El Uso de la Inteligencia Artificial en la Toma de Decisiones Judiciales: Un Estudio a través del Caso COMPAS

Nuria Fernández Piñeiro[1]
Estudiante de Administración y Dirección de
Empresas y Derecho de la Universidade da Coruña

Resumen

En el presente trabajo abordaremos los retos jurídicos que supone el avance tecnológico derivado de la reciente incorporación de la Inteligencia Artificial a nuestras vidas. Concretamente, trataremos el uso de métodos de Inteligencia Artificial en el proceso judicial, tanto desde la perspectiva del "juez IA" como del uso auxiliar de estas tecnologías por jueces y magistrados. Las aplicaciones y beneficios de la Inteligencia Artificial son muchos, ahora bien, también resulta necesario sopesar sus posibles riesgos. Para esta tarea nos centraremos en el caso COMPAS, analizando su funcionamiento, las posibles violaciones de derechos fundamentales generadas por su uso y la opinión jurisprudencial acerca del mismo, creando un supuesto de estudio desde la perspectiva del derecho continental.

Palabras clave: Inteligencia artificial, proceso judicial, Administración de justicia, COMPAS

[1] D. Nuria Fernández Piñeiro, alumna del Programa de Simultaneidad del Grado en Administración y Dirección de Empresas y el Grado en Derecho en la Universidade da Coruña, destaca por su formación académica orientada hacia el Derecho empresarial y los desafíos jurídicos que plantea la Inteligencia Artificial. Ejemplo de ello es su más reciente actividad investigadora de donde nace este estudio junto otros como el artículo "Inteligencia Artificial y Propiedad Industrial: Estudio del Concepto de Inventor a través del Caso DABUS" (2023), donde explora desde la óptica del Derecho comparado el impacto de la IA en el sistema de patentes español.

Abstract

In this paper we will address the legal challenges posed by the technological advances derived from the recent incorporation of Artificial Intelligence into our lives. Specifically, we will deal with the use of Artificial Intelligence methods in the judicial process, both from the perspective of the "AI judge" and the auxiliary use of these technologies by judges and magistrates. The applications and benefits of Artificial Intelligence are many, however, it is also necessary to weigh its possible risks. For this task we will focus on the COMPAS case, analysing its operation, the possible violations of fundamental rights generated by its use and the jurisprudential opinion about it, creating a case study from the perspective of continental law.

Keywords: Artificial intelligence, judicial process, Administration of Justice, COMPAS.

Lista de Abreviaturas y Siglas

IA: Inteligencia Artificial

COMPAS: Correctional Offender Management Profiling for Alternative Sanctions

UE: Unión Europea

RISCANVI: Protocolo de valoración del riesgo

SAVRY: Structured Assessment of Violence Risk in Youth

VioGén: Sistema de Seguimiento Integral en los casos de violencia de género

CDCR: California Department of Corrections and Rehabilitation

INE: Instituto Nacional de Estadística

1. Introducción

El avance tecnológico y científico dado desde principios del siglo XXI ha supuesto un cambio radical en la vida del ser humano y en su entorno. Actualmente, el uso de algoritmos y de métodos de Inteligencia Artificial (en adelante IA) se ha generalizado. Tal es así que podríamos llegar a afirmar que cualquier persona está rodeada de IA desde que se despierta hasta que se acuesta. Un buen ejemplo de esto son los coches dotados de IA, los filtros de *spam* de correo electrónico o las listas de recomendaciones musicales de aplicaciones como *Spotify*.

Todas estas funcionalidades tienen algo en común: el uso de algoritmos y el aprendizaje automático. En la actualidad no existe una definición consensuada de lo que realmente es IA, sin embargo, podemos tomar la definición propuesta en las *Directrices del Grupo Independiente de Expertos de alto nivel en Inteligencia Artificial* creado por la Comisión Europea en junio de 2018: la IA se compone de

> *"sistemas de software (y en algunos casos también de hardware) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno a través de la obtención de datos, la interpretación de los datos estructurados o no estructurados que recopilan, el razonamiento sobre el conocimiento o el procesamiento de la información derivados de esos datos, y decidiendo la acción o acciones óptimas que deben llevar a cabo para lograr el objetivo establecido. Los sistemas de IA pueden utilizar normas simbólicas o aprender un modelo numérico; también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores".*[2]

De esta posible definición extraemos las dos principales características de la IA. Por un lado nos encontramos con la autonomía, esto es la capacidad con la que cuenta el método de IA para ejecutar una o varias tareas en situaciones complejas sin la dirección constante del usuario humano. Por otra parte la adaptabilidad, que se refiere a la capacidad para mejorar la ejecución de las tareas aprendiendo de la experiencia. De esta manera podemos afirmar que una vez elaborado el método concreto de IA, proponiéndole una tarea o meta y suministrándole una serie de datos que sirvan para su entrenamiento este podrá funcionar de manera libre sin la intervención humana.

Este fenómeno se produce principalmente por lo que denominamos aprendizaje automático.[3] En este caso el algoritmo mejora su comportamiento en la ejecución de la tarea gracias a la acumulación de experiencia y datos. Es decir, lo que los ingenieros y científicos han procurado es simular el comportamiento del cerebro humano, que mejora sus procesamientos a través del aprendizaje mediante la reiteración y la búsqueda de nuevas fuentes de información.

La generalización del uso de la IA en nuestra vida cotidiana nos lleva a reflexionar sobre su incorporación a otros ámbitos como podría ser la toma de decisiones judiciales. En un sistema

[2] Grupo Independiente de Expertos de alto nivel sobre Inteligencia artificial. *Directrices Éticas para una IA Fiable*.
[3] María del Mar Maroño, *El Concepto de Inventor en el Derecho de Patentes y los Sistemas de Inteligencia Artificial*. Cuadernos de Derecho Transnacional, vol. 12 (Nº 2), p. 513.

judicial tan colapsado como el español convendría pensar en la agilización de la justicia mediante el empleo de métodos de IA que se ocupasen de la búsqueda de jurisprudencia, de la redacción de autos y sentencias o, mucho más importante, de la propia toma de decisiones, hoy efectuada por jueces y magistrados.

Es por ello por lo que a nuestro parecer resulta interesante dedicar este trabajo al estudio de la posibilidad de la incorporación de la IA al proceso judicial, tanto en el papel del juez o magistrado, como auxiliando a estos en su toma de decisiones, sin perder nunca de vista los posibles peligros que cualquier tecnología puede desentrañar.

2. Aplicación de los Algoritmos a la Toma de Decisiones Judiciales

Ilusos seríamos si dijésemos que la aplicación de la IA en el proceso judicial es un fenómeno totalmente nuevo. Si bien son muy recientes, ya podemos ver manifestaciones del uso de IA en la toma de decisiones por parte de jueces y magistrados.

Su uso se limita básicamente al ámbito penal, donde afloran sistemas sobre la evaluación del riesgo que suponen los reos. Sin ir más lejos, el sistema COMPAS, al que nos dedicaremos con posterioridad, es uno de estos métodos. Así pues, en esta sección de nuestra investigación nos proponemos dilucidar la posibilidad de la aplicación de la IA en los procesos de toma de decisiones judiciales.

1. El "Juez Artificial"

El "juez artificial" es un término que se refiere a la toma de decisiones judiciales de manera autónoma e independiente por medio de un sistema de IA. Es decir, se daría la sustitución del juez o el magistrado humanos por la máquina, lo que podría conllevar notables mejoras en el proceso judicial.

En primer lugar hemos de recordar nuestra premisa; el sistema judicial español se encuentra congestionado. Por poner un ejemplo, según las estimaciones del Consejo General del Poder Judicial,[4] un proceso abreviado en los Juzgados de lo Penal tarda unos 11.1 meses en resolverse (datos 2023).

Una posible solución a esta aglomeración de casos sin que el proceso tenga que verse simplificado o modificado en demasía sería el uso de métodos de IA en la toma de decisiones. El "juez artificial" emplearía una serie de datos que le servirían como base a su conocimiento a partir de los cuales llevaría a cabo un procedimiento de construcción de inferencias para calcular el peso o valor probatorio de un conjunto de pruebas en un proceso judicial.[5] Al "juez artificial" se le darían los datos principales de los escritos iniciadores del proceso, pongamos por caso una

[4] Datos de la página oficial del Consejo General del Poder Judicial. (2025).

[5] Orión Vargas. *Juez Inteligente. Sistema Experto que Asiste al Juez en la Valoración de la Prueba Judicial.* Revista Derecho & Sociedad (Nº. 57), pp. 1-24.

demanda y su contestación. En nuestro ejemplo estos elementos principales son: el *petitum* del demandante, los hechos jurídicamente relevantes conectados favorablemente con esta pretensión y el conjunto de pruebas que lo respalden. A partir de aquí el sistema comenzaría a interrelacionar todos los datos, estableciendo conexiones de condicionalidad y complementariedad entre ellos hasta llegar al resultado.

Por otro lado, el uso de la IA en el proceso judicial también podría aumentar la confianza ciudadana en nuestro sistema judicial puesto que contaríamos con la racionalidad y objetividad máxima. A pesar de que jueces y magistrados han de ser independientes e imparciales, no dejan de ser personas humanas, por lo que el público general puede desconfiar del sistema en su conjunto por la corrupción de una parte del Poder Judicial. Por ello, la incorporación de una máquina sin sentimientos ni intereses podría ayudarnos a lograr una mayor confianza y estabilidad de nuestro sistema.

Pese a la simplicidad aparente del asunto hemos de apuntar que este proceso de toma de decisiones resulta de alta complejidad debido a su posible confrontación con múltiples derechos fundamentales. Y es que no hemos de olvidar que un sistema de IA no deja de ser una máquina y por tanto carece de la racionalidad humana para comprender las consecuencias de sus actos.

1. La Individualización de la Sentencia

Antes de nada hemos de rechazar una idea que nos podría llevar a error en nuestra investigación: la IA es inteligente.[6] Con este propósito hemos de analizar dos casos de estudio, el Test de Turing y el experimento de la habitación china de John Searle.[7]

El Test de Turing es un experimento que pretende comprobar si un método de IA podría suplantar la identidad humana. Para ello se establece una conversación entre tres sujetos; por un lado estarían A y B, A sería una persona corriente, mientras que B sería un método de IA; por otro lado se encontraría el interrogador que sería un humano. A y B se comunicarán con el interrogador mediante mensajes escritos y el interrogador intentará dilucidar cuál de sus dos interlocutores es una máquina. En el caso de que el investigador no fuera capaz de distinguir a la máquina del humano se podría decir que la máquina ha alcanzado una inteligencia equiparable a la de un ser humano.

Este test fue realizado en varias ocasiones con un bot conversacional denominado Eugene Goostman que se hacía pasar por un niño ucraniano de trece años que continuamente se disculpaba con los usuarios por su gramática y su nivel de conocimiento. El resultado del test siempre fue muy parecido en los diferentes experimentos en que se implantó, engañando a unos diez jueces, que pensaron que era una persona real, de los treinta partícipes.

---

[6] Dyango Bonsignore. *Sobre Inteligencia Artificial, Decisiones Judiciales y Vacíos de Argumentación.* Teoría & Derecho (Nº 29), pp. 248-277.

[7] MinnaLearn, Universidad de Helsinki. (2018). Elements of IA. https://www.elementsofai.com/es.

Sin embargo, se han expuesto numerosas críticas al test de Turing que sostienen que el decir cosas inteligentes no te convierte en inteligente. El test de Turing simplemente cae en la imitación de la inteligencia humana, pero no refleja una verdadera inteligencia por parte del sistema de IA. Prueba de esto es el experimento de la habitación china de John Searle.

El experimento consiste en que hay una persona que no sabe chino encerrada en una habitación, pero que cuenta con un manual con todas las instrucciones necesarias para responder a la perfección a una conversación en chino. Al otro lado hay otra persona que puede introducir notas escritas en chino a través de un buzón. La persona que escribe las notas recibe contestaciones en chino sin ningún tipo de error, lo que le hace pensar que está conversando con una persona que comprende y habla su idioma, sin embargo la persona de la habitación no tiene ningún tipo de conocimiento sobre esta lengua.

Lo que se nos intenta hacer ver con este experimento es que que una persona tenga la impresión de estar comunicándose con otra que habla chino no significa que ese individuo realmente sepa chino. De la misma manera, si una máquina se comporta de manera inteligente, no implica que lo sea realmente, simplemente es capaz de realizar una imitación aparentando ser inteligente sin tener que serlo.

El motivo por el que hemos traído a coalición estos dos experimentos es el reflexionar sobre las capacidades de un método de IA y el papel del juez y del magistrado. Este tipo de máquinas funcionan mediante códigos y algoritmos prefijados donde muchas veces es difícil incorporar las sutilezas y los múltiples sentidos del lenguaje humano, lo que nos lleva a plantear el primer problema de este tipo de instrumentos.

Como sabemos la tarea fundamental del juez es la adaptación de la ley al caso concreto. Es el poder legislativo el que crea el sentido literal de la ley, sin embargo la tarea del juez o magistrado no consiste en la mera transcripción de la norma a la hora de dictar sentencia, sino que ha de ser capaz de interpretar su sentido para ver si esta resulta aplicable en el caso concreto, y si así es, la forma más o menos laxa de aplicación. De esta función deriva el derecho a la individualización de la sentencia. A nuestro parecer resulta complicado que un sistema de IA pueda llegar a adaptar la ley al caso concreto pues hemos de recordar que su funcionamiento se basa en interconexiones lógicas y tal rigidez parece hacer imposible la adaptación. Por otra parte, un algoritmo de esta clase está formado por sintaxis,[8] es decir, es capaz de crear estructuras formales alrededor de la operación para resolverla, sin embargo, carece de semántica, por lo que no comprende el significado de las operaciones que está realizando, lo que nos puede llevar a resultados automáticos de la clase si A entonces B que no permitan la interposición de ese margen de discrecionalidad tan necesario con el que cuenta el juez a la hora de interpretar la ley y aplicarla al caso concreto, por lo que pensamos que este derecho a la tutela judicial efectiva consagrado en el

[8] Álvaro Pérez. *Justicia Civil en la Era Digital y Artificial: ¿Hacia una Nueva Identidad?* Revista Chilena de Derecho, vol. *48 (Nº2),* pp. 203-209.

artículo 24 de nuestra Constitución se podría ver malogrado y con él la individualización de la sentencia.

Otra circunstancia que atacaría a esta individualización sería el uso de datos masivos. Para llegar a un resultado la IA se dota de datos sobre multitud de casos, jurisprudencia y estadísticas. De esta manera podemos decir que si en el proceso participase el "juez artificial" no se estaría tratando el caso concreto con sus matices, sino que se realizaría una comparativa entre el presente caso y otros pasados. Evidentemente Juzgados y Tribunales se valen de jurisprudencia anterior para respaldar sus decisiones actuales e incluso buscar sendas resolutivas de estos nuevos casos a través de casos pasados, sin embargo la gran diferencia radica en que el juez o magistrado siempre puede cambiar su criterio al considerar que ha habido un avance en la norma o que este nuevo caso no se llega a identificar totalmente con los anteriores. Una máquina nunca podrá hacer esto pues se basa en unos patrones de búsqueda fijos y en unas interconexiones que necesariamente la han de llevar a un único resultado.

2. **La Cristalización de la Jurisprudencia**

Que un sistema de IA no pueda realizar una adecuada individualización de las sentencias conlleva un nuevo peligro: la cristalización de la jurisprudencia. Como hemos dicho con anterioridad es el Poder Legislativo el encargado de la creación de la ley, pero no podemos olvidar que el marco legal se expande a través de la función judicial, y es que, aunque los jueces y magistrados no tienen una función creativa reconocida, no podemos negar que la jurisprudencia no deja de ser una interpretación de la norma que permite la estabilidad de la misma en el tiempo al modificar su sentido literal según la corriente social y política del momento. De esta forma, si cayésemos en una cristalización jurisprudencial sería mucho más probable que las Cámaras Legislativas tuviesen que aumentar su carga de trabajo para dar a luz nuevas regulaciones que dejasen sin vigencia las normas anteriores y se adaptasen a los nuevos fenómenos sociales, lo que a su vez crearía una situación de inestabilidad permanente por el continuo cambio normativo.

3. **Opacidad del Algoritmo y Principio de Contradicción**

Continuando con la temática de la tutela judicial efectiva hemos de hacer referencia al principio de contradicción y a la falta de transparencia de los algoritmos que componen los sistemas de IA. Con este propósito primero nos detendremos en comprender cuál es el funcionamiento de esta clase de algoritmos, en segundo lugar qué es el aprendizaje automático y por último qué es el aprendizaje profundo.

Gran parte de las formas de IA de hoy en día están basadas en un mecanismo de aprendizaje automático por el cual mediante el procesamiento de los datos el algoritmo es capaz de aprender a reconocer patrones y formular reglas de actuación sin estar programado para ello. Se podría decir que el algoritmo resulta autodidacta. Este tipo de aprendizaje muchas veces se lleva a cabo a través de "redes neuronales" que vienen a funcionar tal y como lo harían nuestras propias neuronas naturales. Así pues, podemos describir la estructura de las redes neuronales de la siguiente manera: en primer lugar nos encontramos con las dendritas a las que se une la neurona,

mediante estas se da una entrada de datos en la neurona donde se procesa la información que acaba saliendo por el axón del sistema que a su vez está conectado con otra dendrita, de modo que la información cada vez más procesada va saltando de capa en capa de neuronas mediante un proceso denominado sinapsis. Podemos imaginarnos este proceso como un gran árbol con múltiples ramificaciones que van ascendiendo hasta el cielo sin que seamos capaces de identificarlas todas ni su final. El uso de estas múltiples capas nos lleva al denominado aprendizaje profundo, una técnica especial de aprendizaje automático que precisamente se caracteriza por la multitud de capas de unidades de procesamiento sencillas (las neuronas) conectadas en red que permite el aumento de sus capacidades de análisis y deducción.

Una de las ventajas o desventajas de estos sistemas de aprendizaje profundo es que, frente al mero aprendizaje automático "en el que el ingeniero tiene que seleccionar las características que han de ser tenidas en cuenta para llegar a un determinado resultado, las redes neuronales profundas pueden determinar de forma automática las características relevantes a partir de datos sin procesar, porque al profundizar en el número de capas, las redes aprenden a reconocer las características que son útiles y cómo utilizarlas para resolver las cuestiones planteadas. Se dice así, que el programa puede aprender a realizar actuaciones inteligentes fuera de la noción programada; puede aprender por sí mismo e irse adaptando en función de los retos planteados".[9]

El aprendizaje profundo puede resultar una desventaja debido a que puede dar lugar a los denominados "algoritmos de caja negra".[10] Los "algoritmos de caja negra" se caracterizan porque, aun conociendo el código fuente o el diseño del algoritmo, resultaría difícil incluso para los expertos mostrar cómo funcionan por la profundidad del sistema.

Debido a lo anterior nos encontramos ante dos escenarios. Por un lado estaría el desconocimiento del código fuente del algoritmo, y es que la mayoría de métodos de IA utilizados hoy en día en los procesos judiciales han sido desarrollados por empresas privadas, las cuales en última instancia pretenden la consecución de un beneficio económico, por lo que protegen estas expectativas mediante el secreto empresarial, ocultando de esta manera el código fuente del algoritmo. En el segundo escenario nos encontramos con los denominados algoritmos de caja negra que también impedirían la comprensión del código fuente ante el desconocimiento de su funcionamiento.

La problemática de la imposible comprensión del código fuente en ambos casos reside en que si el mecanismo de toma de decisiones es opaco el acusado en ningún caso podrá defenderse pues no conoce las variables de entrada ni los pesos que se le dará a cada una de ellas.

---

[9] Maroño, *El Concepto de Inventor en el Derecho de Patentes y los Sistemas de Inteligencia Artificial.* Cuadernos de Derecho Transnacional, vol. 12 (Nº 2), p. 513.

[10] Raúl César Cancio. *¿Sueñan los Jueces con Sentencias Electrónicas?* Análisis Jurídico-Político, volumen 2 (Nº.3), pp. 145-168.

Para resolver el primer escenario nos encontramos con diversas propuestas, como podrían ser la renuncia de las empresas a su secreto profesional (cosa que difícil nos parece si no existen unos incentivos adecuados por parte de las autoridades), el desarrollo por la Administración pública de sistemas algorítmicos de código abierto o la ejecución de mecanismos de certificación y auditoría por parte de agencias especializadas. Esta última opción nos parece la más propicia puesto que permite una solución intermedia que garantice la transparencia del algoritmo sin comprometer los intereses empresariales. De hecho, en la propia Carta Ética Europea para la eficiencia de la justicia se propone esta clase de auditorías para que controlen los métodos de procesamiento y proporcionen un asesoramiento previo a las empresas. [11]

Si bien es cierto que el primer escenario tendría una fácil solución, no podemos decir lo mismo del segundo. Atendiendo a las Directrices éticas para una IA fiable hemos de referirnos al principio ético de la explicabilidad que obliga a que los procesos sean transparentes y explicables (en la medida de lo posible) a las partes que se vean afectadas por sus decisiones, puesto que sin esta información no es posible impugnar adecuadamente una decisión.[12]

Hemos de focalizarnos en el inciso "en la medida de lo posible" y es que en las propias Directrices se tratan los algoritmos de caja negra. Se admite que "no siempre resulta posible explicar por qué un modelo ha generado un resultado o una decisión particular (ni qué combinación de factores contribuyeron a ello)" y continua esclareciendo que

> "*en tales circunstancias, puede ser necesario adoptar otras medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema), siempre y cuando el sistema en su conjunto respete los derechos fundamentales. El grado de necesidad de explicabilidad depende en gran medida del contexto y la gravedad de las consecuencias derivadas de un resultado erróneo o inadecuado*".

Así pues, para que los algoritmos de caja negra respeten los principios para una IA fiable será necesario acudir a otros medios que reflejen su explicabilidad, los cuales dependerán en gran medida de las consecuencias derivadas de los resultados del método. Podemos afirmar que un resultado erróneo en el ámbito judicial sería una manifestación de un gran menoscabo a la fiabilidad de nuestro sistema judicial. Precisamente el principio de explicabilidad ha sido creado para reforzar la confianza de los usuarios en los sistemas de IA, por lo que su aplicación al proceso judicial es extremadamente peligrosa.

4. **La Existencia de Sesgos Discriminatorios en las Resoluciones**

Avanzando en nuestro análisis también hemos de referirnos a un derecho fundamental reconocido en las Directrices éticas para una IA fiable de la UE conectado al principio de equidad de las mismas: la igualdad, no discriminación y solidaridad. La igualdad

[11] Carta Ética Europea sobre el Uso de la Inteligencia Artificial en los Sistemas Judiciales y su Entorno del Consejo de Europa en su sesión plenaria del 3 al 4 de diciembre de 2016.

[12] Grupo Independiente de Expertos de alto nivel sobre Inteligencia artificial. *Directrices Éticas para una IA Fiable.*

"implica que el funcionamiento de este tipo de sistemas no debe generar resultados injustamente sesgados (los datos utilizados para la formación de los sistemas de IA deben ser lo más inclusivos posibles, de forma que estén representados los diferentes grupos de población)".[13]

Conectando con el punto de los datos masivos tratado anteriormente hemos de recordar que los sistemas de IA se basan en conjuntos de información de agrupaciones. Podría darse un caso en que el algoritmo usando estadísticas basadas en el sexo, la etnia o la edad pudiera dar lugar a resultados sesgados puesto que sus datos de entrada podrían presentar "sesgos históricos inadvertidos, lagunas o modelos de gestión incorrecto".

Para apoyar nuestra idea nos gustaría referirnos a los casos Border y Prater, ambos muy parecidos.[14] El primero trata del hurto de una bicicleta y un patinete por dos jóvenes de 18 años valorados en un total de 80 dólares. El segundo se refiere a un hurto valorado en 86.35 dólares por un hombre de 41 años.

La diferencia fundamental entre el caso Borden y Prater reside en la etnia y antecedentes de los delincuentes. Borden era una chica negra que acumulaba en su historial una serie de delitos menores cometidos cuando era menor de edad, mientras que Prater era un hombre blanco con un historial delictivo mayor y más grave integrador de delitos como el robo a mano armada.

En un principio, debido a la entidad de los delitos cometidos con anterioridad, cabría suponer que si sometemos a estas dos personas a un programa de evaluación de riesgo de reincidencia Prater saldría peor parado, sin embargo cuando esta prueba se realizó Border fue calificada de alto riesgo, mientras que Prater recibió una calificación de bajo riesgo.

Dos años después de este análisis Border no fue acusada de nuevos delitos, al contrario que Prater, que fue condenado a ocho años de cárcel por robo de aparatos electrónicos de un almacén por valor de miles de dólares.

El estudio de ProPublica revela que la fiabilidad de estos sistemas no es lo suficientemente alta (solo el 20% de las personas que se preveía que cometieran delitos violentos llegaron a hacerlo, teniendo en cuenta toda la gama de delitos este porcentaje sube al 61%, "algo más preciso que lanzar una moneda al aire").

También se descubrieron los sesgos raciales del sistema. Se señaló erróneamente como futuros delincuentes casi al doble de personas negras que blancas, y los acusados blancos fueron etiquetados erróneamente como de bajo riesgo con más frecuencia que los negros.

Al estar los algoritmos protegidos por el secreto empresarial y, por tanto, no conocer su funcionamiento, no podemos asegurar que datos como la raza o el sexo hayan sido determinantes

[13] Cancio. *¿Sueñan los Jueces con Sentencias Electrónicas?* Análisis Jurídico-Político, volumen 2 (N°.3), pp. 145-168.

[14] Julia Angwin, et al. Jeff Larson, Surya Mattu & Lauren Kirchner. *Machine Bias. There's Software Used across the Country to Predict Future Criminals. And it's biased against Blacks.*

a la hora de que los sistemas de evaluación del riesgo proporcionen un resultado, pero a la luz de los datos claramente podemos decir que sí que existen sesgos en su funcionamiento.

5. Apoyo Normativo

Como hemos visto son múltiples las confrontaciones entre los derechos fundamentales y el uso del "juez artificial", lo que nos llevaría a recomendar el rechazo de su uso. Además podemos sostener nuestra idea a través de tres preceptos jurídicos.

El primero de ellos trata el derecho a no ser objeto de decisiones basadas únicamente en el perfilamiento digital recogido en el artículo 22 del 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).

> *"Artículo 22. Decisiones individuales automatizadas, incluida la elaboración de perfiles.*
>
> 1. *Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar".*

El perfilamiento digital se refiere al uso de datos personales para evaluar aspectos de una persona física. El perfilamiento digital no resulta individual, sino que lo que busca es elaborar patrones con el fin de modelar acciones humanas futuras. Por ello, "si el usuario no es consciente de su elaboración, comparación y uso no deberían ser utilizados para su recolección, tratamiento, comparación y toma de decisiones a partir de ellos".[15] En base a esto podemos descartar el uso del "juez artificial" porque precisamente estos aparatos lo que hacen es extraer un conjunto de datos de los perfiles de los acusados con el objetivo de incorporarlos a sus bases y compararlos estadísticamente con resultados judiciales previos.

Si bien es cierto que en el presente artículo también nos encontramos con una serie de excepciones:

> *"2. El apartado 1 no se aplicará si la decisión:*
>
> *a) es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento;*
>
> *b) está autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, o*

[15] Ricardo Valenciano Muñoz. *Prohibición de la Automatización Decisoria basada en el Perfilamiento Digital.* Novos Estudos Jurídicos (Nº.27), pp. 206-223.

c) se basa en el consentimiento explícito del interesado."

A nuestro parecer el apartado b) es de vital importancia, pues abre la posibilidad del uso de este "juez artificial" de manera general si el Derecho de la Unión o de los Estados miembros lo autoriza, lo que significa que el Gobierno de la Unión Europea así como los Estados individualmente podrían cambiar en cualquier momento su normativa procesal y sustituir el juez persona física por el "juez artificial".

Esta norma ha sido criticada ampliamente en la literatura, pues la doctrina considera que el artículo 22.2 es una norma demasiado holgada,[16] donde caben supuestos de hecho sin límites. En cuanto al apartado b) reiteramos nuestra idea de un "juez artificial" generalizado a mandato de la autoridad política, mientras que el c) ha sido duramente criticado pues se dice que no tiene en cuenta la realidad, un sujeto puede aceptar ser perfilado pero sin entender realmente lo que el perfilamiento conlleva y en qué consiste, por lo que deberíamos replantearnos la validez de este tipo de consentimiento si la población no cuenta con un nivel básico de conocimientos y acercamiento a estas nuevas tecnologías.

Por último hacer referencia al apartado tercero del presente artículo:

> *"3. En los casos a que se refiere el apartado 2, letras a) y c), el responsable del tratamiento adoptará las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión."*

A nuestro parecer en la redacción del Reglamento ya se estaba teniendo en cuenta la necesidad de intervención de la persona física en el proceso de toma de decisiones judiciales mediante perfilado. Creemos que la incorporación de este apartado es un acierto pues permite al juez diferir de la conclusión de la máquina no despersonificando el proceso. Sin embargo, también hemos de apuntar la falta de referencia en este apartado a la letra b) del artículo 22.2. Lo que esto significa es que si los Estados o la Unión Europea quisiesen implementar el "juez artificial" en sus procesos no estarían obligados a contar con un juez humano que revisase las resoluciones del sistema, podrían hacerlo, pero no cuentan con un mandato que les obligue. En nuestra opinión, la falta de referencia al juez persona física en este apartado obliga tanto a la Unión Europea como a sus Estados Miembros a elaborar una regulación lo suficientemente garantista como para que los derechos fundamentales de las personas no se vean menoscabados. Es por ello por lo que el uso del "juez artificial" sin la intervención humana nos resulta difícilmente plausible por todas las violaciones de derechos fundamentales que hemos recogido a lo largo del presente artículo.

Continuando con el segundo precepto jurídico, recogido en numerosos documentos de la Unión Europea como pueden ser su Libro Blanco o sus Directrices para una IA fiable, tenemos la

[16] Pere Simón. *Inteligencia artificial y Administración de Justicia: ¿Quo Vadis, Justitia?* Revista de Internet, Derecho y Política, volumen I (Nº. 33), pp. 1-15.

supervisión humana. [17][18] Parece que desde la promulgación del Reglamento (UE) 2016/679 la Unión Europea ha ido avanzando hasta percatarse de que la exclusión total de la persona física del proceso judicial es improcedente.

Todo sistema de IA ha de verse desde una visión antropocéntrica por lo que hemos de rechazar la idea de que un método de IA pueda resolver un pleito sin ningún tipo de supervisión humana. Por esto es que el resultado del sistema de IA bien "no puede ser efectivo hasta que un humano no lo haya revisado y validado", o bien, puede ser "inmediatamente efectivo pero se garantiza la intervención humana posterior".

A raíz de estas palabras hemos de reflexionar sobre si el "juez artificial" se trata de un verdadero decisor independiente. Desde nuestra perspectiva podemos descartar la idea de que un sistema de IA sea la única autoridad concluyente de un proceso judicial. En el caso de que el resultado dado por el algoritmo haya de ser revisado por el juez para transformarse en efectivo podemos decir que en el fondo la última decisión siempre es tomada por el humano, que puede apartarse del resultado dado por el sistema si lo considera necesario. Por otra parte, en el caso de que las conclusiones proporcionadas por el método de IA sean tomadas como automáticamente efectivas, pero se garantice en todo momento una intervención humana posterior, hemos de decir que el proceso judicial no es llevado plenamente a través de métodos de IA, puesto que para que así fuera las decisiones de un sistema de IA tendrían que ser revisadas por otro sistema de IA que tuviera un funcionamiento interno distinto, de manera que si llegaran a la misma conclusión esta se tendría que dar por firme. Sin embargo aquí con lo que nos encontramos es que la revisión de la sentencia será efectuada por un humano, rompiendo así la cadena del "juez artificial". Por tanto, hemos de concluir que en ambos casos la intervención única del "juez artificial" en el proceso judicial es imposible.

Por último, hemos de traer a coalición el artículo 117.3 de la Constitución española que señala: "*El ejercicio de la potestad jurisdiccional en todo tipo de procesos, juzgando y haciendo ejecutar lo juzgado, corresponde exclusivamente a los Juzgados y Tribunales determinados por las leyes, según las normas de competencia y procedimiento que las mismas establezcan*". Es decir, nuestra norma fundamental admite como detentadores del poder judicial únicamente a los jueces y magistrados que forman parte del mismo. Resulta evidente que nuestra Constitución se refiere a jueces y magistrados personas físicas, por lo que consideramos que nuestra *lex suprema* no permite la sustitución de la persona por la IA como "juez artificial".

2. Medio Auxiliar de Jueces y Magistrados

Una de las más habituales manifestaciones del uso de la IA en el proceso judicial es servir de medio auxiliar a jueces y magistrados en su toma de decisiones. Normalmente es empleada para analizar aspectos accesorios a la sentencia como podrían ser el establecimiento de medidas

[17] Comisión Europea. *Libro Blanco sobre la Inteligencia Artificial. Un Enfoque Europeo Orientado a la Excelencia y la Confianza.*
[18] Grupo Independiente de Expertos de alto nivel sobre Inteligencia artificial. *Directrices Éticas para una IA Fiable.*

cautelares, la concesión de regímenes penitenciarios como son el tercer grado o la libertad condicional o el establecimiento de medidas de seguridad.[19] Sin embargo, también se le han otorgado otras utilidades como puede ser su incorporación al proceso de investigación o, por parte de los despachos de abogados, la predicción del resultado de sentencias, la búsqueda de leyes y jurisprudencia y el perfilamiento de jueces y magistrados (*Legaltech*).

A pesar de estas otras modalidades insistimos en que su uso más frecuente se relaciona con el proceso penal y el estudio del riesgo o peligro que puede suponer el reo. Por mencionar algunos métodos de IA incorporados en nuestro país podemos hablar de RISCANVI o SAVRY, ambos utilizados en la Comunidad Autónoma de Cataluña desde la década de 2010 para evaluar el riesgo de violencia penitenciaria, en el primer caso, y dentro de los centros de menores, en el segundo caso, en Cataluña.[20] Así, con RISCANVI, por poner un ejemplo, las autoridades penitenciarias realizan una evaluación de la conducta de los internos preventivos y penados que sirve como apoyo para la toma de decisiones acerca de la situación interna del penado, auxiliando a la concesión de permisos de salida, reclasificaciones en grados menores y a la concesión de la libertad condicional o impidiéndolas mediante la graduación de las posibilidades que tiene el reo en cometer alguna de las siguientes infracciones: quebrantamiento de condena, violencia interinstitucional, reincidencia violenta y violencia autodirigida.

Más importancia subyace VioGén, una herramienta, no estrictamente IA, utilizada por las autoridades policiales ante denuncias por violencia de género. Con ella se pretende establecer medidas inmediatas de protección hacia la víctima en función de la previsibilidad de que su maltratador pueda volver a cometer un delito e informar al juzgado del riesgo asociado a la víctima (cf.). Para ello, una vez en diligencias policiales se somete a la víctima a un test a cerca de su situación familiar y sentimental con el presunto agresor y, cuando este finaliza, el sistema da un resultado estableciendo un nivel de riesgo permitiendo en todo caso al policía en cuestión subir este nivel el grado que considere necesario en base a las declaraciones de la víctima, pero nunca bajarlo. Una vez fijado el riesgo la autoridad policial toma las medidas correspondientes al mismo, y, lo que es más importante para nuestro análisis, adjunta esta evaluación de riesgo al expediente de la denuncia que será remitido al juez o Tribunal en cuestión.

Como podemos ver, en este caso el sistema de IA no sería el que toma las decisiones, sino que en todo caso se contaría con intervención humana. Sin embargo, muchos de los problemas a los que nos hemos referido en el apartado anterior se reiteran cuando hablamos de la IA como auxiliar del juez. Con ánimo de no ser repetitivos citaremos algunas violaciones a derechos fundamentales comunes al "juez artificial" y nos adentraremos en algunas nuevas consideraciones.

---

[19] Valentina Faggiani. *El Derecho a un Proceso con Todas las Garantías ante los Cambios de Paradigma de la Inteligencia Artificial.* Teoría y Realidad Constitucional (Nº.50), pp. 517-546.

[20] Ana Montesinos. *Los Algoritmos que Valoran el Riesgo de Reincidencia. En especial, el Sistema VioGén.* Revista Aranzadi de Derecho y Proceso Penal (Nº. 64), pp. 19-55.

1. Violaciones Comunes

Principalmente podemos citar tres problemas comunes entre el "juez IA" y el uso de la IA como auxiliar de jueces y magistrados: el cálculo de los resultados a través de datos masivos, el desconocimiento del código fuente del algoritmo por parte del acusado y los sesgos discriminatorios.

En cuanto el uso de datos masivos nos encontramos de nuevo con el problema de individualización de la sentencia, motivado por los resultados de grupo en lugar de adecuar el fallo de la sentencia al caso concreto.

El desconocimiento del código fuente del algoritmo, bien por estar en manos de una empresa privada, bien por tratarse de un algoritmo de caja negra, impiden la defensa eficaz del acusado al carecer del razonamiento que lleva a la motivación del fallo y por tanto no poder emprender un proceso contradictorio.

Por último, los sesgos discriminatorios hacen que ciertas personas por sus características físicas, personales o culturales puedan resultar favorecidas o desfavorecidas respecto a otras sin un pretexto lógico por la utilización de datos de grupo que incidan en su caso de manera perjudicial o no para ellas.

2. Individualización de la Sentencia y Cristalización de la Jurisprudencia

En primer lugar, podemos pensar que el problema de la adaptación de la ley al caso concreto y de la cristalización de la jurisprudencia habría sido resuelto con la incorporación del juez humano al proceso. Como jueces y magistrados pueden apartarse de las resoluciones dadas por los sistemas de IA podemos decir que están capacitados, al contrario que dichos sistemas, para comprender el fondo del caso. El juez siempre podría considerar otras cuestiones que por su naturaleza humana no pueden ser tenidas en cuenta por la máquina, creando una esfera de empatía tanto con la víctima como con el reo que permitiese una mejor individualización de la sentencia. Es por esto mismo por lo que tampoco se daría una cristalización de la jurisprudencia, pues con la apreciación de los pequeños matices de cada caso se conseguiría la creación de precedentes más ricos y abundantes, capaces de adaptarse a los acontecimientos sociales que marcan cada época sin necesidad de tener que caer en continuas reformas legislativas.

Sin embargo, en esta reflexión también hemos de tener en cuenta la racionalidad humana y su predisposición a creer en el método científico. En un mundo donde proporcionamos nuestros datos a grandes plataformas digitales, la IA nos rodea y la confianza en la tecnología es cada vez mayor, resulta muy complicado argumentar el error de una máquina, robot o IA y apartarse de sus razonamientos ("sesgo de automatización").

El sesgo de automatización es una clase de heurístico.[21] Jueces y magistrados cuentan con numerosos casos a conocer diariamente por lo que, al igual que cualquier ser humano, inconscientemente usan atajos mentales que les permitan resolver los casos con mayor agilidad. Por ejemplo, está demostrado que los jueces y magistrados tienden a tener una misma senda resolutiva, de manera que si en el pasado adoptaron una decisión y esta resultó ser correcta lo mismo harán en otros casos similares. De la misma manera, jueces y magistrados tienden a resolver sus asuntos basándose en experiencias próximas, por lo que todas las decisiones que tomen y estén próximas unas de otras seguirán un mismo esquema y solución. O, por último, otra manifestación de los heurísticos judiciales es que si el juez o magistrado tiene una intuición antes de entrar a conocer el caso y las pruebas la refuerzan al final tomará su decisión en base a esa primera intuición.

Recordemos que los resultados de los sistemas de IA se basan en el procesamiento de datos a través de fórmulas matemáticas a priori indiscutibles, mientras que en las decisiones humanas siempre se conjugan otros factores como pueden ser el sentimentalismo o la opinión personal basada en unas ciertas creencias e ideología que pueden conducir a posicionamientos poco imparciales y al error judicial. Es por ello que resultaría natural que el juez o magistrado no se cuestionase la adecuación de la respuesta del sistema pues confía la objetividad del método científico con el que se ha desarrollado el algoritmo y en la credibilidad de sus relaciones.

Un claro ejemplo de este hecho se encuentra en un caso donde tras utilizar la herramienta de evaluación de riesgo VioGén la víctima acabó siendo finalmente asesinada. Sabemos bien que esta herramienta se usa en dependencias policiales pero el supuesto de hecho es perfectamente extrapolable a nuestro caso. El sistema VioGén asignó el nivel más bajo de riesgo a la víctima ("no apreciado") calificando las autoridades policiales el caso de la misma manera a pesar de existir indicios suficientes de que los hechos eran más graves de lo que la aplicación había establecido, por lo que la mujer necesitaba mayores medidas de protección. Esta valoración determinó que el juez denegara la orden de protección solicitada por la víctima conllevando este hecho su asesinato a manos de su marido un mes después de interponer la denuncia.[22]

Claramente los resultados proporcionados por sistemas de IA pueden llevar a una confianza ciega en los mismos, por ello es importante que pongamos en alerta a nuestro Poder Judicial para animar su escepticismo hacia las evaluaciones realizadas por los métodos de IA y que le recordemos las limitaciones de las mismas, puesto que si no acabaríamos cayendo en el uso de estas tecnologías como "juez artificial" en vez de como un auxiliar en la toma de decisiones judiciales, con los riesgos y peligros que este desentraña (expuestos con detalle en el apartado anterior).

3. Juicios Predelictivos

[21] Liced Morales y otros. *Inteligencia Artificial en el Proceso Penal: Análisis a la Luz del Fiscal Watson*. Pensamiento jurídico *(Nº.54)*, pp. 147-164.
[22] Montesinos. *Los Algoritmos que Valoran el Riesgo de Reincidencia. En Especial, el Sistema VioGén.* Revista Aranzadi de Derecho y Proceso Penal (Nº. 64), pp. 19-55.

Un nuevo problema nace con el uso de estos sistemas como auxiliares de jueces y magistrados: el proceso judicial podría enfocarse en un juicio predelictivo en lugar de atender a los acontecimientos presentes. Básicamente esto podría darse en los casos en que herramientas como las evaluaciones de riesgos se destinasen a determinar la graduación de la sentencia. Así pues, el juez o magistrado podría centrarse más en la posibilidad de reiteración de la actividad delictiva en vez de basarse en los hechos que motivaron el proceso.

Como sabemos, en ámbitos como el penal, la reincidencia es un factor que se tiene en cuenta a la hora de determinar la pena a la que se sentenciará al acusado. De esta manera la reincidencia o reincidencia cualificada en su caso funcionan como agravantes de la pena que, en resumidas cuentas pues su explicación no es objeto de esta investigación, llevan a una condena más gravosa. Sin embargo, hemos de destacar que esta reincidencia se refiere a acontecimientos pasados, es decir, la mayor gravedad de la pena se justifica en que la finalidad reeducativa de las anteriores penas no ha sido satisfecha, por lo que resulta necesario la imposición de una pena mayor para reintentar este proceso reeducativo, intimidar al reo y así evitar nuevas infracciones o en su caso inocuizarlo.

Por otra parte el uso de una evaluación del riesgo de reincidencia tiene un fundamento y una finalidad totalmente diferentes. En este caso lo que la motiva es una necesidad de previsión y protección. Ante una mayor posibilidad de que el sujeto vuelva a cometer una infracción similar se le impondrá una sanción mayor para inocuizarlo y reeducarlo. Sin embargo, los posibles hechos no han sucedido ni se sabe si van a suceder. En el caso de la reincidencia como agravante se tiene un pretexto, un hecho firme, pero en este caso no se tiene ninguna certeza, solo se juega con probabilidades, lo que puede suponer la imposición de un castigo mayor sin justificación en el caso de que, de no haberse agravado la sanción, no se reincidiera. En definitiva nos encontramos en un universo paralelo donde la incertidumbre es total.

Desde nuestro punto de vista, el uso de la IA como método de análisis de la posibilidad de reincidencia ataca a principios tales como la presunción de inocencia, no solo inspirador del Derecho Penal, sino de otros como el Derecho Administrativo en cuanto a régimen de infracciones y sanciones se refiere. Tampoco conocemos ninguna posible solución a esta violación, por lo que desechamos la idea de que este tipo de sistemas predelictivos puedan ser de uso general en nuestro sistema judicial.

3. El caso COMPAS

1. Presentación del Caso

Se conoce como caso COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) al suceso litigioso State v. Loomis resuelto en apelación por el Tribunal Supremo de Wisconsin el 13 de julio de 2016.[23]

[23] Estados Unidos. Tribunal Supremo de Wisconsin. Caso 881 N.W.2d 749 (Estado de Wisconsin v. Loomis) de 13 de julio de 2016.

A principios de 2013 Eric Loomis fue acusado de la comisión de cinco delitos por el Estado de Wisconsin en relación con un tiroteo desde un coche en La Crosse. Loomis confesó ser autor del intento de huida de un agente de tráfico y de la conducción de un vehículo de motor sin el consentimiento de su propietario, pero negó su participación en los otros tres delitos en relación con el tiroteo.

Para dictar sentencia el Tribunal se basó tanto en los hechos como en una evaluación del riesgo de reincidencia, para lo que se utilizó el citado sistema COMPAS. COMPAS se basa en un algoritmo que opera con una entrevista con el delincuente en la que se efectúan preguntas como: "¿ha estado alguno de sus padres en la cárcel?", "¿cuántos de tus amigos/conocidos consumen drogas ilegalmente?" o "¿está en acuerdo o en desacuerdo con la siguiente afirmación: una persona hambrienta tiene derecho a robar?"; acompañado de un análisis de sus antecedentes penales. A partir de esta información el sistema proporciona un grado de riesgo sobre tres aspectos: la reincidencia antes del juicio, la reincidencia general y la reincidencia violenta.

Al estar protegido el algoritmo de COMPAS por el secreto empresarial solo se comunica al Tribunal y a los acusados la puntuación obtenida clarificadora del riesgo de reincidencia: bajo, medio o alto. Por lo que, cuando el Tribunal de Primera Instancia condenó a Loomis a seis años de prisión y cinco de supervisión ampliada basándose en parte en esta evaluación, Loomis decidió presentar una moción de amparo argumentando que la confianza del tribunal en la evaluación COMPAS violaba sus derechos al debido proceso por tres razones:

1. En primer lugar, porque se viola el derecho del acusado a ser sentenciado sobre la base de información precisa, pues la propietaria de COMPAS le impide evaluar su exactitud.
2. En segundo lugar, porque se viola el derecho del acusado a una sentencia individualizada.
3. Por último, porque se utiliza indebidamente valoraciones de género en la sentencia.

El Tribunal de Primera Instancia denegó la moción y el Tribunal de Apelación de Wisconsin aprobó la apelación ante el Tribunal Supremo de Wisconsin aceptándola también este último.

2. Pretensiones del Acusado

1. Sentencia sobre la Base de Información Precisa

En cuanto a su primera pretensión Loomis afirma que la empresa creadora del algoritmo le niega el pleno acceso a la información que servirá de fuente para determinar el riesgo concreto, por lo que no puede asegurarse de que se le esté condenando sobre la base de información exacta.

El Tribunal afirma que cualquier sentenciado ha de estarlo en base a información precisa. Este derecho incluye la posibilidad de revisar y verificar la información contenida en la evaluación de riesgos en la que el Tribunal basa su decisión, por lo que no se puede negar su divulgación al acusado. Sin embargo, en este caso no considera que a Loomis se le negara su derecho. Esto es porque, aunque Loomis no puede revisar la forma en que el algoritmo COMPAS calcula el riesgo por pertenecer el algoritmo a una empresa privada, si puede cuestionar los resultados (las puntuaciones de riesgo). Aunque las puntuaciones de riesgo no explican cómo COMPAS utiliza

la información para calcularlas, su empresa creadora, Northpointe, ha publicado una guía (Northpointe's 2015 Practitioner's Guide to COMPAS) que explica la información en qué se basan las puntuaciones, principalmente información estática (antecedentes penales) y algún dato dinámico (arrestos o cargos mientras se disfrutaba de la libertad condicional, vueltas a prisión mientras se disfrutaba de la libertad condicional, abuso de sustancias…).

Concretamente, en el caso Loomis se le realizaron 21 preguntas que el acusado pudo revisar junto a sus respuestas, por lo que, como la evaluación de riesgos se basa en las respuestas a las preguntas que él pudo verificar y en los datos públicamente disponibles sobre sus antecedentes penales, no se considera que se esté violando su derecho a ser sentenciado bajo la base de información precisa pues Loomis tuvo la oportunidad de verificar que las preguntas y respuestas que figuraban en el informe COMPAS eran exactas.

Otro argumento utilizado por Loomis para demostrar la falta de información precisa son los estudios sobre esta herramienta de evaluación de riesgos que han planteado dudas sobre su precisión. Algunos Estados estadounidenses que utilizan el sistema han llevado a cabo estudios sobre la fiabilidad de COMPAS concluyendo que es una herramienta de evaluación de riesgos suficientemente precisa, sin embargo, estudios citados por Loomis, como el realizado por el California Department of Corrections and Rehabilitation (CDCR) en 2007, concluyen que, aunque COMPAS parece evaluar las necesidades criminógenas y el riesgo de reincidencia, hay pocas pruebas de que esto sea lo que COMPAS evalúa realmente. El propio CDCR se contradice en su informe final de 2010 sobre el estudio de validación de COMPAS en California, concluyendo que, aunque no es perfecto, resulta lo suficientemente fiable como para ser utilizado.

A su vez otros estudios inciden en la posibilidad de que estos sistemas, por el uso de datos masivos y sesgados, clasifiquen de forma desproporcionada a los delincuentes pertenecientes a minorías como de mayor riesgo debido a factores que pueden estar fuera de su control como los antecedentes criminales de su familia o su educación. Así, en la sentencia se refleja que un análisis de las puntuaciones de reincidencia de COMPAS con una muestra de 10.000 acusados de delitos penales en el condado de Broward (Florida) concluyó que los acusados negros tenían muchas más posibilidades que los acusados blancos de ser juzgados incorrectamente como de mayor riesgo de reincidencia, así como, los acusados blancos tenían más probabilidades que los acusados negros de ser calificados incorrectamente como de bajo riesgo.

Debido a la diversidad de opiniones en los estudios el Tribunal no establece que se haya violentado el derecho del acusado a ser sentenciado bajo información exacta, pero sí establece ciertas cautelas para jueces y magistrados a la hora de dictar sentencia en base a estas evaluaciones de riesgo para que puedan darle un peso apropiado en su toma de decisiones a las puntuaciones. Estas advertencias tratan sobre la fiabilidad y exactitud de las evaluaciones de riesgo COMPAS y son las siguientes:

a) *"El algoritmo COMPAS tiene un carácter reservado, lo que impide la divulgación de información relativa a cómo se ponderan los factores o cómo se determinan las puntuaciones de riesgo.*

b) *La evaluación de riesgo compara a los acusados con una muestra nacional, pero aún no se ha completado ningún estudio de validación cruzada para una población de Wisconsin.*
c) *Algunos estudios sobre las puntuaciones de la evaluación de riesgos COMPAS han suscitado dudas sobre si clasifican de forma desproporcionada a los delincuentes pertenecientes a minorías con un mayor riesgo de reincidencia.*
d) *Las herramientas de evaluación de riesgos deben ser constantemente supervisadas y renormalizadas para garantizar su precisión debido a los cambios en las poblaciones y subpoblaciones".*

2. Sentencia individualizada

Pasemos ahora a hablar de la segunda denuncia de Loomis. Aunque, según las opiniones expertas, se admite que las evaluaciones de COMPAS están basadas en datos de grupo, el Tribunal niega que se esté violando el derecho de Loomis a una sentencia individualizada puesto que la evaluación de riesgos no es el único dato en el que se basa el Tribunal sentenciador. Así, en la sala se explica que, cuando estos métodos se utilizan en sistemas penitenciarios, se prevé que el personal no esté de acuerdo con la evaluación en aproximadamente un 10% de los casos debido a circunstancias agravantes o atenuantes a las que la evaluación no es sensible, por lo que se les ha de animar a utilizar su juicio profesional y cuestionar estas puntuaciones cuando proceda. De esta forma, el Tribunal espera que jueces y magistrados también cuenten con una actitud crítica ante estos sistemas.

A pesar de reincidir en la advertencia de que sistemas como COMPAS utilizan datos de grupo para dar sus puntuaciones, de manera que identifican grupos de delincuentes de alto riesgo, no un individuo de alto riesgo en particular, también apuntan que la evaluación de riesgos es solo otro instrumento de apoyo más, de manera que, junto con otros factores, es capaz de suministrar al Tribunal tanta información como sea posible para llegar a una sentencia individualizada.

3. Uso del género en las evaluaciones

Debido al uso del género en las evaluaciones de COMPAS el acusado defiende que su derecho a no ser condenado por razón de sexo se ve violentado. Loomis afirma que se desconoce exactamente cómo utiliza COMPAS el género en sus evaluaciones, pero sostiene que COMPAS considera el género como un factor criminógeno. El Estado discrepa aduciendo que la evaluación COMPAS es igual para hombres y mujeres, pero que luego se realiza una comparación de cada delincuente con un grupo "normalizado" de su propio sexo. Sea uno u otro el uso del sexo, Loomis se opone a cualquiera de ellos, pero el Estado defiende la necesidad de su uso para lograr una precisión estadística.

El Estado argumenta que, dado que los hombres y las mujeres tienen diferentes tasas de reincidencia y diferente potencial de rehabilitación, una evaluación de riesgo sin tener en cuenta el género proporcionaría resultados inexactos tanto para los hombres como para las mujeres; de hecho, en el caso estadounidense los hombres por término medio tienen mayores tasas de reincidencia y de comisión de delitos violentos que las mujeres.

El Tribunal considera que el uso del género como factor en la evaluación del riesgo servía al propósito no discriminatorio de promover la precisión y que Loomis no había aportado pruebas suficientes de que el Tribunal de Primera Instancia hubiera tenido realmente en cuenta el género.

3. Fallo del Tribunal

Finalmente, el Tribunal admite en su fallo que el uso de evaluaciones de riesgos como COMPAS en el proceso judicial conlleva la toma de numerosas cautelas para hacerlo seguro, pero que si todas estas se dan no habría problema en su uso.

En primer lugar, recalcan que el algoritmo de COMPAS ha de ser utilizado para: a) desviar a delincuentes de bajo riesgo a alternativas no carcelarias, b) evaluar si un delincuente puede ser supervisado de forma segura y eficaz en comunidad, c) imponer términos y condiciones a su libertad condicional y responder ante posibles violaciones de la misma. En ningún caso ha de usarse para determinar la severidad de una sentencia o si un acusado es encarcelado.

En segundo lugar, establece una serie de limitaciones para garantizar que los jueces ponderen adecuadamente las evaluaciones de riesgos. De esta manera, el Tribunal prescribió tanto la forma en que las evaluaciones deben presentarse a los Tribunales, como cuál es el uso que los magistrados han de darles.

Cuando se suministre una evaluación COMPAS a un Tribunal ha de incluir cinco advertencias escritas para los magistrados: a) la naturaleza patentada del COMPAS impide saber cómo se determinan las puntuaciones de riesgo; b) las puntuaciones no pueden identificar el riesgo de un individuo concreto pues se basan en datos de grupo; c) COMPAS se basa en un amuestra nacional, no ha habido ningún estudio de validación cruzada con la población de Wisconsin, por lo que las ponderaciones podrían cambiar en relación a los resultados nacionales; d) las puntuaciones COMPAS podrían clasificar desproporcionadamente a los delincuentes pertenecientes a minorías como personas con mayor riesgo de reincidencia según estudios; e) el sistema COMPAS no se desarrolló en un principio para ayudar a jueces y magistrados, sino para servir de apoyo al Departamento de Instituciones Penitenciarias para tomar decisiones posteriores a la sentencia.

Por tanto, COMPAS no es una herramienta completamente segura, por lo que se ha de apostar por el escepticismo judicial a la hora de estudiar las puntuaciones suministradas por el sistema, pudiendo los jueces y magistrados basarse en partes de la evaluación y rechazar otras. Aun así, su uso correcto en ningún caso violará el derecho del acusado al debido proceso.

4. Votos Particulares

Hemos de detenernos también en los votos particulares, concretamente en el del juez Abrahamson. El juez Abrahamson estuvo de acuerdo con el fallo, sin embargo manifestó sus preocupaciones en cuanto a la capacidad del Tribunal para entender las evaluaciones algorítmicas de riesgo.

Por un lado criticó la decisión del Tribunal de denegar a Northpointe la oportunidad de presentar un *amicus curiae* donde calmase las aguas recalcando las fortalezas del sistema así como las precauciones que habrían de tomar jueces y magistrados en su uso. Justifica esta decisión en que Northpointe, como desarrolladora del sistema, podría dar una imagen poco imparcial de COMPAS para así asegurar sus beneficios.

Por otra parte, afirma que alentar el escepticismo judicial sobre la exactitud de las evaluaciones de riesgos no es suficiente para indicar a los jueces hasta qué punto deben descartarlas. Actualmente nos encontramos con numerosos estudios tanto a favor como en contra de este tipo de evaluaciones, por lo que es probable que los jueces carezcan de la información necesaria para modular su consideración de la herramienta, de manera que bien podrían dejar de utilizarla por miedo a resultados incorrectos, o bien, adquirir una sobreconfianza en ella debido a la incomprensión de estas evaluaciones.

En resumen el juez Abrahamson afirma que el Tribunal se equivocó al pensar que mientras los jueces estén informados de la posible inexactitud del COMPAS pueden realizar una ponderación de su uso adecuada. Además, dice que la advertencia de su inexactitud probablemente sea ineficaz para cambiar la forma en que los jueces piensan sobre las evaluaciones de riesgo dada la presión dentro del sistema judicial para utilizar estas evaluaciones, así como los heurísticos que apoyan la confianza en sus resultados.

5. Traslado del Caso COMPAS al Sistema Judicial Español

Como sabemos el panorama jurídico continental y el del common law son muy diferentes, es por ello por lo que nos proponemos trasladar el caso State v. Loomis al sistema jurídico español para determinar si las conclusiones serían las mismas, y en su caso realizar una crítica a la sentencia estadounidense. Para ello explicaremos la problemática en tres apartados comunes a las pretensiones tratadas en el punto anterior y añadiremos algún otro que trate otras posibles afectaciones a los derechos fundamentales en referencia al inicio de la investigación.

1. Sentencia sobre la Base de Información Precisa

Desde nuestro punto de vista, que el acusado pueda acceder a las preguntas, respuestas y otra información que servirá de datos de entrada para dar a luz al resultado de la evaluación de riesgos no significa que la sentencia esté basada en información precisa.

Lo importante en este punto es que el acusado conozca las ponderaciones que darán lugar al resultado, es decir, el algoritmo, puesto que una vez que los datos se encuentran dentro del sistema el algoritmo puede realizar distintas interconexiones que lleven a resultados bien diferentes.

Por ello, consideramos que el acusado no habría accedido realmente a la información, de manera que se estaría violando el principio de contradicción al eliminar la posibilidad de defensa.

2. Sentencia Individualizada

La problemática del uso de datos de grupo es reticente en este caso, si bien es cierto que el uso de otros elementos a la hora de la toma de decisiones podrían ayudar a individualizar la sentencia, nos encontramos con un gran sesgo de automatización, apreciado incluso por los jueces estadounidenses.

Compartimos las preocupaciones del juez Abrahamson, y es que por muchas cautelas que se tomen, jueces y magistrados no saben el peso que deberían tomar estas evaluaciones de riesgos como concluyentes del resultado de sus sentencias puesto que los estudios sobre su fiabilidad son diversos.

De esta manera, si el uso de la IA en la toma de decisiones judiciales nos conduce a los mismos heurísticos que si prescindiéramos de ella y aumenta las posibilidades de violar el derecho del acusado a una sentencia individualizada entendemos que su uso en las condiciones actuales no resulta viable.

3. **Uso del Género en las Evaluaciones y Otros Sesgos Discriminatorios**

De nuevo aquí hemos de analizar los datos de entrada que son tomados por cada sistema de IA para aportar el resultado final. Si bien es cierto que históricamente la actividad delictiva se da en mayor parte por hombres que por las mujeres, hay otros aspectos que podrían dar lugar a resultados sesgados en función de los datos introducidos. Para demostrar nuestra teoría nos proponemos realizar una simulación a partir de los datos proporcionados por el Instituto Nacional de Estadística (INE) y del Consejo General del Poder Judicial, así como del estudio de Steven Kemp y Daniel Varona publicado en The British Journal of Criminology.[24]

Según el INE en 2021 había aproximadamente 5,5 millones de extranjeros residentes en nuestro país, representando alrededor del 11.5% de la población[25]. Dadas estas cifras cabe apuntar que en 2021 los presidiarios extranjeros representaban el 29.5% de los presos en España (Tabla 1).[26]

A la luz de los datos podemos pensar que la tasa de criminalidad extranjera resulta mucho más elevada que la española dada su representatividad sobre el total de la población, sin embargo, los estudiosos apuntan a que esta sobrerrepresentación de la población extranjera en las cárceles españolas atiende en parte a prácticas discriminatorias en el sistema de justicia penal.

---

[24] Steven Kemp & Daniel Varona. *Foreign and Dangerous? Unpacking the Role of Judges and Prosecutors in Sentencing Disparities in Spain*. The British Journal of Criminology, vol. XX, pp. 1-19.

[25] Instituto Nacional de Estadística. (2022). *Cifras de Población (CP) a 1 de enero de 2022 Estadística de Migraciones (EM). Año 2021.*

[26] Consejo General del Poder Judicial. (2022). *Distribución de la Población Reclusa Extranjera por Sexo (diciembre 2021). Estadística Penitenciaria, diciembre 2021.*

Existe una abundante literatura a cerca de las desigualdades raciales y étnicas en la imposición de penas,[27] a la que recientemente se suman estudios acerca de la nacionalidad o ciudadanía del delincuente que desvelan una clara discriminación en términos de encarcelamiento.[28]

Esta posible discriminación atiende a los heurísticos utilizados por jueces y magistrados a la hora de dictar sentencia que, como hemos explicado con anterioridad, les llevan a basarse en estereotipos relacionados con las características del acusado como el género, la raza o la edad.

En The British Journal of Criminology se afirma claramente que los extranjeros ingresan más en prisión que los condenados españoles. Esto podemos confirmarlo si miramos las estadísticas del INE, de manera que si introdujésemos estos datos en un método de IA para detectar la posibilidad de reincidencia de un individuo a la hora de establecer medidas cautelares para el mismo sería muy probable que el riesgo de reincidencia en personas extranjeras fuera mucho mayor que en nacionales simplemente por un mal tratamiento de los datos en el entrenamiento del sistema. Y es que, aunque los extranjeros son encarcelados con más frecuencia que los españoles, también hemos de atender a su pirámide poblacional y a la distribución de la población carcelaria en España.

Comenzando por su agrupación por edades podemos apreciar cómo la mayoría de la población extranjera ronda la franja de edad de los 30 a los 44 años (Gráfico 1),[29] coincidente con las edades dónde encontramos más población reclusa (Tabla 2)[30]. Además, reiteramos que en el año 2021 la representatividad de extranjeros en cárceles españolas fue del 29,5%.[31]

---

[27] Darrell Stefensmeier, Jeffery Ulmer and John Kramer. *The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male*, Criminology, 36, *763–98*; Michael Tonry. *Punishing Race: A Continuing American Dilemma.* Studies in Crime and Public Policy. New York: Oxford University Press; Besiki Luka Kutateladze y otros. *Cumulative Disadvantage: Examining Racial and Ethnic Disparity in Prosecution and Sentencing.* Criminology, 52: 514–51; Travis W. Franklin. *The State of Race and Punishment in America: Is Justice Really Blind?* Journal of Criminal Justice, 59*: 18–28*; Marisa Omori and Nick Petersen. *Institutionalizing Inequality in the Courts: Decomposing Racial and Ethnic Disparities in Detention, Conviction, and Sentencing.* Criminology, *58: 678–713*; Alexander Testa and B.D. Johnson. *Paying the Trial Tax: Race, Guilty Pleas, and Disparity in Prosecution*. Criminal Justice Policy Review*, 31: 500–31;* Katherine A. Durante. *County-Level Context and Sentence Lengths for Black, Latinx, and White Individuals Sentenced to Prison: A Multi-Level Assessment*. Criminal Justice Policy Review*, 32: 915–37; Brian D.* Johnson, Cassia Spohn and Anat Kimchi. *Life Lessons: Examining Sources of Racial and Ethnic Disparity in Federal Life without Parole Sentences*. Criminology, 59: 704–37; Ellen A. Donnelly. *Neighborhoods, Criminal Incidents, Race, and Sentencing: Exploring the Racial and Social Context of Disparities in Incarceration Sentences*. The British Journal of Criminology, 62: 145–64.

[28] Scott E. Wolfe, David C. Pyrooz and Cassia Spohn. *Unravelling the Effect of Offender Citizenship Status on Federal Sentencing Outcomes*. Social Science Research*, 40: 349–62;* Michael T. Light. *The New Face of Legal Inequality: Non-citizens and the Long-Term Trends in Sentencing Disparities across U.S. District Courts*, *1992–2009*. Law and Society Review*, 48: 447–78*; Michael T. Light. *The Punishment Consequences of Lacking National Membership in Germany, 1998–2010*. Social Forces*, 94: 1385–408;* Michael T. Light. *Punishing the "Others": Citizenship and State Social Control in the United States and Germany*. European Journal of Sociology, *58: 33–71.*

[29] Instituto Nacional de Estadística. (2022). *Estadística del Padrón Continuo, Total Nacional, Extranjera, Edad (hasta 100 y más), Total, 1 de enero de 2022.*

[30] Ministerio del Interior. (2021). Tabla 4-1-5. *Población reclusa penada a 31-12-2021 por grupo de edad, sexo y administración competente.* En Anuario estadístico del Ministerio del Interior 2021 (p.323).

[31] Consejo General del Poder Judicial. (2022). *Distribución de la Población Reclusa Extranjera por Sexo (diciembre 2021). Estadística Penitenciaria, diciembre 2021.*

A pesar de esto el número de condenados es mayor entre los españoles que entre nacionales de otros continentes (Gráfico 2)[32], lo que nos muestra que el sistema penal resulta más benevolente con nacionales que con extranjeros al reservar sus penas más gravosas para estos últimos.[27]

No podemos concluir que la existencia de más personas extranjeras reclusas conlleve una mayor tasa de criminalidad por dos motivos. En primer lugar, existen más condenados españoles, por lo que necesariamente han cometido más delitos los nacionales que los extranjeros pero por diferentes razones no han sido condenados a penas de prisión. En segundo lugar, la mayor parte de los extranjeros se condensan en una franja de edad donde la comisión de delitos y el consecuente encarcelamiento es más elevada, mientras que los españoles se dispersan sobre todos los grupos de edad a la vez que nos encontramos con una población más envejecida tendente a la menor comisión de delitos.

Así pues, dependiendo de cómo se dé la utilización de estos datos podríamos afirmar que efectivamente la probabilidad de delinquir, y por lo tanto su predicción del riesgo de reincidencia, es superior en extranjeros que en nacionales, o concluir que no contamos con evidencias suficientes para demostrar esta hipótesis.

En cuanto el uso de la IA en estos casos hemos de hacer referencia de nuevo aquí a la diferencia entre la capacidad de sintaxis y de semántica de una IA, y es que, aunque suministrásemos un conjunto de datos suficientemente amplio a la máquina, podría darse que esta no supiese interpretarlos porque los mismos ya apareciesen sesgados o no tuviera en cuenta otras variables interpretativas como en nuestro caso la edad de la población extranjera y su pirámide poblacional.

Nos encontramos ante un claro sesgo injustificado que atenta contra el principio de igualdad por lo que resultaría prudente establecer una auditoría de todos los sistemas de IA utilizados en los procesos judiciales *ex ante* para prevenir y eliminar estos riesgos, que podrían verse incrementados en el caso de algoritmos de caja negra, al no comprender la forma de procesamiento de los datos, y en cualquier tipo de algoritmo de aprendizaje automático, pues al "aprender" a medida que funciona sus resultados y sus fallos no pueden predecirse o apreciarse en la fase de diseño.

4. Conclusiones

Sin duda alguna no podemos negar que el uso de la IA en la toma de decisiones judiciales es efectiva a la hora de agilizar el proceso de decisión, ahorrando costes y atendiendo a la población de una manera más eficaz. Sin embargo, hemos de cuestionarnos la supeditación de los derechos fundamentales de las personas a estos criterios de eficacia y eficiencia.[33]

Como hemos visto, el "juez artificial" resulta imposible en nuestro sistema judicial actual debido a nuestra legislación, pero tampoco recomendaríamos el cambio de la misma en base a la posible

[32] Instituto Nacional de Estadística (2021). *Estadística de Condenados: Adultos, Total, Total, Nacionalidad, 2021*

[33] Simón. *Inteligencia artificial y Administración de Justicia: ¿Quo Vadis, Justitia?* Revista de Internet, Derecho y Política, volumen I (Nº. 33), pp. 1-15.

mejora de las esperas judiciales o de la objetividad dado los numerosos inconvenientes que el uso de la IA, tanto como "juez artificial", como auxiliar del juez humano, en los procesos judiciales conlleva.

Entre nuestras mayores preocupaciones se encuentra la violación del principio de contradicción, ya que, aunque se consiguiera que las empresas compartieran su algoritmo o se creasen algoritmos públicos, nada se podrá hacer en cuanto a los algoritmos de caja negra.

Otro problema muy importante son los sesgos algorítmicos. En nuestra opinión, el ingreso de la IA en el ámbito judicial ha de retrasarse hasta que los datos ingresados en el sistema así como las interconexiones realizadas por el mismo se vean lo suficientemente pulidas.

Por otra parte, aunque el uso de sistemas de IA parece proporcionar una mayor objetividad en la toma de decisiones judiciales por el uso de la ciencia y las matemáticas, están generando nuevos heurísticos. Así que, si lo que pretendíamos era reducir los heurísticos que se dan en el proceso tradicional, ningún sentido tiene crear nuevos. Por no decir que el sesgo de la automatización ataca directamente al derecho a una resolución individualizada pues los resultados proporcionados por los métodos de IA se basan en datos de grupo, así pues, si el juez no se cuestiona la posible adecuación del resultado suministrado por la máquina al caso concreto se dejaría de tratar el caso de una manera personalizada.

Por último, en cuento al problema de los juicios predelictivos, nos encontramos con la perpetuación de los heurísticos, ya que el juez o magistrado podría basarse en los actos pasados del acusado, es decir, tomar un atajo hacia un resultado dado en el pasado, en lugar de evaluar los hechos actuales que son en todo caso en los que se debería basar para dictar sentencia.

Por tanto, tras haber sopesado los posibles riesgos del uso de la IA en cuanto al ataque a los derechos fundamentales de las personas parece razonable limitar su uso. A pesar de que estos métodos faciliten la toma de decisiones agilizando el proceso hemos de recalcar que los derechos fundamentales de las personas en ningún caso pueden supeditarse a criterios de eficiencia o de ahorro en costes, y que, aunque algunos de los inconvenientes de los métodos de IA parecen ser superables, lejos nos encontramos de una IA totalmente segura. Es por ello por lo que nuestra recomendación en este caso va más lejos del establecimiento de cautelas para jueces y magistrados o la imposición de auditorías del sistema, en nuestra opinión, debido a su peligrosidad, los sistemas de IA deberían alejarse del sistema judicial al menos hasta que se dé un mayor avance de la tecnología y su utilización pueda compaginarse con el respeto a las garantías procesales consagradas en nuestro sistema así como a la convivencia pacífica con los derechos fundamentales de las personas.

5. Referencias bibliográficas

1. Fuentes Internacionales

1. Legislación internacional

Comisión Europea. Libro Blanco sobre la Inteligencia Artificial. Un Enfoque Europeo Orientado a la Excelencia y la Confianza (2020) [Español].

Consejo de Europa. 'Carta Ética Europea sobre el uso de la Inteligencia Artificial en los sistemas judiciales y su entorno' Sesión plenaria del 3 al 4 de diciembre de 2016 [Español].

Parlamento Europeo y Consejo Europeo. 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).

2. Jurisprudencia internacional

Estados Unidos. Tribunal Supremo de Wisconsin. Caso 881 N.W.2d 749 (Estado de Wisconsin v. Loomis) de 13 de julio de 2016.

3. Informes internacionales

Angwin J. et al. Larson J., Mattu S. & Kirchner L. (23 de mayo de 2016). Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks.

Grupo Independiente de Expertos de alto nivel sobre Inteligencia artificial. Directrices Éticas para una IA Fiable (2019) [Español].

4. Artículos internacionales

Donnelly E. A. Neighborhoods, Criminal Incidents, Race, and Sentencing: Exploring the Racial and Social Context of Disparities in Incarceration Sentences (2022). The British Journal of Criminology, 62: 145–64.

Durante K. A. County-Level Context and Sentence Lengths for Black, Latinx, and White Individuals Sentenced to Prison: A Multi-Level Assessment (2021). Criminal Justice Policy Review, 32: 915–37.

Franklin T. W. The State of Race and Punishment in America: Is Justice Really Blind? (2018) Journal of Criminal Justice, 59: 18–28.

Johnson B.D., Spohn C. and Kimchi A. Life Lessons: Examining Sources of Racial and Ethnic Disparity in Federal Life without Parole Sentences (2021). Criminology, 59: 704–37.

Kemp S. & Varona D. Foreign and Dangerous? Unpacking the Role of Judges and Prosecutors in Sentencing Disparities in Spain (2022). The British Journal of Criminology, vol. XX.

Kutateladze B. L. y otros. Cumulative Disadvantage: Examining Racial and Ethnic Disparity in Prosecution and Sentencing (2014). Criminology, 52: 514–51.

Light M. T. Punishing the "Others": Citizenship and State Social Control in the United States and Germany (2017). European Journal of Sociology, 58: 33–71.

Light M. T. The New Face of Legal Inequality: Non-citizens and the Long-Term Trends in Sentencing Disparities across U.S. District Courts, 1992–2009 (2014). Law and Society Review, 48: 447–78

Light M. T. The Punishment Consequences of Lacking National Membership in Germany, 1998–2010 (2016). Social Forces, 94: 1385–408

Omori M. and Petersen N. Institutionalizing Inequality in the Courts: Decomposing Racial and Ethnic Disparities in Detention, Conviction, and Sentencing (2020). Criminology, 58: 678–713.

Stefensmeier D., Ulmer J. and Kramer J. The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male (1998). Criminology, 36, 763–98.

Testa A. and Johnson B.D. Paying the Trial Tax: Race, Guilty Pleas, and Disparity in Prosecution (2020). Criminal Justice Policy Review, 31: 500–31.

Tonry M. Punishing Race: A Continuing American Dilemma (2021). Studies in Crime and Public Policy. New York: Oxford University Press.

Wolfe S. E., Pyrooz D. C. and Spohn C. Unravelling the Effect of Offender Citizenship Status on Federal Sentencing Outcomes (2011). Social Science Research, 40: 349–62.

2. Fuentes nacionales

1. Legislación nacional

Constitución Española, de 29 de diciembre de 1978.

2. Artículos

Bonsignore D. Sobre Inteligencia Artificial, Decisiones Judiciales y Vacíos de Argumentación (2021). Teoría & Derecho (Nº 29) [Español].

Cancio R.C. ¿Sueñan los Jueces con Sentencias Electrónicas? (2020) Análisis Jurídico-Político, volumen 2 (Nº.3) [Español].

Faggiani V. El Derecho a un Proceso con Todas las Garantías ante los Cambios de Paradigma de la Inteligencia Artificial (2022). Teoría y Realidad Constitucional (Nº.50) [Español].

Maroño M. M. El Concepto de Inventor en el Derecho de Patentes y los Sistemas de Inteligencia Artificial. (2020) Cuadernos de Derecho Transnacional, vol. 12 (Nº 2) [Español].

Montesinos A. Los Algoritmos que Valoran el Riesgo de Reincidencia. En especial, el Sistema VioGén (2021). Revista Aranzadi de Derecho y Proceso Penal (Nº. 64) [Español].

Morales L. y otros. Inteligencia Artificial en el Proceso Penal: Análisis a la Luz del Fiscal Watson (2021). Pensamiento jurídico (Nº.54) [Español].

Muñoz R. V. Prohibición de la Automatización Decisoria basada en el Perfilamiento Digital (2022). Novos Estudos Jurídicos (Nº.27) [Español].

Pérez A. Justicia Civil en la Era Digital y Artificial: ¿Hacia una Nueva Identidad? (2021). Revista Chilena de Derecho, vol. 48 (Nº2) [Español].

Simón P. Inteligencia artificial y Administración de Justicia: ¿Quo Vadis, Justitia? (2021) Revista de Internet, Derecho y Política, volumen I (Nº. 33) [Español].

Vargas O. Juez Inteligente. Sistema Experto que Asiste al Juez en la Valoración de la Prueba Judicial (2021). Revista Derecho & Sociedad (Nº. 57) [Español].

3. Recursos de Internet

Consejo General del Poder Judicial. (01 de mayo de 2025). Distribución de la población reclusa extranjera por sexo (diciembre 2021). Estadística Penitenciaria, diciembre 2021. https://www.poderjudicial.es/cgpj/es/Temas/Estadistica-Judicial/Estadistica-por-temas/Datos-penales--civiles-y-laborales/Cumplimiento-de-penas/Estadistica-de-la-Poblacion-Reclusa/ [Español].

Consejo General del Poder Judicial. (01 de mayo de 2025). Página oficial del Consejo General del Poder Judicial, estimación de los tiempos medios de duración de los procedimientos judiciales. https://www.poderjudicial.es/cgpj/es/Temas/Transparencia/ch.Estimacion-de-los-tiempos-medios-de-duracion-de-los-procedimientos-judiciales.formato1/?idOrg=23&anio=2023&territorio=Espa%C3%B1a&proc=PROCEDIMIENTOS%20ABREVIADOS [Español].

Instituto Nacional de Estadística. (01 de mayo de 2025). Estadística de Condenados: Adultos, Total, Total, Nacionalidad, 2021. https://www.ine.es/jaxiT3/Datos.htm?t=25698#!tabs-grafico [Español].

Instituto Nacional de Estadística. (2022). Estadística del Padrón Continuo, Total Nacional, Extranjera, Edad (hasta 100 y más), Total, 1 de enero de 2022. https://www.ine.es/jaxiT3/Datos.htm?t=36825#!tabs-grafico [Español].

Ministerio del Interior. (01 de mayo de 2025). Tabla 4-1-5. Población reclusa penada a 31-12-2021 por grupo de edad, sexo y administración competente. En Anuario estadístico del Ministerio del Interior 2021 (p.323). https://www.interior.gob.es/opencms/pdf/archivos-y-documentacion/documentacion-y-publicaciones/anuarios-y-estadisticas/anuarios-estadisticos-anteriores/anuario-estadistico-de-2021/Anuario-Estadistico-2021_web.pdf [Español].

MinnaLearn, Universidad de Helsinki. (01 de mayo de 2025). Elements of IA. https://www.elementsofai.com/es [Español].

3. Tabla de legislación y jurisprudencia

INTERNACIONAL	
Legislación	Comisión Europea. Libro Blanco sobre la Inteligencia Artificial. Un Enfoque Europeo Orientado a la Excelencia y la Confianza (2020) [Español].
	Consejo de Europa. 'Carta Ética Europea sobre el uso de la Inteligencia Artificial en los sistemas judiciales y su entorno' Sesión plenaria del 3 al 4 de diciembre de 2016 [Español].
	Parlamento Europeo y Consejo Europeo. 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).
Jurisprudencia	Estados Unidos. Tribunal Supremo de Wisconsin. Caso 881 N.W.2d 749 (Estado de Wisconsin v. Loomis) de 13 de julio de 2016.
NACIONAL	
Legislación	Constitución Española, de 29 de diciembre de 1978.

6. Anexos

Tabla 1: Distribución de la población reclusa extranjera por sexo (diciembre 2021)

Extranjeros por sexo	Población Reclusa Extranjera			Porcentaje de Extranjeros		
	Hombres	Mujeres	Total	Hombres	Mujeres	Total
Andalucía	2.567	134	2.701	20,7%	16,4%	20,4%
Aragón	611	29	640	36,2%	31,2%	35,9%
Asturias, Principado	87	10	97	9,5%	11,0%	9,7%
Balears, Illes	423	28	451	33,4%	23,0%	32,5%
Canarias	828	52	880	25,8%	25,1%	25,8%
Cantabria	66	6	72	14,4%	15,0%	14,5%
Castilla y León	1.164	46	1.210	36,5%	19,9%	35,4%
Castilla - La Mancha	466	8	474	28,7%	21,6%	28,6%
Cataluña	3.540	198	3.738	48,9%	38,7%	48,3%
Com. Valenciana	1.532	104	1.636	26,7%	18,0%	25,9%
Extremadura	71	7	78	7,8%	11,1%	8,0%
Galicia	443	16	459	16,9%	8,2%	16,3%
Madrid, Comunidad	2.384	277	2.661	39,8%	42,5%	40,1%
Murcia, Región de	453	32	485	28,1%	28,8%	28,1%
Navarra, C. Foral de	75	4	79	23,7%	13,8%	22,8%
País Vasco	390	19	409	30,6%	16,7%	29,5%
Rioja, La	53	1	54	19,3%	5,9%	18,6%
Ceuta	60	4	64	22,0%	30,8%	22,4%
Melilla	80	4	84	43,0%	57,1%	43,5%
TOTAL	15.293	979	16.272	29,9%	24,9%	29,5%

Fuente: Estadística de la Población Reclusa CGPJ

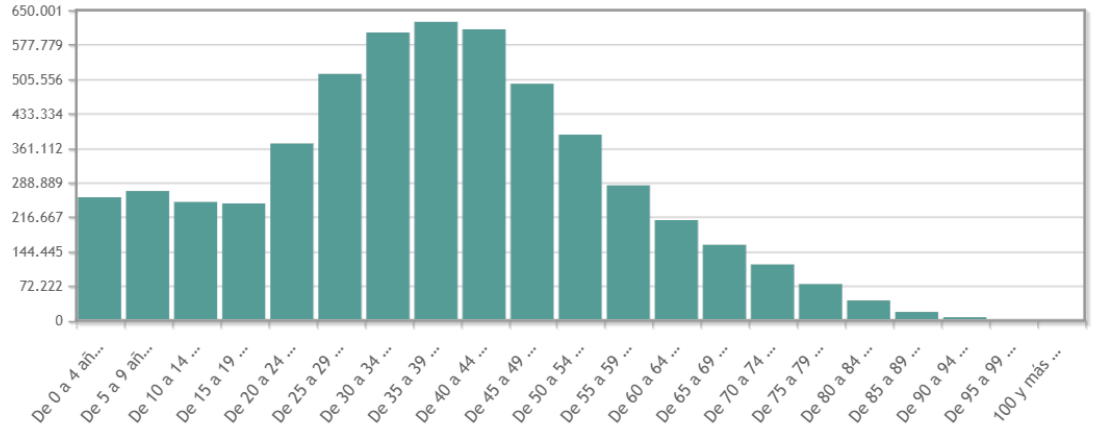
Grupo de edad	Administración General del Estado			Comunidad autónoma de Cataluña*			Comunidad autónoma del País Vasco			Total		
	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total
De 18 a 20 años	239	6	**245**	89	3	**92**	4	0	**4**	**332**	**9**	**341**
De 21 a 25 años	2.307	130	**2.437**	492	41	**533**	79	7	**86**	**2.878**	**178**	**3.056**
De 26 a 30 años	4.080	300	**4.380**	765	52	**817**	129	10	**139**	**4.974**	**362**	**5.336**
De 31 a 40 años	10.654	809	**11.463**	1.805	132	**1.937**	299	28	**327**	**12.758**	**969**	**13.727**
De 41 a 60 años	15.496	1.378	**16.874**	2.375	187	**2.562**	531	58	**589**	**18.402**	**1.623**	**20.025**
De más de 60 años	2.044	163	**2.207**	306	22	**328**	73	2	**75**	**2.423**	**187**	**2.610**
Total	**34.820**	**2.786**	**37.606**	**5.832**	**437**	**6.269**	**1.115**	**105**	**1.220**	**41.767**	**3.328**	**45.095**

* Los datos de la comunidad autónoma de Cataluña incluyen, además de la población reclusa penada, la población reclusa penada con preventivas.

Tabla 2: Población reclusa penada a 31-12-2021 por grupo de edad, sexo y administración competente

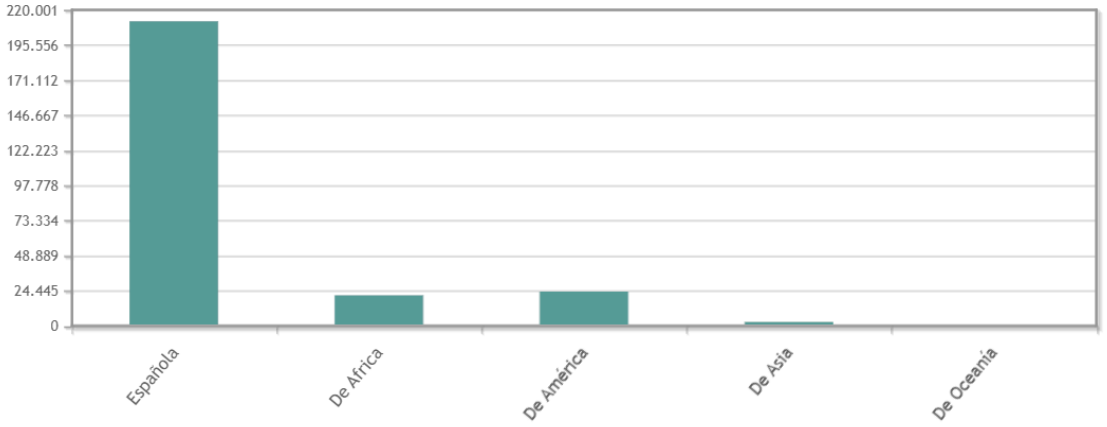
Fuente: Anuario estadístico del Ministerio del Interior 2021

Gráfico 1: Estadística del Padrón Continuo, Total Nacional, Extranjera, Edad (hasta 100 y más), Total, 1 de enero de 2022



Fuente: Instituto nacional de estadística

Gráfico 2: Estadística de Condenados: Adultos, Total, Total, Nacionalidad, 2021



Fuente: Instituto nacional de estadística