado es que el mecanismo sea accesible y comprensible. Junto con ello, también se sugiere otorgar la posibilidad de que los programas sean revisados a través de auditorías externas<sup>62</sup>. La auditoría externa permite conocer en detalle el funcionamiento de la herramienta informática. Esto, a su turno, permite a los intervinientes controvertir los reportes emanados de los sistemas de predicción y, con ello, ejercer adecuadamente el derecho a defenderse en relación con el contenido de dicho antecedente. El ejercicio del nombrado derecho a defensa supone que el acceso al antecedente sea completo, en el sentido que, al examinarse, pueda reproducirse el procedimiento o la manera en que se arribó a un determinado resultado.

## 6. CONCLUSIONES

En la primera parte de este trabajo se revisa el concepto de sistema algorítmico de predicción de riesgos y evaluación. Junto con ello, se hace referencia a algunas funciones que estos pueden cumplir en los procesos judiciales, con especial énfasis en materia penal. En este orden de ideas, se pasa revista de los principales problemas que conlleva el empleo de estas herramientas cuando se utilizan para apoyar la toma de decisiones judiciales. En tal sentido, se constata que las mayores críticas se centran en el sesgo algorítmico, falta de transparencia y de explicabilidad. Posteriormente, se alude a las consecuencias que puede conllevar, esto es, afectaciones al debido proceso de los justiciables.

En la segunda parte, se desarrollan latamente los mencionados cuestionamientos, particularmente los efectuados por algunas defensas de imputados. En este contexto, se manifiestan los principales argumentos proporcionados por la doctrina y la jurisprudencia, a la luz de ciertos casos calificados de emblemáticos durante este estudio. A continuación,

---

<sup>62</sup> Carta Ética Europea sobre el Uso de la Inteligencia Artificial en los Sistemas Judiciales y en su entorno, de diciembre de 2018. p. 9.

se sistematizan las principales afectaciones que los problemas consignados pueden producir al debido proceso, mencionándose los siguientes: 1) afectación del derecho a defensa y a prepararla adecuadamente; 2) tutela judicial efectiva y 3) imparcialidad de juzgador.

Finalmente, se mencionan algunos lineamientos generales para colaborar en la superación de las consecuencias negativas que puede conllevar para los justiciables el uso de estos programas de inteligencia artificial durante el procedimiento penal. Al respecto, se alude a la necesidad de someter a los sistemas de predicción y evaluación de riesgos a auditorías y certificaciones, toda vez que la revisión y control de estos redundaría en una mayor fiabilidad de los datos que les sirven de soporte, con lo que también es factible reducir los riesgos derivados de la discriminación algorítmica. En segundo lugar, se plantea que los reportes o informes que emanan de estos sistemas no puedan ser usados como medios de prueba para ser presentados durante el juicio contradictorio. Ello, porque su conocimiento por parte del órgano encargado de decidir la controversia jurídico penal puede afectar su imparcialidad, al proporcionársele información sobre la historia delictiva del imputado y una prognosis de comportamiento futuro, lo que, claramente, excede los hechos y la participación del individuo en el delito que es objeto del procedimiento. Lo contrario podría favorecer un derecho penal de autor, por sobre uno de acto. En tercer lugar, se enfatiza en la necesidad de permitir a los intervinientes tener acceso al sistema de inteligencia artificial para conocer cómo funciona y cuáles son sus mecanismos y, junto con ello, permitir las auditorías externas para controvertir el funcionamiento del mismo.