

**UNIVERSIDAD DE SAN ANDRÉS (UDES)**  
**MAESTRÍA EN DERECHO PENAL**

**TESIS DE MAESTRÍA**

*Maestranda: Noelia Victoria Mitelli*

*DNI: 33.032.738*

*Tutor: Martín D. Haissiner*

**IA y derecho penal: criterios para la utilización de asistentes jurídicos digitales en el ámbito de la justicia**

**Resumen:**

Este trabajo explorará las problemáticas de aplicación de sistemas predictivos de inteligencia artificial al ámbito de la justicia penal argentina. Una vez identificadas se abordarán posibles soluciones a dichas problemáticas con el fin de delimitar diversos criterios y recomendaciones de aplicación de estos asistentes jurídicos digitales al proceso penal. Todo ello a fin de garantizar que, sin perjuicio de que los magistrados hagan uso de esta tecnología para apoyar sus decisiones, se respeten los derechos fundamentales de los imputados sometidos a proceso.

Con el presente trabajo se busca, en definitiva, realizar un aporte novedoso al área del derecho penal al advertirse que, en materia de inteligencia artificial aplicada al ámbito de la justicia, obran enormes lagunas legales que no hacen sino perjudicar a las personas que se encuentran bajo la mira del derecho penal. A la luz de esta reflexión se pretende brindar criterios de aplicación que guíen a los magistrados a la hora de incorporar este tipo de tecnología a su trabajo jurisdiccional.

**Índice:**

<b>I. – Introducción: conceptos iniciales</b>	Pág. 1
<b>II. Primera parte: los puntos de tensión</b>	Pág. 7
<i>II. A. – Igualdad ante la ley</i>	Pág. 8

<i>II. B. – Imparcialidad</i>	Pág. 12
<i>II. C. – Defensa en juicio, juicio previo y principio de inocencia.</i>	Pág. 16
<i>II. D. – Transparencia</i>	Pág. 21
<i>II. E. – Autoincriminación forzada</i>	Pág. 23
<i>II. F – Principio in dubio pro-reo</i>	Pág. 25
<i>II. G. – Derecho penal de autor</i>	Pág. 27
<i>II. H. – Nulidades</i>	Pág. 28
<b>III. – Segunda parte: la normativa</b>	Pág. 28
<i>III. - Marcos legales</i>	Pág. 29
<i>III. 1 - El Libro Blanco de la UE</i>	Pág. 30
<i>III. 2 - La propuesta del Parlamento Europeo del año 2020</i>	Pág. 31
<i>III. 3 - El proyecto de Ley de Responsabilidad Algorítmica en los EE. UU.</i>	Pág. 32
<b>IV.- Tercera parte: las soluciones</b>	Pag. 33
<i>IV. A. – Criterios para la incorporación de IA al ámbito de la justicia penal</i>	Pág. 33
<i>IV. A. 1.- Transparencia</i>	Pág. 33
<i>IV. A.2.- Ética en el uso de los datos</i>	Pág. 36
<i>IV. A.3.- Sesgos y discriminación algorítmica</i>	Pág. 37
<i>IV. A.4.- Auditorias Algorítmicas</i>	Pág. 44
<i>IV. A.5.- Niveles de riesgo de la IA</i>	Pág. 45
<i>IV. B.- Criterio de incorporación de la IA al proceso penal para el juez</i>	Pág. 46
<b>Conclusión</b>	Pág. 53
<b>Bibliografía</b>	Pág. 54

## **I. – Introducción: conceptos iniciales**

La tecnología evoluciona a gran velocidad y trae consigo desafíos e interrogantes que atraviesan todas las áreas del conocimiento. A ellos se enfrenta el derecho e intenta dar respuesta a los problemas jurídicos surgidos de lo que popularmente se conoce como la era de los datos.

Este nuevo paradigma está marcado por el advenimiento del “big data” y de la Inteligencia Artificial (en adelante IA). Esta es la disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente

humana (como el aprendizaje o el razonamiento lógico)<sup>1</sup>. Son sistemas que manifiestan un comportamiento inteligente pues son capaces de analizar su entorno y pasar a la acción con el fin de alcanzar objetivos específicos<sup>2</sup>.

La IA se alimenta del “big data” – información de alto volumen que es recopilada a muy alta velocidad –la cual es procesada mediante algoritmos que se consolidan como el verdadero corazón de este tipo de tecnología.

Ahora bien, ¿cómo podríamos definir el concepto de algoritmo?. Harari entiende que son un conjunto metódico de pasos que puede emplearse para hacer cálculos, resolver problemas y alcanzar decisiones.<sup>3</sup>

Un algoritmo, en resumidas cuentas, es un conjunto de instrucciones o reglas definidas, no ambiguas, ordenadas y finitas que permite solucionar un problema, realizar un computo, procesar datos y llevar a cabo otras tareas o actividades<sup>4</sup>. Dado un estado inicial de entrada el algoritmo sigue una serie de pasos sucesivos a fin de arribar a un estado final y arrojar la solución a determinado problema.

La IA trabaja sobre grandes volúmenes de datos (“big data”) sobre los que aplica algoritmos y, mediante ellos, identifica patrones de comportamiento, se entrena, aprende de esa información y logra arribar a la predicción de comportamientos futuros. Entonces, podríamos afirmar que los pilares de la IA son los algoritmos, las redes neuronales artificiales y los patrones de razonamiento similares a los humanos.

Se asocia también a este tipo de sistemas con el aprendizaje profundo (“Deep learning”) el cual esta inspirado en la forma en aprende el cerebro humano. De esta manera, al sistema se le brindan datos que pasan por distintas capas o niveles a los cuales se les aplican reglas de aprendizaje a fin de que la maquina evalúe ejemplos e instrucciones y arribe a resultados que se comparan y se ajustan entre si para perfeccionarlo.

Una distinción clásica en materia de IA es la de IA “débil” y “fuerte”. La primera de las categorías se aplica a sistemas mas o menos complejos cuyo objetivo es resolver un problema pero que se encuentran lejos de mostrar cualquier síntoma revelador de un estado cognitivo ya que solo pueden hacer aquello para lo que fueron específicamente programados

---

<sup>1</sup> Diccionario de la Real Academia Española, disponible al 23/11/2020 en <https://dle.rae.es/inteligencia>

<sup>2</sup> COMISIÓN EUROPEA. (2018). *IA para Europa. Comunicación de la Comisión al Parlamento europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones*. COM(2018) 237 final SWD(2018) 137 final. Bruselas, 25.4.2018, p. 1.

<sup>3</sup> HARARI, Yuval Noah, *Homo Deus Breve Historia del Mañana*, Ed. Debate, 2015, Buenos Aires, pág. 100.

<sup>4</sup> URUEÑA, René. Autoridad algorítmica:¿ cómo empezar a pensar la protección de los derechos humanos en la era del “big data”? *Latin American Law Review*, 2019, no 2, p. 99-124.

mediante la codificación de reglas precisas. La segunda categoría alude a la capacidad de un sistema para imitar ciertos procesos cognitivos o capacidades generales del pensamiento humano pudiendo además mejorar continuamente su rendimiento a medida que realizan dichas operaciones y aprender de su propia experiencia.

Ahora bien, ¿Qué relación tiene todo esto con el derecho penal?.

Hace varios años la IA ingreso al ámbito de la justicia y, en la actualidad, es utilizada por algunos tribunales como un asistente jurídico del magistrado. Ello en virtud de que brinda la posibilidad de automatizar procesos manuales y, así, reducir el riesgo de que se produzcan errores materiales. Sumado a ello, permite aumentar la eficacia de los tiempos de trabajo.

Parte de la doctrina se refiere a estos sistemas como verdaderos asistentes del juez que lo auxilian en su tarea de impartir justicia en aquellos actos mecánicos donde se requiera una herramienta predictiva o un instrumento que proponga una solución a determinado conflicto<sup>5</sup>.

Pero los sistemas de IA en el ámbito de la justicia no siempre se circunscriben a meras tareas mecánicas o automáticas sino que existen programas que inciden en el dictado de la sentencia que decidirá el destino de la persona sometida a proceso. Programas como COMPAS o PROMETEA son de uso frecuente por los tribunales y han sido adoptados como herramientas que forman parte del proceso que sigue el magistrado a la hora de dictar sentencia.

Sumado a ello, los abogados no están exentos del avance de la IA. Existen aplicaciones que intentan asistir tanto a la judicatura como al letrado litigante.

Un ejemplo interesante es el software Ross, que permite realizar búsquedas de una manera sumamente sencilla y responde consultas como si fuera el propio profesional. Tiene una copiosa base de datos que incluye doctrina, jurisprudencia y leyes y, además, se actualiza en forma permanente<sup>6</sup>.

Otro caso interesante es el software MarginMatrix, el cual puede analizar las normas de distintas jurisdicciones y, con ello, redactar documentos personalizados. Entre sus funciones se encuentra la de asesorar a entidades bancarias sobre cómo adaptarse a los

---

<sup>5</sup> CABEZA DE VACA HERNANDEZ, Daniel. “*LA aplicada a la justicia*”.

<sup>6</sup>Ross Intelligence, disponible al 14/11/19 en <https://rossintelligence.com/features.html>.

cambios regulatorios que puedan existir en materia contractual. Y por supuesto, garantiza la máxima rapidez y efectividad para realizar dichas tareas<sup>7</sup>.

Otro ejemplo es el software Premonition<sup>8</sup> que se usa en los Estados Unidos para predecir la orientación que van a tener las sentencias de los jueces. Se trata de un sistema donde se cargan diversas variables de datos relativas a criterios judiciales que poseen salas, juzgados o cámaras y, mediante estos datos, el sistema genera una potencial sentencia para orientar como es que podría fallar ese tribunal. Al tener en cuenta los fallos de los juzgados, dependiendo de cómo resulte el sorteo, el abogado sabe con que argumentos tiene mas posibilidades de ganar un litigio<sup>9</sup>.

Aunque en EE. UU. el uso de Premonition es cada vez mayor en Francia la utilización de IA para predecir sentencias judiciales esta prohibida.

Allí no se prohíbe el uso de IA en general sino que se criminaliza la posibilidad de analizar, comparar o hacer predicciones sobre la practica actual o futura de los jueces. La prohibición está ligada a la no neutralidad de las sentencias judiciales debido al uso de este tipo de tecnología<sup>10</sup>.

Ahora bien, tiempo atrás seguramente resultaba impensado que la IA pueda utilizarse como una herramienta de los estrados judiciales pero actualmente, es una idea que se afianza con ahínco.

El Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires y el Laboratorio de Innovación e IA de la Facultad de Derecho de la Universidad de Buenos Aires desarrollaron PROMETEA. Un sistema de IA que combina detección y asistencia inteligente, predicción y automatización. Funciona gracias a técnicas de aprendizaje automático (machine learning) supervisado y de clustering, a partir del etiquetado manual y de máquina<sup>11</sup>.

En sus comienzos, PROMETEA se utilizó en el ámbito del Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires y, a través de un chat (por voz y escrito) interactuaba

---

<sup>7</sup>ALLEN & OVERY NEWS, “Allen & Overy and Deloitte tackle OTC derivatives market challenge”, recuperado de <http://www.allenoverly.com/news/en-gb/articles/Pages/AllenOvery-and-Deloitte-tackle-OTC-derivatives-market-challenge.aspx> citado por Escobar, Felicitas, “La IA como camino hacia un estándar de prueba objetivo”, ponencia ganadora del primer puesto en el concurso de ponencias del Congreso Colombiano de Derecho Procesal XL, realizado en la ciudad de Medellín los días 4, 5 y 6 de septiembre de 2019.

<sup>8</sup>“Los contratos inteligentes ponen en jaque al mundo judicial”, disponible al 21/11/19 en <http://www.telam.com.ar/notas/201610/167456-contratos-inteligente-mundo-judicial.html>

<sup>9</sup>“Los contratos inteligentes ponen en jaque al mundo judicial”, disponible al 21/11/19 en <http://www.telam.com.ar/notas/201610/167456-contratos-inteligente-mundo-judicial.html>

<sup>10</sup> CORVALAN, Juan Gustavo, “The french ban: the use of Artificial Intelligence to trace patterns of judicial behavior”

<sup>11</sup> PROMETEA, disponible al 21/11/2019 en <https://ialab.com.ar/prometea/>.

con el operador jurídico y, luego de la lectura de miles de casos, predecía un dictamen para el caso concreto.

Con el paso del tiempo fue ampliando sus fronteras y actualmente también es utilizado en la Corte Interamericana de Derechos Humanos y en la Corte Constitucional de Colombia.

12

Parte de la doctrina en la materia aborda el tema de la creación de una herramienta de IA que ayude al Juez en el difícil proceso de la valoración de la prueba. Así, en base al acervo probatorio recolectado en un determinado proceso, el sistema podría realizar una predicción en torno a si la hipótesis del magistrado concuerda con aquel o bien, si no lo hace, indicar una nueva<sup>13</sup>.

En base a ello, es posible que llevar un control tan exacto de toda la producción de la prueba ayude al momento de la valoración. De esta manera el juez tendría a su alcance toda la información que necesita para tomar una decisión lo más completa posible. La IA ayudaría a conformar un umbral de certeza objetivo basado en los datos empíricos recopilados a lo largo del proceso<sup>14</sup>.

Sin embargo, sin perjuicio de las ventajas que podría traer aparejada la incorporación de este tipo de tecnología a la justicia lo cierto es que asociados a este tipo de sistemas existen diversas problemáticas que van desde la privacidad algorítmica (esto es la imposibilidad de conocer la forma de razonar del sistema por cuestiones relativas al derecho de propiedad) hasta los sesgos raciales y de género.

El pretenseo científico de este tipo de tecnología otorga al magistrado una falsa sensación de seguridad respecto de los resultados que arroja. En este sentido, el juez tendrá la tendencia de fallar conforme lo sugiere el sistema pues la conclusión que este devuelve tendrá la “garantía” de ser fruto de un análisis de tipo matemático que la dota de un alto grado de precisión.

Sin perjuicio de lo expuesto, lo cierto es que estos sistemas son falibles y cometen errores de análisis que conllevan graves consecuencias para las personas que deben ser juzgadas en procesos que hacen uso de ellos.

---

<sup>12</sup> CORVALÁN, Juan Gustavo. “La primera IA predictiva: Prometea.” Recuperado de: <http://thomsonreuterslatam.com/2017/10/la-primera-inteligencia-artificial-predictiva-al-servicio-de-la-justicia-prometea/>. e “IA al servicio de la justicia penal, contravencional y de faltas. Prometea, en el ámbito de la Ciudad de Buenos Aires”, publicado en la obra de DUPUY Daniela; KIEFER Mariana. Cibercrimen II; Ed. B de F, 2018.

<sup>13</sup> NIEVA FENOLL, Jordi, IA y proceso judicial, Marcial Pons, Madrid, 2018, pág. 100/1.

<sup>14</sup> ESCOBAR, Felicitas, “La IA como camino hacia un estándar de prueba objetivo”, ponencia que obtuvo el primer puesto en el Concurso de Semilleros del XL Congreso Colombiano de Derecho Procesal, Medellín, 4, 5 y 6 septiembre de 2019.

Lo preocupante es que, sin perjuicio de que, tal como se expuso, en ciertos países este tipo de sistemas esta en pleno funcionamiento, ni en el ámbito local ni en el internacional existen criterios definidos desde lo legal que establezcan y delimiten que condiciones deben cumplirse para que el uso y aplicación de asistentes jurídicos digitales en el ámbito de la justicia sea posible sin vulnerar garantías fundamentales de las personas sometidas a proceso.

En otras palabras, con independencia de si el uso de este tipo de sistemas resulta o no beneficioso para la justicia es necesario reflexionar sobre la delimitación de criterios de aplicación al ámbito local a fin de garantizar que su uso no vulnere derechos fundamentales del imputado.

A la luz de lo expuesto hasta este punto, este trabajo explorará las problemáticas de aplicación de sistemas predictivos de IA al ámbito de la justicia penal argentina.

Una vez identificadas se abordarán posibles soluciones a dichas problemáticas con el fin de delimitar criterios y recomendaciones de aplicación de estos asistentes jurídicos digitales al proceso penal. Todo ello a fin de garantizar que, sin perjuicio de que los magistrados hagan uso de esta tecnología para apoyar sus decisiones, se respeten los derechos fundamentales de los imputados sometidos a proceso.

Con el presente trabajo se busca, en definitiva, realizar un aporte novedoso al área del derecho penal al advertirse que, en materia de IA, obran enormes lagunas legales que no hacen sino perjudicar a las personas que se encuentran sometidas a proceso. A la luz de esta reflexión se pretende brindar criterios de aplicación que guíen a los magistrados a la hora de incorporar este tipo de tecnología a su trabajo jurisdiccional.

## **II. Primera parte: los puntos de tensión**

La aplicación de IA al ámbito de la justicia penal genera puntos de tensión con las garantías constitucionales que rigen el proceso. Resulta beneficioso para el análisis de dichos puntos dividir el abordaje en distintos ejes temáticos a fin de visualizar con mayor claridad las problemáticas que la inclusión de este tipo de tecnología al proceso penal trae aparejadas.

A dicho fin, se trazara como línea de acción la identificación de la garantía que se vulneraria mediante la utilización de un sistema de IA en un proceso penal - sin la delimitación de ningún tipo de parámetros o criterios que orienten dicha incorporación - en contraposición con la característica central del sistema de IA que genera dicha vulneración.



## **II. A. – Igualdad ante la ley**

A nivel normativo, el artículo 16 de la Constitución Nacional, la Declaración Universal de los Derechos Humanos y el Pacto Internacional de Derechos Civiles y Políticos delimitan las bases del principio de igualdad al establecer que todos los habitantes son iguales ante la ley.

La igualdad se instala en el ordenamiento jurídico como principio y, a partir de ello, configura un derecho y una garantía.

En su condición de principio constituye una guía de acción y, al mismo tiempo, un mandato que debería conformar una directriz a la hora de aplicar el derecho.

En su condición de derecho habilita a los individuos dentro del sistema a oponerse frente a normas o actos discriminatorios violatorios de aquel principio (estatus negativo) o a exigir algún comportamiento determinado de los poderes públicos para subsanar dicha violación (estatus positivo).

Finalmente, en su condición de garantía constituye un presupuesto para la efectividad de las diversas libertades o derechos anclándose en el sistema jurídico constitucional para maximizar su tutela.

Retomando la primera de las acepciones, el concepto de igualdad es integrado por dos derechos: el de ser tratado como igual y el derecho a igual tratamiento, es decir, igualdad de derecho e igualdad de hecho o, en otras palabras, igualdad formal y material.

La igualdad de derecho (igualdad jurídica o igualdad ante la ley) se identifica con la igualdad formal o igualdad legal e implica que la ley o norma sea aplicada sin distinciones a aquellos a quienes está destinada o dirigida.

La igualdad material se refiere a la formulación del derecho en cuanto al contenido mismo y a las consecuencias. Apunta a la aproximación a las desigualdades reales, a las discriminaciones expresas o encubiertas y tiende a la búsqueda de medidas que garanticen la igualdad real de oportunidades como guía para la igualdad de derechos<sup>15</sup>.

Es que, el principio de igualdad ante la ley posee un carácter abierto en un doble sentido. En primer termino, desde el punto de vista histórico. Los distintos parámetros que

---

<sup>15</sup> Para mas información sobre el tema ver: CAYUSO, S. G. (2009). El principio de igualdad: problemas e interrogantes: el sistema constitucional argentino. *Revista de Derecho Político*, (75-76).



pueden utilizarse para dar un trato igual o desigual a las personas han variado con el tiempo. Tanto su significación concreta, como su significación abstracta, dependen de las estructuras políticas, jurídicas, económicas y sociales existentes, y de los valores socialmente aceptados en una época dada. Por esto es posible afirmar que las normas jurídicas que establecen una desigualdad o, por el contrario, suprimen una y proclaman una determinada igualdad, son la expresión de una desigualdad social, o bien de una reacción debida al cambio de las valoraciones sociales, que es a su vez el resultado, en gran parte, de una transformación de las estructuras sociales.

En segundo lugar, el carácter abierto del principio de igualdad derivado del aspecto relacional, se proyecta en el sentido de que, a partir de él, no es posible enumerar o hacer un listado de los rasgos que han de ser considerados irrelevantes y que, por tanto, no pueden ser tenidos en cuenta para dar un trato diferente a dos o más personas.

Como consecuencia, la introducción de cláusulas que prohíben la discriminación es verdaderamente trascendental en cuanto a la expresión del principio de igualdad. Por ello, en diversos instrumentos se ha incluido una cláusula antidiscriminatoria como una herramienta para proteger la igualdad de las personas<sup>16</sup>.

Ahora bien, este principio no podría tolerar jamás la existencia de sesgos a la hora de juzgar a una persona y es aquí donde entra en tensión con la IA y su uso en el ámbito del derecho penal. En estos sistemas la presencia de sesgos es frecuente en virtud de las bases de datos con las que fueron entrenados.

La IA no está exenta de errores humanos, a pesar de la creencia de que estos sistemas pueden pensar independientemente, aprender solos o ser completamente objetivos por su funcionamiento lógico matemático<sup>17</sup>. Su fuente de conocimiento viene de datos recolectados por seres humanos por lo que esta tecnología no es más objetiva que sus desarrolladores.

Los sesgos en los sistemas de IA se definen como errores, inexactitudes o anomalías que se presentan en los algoritmos con los que esta tecnología opera, pudiendo tener impactos no deseados en individuos y grupos, principalmente en términos de discriminación o

---

<sup>16</sup> FERNANDEZ SOSA, Macarena. EL PRINCIPIO DE IGUALDAD. Disponible en: <https://www.aiddp.com/wp-content/uploads/2021/10/Fernandez-Sosa-Macarena-Universidad-Nacional-de-Cuyo.pdf>

<sup>17</sup> MENDOZA, D. (2020). Racismo y roles de género, conductas perpetuadas en algoritmos de IA. COLOQUIO, (65), 131-135.

exclusión; de ahí que se considere que la existencia de algún tipo de sesgo en un sistema de IA determina si este es o no justo.<sup>18</sup>

Los sesgos en los sistemas de IA pueden darse principalmente por dos razones. La primera de ellas, por suposiciones o prejuicios que se introducen durante el diseño y desarrollo del algoritmo. Ello porque este es realizado por humanos que tienen sus propios sesgos cognitivos y sociales que terminan siendo integrados en los parámetros que se definen para el sistema. La segunda, por suposiciones o prejuicios sociales o históricos, introducidos durante la recolección o etiquetado de los datos que se emplean para entrenar el sistema, o por su falta de representatividad, es decir, la ausencia de datos.<sup>19</sup>

En ese sentido, los errores y sesgos que se podrían presentar en los algoritmos tienen su base en las ideas humanas y en los modelos sociales y prejuicios creados.

Los sesgos raciales y de género presentes en IA representan cómo funciona la sociedad en general y como hay áreas aún dominadas solo por ciertos grupos.

Ahora bien, la automatización de ciertos procesos podría representar un peligro cuando los datos que son utilizados para esto están sesgados; si un algoritmo se entrenó con un sesgo racial, va a seguir incorporando y aplicando este sesgo.

En este punto me pregunto, ¿se respeta el principio de igualdad ante la ley si la justicia utiliza como asistente digital un sistema de IA sesgado?.

Abordare el tema mediante un ejemplo. En los tribunales de Estados Unidos se utiliza el sistema COMPAS. Compas es un acrónimo que puede traducirse como Administración de Perfiles de Criminales para Sanciones Alternativas del Sistema de Prisiones de EE.UU (Correctional Offender Management Profiling for Alternative Sanctions). Este sistema tiene como objetivo determinar el nivel de riesgo de fuga que presenta la persona sometida a proceso. COMPAS fue entrenado a partir de treinta y ocho variables relacionadas con los antecedentes penales, laborales, académicos y habitacionales, incluyendo también aquellos vinculados al consumo problemático de sustancias.

El sistema utiliza los datos que obtiene de un formulario que se les da a las personas al momento de su detención y que contiene las variables detalladas en el párrafo anterior.

---

<sup>18</sup> VÉLEZ, M. I., GÓMEZ SANTAMARÍA, C., & OSORIO SANABRIA, M. A. (2022, June 30). Conceptos fundamentales y uso responsable de la IA en el sector público. Informe 2. Caracas: CAF. Disponible en: <http://cafscioteca.azurewebsites.net/handle/123456789/1921>

<sup>19</sup> VÉLEZ, M. I., GÓMEZ SANTAMARÍA, C., & OSORIO SANABRIA, M. A. (2022, June 30). Conceptos fundamentales y uso responsable de la IA en el sector público. Informe 2. Caracas: CAF. Disponible en: <http://cafscioteca.azurewebsites.net/handle/123456789/1921>

Cada una de las respuestas recibe un puntaje que varía de 1 a 10. El sistema luego genera un “promedio de riesgo” mediante el cual se determina si el imputado puede salir en libertad, debe ir a la cárcel o puede recibir otro tipo de tratamiento.

Quienes están a favor de COMPAS sostienen que mediante su aplicación se coteja toda la información del imputado colaborando en que las decisiones de los jueces sean menos subjetivas mediante la utilización de un algoritmo que permita cruzar la información.

Lo cierto es que diversas investigaciones mostraron que se trata de un sistema poco objetivo y parcial que favorecería a las personas caucásicas y daría un peor trato a las personas afroamericanas. Hablamos de un sesgo que genera que se produzca una discriminación racial agravando la condena para personas afroamericanas.

Sumado a ello, investigaciones llevadas adelante por Julia Dressel (Dartmouth College) demostraron que los algoritmos de COMPAS no son mucho más fiables que la predicción que puede realizar una persona cualquiera. Dressel utilizó la base de datos Pro Publica que posee puntuaciones de COMPAS a diez mil acusados en espera de juicio en Florida y sus registros de arresto para los dos años siguientes. Para su investigación seleccionó al azar mil de esos diez mil y reclutó a cuatrocientas personas a través Amazon Mechanical Turk, un servicio de crowdsourcing en línea para encontrar voluntarios de investigación.

A cada voluntario se le dieron cincuenta perfiles de acusados y se le solicitó que pronostiquen si esas personas volverían a cometer un delito en los próximos dos años (mismo funcionamiento que COMPAS). Los participantes acertaron casi tanto como el sistema de IA. Tanto los humanos como la máquina supusieron que las personas afroamericanas cometerían mayor cantidad de delitos en los próximos dos años que las caucásicas.

Los datos aportados por Pro publica revelaron que el algoritmo identificaba con mayor grado de peligrosidad a personas afroamericanas mientras que cataloga mayoritariamente de menor peligrosidad a personas caucásicas.<sup>20</sup>

Se dice que esta parcialidad se debe a la forma en que el sistema fue entrenado debido a que la IA aprende según los datos que se le brindan. Así, al conformar la base de datos inicial que alimentó a COMPAS se tomó una muestra no representativa de la sociedad, es decir, una

---

<sup>20</sup> La conclusión extraída de la investigación determino, según palabras de la directora de ProPublica que “si comparas a una persona negra y una blanca que tienen el mismo historial, la misma edad, el mismo género, el mismo pasado judicial y el mismo 'futuro criminal' (las posibilidades de cometer un crimen, dos crímenes o ninguno), el acusado negro tiene un 45% más de posibilidades de obtener un puntaje de riesgo que un acusado blanco. “*Machine Bias*” disponible al 21/11/2019 en <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

muestra parcializada. Así, conforme la muestra utilizada para entrenar al sistema, la franja de población afroamericana seleccionada cometió mayor cantidad de delitos que la franja de población caucásica seleccionada, entonces el sistema tiende a evaluar que las personas afroamericanas revisten una peligrosidad mayor que las caucásicas. De esta manera, las preguntas que se le hacen al imputado al momento de su detención y que luego pasan al sistema en forma de datos conducen a la parcialidad.

COMPAS es solo un ejemplo entre muchos pues uno de los principales problemas que poseen los sistemas de IA es la presencia de sesgos que arrastran desde su entrenamiento inicial y que condicionan sus conclusiones.

Conforme lo hasta aquí expuesto, resulta evidente que las personas no se encuentran en pie de igualdad ante la ley si desde los tribunales que se encargan de juzgar sus acciones se utiliza como asistente del magistrado un sistema de IA sesgado que genera a priori una discriminación (en este caso racial) que incidirá en la posibilidad, por ejemplo, de que el imputado transite el proceso en libertad.

## ***II. B. – Imparcialidad***

Ligada al principio de igualdad ante la ley, se halla una garantía que ampara al imputado en el proceso penal, esto es, la imparcialidad judicial.

La palabra “imparcial” refiere a aquel que no es parte en un asunto y debe decidir sin interés personal alguno. Asimismo, el concepto refiere semánticamente a la ausencia de prejuicios a favor o en contra de las personas o de la materia acerca de las cuales debe decidir el magistrado.

Ahora bien, el calificativo imparcialidad aplicado a la tarea de un juez no puede ser tomado como un ideal sino como una serie de previsiones que se han definido con la palabra neutralidad y que se rigen por tres máximas puntuales: la independencia de los jueces de todo poder estatal que pueda influir en la consideración del caso, la llamada imparcialidad del caso (determinada por la relación que establece el juzgador con el caso concreto) que también se denomina temor o sospecha de parcialidad del juez y el principio de juez natural.<sup>21</sup>

En lo que a este trabajo interesa me centrare en la arista concerniente a lo que se denomina temor o sospecha de parcialidad. Esta exigencia tiene dos dimensiones diferentes.

---

<sup>21</sup> MAIER, Julio B. J., *Derecho procesal penal, Tomo I...*, p. 764

La primera de ellas, llamada imparcialidad subjetiva, requiere que el juzgador carezca de un interés privado o personal en el resultado del caso<sup>22</sup>. En cambio, la otra dimensión, denominada imparcialidad objetiva, exige que el juez carezca de una inclinación hacia una de las soluciones posibles por razones relacionadas con el rol desempeñado durante el transcurso del proceso mismo<sup>23</sup>.

Para que se configure un supuesto de falta de imparcialidad subjetiva esta debe ser acreditada por la parte que la alega<sup>24</sup>.

En cambio, la ausencia de imparcialidad objetiva se tiene por configurada con el mero acontecimiento de una intervención en la cual el juzgador haya podido formar un prejuicio sobre el asunto que debe resolver. Esta dimensión de la garantía busca evitar que surja un temor de parcialidad respecto a la capacidad de decisión del juzgador<sup>25</sup>.

Ahora bien, los algoritmos de IA no aprenden solos sino que dependen de la información y bases de datos que se les proporcione, por lo que representarían de cierta manera el pensamiento y la realidad de sus creadores. En este sentido se advierte que la IA desde su inicio no es imparcial sino que presenta un sesgo cognitivo, que condiciona su capacidad de decisión. Este tipo de sesgo incluye estereotipos que existen en la sociedad y en la percepción propia del programador.

Lo hasta aquí expuesto remarca la falta de objetividad que caracteriza a un sistema de IA. Me pregunto entonces que ocurre si este sistema influye a la hora de que un juez resuelva. Me cuestiono que ocurre si, aun no siendo determinante, este sistema ofrece al juzgador o instala en su pensamiento una posible solución al caso que ya desde un inicio no es objetiva porque obedece a sesgos que arrastra de su entrenamiento. Incluso, la sola idea de que un sistema, aun no sesgado, instale en la mente del juez una solución del caso previo al desarrollo del juicio, ¿no es acaso condicionarlo?.

El Código Procesal Penal de la CABA establece como uno de sus principios rectores el de la oralidad. En este sentido, el juicio oral se caracteriza por el hecho de que el magistrado debe tomar una decisión única y exclusivamente con las pruebas que oralmente las partes produzcan al momento del debate (con excepción por supuesto de aquellas que hayan sido

---

<sup>22</sup> FERRAJOLI, Luigi; op. cit., p. 581

<sup>23</sup> MAIER Julio B. J.; *Derecho procesal penal, Tomo I...*, p. 755

<sup>24</sup> PIQUÉ, María Luisa y FERNÁNDEZ VALLE, Mariano; *La garantía de imparcialidad judicial desde la perspectiva de género*, publicado en HERRERA, Marisa, FERNÁNDEZ, Silvia E., DE LA TORRE, Natalia (Directoras generales); *Tratado de géneros, derechos y justicia. Derecho penal y sistema judicial. Tomo I*, 1º edición, Rubinzal Culzoni, Santa Fe, 2020, p. 124

<sup>25</sup> *Ibidem*, p. 125

incorporadas para su lectura). En este sentido, el magistrado que intervenga en la etapa de juicio tiene prohibido tomar contacto con el legajo de investigación producido en el etapa anterior, esto es, en la etapa de investigación penal preparatoria.

En este tipo de sistemas que son de tipo acusatorio y se caracterizan por la plena oralidad, ¿Qué pasaría si se admitiera que el magistrado, al momento del debate, puede ser asistido por un software de IA que prediga la peligrosidad futura del imputado? ¿Podría dicho dictamen ser admitido como prueba en el proceso? ¿Estaría contaminando al magistrado la solución al caso que sugiere un sistema de IA?

Desde mi punto de vista, en un sistema acusatorio como el de la CABA existiría un serio problema si el magistrado previo al debate o incluso durante la sustanciación de este tomara contacto con el dictamen emitido por un sistema de IA. Básicamente porque la conclusión que este sistema emita sobre el imputado condicionará la visión del magistrado quien ya no estará posicionado desde la neutralidad a la hora de escuchar las exposiciones de las partes. Ello fundado en el halo de certeza matemática del que todo sistema informático goza lo cual le da una falsa sensación de credibilidad y podría generar en el magistrado la confianza suficiente como para apoyar su resolución en lo dictaminado por la IA.

La Corte IDH en el caso “Manuela y otros Vs. El Salvador”<sup>26</sup> establece un estándar que puede resultar de utilidad. Dicho organismo dejó asentado que el uso de estereotipos resulta violatorio de la garantía de imparcialidad cuando se recurre a ellos para complementar la evidencia con la que se contaba a fin de llegar a una decisión determinada, por lo que los juzgadores no pueden recurrir a estereotipos sin sustento estadístico para resolver sobre cuestiones de hecho.

El problema, en gran cantidad de ocasiones y particularmente en lo que atañe al derecho penal, justamente radica en que las personas de las cuales se obtuvo la información que compone la base de datos que alimenta el sistema de IA no constituyen una muestra representativa de la población general, dado que sus particularidades demográficas no obedecen a un factor aleatorio sino a la forma en que esos individuos son captados por las agencias judiciales y policiales. Por lo tanto, la herramienta de IA termina arribando a resultados sesgados producto de la selectividad de las agencias del sistema penal.

---

<sup>26</sup> CORTE IDH, caso “Manuela y otros Vs. El Salvador”, (Excepciones Preliminares, Fondo, Reparaciones y Costas), Sentencia del 2 de noviembre de 2021, parr. 155.



Desde la perspectiva de igualdad ante la ley y la garantía de imparcialidad estas deficiencias resultan determinantes ya que los juzgadores, al consultar estas herramientas sesgadas, se expondrían a la información contenida en un pronóstico basado en un estereotipo.

Al tomar contacto con ese resultado el juzgador ya fue expuesto al razonamiento que contiene ese estereotipo afectando de esta manera su capacidad de resolver la controversia libre de prejuicios.

De acuerdo a los términos que desarrollamos, estaríamos ante una afectación a la garantía de imparcialidad en su faz objetiva, ya que la capacidad del juzgador de abordar el objeto de la decisión libre de preconceptos no se ve comprometida por un interés personal en el resultado del proceso, sino por haber sido expuesto a un estereotipo presente en la herramienta de IA que fuera introducida al proceso penal como un asistente del magistrado.

En este caso, el uso de la IA constituye un elemento objetivo que sustenta la posible existencia de un temor de parcialidad por parte del imputado, ya que el juzgador toma contacto anticipado con una conclusión sobre su persona que puede tener una incidencia en una futura decisión que afecte su libertad. Este dato predispone al juez a resolver en base a generalizaciones infundadas sobre ciertos datos sesgados que condicionan su capacidad de decidir el caso de acuerdo a los elementos objetivos que fueron válidamente admitidos al proceso y producidos en el juicio.

Aquí, lo que resulta relevante, entonces, es la predisposición que tendría el juez de resolver de cierta manera respecto de los miembros de ciertos grupos sin que exista una base cierta para efectuar esa diferenciación.

Es claro que la utilización de algoritmos judiciales en el “iter” del proceso como técnica o prueba aportada por la acusación influirá en la convicción del juez respecto del caso concreto que está obligado a valorar y lo hará en un doble sentido.

Por un lado, si el algoritmo presenta sesgos el resultado de su análisis afectará a la imparcialidad del juzgador en el momento de dictar la sentencia porque su convicción judicial estará condicionada al sesgo que presente la herramienta de IA.

Por el contrario, si se consigue que la herramienta de IA que se use tenga un alto grado de fiabilidad y, por lo tanto, sea admisible en el proceso penal, igualmente afectará a la convicción judicial; en este caso, en el momento de valoración de la prueba por la alta



posibilidad de que el juez tenga confianza en el resultado que brinda un sistema que por su forma lógico-matemática de razonar parecería tener certezas absolutas sin margen de error.<sup>27</sup>

Por estas razones es importante determinar criterios de aplicación de este tipo de tecnología al proceso, para garantizar a la persona sometida a el que será juzgada en igualdad de condiciones y sin temor alguno de parcialidad.

### ***II. C. – Defensa en juicio, juicio previo y principio de inocencia.***

El principio de inocencia surge de los artículos 18, 19 y 33 de la Constitución Nacional. Para que un sujeto sea culpable, se debe llevar adelante un juicio previo, basado en una ley anterior al hecho del proceso, debe haber acusación, defensa, prueba y sentencia. El debido proceso tiene como eje a un sujeto inocente.

Este principio también está receptado, en los diversos códigos procesales penales, en el artículo 8.2 de la Convención Americana sobre los Derechos Humanos y 14.2 del Pacto Internacional de Derechos Civiles y Políticos.

En resumidas cuentas, esto significa que toda persona es inocente y así debe ser tratada mientras no se declare en una sentencia judicial su culpabilidad.

De manera evidente, ese juicio debe estar fundado en la ley penal previa al hecho que se juzga (principio de legalidad en materia penal).

La sentencia penal debe ser fundada para ser válida. Se entiende por fundar una sentencia no solo la expresión de las premisas del juicio, las circunstancias de hecho verificadas y las reglas jurídicas aplicables sino también la exposición de las razones de hecho y de derecho que justifican tal decisión.

Entonces, una sentencia está fundada, cuando menciona los elementos de prueba a través de los cuales arriba racionalmente a una determinada conclusión fáctica. Dichos elementos han sido válidamente incorporados al proceso y son aptos para ser valorados (legitimidad de la valoración).

Sumado a ello, la sentencia exterioriza la valoración probatoria, esto es, contiene la explicación del porque de la conclusión siguiendo las leyes del pensamiento humano (principios lógicos de igualdad, contradicción, tercero excluido y razón suficiente) de la

---

<sup>27</sup> Miguel BerIaIn, I., & Pérez Estrada, M. J. (2019). La IA en el proceso penal español: un análisis de su admisibilidad sobre la base de los derechos fundamentales implicados.

experiencia y de la psicología común.<sup>28</sup> De no estar fundada la sentencia podrá ser tachada de arbitraria y ser plausible de ser revocada por los órganos de revisión.

Ahora bien, la ley fundamental supone también un procedimiento previo a una sentencia tal que le procure elementos de decisión al tribunal respecto de la imputación deducida, esto es, los elementos que le permitan construir, sobre todo, la premisa fáctica en la que apoyara su resolución aplicando la ley penal o prescindiendo de su actuación.

En la misma línea, rever la decisión final de un proceso se presenta como la posibilidad de que lo decidido sea sometido a una doble seguridad, como límite al ejercicio del poder estatal en el caso concreto y como garantía de racionalidad y eficacia en la adopción de decisiones jurisdiccionales<sup>29</sup>.

En esta dirección, el derecho a recurrir el fallo de condena debe necesariamente incluir la discusión de los medios de conocimiento que le dieron sustento.

Debemos tener en cuenta entonces, que en nuestro sistema de valoración de la prueba -según la sana crítica racional- el razonamiento se caracteriza porque el juez es quien fija las máximas de la experiencia según las cuales le otorga o no credibilidad a un determinado medio de prueba.

Así, hay que tener en cuenta que la posibilidad de impugnar las resoluciones jurisdiccionales constituye una derivación esencial del derecho de defensa en juicio, dado que implica someter al control de legalidad a diversa cantidad de actos desarrollados por la totalidad de las autoridades públicas que pueden llegar a intervenir dentro de un proceso penal, así como también se pueden valorar la posible afectación o menoscabo de derechos en el desarrollo de la pesquisa.<sup>30</sup>

Asimismo, la ley fundamental impide que se trate como si fuera culpable a la persona a quien se le atribuye un hecho punible hasta tanto el Estado no pronuncie una sentencia penal firme que declare su culpabilidad y lo someta a una pena.

A partir de lo expuesto con anterioridad es interesante analizar que esta ocurriendo con el uso de IA en el proceso penal.

---

<sup>28</sup> MAIER, Julio B. J., *Derecho procesal penal, Tomo I...*, p. 482

<sup>29</sup> CLARIA OLMEDO, JORGE A., *Derecho procesal penal, Tomo I*, Buenos Aires, Rubinzal-Culzoni, 1998, págs. 273 a 275.

<sup>30</sup> La Rosa, Mariano. “Defensa en juicio, debido proceso y juicio por jurados”. Disponible en: <https://www.pensamientopenal.com.ar/system/files/2016/02/doctrina42882.pdf>

En algunos países, desde hace algunos años existe una tendencia en la fase de determinación de la pena. Bajo la denominación de “evidence-based sentencing” o “smart sentencing”, intentan incorporar las valoraciones de riesgo que emiten los sistemas de IA como uno más de los criterios a tener en cuenta a la hora de establecer tanto el tipo como la cuantía de la pena, ello con el objetivo último de reducir la reincidencia.

Este planteamiento se fundamenta en varias premisas: la primera, que el objetivo de evitar la reiteración delictiva es uno de los fines legítimos del castigo penal; la segunda, que la impresión del juez sobre el grado de peligro que presenta un sujeto es intuitiva y por ello arbitraria, subjetiva y sesgada; y por último, que estos sistemas proporcionan conocimientos rigurosos y empíricamente fundamentados acerca de dicho riesgo. En consecuencia, los resultados de la valoración estructurada del riesgo de reincidencia deben ser tenidos en cuenta en las sentencias porque constituyen evidencia científica de mejor calidad sobre la peligrosidad del sujeto que la impresión personal que pueda tener el juez.

Un caso paradigmático que sigue esta línea de pensamiento y refiere al uso de IA en el marco de un proceso penal se dio en el año 2016 cuando Eric Loomis fue condenado a prisión debido a que no paso la evaluación de COMPAS.<sup>31</sup>

Loomis había sido acusado de cinco delitos por su presunta intervención en un tiroteo efectuado desde un vehículo (dos cargos por posesión de armas de fuego, intento de evasión frente a un agente de tráfico, puesta en peligro de la seguridad pública, y conducción de un vehículo ajeno sin autorización). El acusado negó su participación en el tiroteo aunque admitió haber conducido el vehículo con posterioridad a los hechos, y llegó a un acuerdo con la Fiscalía para aceptar los dos últimos cargos y rechazar el resto de acusaciones.

El 13 de julio de 2016 el Tribunal Supremo del Estado de Wisconsin dictó sentencia en el caso *State v. Loomis*. El juez aceptó la conformidad y al dictar la pena tuvo en cuenta, entre otras cosas, el contenido de un “presentence investigation report”, que incluía entre otros datos el resultado de la evaluación del sujeto mediante COMPAS: Loomis había sido identificado como “un individuo de alto riesgo para la comunidad” por el sistema de IA.

Según COMPAS Loomis presentaba un riesgo alto tanto de reincidencia en general como de reincidencia violenta. El juez condenó a Loomis a seis años de prisión y otros 5 de supervisión post penitenciaria, y en la sentencia explicó que a la hora de valorar los distintos

---

<sup>31</sup> *State v. Loomis*. 881 N.W. 2d 747.753, 2016.

factores, excluyo la suspensión del proceso a prueba debido a la gravedad del crimen y porque su historial en materia de supervisiones judiciales, y las herramientas de valoración del riesgo que aplicaron indicaban que Loomis presentaba un riesgo extremadamente alto de reincidir.

Loomis solicitó al tribunal que revisara la cuantía de la pena, argumentando que haberla fundamentado en el resultado de COMPAS había vulnerado su derecho al debido proceso. Pero el tribunal no modificó la condena y explicó además que sólo había utilizado la valoración de riesgo para corroborar las conclusiones a las que ya había llegado a partir del resto de información disponible, añadiendo que habría impuesto exactamente la misma pena aunque no hubiera tomado en consideración el resultado de COMPAS.

Loomis recurrió entonces ante el TSW alegando que el uso de COMPAS para determinar la pena había vulnerado su derecho a un proceso con todas las garantías por tres razones: en primer lugar porque se había violado su derecho a ser condenado sobre la base de información fiable y precisa, ya que la naturaleza de secreto comercial del algoritmo con el que funciona COMPAS le había impedido conocer cómo se habían calculado las estimaciones de riesgo y por tanto refutar la validez científica del nivel de riesgo que dicha herramienta le había asignado; en segundo lugar porque se había vulnerado su derecho a obtener una sentencia individualizada, y en tercer lugar porque se había tenido en cuenta indebidamente su género para determinar la pena.<sup>32</sup>

La Corte Suprema de Wisconsin determinó que el uso de COMPAS de forma adecuada el sistema no viola los derechos de los acusados. Resolvió, sin embargo, que “en el futuro en el análisis de riesgo que realiza el sistema deberán aclararse diversas cuestiones respecto de sus limitaciones: detallar que el algoritmo será mantenido en secreto y que el puntaje de riesgo está basado en cómo las personas en ciertos ambientes tienden a comportarse en general y no en una persona en particular”.<sup>33</sup>

Ahora bien, a la luz de este caso es interesante visualizar como la aplicación de este mismo supuesto a nuestro ordenamiento jurídico traería aparejada la violación los principios de defensa en juicio, juicio previo y de inocencia.

---

<sup>32</sup> Martínez Garay, L. (2018). Peligrosidad, algoritmos y due process: el caso State v Loomis. *Revista de Derecho Penal y Criminología (uned)*, 2018, num. 20, p. 485-502.

<sup>33</sup> “¿Cómo en Estados Unidos las matemáticas te pueden meter en prisión?” Disponible al 21/11/2019 en <https://www.bbc.com/mundo/noticias-37679463>

Tal como puede advertirse la IA arroja una predicción futura sobre el potencial accionar del sujeto activo en el tiempo a partir de diversos datos que le fueron aportados y de los cuales extrae conclusiones a futuro. Esto es, sin que se haya concretizado ninguna acción en particular que implique la violación a la norma penal, el sistema determina que el imputado, por ejemplo, podría ser reincidente, tiene un grado de peligrosidad alta para la sociedad, puede darse a la fuga o entorpecer la investigación. Todo esto es una aplicación ex ante del derecho penal que genera en el juzgador su ingreso al proceso con un prejuicio que viola el principio de inocencia, esto es, el sistema le esta advirtiendo al juez que el imputado es culpable, o por lo menos peligroso, antes de que se lo haya juzgado.

En otras palabras, a partir de una acción lesiva que el imputado podría haber realizado (o no, ya que aun no ha sido juzgado por ello) este es sometido a proceso y, mediante datos que se le brindan a un sistema de IA, previo a la realización del juicio, este arroja una conclusión futura sobre como se comportara el sujeto y se la aporta al juez. Resulta mas que claro, que el magistrado ingresa al proceso con una idea en su mente: la presunción de que el imputado es peligroso.

Por otra parte, la sentencia que condena al sujeto tendría problemas en su fundamentación. Esto en virtud de que basaría sus conclusiones en un dictamen que le aporsto el sistema de IA cuyo algoritmo no puede ser revelado.

Justamente esto es una particularidad de COMPAS, sus creadores se niegan a revelar la naturaleza y composición de sus algoritmos bajo el argumento de que es un proyecto privado y que los mismos están protegidos bajo el derecho de autor. Entonces, el algoritmo de donde el sistema obtiene el puntaje es guardado en secreto. Debido a este secreto comercial no se sabe cómo es que COMPAS establece el puntaje. Si el imputado quisiera apelar su puntaje no podría hacerlo pues no sabría cómo.

El problema central reside en que el imputado y su defensa no pueden conocer como es que el algoritmo razona, es decir, mediante que proceso lógico arribo a la conclusión que efectuó. Esto equivaldría a desconocer como fue el razonamiento que llevo a fallar de determinada manera. Si se desconoce dicho razonamiento difícilmente pueda operarse una defensa efectiva respecto del dictamen emitido por el sistema.

Lo hasta aquí expuesto conlleva un grave problema a la hora de fundar la sentencia en virtud de que gran parte de ella esta apoyada en la valoración que efectuó compas cuyo

razonamiento no puede ser conocido en virtud de que no puede accederse al algoritmo que permitiría entender como piensa el sistema. Sin perjuicio de que el imputado pueda acceder a la información de entrada (los datos con los que fue evaluado) y la información de salida (los resultados a los que arribo el sistema) si no puede entenderse como es que el sistema arribo a la conclusión de que un sujeto es peligroso y plausible de reincidir difícilmente pueda fundarse la sentencia que lo declara culpable basándose en el dictamen emitido por la IA.

## ***II. D. – Transparencia***

Retomando algunas ideas del acápite anterior puedo afirmar que la forma de funcionar de un algoritmo es extremadamente similar al razonamiento que efectúa un magistrado a la hora de dictar sentencia.

Para realizar esta afirmación me baso en el hecho de que el Juez evalúa todo el acervo probatorio que compone un expediente y mediante la sana crítica arriba a diversas conclusiones que lo llevarán a la absolución o a la condena de una persona.

Todos los razonamientos que este funcionario efectúa están plasmados en los fundamentos de la sentencia y son la explicitación de como es que llego a la conclusión de que, por ejemplo, puede tenerse por acreditada la materialidad de una conducta y que dicho ilícito es atribuible a un sujeto en particular a quien se le aplicara por el disvalor de su acción y por la defraudación de expectativas normativas una pena.

De la misma manera funciona un algoritmo. Al sistema se le aportan datos, cuya analogía en nuestra comparación serían las pruebas que aportan las partes al juez, y se le solicita que con esos datos siga una serie de pasos lógicos mediante los cuales deberá extraer una conclusión a fin de dar solución a un conflicto.

Tal como puede observarse, el juez razona mediante las pruebas que le son aportadas y falla sobre la culpabilidad de la persona sometida a proceso. El algoritmo razona mediante los datos que le son dados y arriba a determinada solución a un problema. Es evidente que ambos funcionan de manera similar.

¿Qué los diferencia? Entre otras cosas, la transparencia. Por “transparencia” se debe entender que todas las decisiones de los jueces deben estar debidamente fundadas en el derecho, en la correcta valoración de los elementos de prueba de los hechos y en la lógica, como garantía del debido proceso.



El juez tiene el deber de expresar los motivos o fundamentos de hecho y de derecho en los que se apoya la solución que debe ser acorde a las cuestiones planteadas en el proceso. Este deber es impuesto como garantía de los individuos para el ejercicio de los controles de legalidad y legitimidad del acto jurisdiccional.

Entonces, la sentencia estará motivada cuando el órgano judicial exteriorice el razonamiento que justifica la decisión.

El deber de motivar las sentencias encuentra dos fundamentos. En primer lugar, como instrumento procesal facilita a las partes la impugnación. Por otro lado, opera como función de garantía.

El proceso formativo de la sentencia apareja ineludiblemente un momento valorativo. La mera confrontación de hechos y normas resulta insuficiente para solucionar el conflicto por la vía deductiva.

El Juez debe expresar y justificar plenamente su labor selectiva tanto en la aprehensión y valoración de los hechos y pruebas como de las normas jurídicas.

Las pautas rectoras del deber de motivación son: a) racionalidad: argumentación coherente, sin contradicciones, completa y constringente; b) Congruencia: conformidad entre la sentencia, la pretensión y la oposición; c) Integración: Los fundamentos deben integrarse a la sentencia descartándose la motivación oculta o implícita y las remisiones y d) Controlabilidad: por las partes, órganos superiores y ciudadanos. A tal fin la motivación deberá ser pública, accesible y comprensible.<sup>34</sup>

Ahora bien, el proceso que recorre un asistente jurídico digital para adoptar una decisión respecto de la persona sometida a proceso, está revestido de una opacidad que le resulta inherente y en la que probablemente residen las principales fuentes de riesgo para los derechos de las personas sometidas a proceso.

Cuando hablo de opacidad me refiero a la imposibilidad de conocer la forma en que el sistema razona y los fundamentos que llevaron a la IA a determinar la conclusión que arroja luego de evaluar las variables que le fueron ingresadas.

---

<sup>34</sup> TESSONE, Alberto J.: «El deber de motivación de las sentencias», en Revista Jurisprudencia Argentina N.º 5706, p. 5. Disponible en: [http://www.saij.gob.ar/doctrina/daca900130-tessone-deber\\_motivacion\\_las\\_sentencias.htm](http://www.saij.gob.ar/doctrina/daca900130-tessone-deber_motivacion_las_sentencias.htm).



Esta opacidad no solo fluye de la complejidad propia de la ciencia detrás de aquella, o de la enorme cantidad de datos que se procesan mediante su uso, sino también de aspectos relacionados con la convivencia de intereses corporativos resguardados por el ordenamiento.

El tratamiento de grandes volúmenes de datos que realizan los algoritmos de IA resulta complejo y difícil de entender para la mente humana. Cuanto más sofisticado es el modelo usado, menores son las posibilidades de participación y entendimiento. Algoritmos de aprendizaje automático, como las redes bayesianas y los árboles de decisión, son relativamente más comprensibles que, por ejemplo, las redes neuronales profundas o los algoritmos genéticos, cuyo proceso de aprendizaje es tan autónomo que resulta realmente difícil determinar qué parámetros se utilizan para tomar decisiones.<sup>35</sup>

Esta dificultad para entender cómo y por qué un sistema de IA genera un resultado o toma una decisión los convierte en una “caja negra”, cuyo contenido es desconocido hasta para sus programadores.

Cuando una decisión tomada o informada por un sistema de IA tiene implicaciones en la vida de las personas la necesidad de entender las razones que la generaron y, con ello, permitir su refutación en caso de considerarse equivocada o injusta, exige que el proceso de toma de decisión sea transparente y explicable.

Ahora bien, ciertas compañías se niegan a revelar como funciona su algoritmo amparadas en su derecho de propiedad intelectual. Esto genera un gran problema en materia penal en virtud de que, tal como se abordó, al no conocer como es que el sistema razona no es posible entender como es que arribo a la decisión que arribo.

De esta manera, si no se entiende porque el sistema plantea que un individuo es peligroso difícilmente la defensa técnica de este imputado pueda construir una estrategia contra este veredicto. Esto claramente resultaría violatorio del derecho de defensa del imputado.

## ***II. E. – Autoincriminación forzada***

Maier postula que, el procedimiento penal actual toma como meta la persecución penal estatal y la averiguación de la verdad histórica manteniendo las bases sentadas por la

---

<sup>35</sup> Vélez, M. I., Gómez Santamaría, C., & Osorio Sanabria, M. A. (2022, June 30). Conceptos fundamentales y uso responsable de la IA en el sector público. Informe 2. Caracas: CAF. Disponible en: <http://cafscioteca.azurewebsites.net/handle/123456789/1921>

Inquisición. Sin embargo, dichas metas se ven modificadas por una serie de límites que refieren a la dignidad del ser humano.<sup>36</sup>

En nuestro derecho penal rige el principio de libertad de prueba que enuncia que en el proceso penal todo objeto de prueba puede ser probado y por cualquier medio.<sup>37</sup>

Sin embargo, este principio se ve limitado por excepciones o prohibiciones absolutas y relativas.

Las prohibiciones absolutas refieren al objeto de la prueba o al tema sobre el que versa la prueba, mientras que las prohibiciones relativas impiden la verificación de un objeto de prueba por ciertos medios o restringen la posibilidad de probarlo a determinados medios de prueba.

Esta última categoría se vincula a las limitaciones de origen constitucional cuya raíz reside en la protección que se otorga a las personas en un Estado de Derecho por razón de su propia dignidad.

Siguiendo lo analizado por Maier, a nivel general, los límites, establecidos en el artículo 18 de nuestra Constitución Nacional, son los siguientes: la imposibilidad de obligar al imputado a declarar contra sí mismo, la supresión de los tormentos y la inviolabilidad del domicilio, las comunicaciones y papeles privados.

La garantía del *nemo tenetur* surge de la interpretación de los artículos 18 de la Constitución Nacional, 8.2 de la CADH y 14.3 del PIDCP. Refiere a que nadie puede ser obligado a auto incriminarse. Su alcance depende de su fundamento, pero su contenido mínimo recae en la imposibilidad de extraer información incriminante de la declaración del imputado.

Existen tres interpretaciones principales en referencia a la garantía mencionada.

Una interpretación nuclear que centra su atención en la libertad de declarar o de no hacerlo que posee el imputado. Como contrapartida, postula que el Estado no puede exigir al acusado una declaración ni emplear coerción para obtenerla.<sup>38</sup>

---

<sup>36</sup> Maier, Julio B.J. *Derecho Procesal Penal. Tomo I, Fundamentos*. Ad-Hoc Buenos Aires. 2016. Pag. 664 y sgtes. Se recomienda para ampliar también ver: Binder, Alberto M. *Introducción al derecho procesal penal*. Ad-Hoc, Buenos Aires, 2009.

<sup>37</sup> Florian, Eugene. *Elementos de derecho procesal penal*. Bosch, Barcelona.1933.

<sup>38</sup> Maier, Julio B. J, *Derecho procesal penal*, t. 1, Parte general, 2ª ed., 3ª reimpresión, Buenos Aires, Editores del Puerto, Buenos Aires, 2004, § 6, D, pp. 664 y665.

Una interpretación intermedia que focaliza en la voluntariedad de las declaraciones. La Constitución garantiza que toda comunicación del acusado sobre los cargos sea voluntaria. Protege al imputado frente a la coerción y el engaño.<sup>39</sup>

Una interpretación expansiva que extiende el ámbito de protección a comunicaciones no verbales. Refiere a la distinción que puntualiza Maier entre sujeto y objeto de prueba estableciendo que el imputado no puede ser obligado a ser instrumento de su propia incriminación. Bajo un criterio de actividad/pasividad no puede ser obligado a ningún “hacer”, pero, en ciertos casos, si puede ser obligado a un soportar.<sup>40</sup>

Los límites referidos a la autoincriminación forzada refieren a la consideración del imputado como órgano de prueba.

La garantía del artículo 18 establece que nadie puede forzar al imputado a que declare en contra de su voluntad, es necesario su consentimiento expresado de manera libre y sin coacción.

Esto es de vital importancia para la posibilidad de la aplicación de sistemas de IA al proceso penal. En virtud del principio de libertad probatoria quedaría habilitado su uso pero considerando que mediante la utilización de este tipo de sistemas el imputado aporta un relato al proceso que tiene contenido declarativo para que su incorporación sea válida deberán seguirse ciertos requisitos que se abordarán al final del presente trabajo y sin los cuales no podría ser introducida este tipo de tecnología sin resultar violatoria de garantías de tipo constitucional.

## ***II. F – Principio in dubio pro-reo***

El principio in dubio pro reo viene atado a la presunción de inocencia y es una consecuencia del principio pro homine en materia interpretativa. Su contenido es claro, refiere a la exigencia de que la sentencia de condena y la aplicación de una pena sólo puede estar fundada en la certeza del tribunal que falla acerca de la existencia de un hecho punible

---

<sup>39</sup> Maier plantea que la “libertad de decisión del imputado durante su declaración no puede ser coartada por ningún acto o situación de coacción física o moral (tortura o tormento, amenaza, juramento, cansancio, pérdida de la serenidad, cargos o reconveniones, respuestas instadas perentoriamente tendientes a obtener alguna confesión) por la promesa ilegítima de una ventaja o por el engaño (preguntas capciosas o sugestivas)” ver Maier, Julio B. J., Derecho procesal penal, t. 1, Parte general, 2ª ed., 3ª reimpresión, Buenos Aires, Editores del Puerto, Buenos Aires, 2004, § 6, D, pp. 666.

<sup>40</sup> Maier puntualiza que “la garantía solo ampara a una persona como sujeto u órgano de prueba, esto es, como quien, con su relato incorpora al procedimiento un conocimiento cierto o probable sobre un objeto de prueba. No la ampara, en cambio, cuando ella misma es objeto de prueba, esto es, cuando es objeto investigado, como cuando, por ejemplo, se extrae una muestra de sangre o de piel.” ver Maier, Julio B. J., Derecho procesal penal, t. 1, Parte general, 2ª ed., 3ª reimpresión, Buenos Aires, Editores del Puerto, Buenos Aires, 2004, § 6, D, pp. 675.

atribuible al acusado. La falta de certeza representa la imposibilidad del Estado de destruir la situación de inocencia construida por la ley que ampara al imputado, razón por la cual ella conduce a la absolución. Cualquier otra posición del juez respecto de la verdad, la duda o la probabilidad, impiden la condena y desembocan en la absolución.

El principio asegura que el estado de duda implica siempre una decisión de no punibilidad. Solo la certeza de culpabilidad emanada de las autoridades legítimas para pronunciarla puede modificar la situación de inocencia reconocida constitucionalmente.<sup>41</sup>

Entonces, cuando nos referimos al *in dubio pro reo* hablamos de una búsqueda de la verdad que realiza el juez a fin de arribar a un grado de certeza suficiente como para condenar o absolver a la persona que se encuentra sometida a proceso y, en caso de duda, esta jugará a favor del imputado y conducirá a su absolución.

Así, en el caso de las máquinas éstas poseen más certeras respuestas que el ser humano, porque es más difícil que cometan errores procedimentales pero el resultado de sus juicios pueden no ser verdaderos, dependiendo de los datos y programas que le ha cargado el hombre.

Pero, ¿es idéntico el sistema de una computadora a la estructura intelectual humana? Debe decirse que no, que no podría serlo, dado que para constituir el sistema de computadoras, que fue el primer claro representante de la IA, no se usó la totalidad de las capacidades intelectuales humanas, sino exclusivamente la lógico-matemática, es decir un sistema donde se aplicó el esquema lógico a un sistema numérico.

Sin perjuicio de que la IA no es igual a la inteligencia humana lo cierto es que este rigor matemático de la cual esta dotada la primera genera que sus predicciones y resultados carezcan de un margen de error como el que podría cometer un humano al manejar la gran cantidad de información que procesa la máquina.

Aun así, no olvidemos que los sistemas de IA poseen sesgos, razón por la cual, el cálculo matemático podrá ser preciso pero al estar sesgado desde sus albores puede llegar a ser injusto e incorrecto.

A pesar de ello, esta pretensa exactitud matemática de la cual están dotadas las máquinas genera en el juzgador una sensación de certeza que atentaría contra el principio *in dubio pro reo* ya que, aparentemente, eliminaría la posibilidad de duda pero ello basado en

---

<sup>41</sup> RUSCONI, Maximiliano Adolfo. Principio de inocencia e "in dubio pro reo". *Jueces para la Democracia*, 1998, no 33, p. 44-68.

conclusiones que parten de premisas sesgadas, conclusiones que podrían ser un perjuicio para el imputado.

Así, y unido a lo expuesto en el párrafo anterior, el sistema arroja una predicción respecto de un sujeto sometido a proceso previo a que este sea juzgado que condiciona la visión del juzgador. Esta predicción se apoya en la idea de la perfección matemática que rodea a la tecnología y que deja muy poco margen a la duda pero esta tecnología no esta exenta de error y la pretensa exactitud matemática puede no ser mas que una ilusión que arroje un resultado sesgado que derive en una condena cuando, tal vez, sin la utilización de este sistema el margen de duda hubiera favorecido al imputado. De esta manera, la aplicación de la IA sin criterios orientadores vulneraría el principio mencionado al ser el imputado condenado antes de ser juzgado en base una predicción matemática sesgada.

## ***II. G. – Derecho penal de autor***

Se habla de un derecho penal de acto cuando las normas punitivas se dirigen a lo que el hombre hace y no a lo que es. Aquí, el hecho es la causa de la pena y, a la vez, el criterio más importante para su medición.<sup>42</sup>

Por otro lado, como derecho penal de autor podría definirse un derecho penal que pone en primer plano la peculiaridad del autor y que también proporciona el criterio fundamental para graduar la pena.

El principio del acto es el presupuesto fundamental del principio de culpabilidad que concibe al delito como un conflicto que produce una lesión jurídica provocada por un acto humano como decisión autónoma de un ente responsable a la que se le puede redirigir un reproche penal.

Desde la óptica del derecho penal de acto, el objeto se agota en el acto mismo, esto es, la lesión al bien jurídico, mientras que para el derecho penal de autor la acción lesiva es solo una lente que permite ver algo en lo que se deposita el desvalor y que se halla en una característica del autor. Así la esencia del delito radica en una característica del autor que explica la pena.

Ahora bien, a fin de respetar los derechos del imputado en el proceso, los organismos estatales deberían circunscribirse a aplicar un derecho penal de acto y no de autor. Esto es,

---

<sup>42</sup> RAMÍREZ, Nola Gómez. Análisis de los principios del derecho penal. 2004.

juzgar, y en su caso aplicar una pena, a quien cometió una acción lesiva en contra de la norma. Es decir, la reprochabilidad penal no debería estar basada en características personales del autor sino en la concreta acción que cometió.

Si pensamos en los sistemas de IA que asisten al magistrado, muchos de ellos toman características personales del autor y con ello elaboran conclusiones. El sistema COMPAS, por ejemplo, genera un promedio de riesgo que incidirá en las posibilidades del imputado de transitar el proceso en libertad basando sus conclusiones en preguntas que se le hacen al sujeto sometido a proceso al momento de su detención. Muchas de estas preguntas son de índole personal y refieren a su vida y su contexto. Tal como puede advertirse, se trataría de una aplicación del derecho penal que tiene más que ver con un derecho penal de autor y que implica una reducción de las posibilidades de libertad del imputado que está sujeta a que el sistema de IA apruebe su vida sin atender a la acción lesiva presuntamente cometida.

## ***II. H. – Nulidades***

En materia de nulidades prima un criterio de interpretación restrictiva y se exige, como presupuesto esencial, que el acto impugnado tenga trascendencia sobre la garantía de la defensa en juicio o se traduzca en la restricción de algún otro derecho.

Conforme lo analizado hasta este punto, el uso de los sistemas de IA importaría la afectación del derecho de defensa en juicio y de las garantías constitucionales mencionadas lo cual podría importar la nulidad del dictamen de IA y lo actuado en consecuencia.

Sumado a ello, la parte interesada también podría invocar la nulidad de dicha prueba por considerarla prueba no científica y, por tanto, no válida para ser incorporada al proceso judicial.

En atención a ello, resulta imperioso desarrollar soluciones a los problemas reseñados a fin de poder incorporar este elemento tecnológico al proceso penal sin vulnerar garantías constitucionales que amparan al ciudadano.

## **III. – Segunda parte: la normativa**

Advertidos los problemas que representa el uso de IA como asistente del magistrado en el ámbito penal intentaré delinear soluciones a dichos conflictos.

A tal fin, dividiré el planteo en tres ejes. En primer lugar, en esta segunda parte, en atención a que los criterios que se desarrollaran deberán servir como principios rectores a la hora de incorporar IA al ámbito de la justicia penal tomare como base para su delimitación lo desarrollado en otros países a nivel normativo.

Se observará en el desarrollo de este apartado que la mayor parte de la exposición se basa en legislación desarrollada en países europeos o en EE. UU. Esto no es una elección aleatoria, sino que tiene que ver con la laguna a nivel legislación que este tipo de tecnología posee en Argentina. Por ello, se utilizará lo desarrollado por otros países a fin de tomar en cuenta sus principales postulados y delimitar principios de aplicación al marco legal local.

En segundo lugar, en la tercera parte del trabajo, abordaré lo que denomino criterios para la incorporación de IA en el ámbito de la justicia penal y desarrollaré principios rectores que deben ser tenidos en cuenta a la hora de incorporar estos sistema en el Poder Judicial considerado a nivel macro. En este acápite abordare la cuestión desde lo que se denomina “ética algorítmica” la cual, a mi criterio, debe prevalecer como máxima a la hora de utilizar IA en un proceso judicial.

Luego, en tercer lugar, abordaré la problemática desde lo que denomino criterios de incorporación de la IA al ámbito penal en materia de valoración de la prueba. Aquí centrare mi atención en la figura del juez y delineare principios a seguir a fin de que la incorporación de los dictámenes que la IA emita pueda ser valida en el marco de un proceso penal.

Este tercer eje esta pensado desde un nivel micro, es decir, exclusivamente desde la figura del juez.

En virtud de ello, considero apropiado comenzar el desarrollo por el abordaje de los marcos legales existentes en la materia a nivel internacional y luego esbozar los criterios de incorporación de IA al Poder Judicial a nivel macro, esto es, al ámbito de la justicia penal a nivel general. Ello para, posteriormente, desarrollar los principios de incorporación de IA como prueba, es decir, a nivel micro, criterios que deberá seguir el magistrado a la hora de incorporar este tipo de elementos al proceso penal.

### ***III. - Marcos legales***



### **III.1 - El Libro Blanco de la UE** <sup>43</sup>

El Libro Blanco sobre IA, publicado por la Comisión Europea de DDHH en febrero del año 2020, propicia que los marcos regulatorios que se dicten, por un lado, promuevan la adopción de la IA, y por otro, aborden adecuadamente los riesgos asociados con ciertos usos de esta nueva tecnología, partiendo de la base de que las características particulares de algunas técnicas de aprendizaje automático, como su opacidad, complejidad, imprevisibilidad y un comportamiento parcialmente autónomo, pueden dificultar la comprobación del cumplimiento de la regulación protectora de los derechos fundamentales e impedir su efectivo cumplimiento.

Para ello, propicia que el marco regulador a dictar se centre en cómo minimizar los riesgos que suponen los sistemas con IA para los derechos fundamentales, especialmente la protección de datos personales, la privacidad y la no discriminación.

El Libro Blanco señala que en el diseño de normas que regulen de forma directa las aplicaciones de IA, debe ser cuidadoso y considerar los siguientes aspectos relevantes: 1) la regulación debe ser flexible a fin de lograr su objetivo sin desalentar el desarrollo tecnológico, 2) la aplicación de IA debe ser diseñada de modo transdisciplinario con profesionales que conozcan de IA, machine learning, Deep learning, data science y propiedad intelectual de modo tal que se arribe a una regulación concreta y efectivamente aplicable, 3) se debe considerar que las normas demasiado precisas o específicas pueden caer en desuso en pocos años ya que esta es una tecnológica en constante evolución, 4) la regulación excesivamente prescriptiva podría generar una carga desproporcionada para las pequeñas y medianas industrias en virtud de lo cual se debe evitar dicha característica sin sacrificar la efectividad de la norma.

Además, en el Libro Blanco se sugiere regular a partir de considerar el mayor o menor riesgo que conlleven las aplicaciones de IA, para lo cual deberán considerarse, por un lado, los derechos que estén en juego, y por otro, el sector y el uso para el que esas aplicaciones fueron previstas.

---

<sup>43</sup> El presente acápite fue desarrollado conforme lo abordado en: Sánchez Caparrós, M. Prevenir y controlar la discriminación algorítmica disponible en: <https://www.justierradelfuego.gov.ar/wp-content/uploads/2021/07/Sánchez-Caparrós.-Prevenir-y-controlar-la-discriminación-algor%C3%ADtmica.pdf>

En ese esquema, una aplicación de IA, sin dudas, será de riesgo elevado si se la utiliza en un sector respecto del cual es previsible que existan riesgos significativos o cuando se la emplee en una forma que pueda importar riesgos significativos.

Por otra parte, en lo que es de relevancia en materia de prevención y mitigación del sesgo algorítmico, el marco regulatorio según el Libro Blanco debería establecer requisitos con relación a: 1) los datos de entrenamiento del sistema, procurando que el uso posterior del sistema de IA no genere resultados que conlleven discriminación, para lo cual debería establecerse la obligación de utilizar conjuntos de datos de entrenamiento suficientemente representativos, garantizando que todas las dimensiones de género, etnicidad y otros posibles motivos de discriminación queden reflejadas en esos datos, 2) la conservación de registros y datos que se utilizaron para programar los algoritmos, para entrenar y testear sistemas de riesgo elevado; de modo tal de poder conocer las características del conjunto de datos y el modo en que se los escogió, o incluso el propio conjunto de datos en ciertos supuestos; o bien las metodologías de programación, entrenamiento, validación y testeo de los algoritmos, 3) el acceso a la información con respecto a las posibilidades y limitaciones de la IA, los objetivos para los que fue programada, la aplicación y el nivel de exactitud que se estima conseguir en el logro de ese objetivo, tanto para los implementadores, como para las autoridades competentes y partes afectadas, según el caso, 4) el conocimiento del usuario acerca de cuándo está siendo sujeto de una decisión automatizada y cuando, en cambio, existe interacción con, o intervención de, un ser humano, 5) la reproducibilidad de los resultados de la aplicación del sistema de IA utilizado, 6) la supervisión humana constante, que garantice una IA fiable, que no provoque efectos adversos con posibilidad de desactivar el sistema si se vuelve inseguro.

### ***III.2 - La propuesta del Parlamento Europeo del año 2020***

El 20 de octubre del año 2020 el Parlamento Europeo aprobó tres propuestas de regulación sobre inteligencia artificial, referidas a tres áreas específicas: las normas éticas que deben regir la IA; las reglas sobre la responsabilidad derivada del uso de sistemas de IA y las reglas sobre derechos de propiedad intelectual vinculadas con el uso de sistemas de IA.

La propuesta para un marco ético para la inteligencia artificial insta a la Comisión Europea a diseñar un nuevo marco jurídico que desarrolle los principios éticos y obligaciones

jurídicas vinculados con el desarrollo, despliegue y empleo de la IA, la robótica y tecnologías relacionadas en el ámbito de la UE.

Destaca el Parlamento Europeo en su propuesta que resulta necesario contar con un marco regulatorio armónico basado en el Derecho de la UE, la Carta Europea y el Derecho internacional de Derechos Humanos, con la finalidad de establecer normas que sean iguales para toda la UE y protejan eficazmente sus valores; que persigan el respeto de la dignidad humana, la autonomía y la autodeterminación de las personas; la prevención del daño y la eliminación de los sesgos y la discriminación, entre otros objetivos destacados.

Puntualmente, respecto de la temática atinente a los sesgos, reconoce el Parlamento Europeo que la IA, según su desarrollo y los usos que se le asignen, puede crear y reforzar sesgos, y también discriminar como consecuencia de los sesgos que son inherentes a los conjuntos de datos empleados para entrenar los sistemas lo que demanda requerir a la Comisión Europea y a los Estados miembros que adopten las medidas pertinentes a los fines de evitarlos y garantizar la plena protección de los derechos fundamentales.

En línea con ello, manifiesta su preocupación por el peligro que pueden significar los sesgos y la discriminación en el desarrollo y uso de la IA, y sostiene que los valores éticos de la equidad, la exactitud, la confidencialidad y la transparencia deben ser la base de esta tecnología, lo que implica en este contexto que sus operaciones deban concebirse de un modo tal que garantice la igualdad de trato y la no discriminación.

Finalmente, remarca la relevancia de garantizar que las personas cuenten con recursos efectivos, y que los Estados velen por el establecimiento de procedimientos y mecanismos de revisión accesibles, asequibles, independientes y eficaces para garantizar la revisión imparcial, efectuada por seres humanos, de las denuncias de vulneración de los derechos de ciudadanos realizada por sistemas algorítmicos.

### ***III.3 - El proyecto de Ley de Responsabilidad Algorítmica en los EE. UU.***

En el año 2019 se presentó en el Congreso de los EE.UU. un proyecto de ley de responsabilidad algorítmica con el objeto de exigir a las empresas que evalúen periódicamente los sistemas de IA que utilizan, a fin de determinar la precisión, el sesgo y la discriminación en que esos sistemas pueden incurrir.

En el proyecto, un sistema se considera de alto riesgo si cumple alguna de estas condiciones: 1) plantee un riesgo significativo para la privacidad o seguridad de la información personal, 2) pueda contribuir en decisiones injustas, inexactas, sesgadas o discriminatorias que afecten a la ciudadanía, 3) adopte decisiones o facilite la toma de decisiones, basado en una evaluación extensiva y sistemática de los ciudadanos, incluyendo intentos de analizar o predecir aspectos sensibles de su vida que alteren derechos o impacten significativamente en ellos, 4) involucre información personal de un número significativo de ciudadanos, 5) monitoree lugares de acceso público y grandes dimensiones, 6) cumpla con cualquier otra condición que se incluya por vía reglamentaria.

Además, de acuerdo con el proyecto, la evaluación de impacto del sistema, cuando sea razonable, deberá ser realizada con la intervención de terceros externos, independientes, expertos en tecnología o auditores; de forma previa cuando se trate de sistemas nuevos y con tanta frecuencia como la autoridad lo determine cuando se trate de sistemas ya existentes.

Por otro lado, la propuesta determina que la evaluación de impacto del sistema de alto riesgo deberá analizar tanto al sistema como al proceso de desarrollo del sistema de decisión, incluyendo el diseño y el conjunto de datos de entrenamiento, en cuanto a sus impactos en materia de precisión, equidad, sesgo discriminación, privacidad y seguridad.

El proyecto exige, asimismo, que las compañías concreten una evaluación de impacto de protección de datos para sus sistemas de información de alto riesgo, que evalúe hasta dónde el sistema protege la seguridad y privacidad de la información personal que procesa.

#### ***IV. Tercera parte: las soluciones***

##### ***IV. A. – Criterios para la incorporación de IA al ámbito de la justicia penal***

###### ***IV. A. 1.- Transparencia<sup>44</sup>***

A fin de evitar lo enunciado en acápites anteriores, en primer lugar, es necesario garantizar la transparencia algorítmica.

La Carta de Derechos Digitales de España establece que en materia de derechos ante la inteligencia artificial se deberá asegurar la transparencia, auditabilidad, explicabilidad y trazabilidad. Señala que los procesos de transformación digital, el desarrollo y el uso de la

---

<sup>44</sup> El presente acápite se desarrolló conforme lo postulado por Gabriele Vestri en la obra: Vestri, G. (2021). La inteligencia artificial ante el desafío de la transparencia algorítmica: Una aproximación desde la perspectiva jurídico-administrativa. *Revista Aragonesa de Administración Pública*, (56), 368-398.

tecnología digital, así como cualquier proceso de investigación científica y técnica relacionado con ellos o que los utilice instrumentalmente, deberán tener presente la exigencia de garantizar la dignidad humana, los derechos fundamentales, el libre desarrollo de la personalidad y ordenarse al logro del bien común. En definitiva, se entiende a la transparencia de la IA como reflejo de los Derechos Humanos en el seno de un uso masivo de datos.

La transparencia algorítmica hay que entenderla de manera bifocal. Por un lado, resulta necesario verificar que la elección del algoritmo por parte del Poder Judicial sea transparente y, por otro lado, también resulta necesario verificar la transparencia del algoritmo cuando este se encuentre funcionando en el Poder Judicial a fin de garantizar la accesibilidad del interesado a conocer como el algoritmo adopta la decisión final.<sup>45</sup>

Así, un algoritmo será transparente cuando: 1) publique periódicamente y donde corresponda información relevante acerca de su funcionamiento, 2) ponga a disposición de la ciudadanía los documentos y el contenido técnico del algoritmo, 3) pueda demostrar que el algoritmo actúa en beneficio exclusivo del interés público, 4) este dotado de una base inquebrantable que permita su control.

En este punto, es necesario, tal como lo plantea Vestri, analizar la transparencia desde dos perspectivas: a nivel *ex ante* y a nivel *ex post*.

### ***De la transparencia a nivel ex ante: cesión de derechos de propiedad intelectual***

El Poder Judicial no es creador de IA, es decir, no inventan sus propios sistemas de IA sino que se apoyan en desarrolladores y empresas que les suministran las herramientas necesarias. A partir de ello, es necesario observar que podrían surgir problemas en la IA utilizada por el sector público y los derechos de propiedad intelectual de los productos objeto del contrato.

De esta manera, considero que los contratos que tengan por objeto el desarrollo y la puesta a disposición de productos protegidos por el derecho de propiedad intelectual deberán necesariamente llevar aparejada la cesión de estos derechos al Poder Judicial contratante. Ello a fin de evitar conflictos como el desarrollado con el sistema COMPAS.

Así, la empresa que desarrolle un sistema de IA para el Poder Judicial está obligada a transmitir a la entidad pública los derechos de propiedad intelectual y, por ende, deberá

---

<sup>45</sup> Vestri, G. (2021). La inteligencia artificial ante el desafío de la transparencia algorítmica: Una aproximación desde la perspectiva jurídico-administrativa. *Revista Aragonesa de Administración Pública*, (56), 368-398.

transferir los códigos fuente que serán puestos a disposición de quien quiera acceder a la información de como funciona dicho algoritmo.

Ahora bien, a veces, las organizaciones públicas suelen requerir una IA ya existente que pueda ser adaptada a las exigencias judiciales. Aquí no podrá articularse una total cesión de derechos de propiedad intelectual ya que la IA base se utilizará en otras actividades que desarrolle la compañía y no exclusivamente para el ámbito de la justicia. En estos casos, el desarrollador deberá ceder los derechos sobre los códigos fuente que posibilitan la adaptación del sistema de IA a las necesidades del organismo judicial.

Se plantea entonces una partición del sistema de IA en el que finalmente solo la parte correspondiente a la especificidad del servicio público prestado sería sometida a la debida transparencia pública.

### ***De la transparencia ex post: inteligibilidad del algoritmo***

En este punto, asume un papel importante la transparencia de la labor del sistema cuando este ya se encuentre instalado y el interesado quiera conocer como el algoritmo adopto determinada decisión.

En este sentido, el sistema es transparente cuando es conocible y descifrable. Hablamos del derecho que tiene el interesado a saber que el organismo judicial adoptará determinadas decisiones mediante un algoritmo que deberá ser descifrable y, en consecuencia, transparente.

De esto nos da cuenta la Carta de Derechos Digitales que introduce términos como: explicabilidad, trazabilidad y garantía de la accesibilidad, usabilidad y fiabilidad.

Por otra parte, el principio de transparencia debe operar tanto para la administración pública titular del poder para cuyo ejercicio se prevé el uso de la herramienta del algoritmo, como para los sujetos involucrados por el poder mismo. El algoritmo es conocible cuando es posible saber todos los diferentes aspectos que lo rodean. Para lograr este grado de transparencia, los jueces no deben aceptar que las empresas que suministran estos tipos de servicios informáticos invoquen las cláusulas de confidencialidad.

En definitiva, la transparencia algorítmica debe asentarse sobre dos principios fundamentales: el conocimiento omnicomprendido del algoritmo, que incluye la inteligibilidad de este, y la no confidencialidad y exclusividad de la decisión algorítmica,



particularmente cuando la decisión automatizada produce efectos significativos sobre el interesado.

### ***Principios para garantizar la transparencia algorítmica***

Conforme lo analizado, exigir mayor transparencia en el funcionamiento de los algoritmos, implica dar a conocer el criterio y proceso lógico desarrollado para la obtención de los resultados. A tal fin, utilizar los siete principios pronunciados por la Association for Computer Machinery<sup>46</sup> puede ser de gran utilidad para los organismos judiciales a fin de garantizar la transparencia algorítmica: 1) Conciencia: quienes utilicen algoritmos deben ser conscientes de las consecuencias de su uso en función de los sesgos en que puedan incurrir, 2) Acceso y corrección: las entidades reguladoras deben poseer mecanismos para corregir los perjuicios individuales causados por los sesgos, 3) Responsabilidad: las instituciones son responsables de las decisiones tomadas por los algoritmos que utilizan, aunque no sean capaces de explicar cómo se tomaron esas decisiones, 4) Explicación: las instituciones públicas que utilizan algoritmos para la toma de decisiones deben explicar los procedimientos seguidos por los algoritmos y las decisiones adoptadas en su mérito, 5) Procedencia de los datos: los programadores de los algoritmos deben explicar las características de los datos de entrenamiento así como ofrecer una exploración de los posibles sesgos del algoritmo debido al uso de esos datos, 6) Auditoría: se deben guardar los modelos, los algoritmos, los datos y las decisiones para auditorías futuras, 7) Validación y testeo: las instituciones deben instaurar métodos rigurosos para la validación de los modelos y documentarlos. Asimismo, deben realizar testeos específicos para detectar posibles sesgos del algoritmo. Los resultados de estos testeos deben ser públicos.

#### ***IV.A.2.- Ética en el uso de los datos***

En materia de ética digital el Consejo de Europa<sup>47</sup> ha propuesto un conjunto de recomendaciones tendientes a la constitución de equipos éticos que valoren el tratamiento de información personal.

---

<sup>46</sup> <https://www.acm.org/about-acm/mission-vision-values-goals>

<sup>47</sup> [https://www.gdprsummary.com/?gclid=EAJaIQobChMIxomGiv349gIVwwiRCh1F\\_wv0EAMYAyAAEgLASfD\\_BwE](https://www.gdprsummary.com/?gclid=EAJaIQobChMIxomGiv349gIVwwiRCh1F_wv0EAMYAyAAEgLASfD_BwE)

En concreto, en materia de transparencia algorítmica es necesario analizar ciertas cuestiones relacionadas con el tratamiento de los datos personales que en el escenario digital y automatizado serán tratados por un algoritmo.

En este sentido, cabe destacar que el Parlamento Europeo dictó el Reglamento general de protección de datos (UE) 2016/679 (RGPD)<sup>48</sup>, el cual tiene por objetivo regular la protección y libre circulación de datos personales de personas físicas.

Ahora bien, en lo que a este trabajo le importa, las principales normas de esta regulación están representadas por los artículos 13 y 15 del RGPD, que establecen el derecho del interesado a conocer la existencia de los procesos de decisión automatizados que le conciernen y a recibir información sobre la lógica utilizada para arribar a dicha decisión.

Someramente, las dos normas señaladas introducen ciertas características que finalmente recalcan cuestiones centrales sobre transparencia algorítmica.

En primer lugar, el derecho a recabar informaciones mediante la existencia de un proceso decisional automatizado. Esta cuestión concierne al conocimiento del algoritmo, es decir, al derecho del interesado a ser informado de la existencia de un proceso de decisión automatizado de tratamiento de sus datos personales.

En segundo lugar el derecho a obtener informaciones acerca del funcionamiento del sistema automatizado.

Por último, cabe referirse, al elemento de la comprensibilidad del algoritmo. El interesado tiene el derecho a obtener información sobre el modo en que el sistema automatizado ha tratado sus datos.

#### ***IV.A.3.- Sesgos y discriminación algorítmica***<sup>49</sup>

La propuesta de reglamento incluye un artículo específico dirigido a regular lo atinente a la cuestión del sesgo y la discriminación.

En concreto, prevé que los programas informáticos, los algoritmos o los datos utilizados o producidos por la IA, la robótica y las tecnologías conexas desarrolladas, deberán

---

<sup>48</sup> REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). Disponible en: <https://www.boe.es/doue/2016/119/L00001-00088.pdf>

<sup>49</sup> El acápite relativo a sesgos y discriminación algorítmica fue desarrollado en base a lo abordado, investigado y postulado por la autora Mariana Sánchez Caparros (postulados a los que adhiero por completo) en la obra: Sánchez Caparros, M. Prevenir y controlar la discriminación algorítmica disponible en <https://www.justierradelfuego.gov.ar/wp-content/uploads/2021/07/Sánchez-Caparros.-Prevenir-y-controlar-la-discriminación-algor%C3%ADtmica.pdf>

ser imparciales y no podrán discriminar por motivos de raza, sexo, orientación sexual, embarazo, discapacidad, características físicas o genéticas, edad, minoría nacional, origen étnico o social, lengua, religión o creencias, opiniones políticas o participación cívica, nacionalidad, estado civil o económico, educación o antecedentes penales. Sólo podrá justificarse el trato diferenciado entre personas o grupos de personas cuando exista una finalidad objetiva, razonable, legítima, proporcionada y necesaria, en la medida en que no exista otra alternativa que interfiera menos con el principio de igualdad de trato.

Ahora bien, el algoritmo aprende a través de los inputs que se cargan en él. La inserción de una información discriminatoria produce un aprendizaje discriminatorio. Así, la intervención del Estado es trascendental para evitar el sesgo algorítmico. Dicha intervención, debe abogar por la aplicación de los principios éticos de la IA que pueden resumirse en: el principio del respeto de la autonomía humana, el de equidad, de prevención del daño y de explicabilidad.<sup>50</sup>

En este sentido, y siguiendo para el desarrollo del presente acápite lo postulado por la autora Mariana Sánchez Caparros a cuya teoría adhiero por completo, se deben generar estructuras normativas que resguarden y/o eviten la discriminación algorítmica considerando las desigualdades sociales existentes.

Asimismo, se debe promover su transparencia y el hacerlos comprensibles de manera de no generar asimetrías de información. Todo lo anterior, en la búsqueda de limitar espacios de irregularidades en el espacio público y un uso indebido en el ámbito privado.

En primer lugar, hay que preocuparse, desde el origen, de producir y tratar datos de calidad, con la debida desagregación y estandarización, de modo tal de garantizar una correcta lectura de éstos por parte de los sistemas de Inteligencia Artificial.

En esta línea, es fundamental la creación de controles para supervisar los procesos de recolección y verificación de datos, hecha por personas con criterios de no discriminación. Es importante también, generar instancias de auditoría independiente, certificación y monitoreo constante para testear sesgos y otros perjuicios.

Asimismo, es relevante desarrollar marcos legales robustos para promover la transparencia en la construcción y uso de algoritmos para tomar decisiones.

---

<sup>50</sup> Vestri, G. (2021). La inteligencia artificial ante el desafío de la transparencia algorítmica: Una aproximación desde la perspectiva jurídico-administrativa. *Revista Aragonesa de Administración Pública*, (56), 368-398.

Partiendo de allí, y considerando que estamos frente a una tecnología en pleno desarrollo, que además es muy dinámica, pues cambia constantemente, el marco jurídico que se proponga debería basarse en el respeto de los derechos y en el establecimiento de una serie de principios básicos y fundamentales que sienten las bases para una IA ética, robusta, fiable y centrada en el ser humano.

Estos principios son especialmente relevantes durante la fase de diseño del sistema, en tanto es en este momento en el que se configuran y programan sus funcionalidades y objetivos.

Es durante la fase inicial que pueden abordarse de manera más efectiva algunos de los principales riesgos que plantea la IA, y es por ello que frecuentemente se habla de ética por diseño, que considera la idea de que estos principios éticos se incorporen expresamente en esta primera etapa crítica. Ello sin perjuicio de que también alcancen las fases de desarrollo, implementación y adopción del sistema.

En el ámbito de la Comunidad Europea, inspirados en los derechos fundamentales consagrados en diversos Tratados de Derechos Humanos, se han identificado cuatro principios imperativos éticos que deberían observarse en todo momento y fundamentalmente, entiendo, en la etapa de diseño del sistema: 1) *Respeto de la autonomía humana*: Los sistemas de IA deben garantizar el respeto de la libertad y la autonomía humana, evitando subordinar, manipular, condicionar o dirigir injustificadamente a los seres humanos lo que exige una supervisión humana constante; 2) *Prevención del daño*: los sistemas de IA no deberían provocar daños o agravar los existentes, ni perjudicar de cualquier otro modo a los seres humanos. Deben ser seguros y robustos desde el punto de vista técnico, garantizándose que no pueden ser objeto de un uso malintencionado. Se debe prestar atención a los efectos adversos que pueden venir dados por la asimetría de las relaciones entre grupos de personas; 3) *Equidad y no discriminación algorítmica*: El desarrollo, despliegue y utilización de sistemas de IA debe ser equitativo, garantizando una distribución justa e igualitaria de los costos y beneficios, y asegurar que las personas no sufran discriminación injusta. Las personas deben poder oponerse a las decisiones adoptadas por los sistemas de IA; 4) *Explicabilidad*. Los procesos de desarrollo, despliegue y uso de IA deben ser transparentes; debe comunicarse abiertamente la capacidad y finalidad de los sistemas de IA, y las decisiones adoptadas por ellos o por las personas que los emplean deben ser explicables.

La equidad y la explicabilidad, entiendo, son dos principios clave cuando se trata de prevenir, mitigar y cuestionar el sesgo presente en los algoritmos de los sistemas de IA.

Podríamos sumar a ellos, la necesidad de garantizar la inclusión y la diversidad a lo largo de todo el ciclo de vida de los sistemas de IA, y la igualdad de acceso mediante procesos de diseño inclusivo.

Además, siempre que sea posible, los sesgos identificables y discriminatorios deberían eliminarse en la fase de recopilación de datos, y de no ser posible esto último, debería velarse por una supervisión constante que permita analizar y abordar el propósito, las restricciones, los requisitos y las decisiones del sistema de un modo claro y transparente.

En ese marco, la IA requiere que los Estados tomen parte en el establecimiento de un marco regulatorio dirigido a asegurar que esta tecnología sea segura y contribuya al bienestar general de la sociedad, y especialmente, a evitar los efectos no deseados que su uso creciente puede provocar, entre los que sin dudas se encuentra el de la discriminación algorítmica lo cual conlleva una enorme gravedad cuando hablamos de procesos judiciales como el penal donde se puede ver afectada la libertad de una persona.

Hasta tanto existan normas que regulen la cuestión o surjan otras herramientas específicas de control, dado que la IA ya es presente y no sólo futuro, debemos bucear en el conjunto de preceptos y teorías que tenemos disponibles para dar con alguna herramienta que nos permita encauzar cuestiones que he desarrollado en la primera parte de este trabajo, en concreto, me refiero a la impugnación de la decisión automatizada que provoca un tratamiento discriminatorio.

El primer aspecto a tener en cuenta es que en cualquier caso, aun cuando no haya un marco que regule su empleo, las aplicaciones de IA deben respetar los derechos fundamentales consagrados en los tratados de Derechos.

En la actualidad, siguiendo los postulados de la Dra. Sánchez Caparros, ese marco jurídico nos provee de una teoría que puede ser de suma utilidad para plantear y decidir un caso cuando se cuestiona una decisión automatizada porque presuntamente discrimina.

Hablamos de la doctrina de las categorías sospechosas que fue adoptada en Argentina con el precedente Hooft de la Corte Suprema de Justicia de la Nación en el año 2005 también ha sido receptada por la CIDH que la ha empleado para decidir distintos casos vinculados con el derecho a la igualdad y a la no discriminación.

En el ámbito latinoamericano, recordemos, el art. 1 de la CADH prohíbe discriminar por motivos de raza, color, sexo, idioma, religión, opiniones políticas o de cualquier otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social.

A la luz de este precepto, la CIDH ha señalado que existe un vínculo indisoluble entre la obligación de respetar y garantizar los derechos humanos, y el principio de igualdad y no discriminación, lo que determina el deber de los Estados de respetar y garantizar sin discriminación los derechos contenidos en la CADH.

Bajo esta premisa, ha receptado la doctrina de las categorías sospechosas en los precedentes *Atala Riffo y niñas vs. Chile* y *Norín Catrimán y otros vs. Chile* concluyendo que cuando estamos frente una distinción realizada por alguno de los motivos vedados por el art. 1 de la Convención Americana sobre Derechos Humanos, estaremos frente a lo que se conoce como una categoría sospechosa de discriminación.

Como consecuencia de ello, en estos casos se disparará un control judicial de excepción, que exigirá al Estado una fundamentación rigurosa, trasladándose a su cargo el peso de probar que la decisión no tenía un propósito, ni un efecto discriminatorio.

En otras palabras, la presencia de un motivo prohibido o categoría sospechosa exigirá una fundamentación rigurosa y de mucho peso a quien pretenda sostener la diferencia de trato, y en virtud de la inversión de la carga probatoria que trae su presencia, le demandará al acusado demostrar que su decisión no tenía un propósito ni un efecto discriminatorio.

Considerando lo expuesto, para la adecuada identificación de las categorías sospechosas se necesita, primeramente, corroborar que la diferencia de trato alegada como discriminatoria se basa en alguno de los motivos de discriminación prohibidos por la CADH, en cuyo caso estaremos en presencia de una categoría potencialmente sospechosa.

Luego, la naturaleza "sospechosa" la distinción se confirmará si la diferencia de trato afecta a un grupo minoritario que carece de herramientas para solucionar la situación de desventaja en la que se encuentra mediante una participación efectiva en el proceso que no la protege del mismo modo en que resguarda a los grupos predominantes, que sí están bien representados en dicho sistema.

Para identificar estos grupos existen instrumentos específicos, que se han bautizado bajo el nombre de "indicios" o "factores" de sospecha, que van a estar presentes con relación



al grupo perjudicado por la diferencia de trato, y que serán los que van a permitir confirmar el carácter sospechoso de la distinción.

Y en estos casos, se disparará la alternativa de que distinción que los emplea se examine bajo un test severo, que va a determinar su presunción de inconstitucionalidad e invertirá la carga de la prueba, exigiendo a quien pretenda sostener su validez una fundamentación rigurosa y de mucho peso, así como demostrar que su decisión no tenía un propósito ni un efecto discriminatorio.

En este marco, si un sistema de IA es entrenado con datos que presenten algún tipo de sesgo los resultados generados por ese sistema no sólo van a reproducir el mismo sesgo, sino que es probable que lo amplíen y lo institucionalicen en su ámbito de aplicación.

El potencial de la IA de establecer y perpetuar sesgos y discriminación demanda un gran trabajo interdisciplinario, en el que operadores informáticos y jurídicos deberán colaborar para desarrollar herramientas que permitan identificar los sesgos para prevenir sus resultados indeseados.

Desde el derecho se ha postulado la vigencia de una serie de principios jurídicos generales dirigidos a lograr una IA que sea compatible con el Estado constitucional de Derecho y con los Derechos Humanos. Entre ellos se destaca, sin dudas, el principio de no discriminación algorítmica.

En virtud de este principio general, debe exigirse que el diseño y la implementación de sistemas de IA respeten el principio de no discriminación, lo que impide aceptar que estos sistemas procesen la información o los datos de las personas humanas sobre la base de criterios de distinción prohibidos.

En línea con ello, si el resultado de la aplicación de un sistema de IA se explica por categorías prohibidas nos encontraremos frente a un supuesto de discriminación algorítmica y el sujeto afectado podrá acudir a la justicia para solicitar que se deje sin efecto el acto o decisión discriminatorio.

Ahora bien, la alternativa de demandar eficazmente el cese del acto discriminatorio se encuentra directamente relacionada con el principio de explicabilidad algorítmica cuya efectiva vigencia será sustancial para que el sujeto afectado por la decisión asociada a un sistema de IA pueda ejercer su derecho de defensa en oportunidad de cuestionarla.

Considerando este extremo, debemos tener en cuenta que los diversos enfoques de aprendizaje automático presentan enormes diferencias en orden a su explicabilidad, y estas diferencias deben ponderarse al tiempo de diseñar la estrategia defensiva dirigida a impugnar una decisión automatizada por presumirla discriminatoria.

Es que el resultado de un árbol de decisiones, que es un típico ejemplo de IA de caja blanca, posiblemente se pueda explicar enteramente, lo que no sólo va a facilitar la tarea de formular la impugnación al afectado, sino la del juez de corroborar si la decisión asociada al sistema de IA es o no discriminatoria.

En cambio, cuando se trata de resultados provenientes de la aplicación de sistemas de caja negra, por la opacidad propia de la técnica, la tarea de identificar y probar que un determinado resultado obedece a un distingo prohibido se volverá ciertamente más ardua.

No sólo porque estos sistemas suelen estar protegidos por regulaciones de propiedad intelectual que impiden el acceso al código sino además porque por la propia dinámica de funcionamiento de esta clase de redes, no es posible explicar totalmente el paso a paso que permita comprender cómo se han sopesado los diversos atributos y la relevancia que se ha asignado a cada dato e información para arribar a un resultado.

Sumado a ello, existen algoritmos que aprenden y se refinan con el tiempo, lo que lleva a que los parámetros que emplea el algoritmo y la forma de arribar a los resultados varíe de un momento a otro, tornando aún más complejo resolver la problemática vinculada a su explicabilidad

Dada esta realidad, cuando una persona se ve perjudicada por una decisión y sospecha que ella puede ser el resultado de la aplicación de un sistema de IA que ha utilizado motivos de distinción prohibidos, sea directa o indirectamente, dando lugar a un resultado presuntamente discriminatorio o injusto, se justificaría apelar a la doctrina de las categorías sospechosas, que puede convertirse en una gran aliada para controlar y juzgar el sesgo algorítmico.

De aceptarse esta propuesta, para el sujeto afectado negativamente por el resultado de la aplicación de un sistema de IA, debería bastar con alegar que existen indicios serios de que el trato discriminatorio puede basarse en una categoría vedada y que pertenece a una clase vulnerable en los términos señalados más arriba, solicitando el análisis de su caso bajo los parámetros de esta doctrina.

Como consecuencia de ello, se trasladará a quien ha elegido valerse del sistema de IA, la carga de probar que la decisión o el resultado se basa en distinciones lícitas o bien, que la diferencia de trato se justifica en razones de mucho peso y que no tenía un propósito ni un efecto discriminatorio.

Ello independientemente de que quien utilice el sistema de IA lo emplee como soporte o que le hubiera delegado directamente la decisión, pues insisto, es quien en definitiva lo ha elegido y el que más aprovecha de los beneficios que estos sistemas suelen reportar.

#### ***IV.A.4.- Auditorias Algorítmicas<sup>51</sup>***

En esta breve aproximación a la transparencia algorítmica que en definitiva representa la transparencia del sistema de IA es necesaria la previsión de un procedimiento de auditoría algorítmica. Esto tiene su razón en la necesidad de individualizar quién controla que el algoritmo sea transparente y que lo sea en todos sus elementos.

Así, el organismo judicial que usa un sistema de IA debe garantizar su propio control y que éste resulte transparente (auditoría interna).

Sumado a ello, al ser un organismo público deberá contemplar la posibilidad de que la mencionada actividad de control sea encargada a agencias públicas independientes o a empresas privadas independientes que en todo caso permanecerían sujetas a la obligación de la confidencialidad (auditoría externa).

En ambos casos —agencias o empresas—, la auditoría podría consistir tanto en la evaluación de todo el proceso de uso de la IA como de la estimación de las consecuencias que produce el sistema de IA en las operaciones del tribunal.

En la misma línea, es necesario pensar si las indicaciones de quien audita y controla se deben considerar meras recomendaciones o si, por lo contrario, deben resultar vinculantes para el organismo judicial.

Considero que la actividad de las agencias de control debería ser recomendatoria para que, precisamente, el Poder Judicial, en beneficio de la transparencia y el interés público, pueda rectificar la acción algorítmica y que los resultados del control deberán ser puestos en conocimiento de la ciudadanía a fin de brindar a la persona sujeta a proceso garantías de que

---

<sup>51</sup> El presente acápite fue desarrollado conforme lo postulado en la siguiente obra: Vestri, G. (2021). La inteligencia artificial ante el desafío de la transparencia algorítmica: Una aproximación desde la perspectiva jurídico-administrativa. *Revista Aragonesa de Administración Pública*, (56), 368-398.

el algoritmo que funciona como asistente del magistrado lo hace libre, por ejemplo, de sesgos y parcialidades.

#### ***IV.A.5.- Niveles de riesgo de la IA***

La Comisión Europea presentó el Proyecto de ley de inteligencia artificial de la UE<sup>52</sup>, cuyo objetivo es regular el uso de sistemas de IA garantizando la protección de los derechos fundamentales y la seguridad de los usuarios.

El Proyecto propone un planteamiento basado en cuatro niveles de riesgo: 1) Riesgo inadmisibles: Se prohibirá un conjunto muy limitado de usos especialmente nocivos de la IA que violen los derechos fundamentales protegidos por la Carta. Por poner algunos ejemplos citados por la propia Comisión, la puntuación social por parte de los Gobiernos, la explotación de los puntos débiles de los niños, el uso de técnicas subliminales y, salvo contadas excepciones, determinados sistemas de identificación biométrica remota en directo en espacios públicos con fines policiales; 2) alto riesgo: Se considera de alto riesgo un número limitado de sistemas de IA definidos en el Proyecto y que tienen un impacto negativo en la seguridad de las personas o en sus derechos fundamentales. El Proyecto incluye un listado de sistemas de IA de alto riesgo, que se irá adaptando a la evolución de los casos de uso de la IA. En cualquier caso, siempre se considerarán de alto riesgo cuando estén sujetos a una evaluación de la conformidad por parte de terceros con arreglo a la legislación sectorial; 3) riesgo limitado: Para determinados sistemas de IA se imponen obligaciones específicas de transparencia, por ejemplo, cuando exista un riesgo claro de manipulación, ya que se considera que los usuarios deben ser conscientes de que están interactuando con una máquina, 4) riesgo mínimo: Todos los demás sistemas de IA pueden desarrollarse y utilizarse con arreglo a la legislación vigente sin obligaciones jurídicas adicionales. La mayoría de los sistemas de IA utilizados actualmente en la UE pertenecen a esta categoría. De forma voluntaria, los proveedores de estos sistemas pueden optar por aplicar los requisitos de una IA digna de confianza y adherirse a códigos de conducta voluntarios.

---

<sup>52</sup> Comisión Europea. Proyecto de ley de inteligencia artificial de la UE. Disponible en: [https://ec.europa.eu/commission/presscorner/detail/es/QANDA\\_21\\_1683](https://ec.europa.eu/commission/presscorner/detail/es/QANDA_21_1683)

A los fines de la incorporación de esta tecnología a la justicia penal resulta necesario delimitar niveles de riesgo y establecer en que categoría ingresa el sistema a los fines de conocer que nivel de afectación representa para los derechos de los ciudadanos.

#### **IV. B.- Criterio de incorporación de la IA al proceso penal para el juez**

Los tribunales deberían establecer la verdad de los hechos en litigio y la verdad debería ser determinada tomando como base los medios de prueba relevantes y admisibles. Por consiguiente, los elementos de prueba se deberían concebir como el medio a ser usado para establecer la verdad.<sup>53</sup>

Aun así, ninguna verdad absoluta es alcanzable en el ámbito judicial, solo se puede alcanzar una verdad relativa, es decir, un grado de conocimiento que supuestamente se ha de situar en un punto entre 0 (ningún conocimiento fiable) y 1 (que indica certeza o verdad absoluta). Se supone que el “conocimiento intermedio” es incierto pero fiable.<sup>54</sup>

La aparición de métodos de investigación científico-tecnológica cada vez más sofisticados como los detallados y el aumento constante de los tipos de saber han generado un problema en lo relativo a la comparación de los hechos en un juicio y en la determinación de su valor como elemento probatorio. El problema es que no todo conocimiento presentado como científico merece de ser usado como prueba. Se trata, pues, de distinguir la “buena ciencia”, válida y fiable, de la “mala ciencia” o “junk science”.

Se considera “Buena Ciencia” a aquella que se encuentra validada y “mala ciencia” o “Junk Science” a aquella que no posee el nivel de validación requerido.

De la última encontramos dos categorías: aquella que es totalmente descartable y aquella que no es fiable y hay que tomar serios recaudos para su implementación.<sup>55</sup>

Asimismo, un buen método científico que es válido y correcto en sí mismo puede ser aplicado de manera incorrecta y, por ende, generar resultados carentes de valor cognoscitivo y probatorio.<sup>56</sup>

El desafío de la posible presencia de “junk science” en el proceso judicial ha sido afrontado en el sistema jurídico norteamericano. En concreto, se pretende dejar fuera del

---

<sup>53</sup> Taruffo, Michele: La prueba, Madrid et. Al., Marcial Pons, 2008, pp. 20.

<sup>54</sup> Taruffo, Michele: La prueba, Madrid et. Al., Marcial Pons, 2008, pp. 30.

<sup>55</sup> Métodos muy cuestionados son: el análisis microscópico del cabello, la prueba de incendio, el análisis comparativo de plomo de balas, odontología forense, pericia caligráfica, odología forense, huellas de zapatos o neumáticos y el polígrafo.

<sup>56</sup> Taruffo, Michele: La prueba, Madrid et. Al., Marcial Pons, 2008, pp. 282.

ámbito probatorio el uso de conocimientos que se presentan como científicos, pero que no corresponden a paradigmas de validez científica compartidos. Puntualmente, la Corte Suprema de los Estados Unidos desarrollo una serie de criterios a seguir en materia de admisibilidad de prueba que refieren directamente a este tema.

Si bien fue en el caso Daubert donde la Corte Suprema estadounidense afirmó que la “fiabilidad probatoria” era la “cientificidad de la prueba”, la enunciación judicial de ciertos parámetros para valorar las pruebas periciales tiene sus antecedentes en el caso Frye, en el que se planteó uno de los factores que luego se recogió en Daubert: gozar de la aceptación de la comunidad científica. Sin embargo, la discusión no se ciñe sólo a esos casos. Con posterioridad a Daubert, la Corte modificó sus criterios en otros dos casos, Joiner y Kumho, que conforman la conocida “trilogía Daubert”, y que no sólo generaron una reforma legislativa en los Estados Unidos, sino que han replanteado el análisis de la prueba pericial más allá de tales fronteras.<sup>57</sup>

El holding del caso Frye plantea que solo es admisible la prueba científica aportada si la ciencia en la que se basa ha ganado aceptación general en el área relevante. Ahora bien, el momento exacto en el cual un descubrimiento o principio científicos cruza la línea entre la fase experimental y la demostrable es difícil de definir. En algún lugar de esta zona de penumbra la fuerza probatoria del principio debe ser reconocida y aunque los tribunales contribuirán mucho al admitir testimonio experto derivado de un descubrimiento o principio científico bien establecido la cosa de la cual se deriva debe estar suficientemente establecida como para haber ganado aceptación general en el campo particular al que pertenece.

El criterio Frye fue considerado adecuado por parte de la judicatura y la doctrina porque podía ser fácilmente aplicado y sus resultados determinados también con relativa facilidad, sin necesidad de conocimientos especializados por parte del juzgador de los hechos.<sup>58</sup> Sin embargo, el mayor defecto de este criterio fue oscurecer la identificación del área de conocimiento relevante. En efecto, las críticas afirman que resulta precario y demasiado taxativo y giran en torno al siguiente interrogante: ¿qué quiere decir aceptación general en el área relevante?

---

<sup>57</sup> Vázquez, Carmen, De la prueba científica a la prueba pericial, Marcial Pons, 2015, pp. 85.

<sup>58</sup> Vázquez, Carmen, De la prueba científica a la prueba pericial, Marcial Pons, 2015, pp. 94.



Asimismo, el criterio Frye admite dos versiones: una amplia y una estricta. En la amplia se debería prestar atención al fundamento de la aceptación general en el área relevante, pues éste dice que los principios subyacentes deben estar lo suficientemente establecidos para ello (versión epistemológica del criterio). Mientras que, la estricta alude a la aceptación general en sí misma en el área relevante, sin hacer ninguna referencia a las razones de la aceptación (interpretación sociológica). En la versión epistemológica la aceptación podría ser contextual, de modo que algo puede ser aceptado o no dependiendo del propósito o la aplicación concreta.<sup>59</sup>

Otro problema es que este criterio se utilizó para analizar todo tipo de prueba pericial, incluso para las surgidas de las denominadas “ciencias novel” las cuales aun no eran admitidas dentro de la comunidad científica en general lo cual produjo serios problemas y genero que los tribunales dejaran de aplicarlo.

Posteriormente, el Caso Daubert agrega criterios nuevos y los hace no taxativos. Se trata de puntos a los que el juez debería atenerse para admitir o excluir los medios de prueba científica presentados por las partes. Hablamos de cuatro criterios: a) La controlabilidad y falseabilidad de la teoría científica o la técnica en la que se fundamenta la prueba<sup>60</sup>; b) El porcentaje de error conocido o potencial, así como el cumplimiento de los estándares correspondientes a la técnica empleada; c) La publicación en revistas sometidas al control de otros expertos, de la teoría o la técnica en cuestión; d) La existencia de un consenso general de la comunidad científica interesada.

Aunado a esto, el conocimiento científico en cuestión debe ser relevante y concretamente útil para decidir sobre los hechos del caso.<sup>61</sup>

Por otra parte, los jueces cumplen una función de arqueros al evaluar si el razonamiento y la metodología en el que se basa el testimonio es científicamente valido y si pueden ser adecuadamente aplicados a los casos en cuestión. Es decir, el juez es el custodio y debe conocer sobre el método de la prueba científica que se introducirá al proceso.

---

<sup>59</sup> Vázquez, Carmen, De la prueba científica a la prueba pericial, Marcial Pons, 2015, pp. 94.

<sup>60</sup> Validez científica del método (controlabilidad / testeada). Refiere a si la prueba ha sido o puede ser sometida a controles. El requisito es la no refutación (intentos por destruir el método que no han prosperado). Asimismo, la admisibilidad de la prueba científica debe basarse no solo en la relevancia sino también en la fiabilidad. La fiabilidad depende de la validez científica de la prueba que a su vez depende de la metodología. Para analizar la fiabilidad el juez debe considerar una serie de factores y no solo la aceptación general en el área relevante.

<sup>61</sup> Taruffo, Michele: La prueba, Madrid et. Al., Marcial Pons, 2008, pp. 283

Entonces se diferencian tres decisiones en el caso Daubert: a) El rol de *gatekeeper* de los jueces que deciden la admisión, b) La científicidad y/o la fiabilidad como criterio de admisión, c) la previsión de una etapa de admisión rígida.

Respecto al primer punto, es necesario reparar en que, en el sistema estadounidense, dada la presencia del jurado lego como juzgador de los hechos, se otorga importancia al reparto de facultades entre éste y el juez profesional que decide la admisión de las pruebas. Al juzgador de los hechos le corresponde la determinación de las cuestiones fácticas, el juez debe decidir todo aquello que es considerado como cuestión de derecho. Asimismo, de la mano del rol vigilante del juez está la previsión de una etapa de admisión restrictiva fundada en cierta desconfianza hacia el jurado lego bajo el argumento de que es fácilmente persuadido por aquellas pruebas que se presentan como científicas. Resulta entonces indispensable adoptar un paternalismo epistémico para proteger al juzgador de la formación de creencias equivocadas.<sup>62</sup>

Posteriormente, en lo que se conoce como la trilogía Daubert, dos casos ampliaron los criterios allí establecidos.

El caso Joiner estableció que no solo hay que analizar la metodología sino también el razonamiento de los peritos. Dicha jurisprudencia, fija los estándares de revisión manifestando que el criterio aplicable es el abuso de discreción del juez de primera instancia, quien goza de una amplia discreción para admitir o excluir pruebas. El revisor sólo puede revocar tales decisiones cuando se trata de un error manifiesto.

Por otra parte, el caso Kumho, extiende los criterios Daubert planteando que estos no rigen solo para la prueba científica sino para toda la prueba pericial.<sup>63</sup>

De este último fallo se destacan dos cuestiones: el abandono de la demarcación entre la ciencia y la no-ciencia, y las particularidades del caso como determinantes de los criterios para evaluar la fiabilidad de la prueba pericial.<sup>64</sup>

Los dictámenes emitidos por un sistema de IA pueden asimilarse a un peritaje realizado por un experto. En ese sentido, este sería el carácter con el cual deberían ingresar al proceso, como una prueba científica de tipo pericial.

---

<sup>62</sup> Por otra parte, una de las críticas más recurrentes a la Sentencia Daubert fue haber usado fiabilidad cuando en realidad estaba haciendo referencia a la validez. En todo caso, la Corte explicó que con fiabilidad probatoria estaba haciendo referencia a la confianza o la credibilidad (trustworthiness). Taruffo, Michele: La prueba, Madrid et. Al., Marcial Pons, 2008, pp. 100 y 101.

<sup>63</sup> En virtud de ello, y debido a que las pruebas pueden ser diferentes el juez debe tener una discrecionalidad enorme para decidir sobre la validez científica del método ya que habrá pruebas científicas, no científicas, conocimiento no especializado o basado en la experiencia, etc.

<sup>64</sup> Vázquez, Carmen, De la prueba científica a la prueba pericial, Marcial Pons, 2015, pp. 136.

Ahora bien, las posibilidades de incorporar este tipo de prueba científica al proceso son claras y existen estándares de admisibilidad, pero de ser ingresadas, ¿se viola alguna garantía constitucional?

En acápites anteriores ya quedo establecido que el imputado aporta un relato al proceso mediante los datos que se utilizan para entrenar a los sistemas de IA y que este relato tiene contenido declarativo. En este sentido, tomamos al imputado como órgano de prueba, para proceder al uso de métodos que puedan procesar gran cantidad de información sensible y personal del individuo necesitaremos como requisito su consentimiento libre, expreso y sin coacción.

Asimismo, el imputado debe estar plenamente informado del funcionamiento del sistema y de sus consecuencias, así como también, debidamente asesorado por su defensor para saber qué consecuencias podría tener su uso y considerar todas las opciones para garantizar su derecho de defensa.

Ahora bien, tal como se ha expuesto anteriormente, la garantía de la no autoincriminación forzada no surge pura y exclusivamente de nuestra Constitución, también esta presente en la declaración de Virginia<sup>65</sup>, en la Convención Americana sobre Derechos Humanos, en el artículo 8.2<sup>66</sup> y en el Pacto Internacional de Derechos Civiles y Políticos, en el artículo 14<sup>67</sup>.

Todos estos instrumentos internacionales fueron ratificados por nuestro país e incorporados a nuestra Constitución Nacional mediante el artículo 75 inciso 22.

A la luz de lo normado a nivel nacional e internacional, el imputado adquiere el derecho de: 1) Abstenerse de declarar (no pudiendo considerarse su negativa como una presunción en su contra), 2) Que su declaración sea voluntaria. Dicha voluntad no puede ser eliminada o menoscabada por ningún medio que pueda excluirla, por lo cual los instrumentos que registren reacciones inconscientes o reflejos condicionados están catalogados como métodos prohibidos ya que afectan la libertad de decidir aquello que se informa, 3) Tener libertad de decisión durante su declaración, la cual no puede ser en ningún caso o situación

---

<sup>65</sup> Virginia Declaration of Rights, 1776. Recuperado de: <https://www.archives.gov/founding-docs/virginia-declaration-of-rights>.

<sup>66</sup> Convención Americana sobre Derechos Humanos, 1969. Recuperado de: [http://www.oas.org/dil/esp/tratados\\_B-32\\_Convencion\\_Americana\\_sobre\\_Derechos\\_Humanos.htm](http://www.oas.org/dil/esp/tratados_B-32_Convencion_Americana_sobre_Derechos_Humanos.htm).

<sup>67</sup> Pacto Internacional de Derechos Civiles y Políticos, 1966. Recuperado de: <http://copredch.gob.gt/wp-content/uploads/Pacto-Internacional-de-Derechos-Civiles-y-Politicos.pdf>.

menoscabada ya sea por coacción física o moral, por la promesa ilegítima de una ventaja o por el engaño.

Solamente la declaración del imputado prestada bajo estas reglas podrá ser valorada por el juez para fundar su decisión. Sin embargo, si la declaración adolece de algún vicio por inobservancia de las reglas mencionadas, pero favorece al imputado y, como consecuencia de ella, la decisión es favorable a él (absolución), la valoración es inobjetable<sup>68</sup>. Esto es así debido a que ninguna garantía opera en perjuicio del imputado.

Otra exigencia fundamental es la presencia y asistencia del defensor en su declaración<sup>69</sup>.

Ahora bien, conforme lo analizado hasta el momento, ¿se encontrarían prohibidos los medios probatorios que aporta la IA al carecer el imputado de un control sobre la información que aporta al proceso? Desde mi punto de vista no estarían dentro del catálogo de medios probatorios prohibidos, pero, para que esto sea así, es necesario realizar ciertas aclaraciones: el imputado debe, como ya se dijo, prestar un consentimiento expreso, libre e informado y contar con la asistencia de su defensor. Además, con posterioridad al dictamen que efectúa el sistema, son ellos quienes decidirán si aportan o no la prueba obtenida al proceso, ello considerando que las garantías operan a favor del imputado y que, de ser perjudicial hacia su situación procesal debe tener la posibilidad de no aportar dicha prueba, sino se estaría violando la garantía que protege al sujeto contra la autoincriminación forzada.

En referencia al tema, el principio protege a la persona en tanto proporciona datos que operan en su contra en el procedimiento. Aun aplicado el método de manera inadmisibles, si el resultado es un dato que lo favorece, el elemento de prueba que de allí surge no solo puede, sino que debe ser valorado para su beneficio. Las garantías no pueden ser utilizadas en perjuicio del imputado. Si este pide su aplicación expresamente, como único modo de defenderse de una imputación cuyos signos exteriores no puede contrarrestar de otro modo, esta debe ser incorporada al proceso<sup>70</sup>.

Maier establece como punto de conflicto el consentimiento del imputado. Plantea como solución el admitir la utilización de estos métodos siempre que el propio imputado lo

---

<sup>68</sup> Maier, Julio B.J. *Derecho Procesal Penal. Tomo I, Fundamentos*. Ad-Hoc Buenos Aires. 2016. Pag. 664 y sgtes.

<sup>69</sup> “Francomano, Alberto José y otros s/ inf. ley 20.840.”, 1987, CSJN. Recuperado de: <http://www.sajj.gob.ar/corte-suprema-justicia-nacion-federal-ciudad-autonoma-buenos-aires-francomano-alberto-jose-otros-inf-ley-20840-fa87000462-1987-11-19/123456789-264-0007-8ots-eupmocsollaf>.

<sup>70</sup> Liszt, Franz von; Jimenez de Asua. *Tratado de derecho penal*. Reus, Madrid, 1927.

requiera, con información suficiente sobre sus derechos y asistencia técnica previa obligatoria (deliberación con su defensor).

Asimismo, el método debe estar autorizado científicamente bajo los criterios que se desarrollaron en la primera parte del presente acápite y debe ser ordinariamente utilizado en el tráfico social (permitido).

Por otra parte, el acto, sujeto a autorización judicial, debe perseguir como único fin el tornar viable la defensa del imputado que se frustraría de evitarse la aplicación de esa técnica. Por ello solo consciente como resultado una valoración única a favor del imputado prescindiendo de lo que le sea desfavorable. El acto se califica como una peritación y no reúne las características de declaración del imputado.

Los resultados serán presentados al defensor y su imputado y ellos decidirán si estos son incorporados o no al proceso<sup>71</sup>.

Por último, el Pacto Internacional de Derechos Civiles y políticos no prohíbe recurrir a estos métodos si se cuenta con el libre consentimiento del interesado<sup>72</sup>.

En virtud de lo expuesto la utilización de sistemas de IA se encontraría habilitada bajo las limitaciones, condiciones y consideraciones formales expuestas.

Ahora bien, la garantía solo ampara a una persona como sujeto u órgano de prueba, es decir, como quien con su relato incorpora al procedimiento un conocimiento cierto o probable sobre un objeto de prueba. No lo ampara cuando el sujeto es objeto de prueba, es decir, cuando es objeto investigado, por ejemplo, ante la extracción de sangre o de piel, actos que no consisten en proporcionar información referida al relato de los hechos. En estos casos, no es necesario el consentimiento de la persona afectada y puede ser forzada al examen<sup>73</sup>.

¿Por qué los sistemas de IA no se consideran como parte de este último grupo de métodos de examinación? A mi juicio, debido a que mediante ellas se extraen datos del sujeto que aportan directamente al relato de los hechos. No es una simple muestra de material biológico lo que se extrae, sino que se realiza un relato, por ejemplo, de factores de comportamiento, de elementos contextuales que llevarían al sujeto a actuar de determinada manera, se valoran datos que potencialmente podrían haber generado que el individuo se

---

<sup>71</sup> Maier, Julio B.J. *Derecho Procesal Penal. Tomo I, Fundamentos*. Ad-Hoc Buenos Aires. 2016. Pag. 672.

<sup>72</sup> Pacto Internacional de Derechos Civiles y Políticos, 1966. Recuperado de: <http://copredeh.gob.gt/wp-content/uploads/Pacto-Internacional-de-Derechos-Civiles-y-Políticos.pdf>.

<sup>73</sup> Maier, Julio B.J. *Derecho Procesal Penal. Tomo I, Fundamentos*. Ad-Hoc Buenos Aires. 2016. Pag. 675.

comporte como lo hizo, entre otros. No se trata de un simple muestreo sino de una técnica que refiere a un relato mediante datos extraídos de información personal del imputado con lo cual aquí el sujeto es órgano y no objeto de prueba.

En conclusión, queda claro que, por el principio de amplitud probatoria, está habilitado el uso de prueba científica, en este caso dictámenes de sistemas de IA que se asemejarían a un peritaje, en un proceso penal. Pero para que dicha prueba sea admisible es necesario que se cumplan con los criterios que se han desarrollado en materia de admisibilidad de prueba y con los límites y las condiciones que imponen las garantías constitucionales.

### **Conclusión**

El presente trabajo se propuso realizar un aporte novedoso al área del derecho penal mediante la articulación de criterios de aplicación que guíen a los magistrados a la hora de incorporar este tipo de tecnología a su trabajo jurisdiccional.

En la primera parte se establecieron puntos de tensión entre diversas garantías constitucionales del proceso penal y problemas surgidos de la incorporación de inteligencia artificial al ámbito de la justicia. Estos puntos de tensión versaron fundamentalmente sobre dos problemas concretos: la opacidad algorítmica y los sesgos que presenta la IA. Estos dos grandes tópicos se trabajaron desde diversos ángulos conforme la garantía que se tomó como base de análisis lo que permitió visualizar desde distintas perspectivas la tensión entre este tipo de tecnología y el derecho penal.

En la segunda parte del trabajo se abordó lo desarrollado a nivel normativo en el ámbito internacional respecto de la IA y su aplicación a nivel mundial. Estos preceptos legales se utilizaron como base normativa para el desarrollo de la tercera parte del trabajo y también permitieron advertir las lagunas legales que nuestro país posee en la materia.

En la tercera parte de esta investigación se delinearon criterios de aplicación de la IA al proceso penal. Para ello se buscaron posibles soluciones a los puntos de tensión advertidos en el primer apartado. Estas soluciones versaron sobre cuatro grandes categorías: la ética, la transparencia y las auditorías algorítmicas y la delimitación de diversos niveles de riesgo que esta tecnología presenta en su aplicación práctica.



Así se propusieron diversos criterios rectores a la hora de proceder a la utilización de IA en el proceso y se optó por dividir esta delimitación en dos grandes grupos: criterios para su incorporación en la justicia a nivel general y criterios para la incorporación de IA referidos al trabajo del juez en concreto. Ello en virtud de que se advirtió que era necesaria la incorporación de criterios generales que versaran sobre ética algorítmica y criterios específicos que giraran en torno a la valoración probatoria. De esta manera el problema inicial se abordó de manera global y se propusieron soluciones que abarcan las tensiones advertidas al comienzo de la investigación.

A través de lo analizado en el presente trabajo no se pretende brindar una solución al enorme problema legal que implican las lagunas existentes en el ámbito del derecho en materia de inteligencia artificial y derecho penal. Simplemente se pretende debatir sobre el tema y proponer ciertos puntos a tener en consideración a la hora de incorporar este tipo de tecnología en la justicia. En definitiva, se intenta brindar luz mediante la investigación académica a un área poco explorada a fin de aportar nuevos enfoques a desafíos surgidos del ámbito de la tecnología en su relación con el derecho penal invitando a que el presente trabajo favorezca a la apertura de más investigaciones en materia de IA y derecho que profundicen lo desarrollado y brinden mayor desarrollo doctrinal en lo relativo a IA y derecho a fin de garantizar a la persona sometida a proceso que sus derechos no se verán vulnerados por la aplicación de este tipo de tecnología al proceso.

La IA es una realidad, su desarrollo es veloz y se plasma en distintos ámbitos sociales. El derecho es uno de ellos pero, generalmente, su velocidad no es la misma que la de la tecnología. En virtud de ello, este tipo de trabajos devienen necesarios a fin de abordar esta materia que está en constante cambio y que se utiliza en procesos tan complejos como el penal. Se espera que este trabajo sea un aporte más a la materia que brinde un poco de luz sobre un tema tan complejo y actual como lo es la IA en materia penal.

### **Bibliografía**

- “¿Cómo en Estados Unidos las matemáticas te pueden meter en prisión?”  
Disponibile en <https://www.bbc.com/mundo/noticias-37679463>

- “*Francomano, Alberto José y otros s/ inf. ley 20.840.*”, 1987, CSJN. Recuperado de: <http://www.saij.gob.ar/corte-suprema-justicia-nacion-federal-ciudad-autonoma-buenos-aires-francomano-alberto-jose-otros-inf-ley-20840-fa87000462-1987-11-19/123456789-264-0007-8ots-eupmocsollaf>.
- “*Los contratos inteligentes ponen en jaque al mundo judicial*”, disponible en <http://www.telam.com.ar/notas/201610/167456-contratos-inteligente-mundo-judicial.html>
- “*Machine Bias*” disponible en <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Gil Domínguez Andrés. *Inteligencia Artificial y Derecho. Capítulo 1, El poshumanismo y los derechos*, Rubinzal Culzoni Editores, noviembre 2019
- MIT Technology Review. *Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo*, por Karen Hao, 8 de febrero de 2019
- MIT Technology Review. *Investigadores de IA dicen que los editores científicos ayudan a perpetuar algoritmos racistas*, 23 de junio de 2020
- Revista Science. *Incluso la inteligencia artificial puede adquirir prejuicios contra la raza y el género*, 13 de abril de 2017
- The Technolawgist. *Sesgo en la toma de decisiones judiciales de IA*, 31 de mayo de 2019
- Allen & Overy News, “*Allen & Overy and Deloitte tackle OTC derivatives market challenge*”, recuperado de <http://www.allenoverly.com/news/en-gb/articles/Pages/AllenOvery-and-Deloitte-tackle-OTC-derivatives-market-challenge.aspx> citado por Escobar, Felicitas, “*La IA como camino hacia un estándar de prueba objetivo*”, ponencia ganadora del primer puesto en el concurso de ponencias del Congreso Colombiano de Derecho Procesal XL, realizado en la ciudad de Medellín los días 4, 5 y 6 de septiembre de 2019.
- BBC Mundo: “*¿Es posible predecir un delito?: el controversial algoritmo usado por la policía de Chicago, la ciudad con más criminalidad de Estados Unidos*”, BBC, 2017.
- Binder, Alberto M. *Introducción al derecho procesal Penal*. Buenos Aires, Ad Hoc, 2013.

- Bourcier, D. *Inteligencia artificial y derecho*. Barcelona, UOC, 2003.
- Buchanan, B. y Headrick, T: “*Some speculation about artificial intelligence and legal reasoning*”. California, Stanford Law Review, vol. 23, 1970.
- Bujosa Vedell, L. *La valoración de la prueba electrónica, estudios sobre nuevas tecnologías y justicia*. Comares, Granada, 2015.
- Buranyi, S: “*Rise of the racist robots – how AI is learning all our worst impulses*”, The guardian, 2017.
- Cabeza De Vaca Hernández, Daniel. “*IA aplicada a la justicia*”.
- Cayuso, S. G. (2009). *El principio de igualdad: problemas e interrogantes: el sistema constitucional argentino*. Revista de Derecho Político.
- Chirino Sánchez, A. *Las tecnologías de la información y el proceso penal*. NEXOS Costa Rica, 2015.
- Claria Olmedo, Jorge A., *Derecho procesal penal, Tomo I*, Buenos Aires, Rubinzal-Culzoni, 1998, págs. 273 a 275.
- Comisión Europea: “*Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre las órdenes europeas de entrega y conservación de pruebas electrónicas a efectos de enjuiciamiento penal*”, 2018.
- Comisión Europea. *Estudio sobre el uso de tecnologías innovadoras en el campo de la Justicia*, septiembre 2020.
- Comisión Europea. *Informe del grupo de expertos sobre responsabilidad y nuevas tecnologías. Responsabilidad por la inteligencia artificial y otras tecnologías digitales emergentes*.
- Consejo de Europa. *Carta ética europea sobre el uso de inteligencia artificial en los sistemas de justicia y su entorno*, 3-4 de diciembre de 2018.
- Convención Americana sobre Derechos Humanos, 1969. Recuperado de: [http://www.oas.org/dil/esp/tratados\\_B-32\\_Convencion\\_Americana\\_sobre\\_Derechos\\_Humanos.htm](http://www.oas.org/dil/esp/tratados_B-32_Convencion_Americana_sobre_Derechos_Humanos.htm).
- CORTE IDH, caso “*Manuela y otros Vs. El Salvador*”, (Excepciones Preliminares, Fondo, Reparaciones y Costas), Sentencia del 2 de noviembre de 2021, parr. 155.
- Corvalan, Juan G. “*Inteligencia artificial al servicio de la justicia penal, contravencional y de faltas. Prometea, en el ámbito de la Ciudad de Buenos Aires*”,

publicado en la obra de DUPUY Daniela; KIEFER Mariana. *Cibercrimen II*; B de F, Buenos Aires, 2018.

- Corvalan, Juan; Dupuy, Daniela; Kiefer, Mariana. *Cibercrimen III*. Ed. B de F. Buenos Aires, 2020.
- Corvalan, Juan. “*El peligro de la inteligencia artificial como oráculo del sistema penal.*” Consultado en: <https://www.infobae.com/opinion/2017/08/30/el-peligro-de-la-inteligencia-artificial-como-oraculo-del-sistema-penal/>
- Danesi Cecilia, Mitelli Noelia. *Asistentes jurídicos digitales: el impacto de la inteligencia artificial en el proceso.*
- Danziger, S. “*Extraneous factors in judicial decisions*”, PNAS, 2011. Consultado en: <https://www.pnas.org/content/108/17/6889>
- Davey, M. “*Chicago police try to predict who may shoot or be shot*”, The New York Times, 2016.
- Delgado, M. *La inteligencia artificial. Realidad de un mito moderno*, Granada, Servicio de Publicaciones de la Universidad de Granada, 1996.
- *Diccionario de la Real Academia Española*, disponible al 23/11/2020 en <https://dle.rae.es/inteligencia>
- Dieterich, W. “*COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*”, Northpointe, 2016. Consultado en: [https://go.volarisgroup.com/rs/430MBX989/images/ProPublica\\_Commentary\\_Final\\_070616.pdf](https://go.volarisgroup.com/rs/430MBX989/images/ProPublica_Commentary_Final_070616.pdf)
- EQUIVANT, *Practitioner’s Guide to COMPAS Core*, 2019, p. 31, disponible en <https://www.equivant.com/practitioners-guide-to-compascore/#:~:text=The%20Practitioner's%20Guide%20provides%20an,system%20for%20criminal%20justice%20practitioners.>
- ESCOBAR, Felicitas, “*La IA como camino hacia un estándar de prueba objetivo*”, ponencia que obtuvo el primer puesto en el Concurso de Semilleros del XL Congreso Colombiano de Derecho Procesal, Medellín, 4, 5 y 6 septiembre de 2019.
- Fernandez Sosa, Macarena. *El Principio De Igualdad*. Disponible en: <https://www.aiddp.com/wp-content/uploads/2021/10/Fernandez-Sosa-Macarena.-Universidad-Nacional-de-Cuyo.pdf>

- Florian, Eugene. *Elementos de derecho procesal penal*. Bosch, Barcelona.1933.
- Gil Domínguez Andrés. *Inteligencia Artificial y Derecho. Capítulo 1, El poshumanismo y los derechos*, Rubinzal Culzoni Editores, noviembre 2019.
- Güemes María Belén. *Tiempos de cambios en los procesos judiciales*, La Ley, 31 de julio de 2020
- Harari, Yuval Noah, *Homo Deus Breve Historia del Mañana*, Ed. Debate, 2015, Buenos Aires, pág. 100.
- <https://morethanlaw.es/blog/abogacia-innovacion-y-el-futuro-de-laprofesion>
- <https://www.acm.org/about-acm/mission-vision-values-goals>
- [https://www.gdprsummary.com/?gclid=EAIaIQobChMIxomGiv349gIVwwiRCh1F\\_wv0EAMYAyAAEgLASfD\\_BwE](https://www.gdprsummary.com/?gclid=EAIaIQobChMIxomGiv349gIVwwiRCh1F_wv0EAMYAyAAEgLASfD_BwE)
- Klein, Konstantin. “*Inteligencia artificial en los tribunales: ¿cómo funciona?*” Consultado en: [https://www.dw.com/es/inteligenciaartificial-en-los-tribunales cómo-funciona/a-47185445](https://www.dw.com/es/inteligenciaartificial-en-los-tribunales-cómo-funciona/a-47185445)
- La Rosa, Mariano. “*Defensa en juicio, debido proceso y juicio por jurados*”. Disponible en: <https://www.pensamientopenal.com.ar/system/files/2016/02/doctrina42882.pdf>
- Liszt, Franz von; Jimenez de Asua. *Tratado de derecho penal*. Reus, Madrid, 1927.
- Maier, Julio B.J; *Derecho Procesal Penal. Tomo I, Fundamentos*. Buenos Aires, Editores del Puerto s.r.l, 2004.
- Maldonado, Gabriela Leites. *La subjetividad humana frente al desafío de la ia*. Disponible en: [http://www.razonypalabra.org.mx/Razonarte/N82/Leites\\_Columna\\_RA82.pdf](http://www.razonypalabra.org.mx/Razonarte/N82/Leites_Columna_RA82.pdf)
- Martínez Garay, L. (2018). *Peligrosidad, algoritmos y due process: el caso State v Loomis*. Revista de Derecho Penal y Criminología (uned), 2018, num. 20, p. 485-502.
- Maybin, Simon. “*¿Cómo en Estados Unidos las matemáticas te pueden meter en prisión?*”. Consultado en: <https://www.bbc.com/mundo/noticias-37679463>
- Mendoza, D. (2020). *Racismo y roles de género, conductas perpetuadas en algoritmos de IA*. COLOQUIO, 131-135.

- Miguel BerlaIn, I., & Pérez Estrada, M. J. (2019). *La IA en el proceso penal español: un análisis de su admisibilidad sobre la base de los derechos fundamentales implicados*.
- Nieva Fenoll, Jordi, *IA y proceso judicial*, Marcial Pons, Madrid, 2018, pág. 100/1.
- Oliva León, Ricardo. “*La inteligencia artificial en el sector legal*”. Consultado en: <https://www.algoritmolegal.com/tecnologias-disruptivas/la-inteligencia-artificial-en-el-sector-legal/>
- Pacto Internacional de Derechos Civiles y Políticos, 1966. Recuperado de: <http://copredeb.gob.gt/wp-content/uploads/Pacto-Internacional-de-Derechos-Civiles-y-Politicos.pdf>.
- Pardo P. “*Por qué la Inteligencia Artificial necesita clases de ética: “Hemos creado un Frankenstein”*”, El Mundo, 2018.
- Peralta, José, “*Nemo tenetur y derecho procesal penal preventivo, acerca de la posible relevancia epistémica del derecho a no auto incriminarse*”, en Ambos y Malarino (dirs.) *Prevención e imputación*, Hammurabi, Buenos Aires.
- Piqué, María Luisa y Fernández Valle, Mariano; *La garantía de imparcialidad judicial desde la perspectiva de género*, publicado en Herrera, Marisa, Fernandez, Silvia E., De La Torre, Natalia (Directoras generales); *Tratado de géneros, derechos y justicia. Derecho penal y sistema judicial. Tomo I*, 1º edición, Rubinzal Culzoni, Santa Fe, 2020, p. 124
- Ramírez, Nola Gómez. *Análisis de los principios del derecho penal*. 2004.
- Reglamento (Ue) 2016/679 Del Parlamento Europeo y Del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). Disponible en: <https://www.boe.es/doue/2016/119/L00001-00088.pdf>
- Rusconi, Maximiliano Adolfo. *Principio de inocencia e" in dubio pro reo"*. *Jueces para la Democracia*, 1998, no 33, p. 44-68.
- Salas, Javier. “*El algoritmo que ‘adivina’ los delitos futuros falla tanto como un humano*”. Consultado en: [https://elpais.com/elpais/2018/01/17/ciencia/1516194073\\_122982.html](https://elpais.com/elpais/2018/01/17/ciencia/1516194073_122982.html)



- Sánchez Caparrós, M. *Prevenir y controlar la discriminación algorítmica* disponible en: <https://www.justierradelfuego.gov.ar/wp-content/uploads/2021/07/Sánchez-Caparrós.-Prevenir-y-controlar-la-discriminación-algor%C3%ADtmica.pdf>
- Santos González, M. “*Regulación legal de la robótica y la inteligencia artificial: retos de futuro*”, Revista Jurídica de la Universidad de León, vol. 4, 2017.
- Stanford Law School Policy Lab, *Risk Assesment Factsheet. Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) Pretrial Release Risk Scale - II (PRRS-II)*, 2019, disponible en <https://law.stanford.edu/pretrial-risk-assessment-tools-factsheet-project/>
- State v. Loomis. 881 N.W. 2d 747.753, 2016.
- Taruffo, Michele. *La prueba*, Madrid et. Al., Marcial Pons, 2008
- Tessone, Alberto J.: «El deber de motivación de las sentencias», en Revista Jurisprudencia Argentina N.º 5706, p. 5. Disponible en: [http://www.saij.gob.ar/doctrina/daca900130-tessone-deber\\_motivacion\\_las\\_sentencias.htm](http://www.saij.gob.ar/doctrina/daca900130-tessone-deber_motivacion_las_sentencias.htm).
- Urueña, René. *Autoridad algorítmica: ¿ cómo empezar a pensar la protección de los derechos humanos en la era del “big data”?*. Latin American Law Review, 2019, no 2, p. 99-124.
- Valls Prieto, J. *Problemas jurídico penales asociados a las nuevas técnicas de prevención y persecución del crimen mediante inteligencia artificial*. Dykinson, Madrid, 2017.
- Vázquez, Carmen, *De la prueba científica a la prueba pericial*, Marcial Pons, 2015
- Vélez, M. I., Gómez Santamaría, C., & Osorio Sanabria, M. A. (2022, June 30). *Conceptos fundamentales y uso responsable de la IA en el sector público*. Informe 2. Caracas: CAF. Disponible en: <http://cafscioteca.azurewebsites.net/handle/123456789/1921>
- Vestri, G. (2021). La inteligencia artificial ante el desafío de la transparencia algorítmica: Una aproximación desde la perspectiva jurídico-administrativa. *Revista Aragonesa de Administración Pública*, (56), 368-398.
- Virginia Declaration of Rights, 1776. Recuperado de: <https://www.archives.gov/founding-docs/virginia-declaration-of-rights>