



El sesgo de la máquina en la toma de decisiones en el proceso penal

MACHINE BIAS IN CRIMINAL PROCEDURE

Raquel Borges Blázquez

Universidad de Valencia

raquel.borges@uv.es  0000-0003-4899-406X

Recibido: 07 de noviembre 2020 | Aceptado: 14 de diciembre 2020

RESUMEN

La inteligencia artificial tiene innumerables ventajas en nuestras vidas. La capacidad de almacenar y conectar datos que tiene un ordenador es muy superior a la capacidad humana. Pero esta "inteligencia" comporta también problemas de hondo calado ético que el derecho deberá responder. Digo "inteligencia" porque a día de hoy las máquinas no son inteligentes. Las máquinas solo utilizan aquellos datos que, previamente, un humano le ha ofrecido como ciertos. La verdad es relativa y los datos van a tener los mismos sesgos y prejuicios que tiene el humano que programa la máquina. En otras palabras, las máquinas van a ser racistas, sexistas y clasistas si lo son sus programadores. A todo esto se le suma un nuevo problema, la dificultad de comprender el algoritmo por parte de los aplicadores del derecho. Esto puede atacar derechos fundamentales básicos en el sistema de justicia penal como son el derecho a la publicidad de las actuaciones y el derecho a la motivación de las decisiones judiciales. Esta situación obliga a repensar el proceso penal tan y como hoy lo conocemos, incluyendo la inteligencia artificial e hilando muy fino indicando cómo, cuándo, por qué y en qué supuestos podemos hacer uso de la inteligencia artificial y, sobretodo, quién va a programarla. Porque, tal y como indica Silvia Barona, tal vez la pregunta debería ser: ¿quién va a controlar el pensamiento jurídico global?

ABSTRACT

Artificial intelligence has countless advantages in our lives. On the one hand, computer's capacity to store and connect data is far superior to human capacity. On the other hand, its "intelligence" also involves deep ethical problems that the law must respond to. I say "intelligence" because nowadays machines are not intelligent. Machines only use the data that a human being has previously offered as true. The truth is relative and the data will have the same biases and prejudices as the human who programs the machine. In other words, machines will be racist, sexist and classist if their programmers

PALABRAS CLAVE

Inteligencia Artificial
Derechos Humanos
Proceso Penal

KEYWORDS

Artificial Intelligence
Human Rights
Criminal Procedure

are. Furthermore, we are facing a new problem: the difficulty to understand the algorithm of those who apply the law. This situation forces us to rethink the criminal process, including artificial intelligence and spinning very thinly indicating how, when, why and under what assumptions we can make use of artificial intelligence and, above all, who is going to program it. At the end of the day, as Silvia Barona indicates, perhaps the question should be: who is going to control global legal thinking?

I. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL?

Cuando pensamos en Inteligencia Artificial (en adelante, IA) imaginamos un robot que obvia su programación para revelarse contra la humanidad y causar daños a las personas. La ciencia ficción ha contribuido a ello: la IA, al adquirir inteligencia propia, decide en sus fríos cálculos que lo mejor que le puede pasar a la tierra es que desaparezca la humanidad. Es éste precisamente el argumento de *Terminator*. Hablar de IA nos hace pensar en robótica, pese a ser una disciplina distinta. Con orígenes distintos, aunque convergentes en el presente, y probablemente en el futuro.¹

Utilizamos el término IA para referirnos a los sistemas que manifiestan un comportamiento “inteligente”. Que son capaces de analizar su entorno y, con cierto grado de autonomía, tomar decisiones para alcanzar objetivos específicos. Estos sistemas pueden ser un simple programa informático (los asistentes de voz, los programas de análisis de imágenes, los motores de búsqueda y los sistemas de reconocimiento entre otros), o la IA puede estar incorporada en dispositivos hardware (los drones, los coches autónomos, el internet de las cosas...)²

Actualmente no existe un consenso a propósito de la definición de la IA ni siquiera a nivel europeo, menos aún a nivel internacional. De hecho, es un término que se usa a modo de cajón de sastre para referirse a un conjunto de ciencias, teorías y técnicas dedicadas a mejorar la habilidad de las máquinas en la realización de tareas que requieren de inteligencia. Un sistema de IA es un algoritmo que hace recomendaciones, predicciones o decisiones para un conjunto de objetivos previamente marcados. Para llevar a cabo su propósito, el sistema de IA 1) utiliza información proporcionada por humanos o por otra máquina para, por medio de estos, percibir entornos reales o virtuales, 2) abstrae esas percepciones y genera modelos (manual o automáticamente) y 3) genera resultados desde estos modelos en forma de recomendaciones, predicciones o decisiones.³

1. MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, *Revista de Derecho Penal y Criminología*, núm. 20, 2018, p. 89.

2. EUROPEAN COMMISSION, Brussels, 25.4.2018 COM (2018) 237 final, *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe*, {SWD(2018) 137 final}, 2018, p. 1

3. Traducción libre: “Actualmente no hay una definición acordada de “Inteligencia Artificial”. Sin embargo, a los efectos de esta Recomendación, la IA se utiliza como un término general para refe-

El término IA es “una mezcla de eufemismo y desiderátum”.⁴ Es eufemismo porque lo correcto sería denominarla “sistema para el tratamiento y análisis automático de información”, denominación mucho menos atractiva. Y es deseo “porque detrás de estos sistemas existe una voluntad de que sean inteligentes, que sean capaces de tener o imitar procesos cognitivos de los seres humanos”.⁵ Aunque no faltan los que defienden que en un futuro próximo se podrán elaborar sistemas artificiales que sean capaces de asimilar todos los procesos mentales humanos o crear otros diferentes, la realidad es que en la actualidad las IA son modelos matemáticos y la única similitud con la inteligencia humana es su capacidad de cálculo.⁶

Las máquinas no son capaces –por ahora– de interpretar conceptos complejos e introducir variables nuevas. Toman decisiones a partir de premisas lógicas básicas que previamente han sido introducidas o consideradas por un humano. La idea de conciencia mediante la cual los sistemas serán capaces de aprender autónomamente del entorno y adaptarse a él a día de hoy es una realidad utópica.⁷ Así, frente al mito de la anticipación al futuro de *Minority Report*, en realidad la IA solo hace mejores estimaciones a partir de los datos que nosotros le hemos ofrecido (cuanto más y de mayor calidad, mejor será la estimación) siendo, además, mucho más rápida que el humano.⁸

La IA predictiva se apoya en técnicas matemáticas tan complejas que lo verdaderamente importante para que sea efectiva será disponer de una gran cantidad de números, más que de buena información. Esto se debe a que las inteligencias artificiales son algoritmos matemáticos, cada día más sofisticados. La sofisticación de estas complejas técnicas matemáticas busca encontrar los mejores análisis. En la actualidad empiezan a destacar unos modelos lo suficientemente complejos como para que el propio sistema

rirse a un conjunto de ciencias, teorías y técnicas dedicadas a mejorar la capacidad de las máquinas para hacer cosas que requieren inteligencia. Un sistema de IA es un sistema basado en máquinas que hace recomendaciones, predicciones o decisiones para un conjunto determinado de objetivos. Lo hace mediante: (i) utilizar la máquina y/o los insumos basados en humanos para percibir entornos reales y/o virtuales; (ii) abstraer dichas percepciones en modelos de forma manual o automática; y (iii) derivar los resultados de estos modelos, ya sea por medios humanos o automatizados, en forma de recomendaciones, predicciones o decisiones”, “COUNCIL OF EUROPE COMMISSIONER FOR HUMAN RIGHTS, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, 2019, p. 5

4. MIRÓ LLINARES, F., “El modelo policial que viene: mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen” en: *Libro Blanco de la Prevención y Seguridad Local Valenciana*, 2019, p. 100.

5. MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, *op. cit.*, pp. 90-91.

6. “Machines ‘think’ differently to humans; ‘When a computer learns and consequently builds its own representation of a classification decision, it does so without regard for human comprehension”. OSWALD, M.; GRACE, J.; URWIN S.; BARNES, G. C., “Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality”, *Information & Communications Technology Law*, vol. 27, núm. 2, 2018, p. 226.

7. MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, *op. cit.*, p. 92.

8. MIRÓ LLINARES, F., “El modelo policial que viene: mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen”, *op. cit.*, p. 100.

informático aprenda y modifique sus esquemas conforme a sus resultados, autocorrigiéndose a sí mismo y purgando los datos erróneos.⁹

1. El análisis predictivo y la trampa de la neutralidad

Existe la creencia de que la IA se alimenta única y exclusivamente de cifras y que los números solo pueden ser tratados de forma matemática alimentando con esto el mito de su matematización, dejando la herramienta desprovista de cualquier atisbo de socialización. Pero con este mito obviamos que no nos estamos apoyando en números. La realidad es que la IA se alimenta de datos que se corresponden con variables y con la información social que hay dentro de ellos. Nos apoyamos en datos, en información de la vida real. Información relacionada con variables concretas que interactúan entre sí por medio de modelos teóricos que, en el caso de que tengan que ver con las personas, solo van a poder construirse sobre la base de ciencias sociales, es decir, de la psicología, la sociología o la criminología. Como brillantemente defiende MIRÓ LLINARES, aunque la IA requiera de análisis matemáticos, estos se realizan sobre variables sociales y son los científicos sociales aquellos que tendrán que configurar unos marcos explicativos adecuados para su función.¹⁰

Muchos de los dilemas que están detrás de las denominadas decisiones ético-críticas que se aplican a la conducción autónoma y a otras muchas otras situaciones de riesgo que puedan generar las máquinas inteligentes han dado lugar a importantes debates de calado ético. Estos debates, en determinadas situaciones, se relacionan con la atribución de la responsabilidad penal. En este sentido, COCA VILA¹¹ refiere que el sistema de la teoría del delito, todavía arraigado a la tradición filosófica y su eterna lucha entre el consecuencialismo y el deontologismo, puede ser un aliado para alcanzar las decisiones éticas más óptimas en la construcción de la IA.

Pero, como apuntábamos, la IA despliega su eficacia en diversas áreas del saber y el derecho procesal penal requiere también de atención. Debemos ser cautos y estar atentos a los diversos contratiempos que puede traer consigo la utilización de la IA para la salvaguarda de determinados intereses esenciales como la privacidad o el propio funcionamiento del sistema penal. El uso de herramientas inteligentes de valoración del riesgo es ya una realidad en la actuación policial para la prevención e investigación de los delitos y, poco a poco, está empezando a serlo también en la determinación judicial y en el tratamiento penitenciario. Este uso debe respetar los principios propios no solo de derecho penal sustantivo, también del procesal.¹²

9. Íbidem, p. 103.

10. Ídem.

11. COCA VILA, I., "Self-driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law", *Criminal Law and Philosophy*, vol. 12, núm., 1, 2018, pp. 59-82.

12. MIRÓ LLINARES, F., "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *op. cit.*, pp. 96-97.

2. ¿Cómo funcionan los algoritmos?

Los algoritmos son series de ecuaciones matemáticas que se van entrelazando para proporcionarnos un resultado.¹³ Su finalidad no es certificar la veracidad de una hipótesis planteada, lo que hacen es buscar correlaciones deterministas entre unos datos y otros. El algoritmo utiliza una gran base de datos ordenados de una manera comprensible (*smart data*). Más tarde, un modelo matemático irá utilizando esos datos de manera aleatoria hasta establecer patrones de correlación entre ellos que le permitirá realizar apreciaciones con exactitud.¹⁴ Existen algoritmos más avanzados, los denominados algoritmos de aprendizaje o *machine learning*. Estos algoritmos son más eficaces para anticipar comportamientos, pero traen consigo un problema de difícil solución: su funcionamiento, llegados a cierto punto de inclusión de datos y variables, resulta incomprensible para la mente humana.¹⁵

Uno podría pensar que los programadores deberían poder destripar el algoritmo y comprender las decisiones que va tomando mientras se edita a sí mismo. No es el caso. Los algoritmos de aprendizaje de datos cuánto más se alimentan de datos, más autónomos se vuelven. A menudo se basan en miles o millones de factores y los relacionan de una manera que resulta indescifrable para los programadores humanos. Estos algoritmos identifican qué variables conducirán a predicciones más exitosas. Sus resultados solo nos muestran sus tasas de éxito y los códigos que le introdujeron sus programadores. Pero no nos muestra las variables, muy cambiantes, que ha utilizado para llegar al resultado. Además, es poco probable que el programador sea capaz de determinar cuáles son estas variables o cuánto importan para la evolución algoritmo.

Conforme va adquiriendo complejidad, el algoritmo comienza a apoyarse en nuevas variables basadas en contexto y comienza a asignar diferentes pesos a cada variable a medida que analiza cuánto importa esa variable en los patrones que percibe.¹⁶ No

13. Algorithm: A finite suite of formal rules/commands, usually in the form of a mathematical logic, that allows for a result to be obtained from input elements. COUNCIL OF EUROPE COMMISSIONER FOR HUMAN RIGHTS, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, 2019, p. 24

14. Por lo tanto, para estimar la característica clave, un tomador de decisiones debe utilizar un algoritmo –un conjunto de reglas– aplicado a los proxies que el tomador de decisiones cree que se han correlacionado con la característica clave en el pasado y es de esperar que continúe prediciendo la característica clave en el futuro. Con un ejemplo se entenderá más claro. La característica clave a analizar es la solvencia financiera y la proxy el tiempo en pagar las deudas o el contrato de trabajo que tiene. BAMBAUER, J.; ROGERS, J. E., “The Algorithm Game”, *Notre Dame Law Review*, núm. 1, 2018, p. 7.

15. “Even with these steps and the highest dedication to transparency, however, opacity seems difficult to avoid. The HART model contains over 4.2 million decision points, all of which are highly interdependent on the ones that precede them within the tree structure. These details could be made freely available to the public, but would require a huge amount of time and effort to fully understand. It is becoming increasingly difficult to explain to non-computer scientists and non-statisticians how a machine learning forecasting model arrives at its outcomes, and the potential for misunderstanding and even intentional misrepresentation is vast.” OSWALD, M.; GRACE, J.; URWIN S.; BARNES, G. C., “Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality”, *op. cit.*, p. 234

16. Traducción libre: O’DONNELL, R. M., “Challenging racist predictive policing algorithms under the equal protection clause”, 2019, *New York University Law Review*, vol. 94, núm. 3, p. 551.

somos capaces de reconocer los patrones que ha hecho la máquina para llegar a una conclusión. Y, de serlo, tardaríamos meses, semanas o incluso años en explicar algunas de sus decisiones. Y esto dificulta la legitimidad de las decisiones que se adoptan.¹⁷

Es innegable que una parte del razonamiento humano se encuentra en nuestro subconsciente y, por lo tanto, es opaco hasta cierto punto. Nunca llegaremos a saber las razones exactas por las que un juez puso más pena a X que a Y cuando el delito cometido es el mismo. En su sentencia explicará los atenuantes y las agravantes, pero la impresión que le generaron ambos al verlos, los sentimientos que sus declaraciones despertaron, etc. es algo que nunca nos dirá, porque, tal vez, ni siquiera él lo sepa. El aprendizaje automático introduce un tipo diferente de opacidad.¹⁸ MIRÓ LLINARES expone una anécdota muy curiosa que muestra perfectamente cómo algunos de los cálculos son inentendibles para nosotros. Walmart, la mayor cadena de supermercados norteamericana, sabe que existe una correlación estadística entre la cercanía de un huracán y la demanda de tartas rellenas de arándanos. Desconoce la razón por la que se da esa correlación estadística, pero el análisis de los 2,5 petabytes¹⁹ de datos que llegan cada hora a sus servidores sobre sus clientes ha arrojado ese curioso resultado.²⁰

Es aquí precisamente donde empieza la distopía. Si el humano pierde el control sobre la máquina, ¿cómo vamos a obtener una explicación del por qué la máquina ha arrojado determinado resultado? “Estos algoritmos predicen, pero no explican, son capaces de reorganizar sus propias variables y, lo que puede ser más problemático, pueden encontrar relaciones entre el resultado final que se quiere medir (la veracidad de la declaración de un testigo o un acusado) y datos como la edad, género, nacionalidad, creencias religiosas, ideológicas, etc. e, incluso, emplear datos sin haberlos pedidos: es lo que se llama sesgo de los algoritmos”. En definitiva, generan correlaciones entre los datos que manejan, no cadenas de causalidad.²¹

Dentro de los sistemas de IA podríamos establecer dos grandes grupos. El primero, formado por sistemas deterministas o expertos. En este caso, las reglas que procesan la información de los algoritmos ya se encuentran totalmente definidas. El segundo grupo estaría formado por los sistemas no deterministas. En este caso, las reglas no están completamente marcadas y el algoritmo tiene autonomía para procesar la información y producir resultados. Ambos sistemas presentan ventajas e inconvenientes. Los sistemas deterministas tienen la ventaja de su interpretabilidad porque todas sus actuaciones han

17. PÉREZ ESTRADA, M. J., “Capítulo XI. El uso de algoritmos en el proceso penal y el derecho a un proceso con todas las garantías”, en: BARONA VILAR, Silvia, *Claves de la Justicia Penal. Feminización, Inteligencia Artificial, Supranacionalidad y Seguridad*, Tirant Lo Blanch, 2019, pp. 238-239.

18. OSWALD, M.; GRACE, J.; URWIN S.; BARNES, G. C., “Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality”, *op. cit.*, p. 241.

19. Un petabyte es una unidad de almacenamiento de información. Su símbolo es PB, y equivale a 10^{15} bytes = 1 000 000 000 000 000 de bytes. El prefijo peta viene del griego πέντε (cinco).

20. MIRÓ LLINARES, F., “El modelo policial que viene: mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen” *op. cit.*, p. 101.

21. PÉREZ ESTRADA, M. J., “Capítulo XI. El uso de algoritmos en el proceso penal y el derecho a un proceso con todas las garantías”, *op. cit.*, p. 239.

sido predefinidas por sus desarrolladores. En cambio, tienen el inconveniente de que en entornos excesivamente complejos los desarrolladores no pueden prever de antemano todos y cada uno de los escenarios a los que se enfrentará el sistema y, consecuentemente, dejan de ser operativos. Un algoritmo no determinista es capaz de aprender de los datos para ir adecuándose a necesidades que, en el momento de su diseño, eran posibles –incluso probables– pero impredecibles. La desventaja es la baja interpretabilidad de los resultados al perder los diseñadores la capacidad de control sobre el sistema.²²

II. EL USO DE LA INTELIGENCIA ARTIFICIAL EN EL PROCESO PENAL: RACISMO, CLASISMO Y MACHISMO

El uso de las herramientas de gestión del riesgo en la justicia penal a día de hoy está siendo estudiado con lupa por su potencial y la posible afectación a derechos fundamentales. La cuestión es cuánto mejor puede llegar a ser la máquina que nosotros. Se dice que cuando las relaciones entre los predictores y la respuesta son complejas, los sistemas de *machine learning* lo pueden hacer mucho mejor que nosotros. Y cuando las relaciones entre los predictores y la respuesta son simples, su exactitud es bastante similar a los medios tradicionales.²³ La IA puede ser un instrumento muy útil porque es capaz de comparar en un tiempo récord una cantidad de sentencias y datos que una persona no podría hacer en un año (o una vida), pero su uso tiene que ser evaluado con cautelas y no depositar en la máquina funciones que solo competen al juez. Porque entonces corremos el riesgo de pasar de la utopía a la distopía sin darnos cuenta.

La satírica historia del *Juztizklavier* (piano de la justicia, traducido literalmente del alemán) escrita en 1906 por el austriaco Alexander Roda describe a la perfección esta distopía. Un ingeniero norteamericano desarrolló un piano en el que las teclas negras reflejaban tipos delictivos y circunstancias incriminatorias. En cambio, las blancas referían los elementos de descargo, las eximentes y las atenuantes. El veredicto se obtenía automáticamente pulsando las teclas adecuadas. El sueño dorado de la IA: una justicia “limpia, rápida e igual para todos”. El *Juztizklavier* sería capaz de convencer a la gente de la independencia de la administración de justicia. Así que decidió vendérselo a un gobernador centroeuropeo, consciente del descrédito de la justicia en ese país. El gobernador aceptó, pero con una condición: que incluyese los pedales que todo piano lleva. Es decir, un pedal de *piano* y un pedal de *forte*. Así, en el momento de poner la sentencia, el juzgador podría decidir el pedal de *piano* para el partido gobernante y simpatizantes y el pedal de *forte* para la oposición y sus simpatizantes.²⁴ Aunque la

22. PALMA ARTIGOSA, A., “La inteligencia artificial en la toma de decisiones automatizadas. Análisis técnico de las fases que comprenden el desarrollo e implantación de los sistemas inteligentes”, en prensa.

23. BERK, R.; HEIDARIC, H.; JABBARIC, S., KEARNSC, M.; ROTH; A., “Fairness in Criminal Justice Risk Assessments: The State of the Art”, *Berk*, 2017, p. 2

24. DE LA OLIVA SANTOS, A. “Justicia predictiva, interpretación matemática de las normas, sentencias robóticas y la vieja historia del ‘juztizklavier’”, *El Cronista del Estado Social y Democrático de Derecho*, núm. 50, pp. 31-32.

historia del *Juztizklavier* es solo eso, una historia, algunas herramientas de IA sí que parecen discriminar entre *piano* y *forte* dependiendo de a quién juzguen haciendo de la neutralidad un *desiderátum* y una trampa.

El uso de herramientas inteligentes de valoración del riesgo es ya una realidad en la actuación policial para la prevención e investigación de los delitos y, poco a poco, está empezando a serlo también en la determinación judicial y en el tratamiento penitenciario. Este uso debe respetar los principios propios no solo de derecho penal sustantivo, también del procesal.²⁵ Los algoritmos predictivos ayudan al juez en la toma de decisiones judiciales. Estos mecanismos no están exentos de polémica. Su uso puede llegar a afectar e incluso a vulnerar diversos derechos fundamentales como el derecho a un juicio justo, la prohibición de discriminación, la privacidad, la libertad de expresión y el derecho a un remedio efectivo.²⁶

1. A propósito del caso Loomis: los algoritmos predictivos en la toma de decisiones judiciales

Los algoritmos predictivos ayudan al juez en la toma de decisiones judiciales. Estos mecanismos no están exentos de polémica. Su uso puede llegar a afectar e incluso a vulnerar diversos derechos fundamentales como el derecho a un juicio justo, la prohibición de discriminación, la privacidad, la libertad de expresión y el derecho a un remedio efectivo.²⁷ Podemos situar el punto de inflexión el 13 de julio de 2016 con el recurso resuelto por la Corte Suprema de apelaciones del Estado de Wisconsin.²⁸ El caso Loomis y el uso del sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) es el ejemplo más conocido y paradigmático del uso de la IA en el sistema judicial. Creado por la empresa *Northpoint*, COMPAS es un sistema de IA que, haciendo uso de un algoritmo no público –y protegido por la ley de propiedad intelectual– combina en su análisis datos de diferente naturaleza para evaluar el riesgo de que una persona cometa un delito en diversos escenarios: cuando está siendo juzgada y se plantea la prisión provisional o cuando ya ha sido sancionada y tiene que determinarse el castigo concreto que se le aplicará y su forma de ejecutarlo.²⁹ A este respecto, sus creadores se mostraron contrarios a que el sistema se utilizase en sede judicial, pero en la práctica cada vez más jueces estadounidenses lo utilizan.³⁰

25. MIRÓ LLINARES, F., "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *op. cit.*, pp. 96-97.

26. FRA, *BigData: Discrimination in data-supported decision making*, 2018, p. 2.

27. Ídem,

28. PÉREZ ESTRADA, M. J., "Capítulo XI. El uso de algoritmos en el proceso penal y el derecho a un proceso con todas las garantías", *op. cit.*, p. 236.

29. MIRÓ LLINARES, F., "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *op. cit.*, pp. 108-109.

30. "After Zilly was scored as a high risk for violent recidivism and sent to prison, a public defender appealed the sentence and called the score's creator, Brennan, as a witness. Brennan testified that he didn't design his software to be used in sentencing. "I wanted to stay away from the courts," Brennan

Refiere PÉREZ ESTRADA que, si queremos hacer uso de algoritmos en el sistema de justicia penal, debemos preguntarnos qué nivel de transparencia tiene el conocimiento del resultado del algoritmo, esto es, cómo ha manejado los datos y qué valor ha dado a cada variable. COMPAS es un *software* que evalúa el riesgo de reincidencia de un acusado y, tras ello, aconseja a los jueces el tipo de pena y los años a imponerle. Este programa matemático ha sido desarrollado por una empresa privada y es “totalmente opaco”.³¹ Esta aplicación debe su fama al caso *State vs. Loomis* del Tribunal Supremo de Wisconsin.³²

Eric Loomis fue acusado de cinco delitos por su presunta intervención en un tiroteo desde un coche. En total: dos cargos por posesión de armas de fuego, un intento de evasión frente a un agente de tráfico, puesta en peligro de la seguridad pública y conducción de un vehículo ajeno sin autorización. Eric negó haber participado en el tiroteo, pero reconoció haber conducido un vehículo ajeno sin autorización y llegó a un acuerdo con fiscalía para aceptar los dos últimos cargos. El juez aceptó la conformidad y para concretar la pena tuvo en cuenta la evaluación del riesgo del sujeto con la herramienta COMPAS. COMPAS indicó que Eric Loomis presentaba un riesgo alto de reincidencia. Loomis fue condenado a seis años de prisión y cinco años de supervisión tras el cumplimiento de la condena. La suspensión de condena fue excluida por la gravedad del crimen, por su historial en materia de supervisiones judiciales y porque las herramientas de valoración del riesgo aplicadas le otorgaban un riesgo extremadamente alto de reincidencia.³³

Loomis solicitó al tribunal la revisión de la pena argumentando, entre otras razones, que el resultado COMPAS vulneró su derecho al *Due process*. El tribunal no modificó en un ápice la condena y explicó que el sistema de valoración del riesgo COMPAS solo le sirvió para corroborar las conclusiones a las que previamente había llegado. Por tanto, la pena habría sido la misma con independencia del uso, o no, del sistema COMPAS. Loomis apeló entonces ante el Tribunal Supremo de Wisconsin alegando que el uso de COMPAS para la determinación de la pena vulneró su derecho a un proceso con todas las garantías (*right to a due process*) y lo fundamentó en tres razones: 1) Que no había sido condenado sobre la base de información fiable y precisa. El secreto comercial del algoritmo con el que funciona COMPAS le impidió conocer cómo se calcularon las estimaciones de

said, explaining that his focus was on reducing crime rather than punishment. “But as time went on I started realizing that so many decisions are made, you know, in the courts. So I gradually softened on whether this could be used in the courts or not.” Still, Brennan testified, “I don’t like the idea myself of COMPAS being the sole evidence that a decision would be based upon.” ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L., “Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks”, en *ProPublica*, 2016, pp. 12-13.

31. PÉREZ ESTRADA, M. J., “Capítulo XI. El uso de algoritmos en el proceso penal y el derecho a un proceso con todas las garantías”, *op. cit.*, p. 237.

32. Case No. 2015AP157-CR. State of Wisconsin v. Eric L. Loomis. July 13, 2016. Disponible en: <https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690> (última consulta: 15-07-2020).

33. MARTÍNEZ GARAY, L., “Peligrosidad, Algoritmos y Due Process: el Caso State v Loomis”, en *Revista de Derecho Penal Y Criminología*, núm. 20, 2018, p. 490.

riesgo. Sin ellas, su derecho a la defensa quedó mermado. 2) Que se vulneró su derecho a obtener una sentencia individualizada. En su sentencia se tuvieron en cuenta los datos de casos pasados incluidos en un algoritmo. En definitiva, estudiar el pasado para así predecir el futuro. Y 3) que su género se tuvo en cuenta de manera indebida para determinar la pena. Ninguno de estos motivos de recurso fueron admitidos. Sin embargo, la importancia del caso reside en que el Tribunal Supremo estableció algunas limitaciones sobre cómo y para qué pueden ser usadas las evaluaciones de riesgo en la fase de determinación de la pena.³⁴

El Tribunal Supremo no vio vulneración en el derecho a la información porque entendió que el secreto de empresa resulta esencial para el negocio.³⁵ Loomis tuvo acceso tanto a los resultados como a la información con la que se calcularon: datos objetivos sobre el historial criminal del acusado. Tampoco vio vulneración en el derecho a una sentencia individualizada porque la sentencia no solo se basó en los resultados de COMPAS. Ni una utilización indebida de su género. Tanto Fiscalía como la defensa aceptaron en el juicio que “toda la evidencia científica disponible indica que los varones tienen unas tasas de reincidencia y de delincuencia violenta superiores a las de las mujeres”.³⁶ Por tanto, prescindir del factor género disminuiría sensiblemente el acierto de las estimaciones, aumentando las tasas de error tanto en hombres como en mujeres.³⁷ Es más, no está probado que estas tasas de reincidencia perjudiquen más a los varones que a las mujeres. Dado que los varones tienen un mayor riesgo de reincidencia violenta que las mujeres, una misma puntuación bruta puede encontrarse en los deciles bajos de la escala de riesgo de hombres y, en cambio, en los deciles altos de la escala de riesgo de las mujeres.³⁸

A pesar de ello, el Tribunal impuso cautelas que deberán ser tenidas en cuenta por los jueces americanos que utilicen sistemas de valoración del riesgo. Las valoraciones del riesgo no pueden utilizarse para decidir el encarcelamiento del sujeto ni para determinar la severidad de la sentencia. En consecuencia, solo pueden utilizarse para 1) imponer a delincuentes de bajo riesgo medidas alternativas a la prisión, 2) valorar si un

34. Ibidem, p. 491.

35. Salvo un voto particular que subraya este problema: el Tribunal no entiende cómo funciona COMPAS, y pese a haber preguntado en repetidas ocasiones, obtuvo pocas respuestas. *State v. Loomis*, 114, voto particular de la juez Shirley S. Abrahamson.

36. MARTÍNEZ GARAY, L., “Peligrosidad, Algoritmos y Due Process: el Caso *State v Loomis*”, *op. cit.*, p. 492.

37. A este respecto: BERK, R.; HEIDARIC, H.; JABBARIC, S.; KEARNSC, M.; ROTH; A., “Fairness in Criminal Justice Risk Assessments: The State of the Art”, *op. cit.*, pp. 1-42.

38. “COMPAS tiene escalas de riesgo diferentes para varones y para mujeres, construidas con grupos de referencia separados por sexo. Como los varones tienen un mayor riesgo de reincidencia y de reincidencia violenta que las mujeres, un varón con una puntuación X puede ser clasificado en un decil bajo de la escala correspondiente a su grupo normativo de varones, mientras que una mujer con la misma puntuación bruta puede entrar en uno de los deciles altos de su escala referenciada a mujeres, en la que como la media de riesgo es más baja, la misma puntuación X corresponde a un decil (nivel de riesgo) superior. De modo que la consideración del sexo en este caso más bien perjudicaría a la mujer”, MARTÍNEZ GARAY, L., “Peligrosidad, Algoritmos y Due Process: el Caso *State v Loomis*”, *op. cit.*, p. 502.

delincuente puede ser supervisado de una manera efectiva y segura por la comunidad y 3) determinar plazos y condiciones de la suspensión de la pena y de la libertad condicional, así como las consecuencias de su quebrantamiento. Además, no pueden ser el elemento determinante para decidir si el sujeto elude, o no, la condena a prisión.³⁹ Los tribunales deben basarse también en otros factores que han tenido en cuenta y explicarlos en su sentencia.⁴⁰

Tras este caso, se consolida la admisibilidad de los algoritmos en Estados Unidos. Diversos Tribunales norteamericanos han ido expresando una voluntad creciente por hacer uso de las herramientas de IA para mejorar el funcionamiento de la justicia. Si descendemos a la realidad española, el uso de la IA en el proceso penal despierta sentimientos contrapuestos. Surge la duda de si la IA va a reforzar el proceso penal garantista o ayudarnos a liquidarlo.⁴¹ Por un lado, hay que aceptar que algunas de sus aplicaciones pueden ser de gran utilidad para “cubrir múltiples carencias que, ahora mismo, tiene que afrontar el sistema jurisdiccional sin las herramientas adecuadas”. Pero, por otro lado, su uso puede traer consigo vulneraciones “más o menos encubiertas” de derechos fundamentales.⁴²

Tras el caso Loomis, ProPublica realizó un reportaje con el siguiente título: “*Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks*”.⁴³ Este artículo analizaba la fiabilidad de las predicciones de COMPAS y ponía en duda la equidad de un sistema que parecía discriminar de manera sistemática a un grupo de la población.⁴⁴ Cita diversos casos en los que el algoritmo da una mayor puntuación de riesgo de reincidencia a personas negras que a personas blancas. Y como ejemplo, un botón: la historia de Brisha Borden.

Brisha Borden y una amiga llegaban tarde a recoger a la hermana de Brisha del colegio cuando vieron una bici y una scooter sin candar en la calle. Cogieron la bici y la scooter para ir al colegio de la hermana pequeña de Brisha, pero desistieron. En el momento en el que las dos adolescentes se dieron cuenta de que eran muy grandes para las bicis (que pertenecían a un niño de seis años) una mujer apareció corriendo detrás de ellas y gritándoles que eran las cosas de su hijo. Ambas dejaron las bicis y siguieron andando.

39. Se pronuncia en los mismos términos el Parlamento Europeo. “Therefore, ‘assessing the risk of future crime plays no role in sentencing decisions’. In contrast, risk assessment is a key instrument to implement the utilitarian approach, in which punishment is justified by the ability to decrease the probability of future crimes”. EUROPEAN PARLIAMENT, *Understanding algorithmic decision-making: Opportunities and challenges*, 2019, p. 25.

40. MARTÍNEZ GARAY, L., “Peligrosidad, Algoritmos y Due Process: el Caso State v Loomis”, *op. cit.*, pp. 492-493.

41. BARONA VILAR, S., “Inteligencia artificial o la algoritmización de la vida y de la justicia: ¿solución o problema?”, *Rev. Boliv. de Derecho*, núm. 28, 2019, p. 45.

42. PÉREZ ESTRADA, M. J., “Capítulo XI. El uso de algoritmos en el proceso penal y el derecho a un proceso con todas las garantías”, *op. cit.*, pp. 250-251.

43. ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L., “Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks”, *op. cit.*

44. MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, *op. cit.*, p. 109.

Pero era demasiado tarde porque un vecino ya había llamado a la policía. Ambas fueron arrestadas y acusadas por robo y hurto (*burglary and petty theft*) por la suma de 80 dólares. Un año antes, un hombre de 41 años llamado Vernon Prater fue detenido por robar en una tienda enseres por valor de 86,35 dólares, una cantidad ligeramente superior al valor de la bici y la scooter. Prater ya había sido condenado por robo con armas y por intentos de robo con armas. Por ello, pasó 5 años en prisión. Borden también tenía historial delictivo, pero era por delitos menores y cuando era menor de edad. Un programa de ordenador predijo la posibilidad que ambos tenían de reincidir. Borden, que es negra, fue clasificada con un riesgo alto. Prater, que es blanco, fue clasificado con un riesgo bajo. Dos años después, sabemos que el programa estaba equivocado. Borden no volvió a cometer nuevos delitos y Prater pasó ocho años en prisión por un robo posterior.⁴⁵ Esto, lejos de ser un caso aislado parece ser un patrón que se repite. El motivo probablemente sea que los datos de los que dispone la máquina muestran una realidad sesgada. Es imposible predecir el crimen porque nos faltan datos. No conocemos todos los crímenes que sucedieron, ende, los datos están sesgados. En lenguaje informático: *Bias In, Bias Out*.⁴⁶

Tabla 1. Puntuaciones de riesgo de acusados negros.

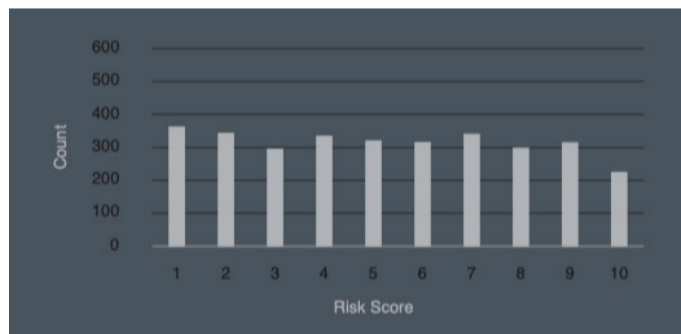
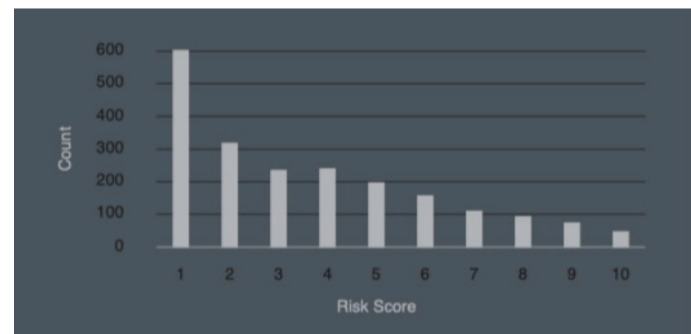


Tabla 2. Puntuaciones de riesgo de acusados blancos.



Fuente: *Machine Bias. ProPublica*

Con una puntuación de 10 como riesgo muy alto y una puntuación de 1 como riesgo inexistente, vemos como el porcentaje de acusados negros se mantiene en los mismos valores a lo largo de la escala de riesgo. En cambio, la gran mayoría de acusados blancos tienen riesgos muy bajos, siendo que los grupos más reducidos se encuentran en los riesgos más altos. Pero lejos de quedarnos en la lectura pesimista, tenemos que aprender del pasado y ser capaces de corregirlo. Sin intervención, el futuro repetirá los datos del pasado. El análisis predictivo es un espejo del pasado que se proyecta en el futuro. Predecir el futuro bajo las condiciones *status quo* es simplemente proyectar el historial hacia adelante. Debido a la naturaleza de la predicción, un pasado con desigualdades proyectará un futuro desigual. Para ser más específicos, si lo que nos comprometemos a predecir, digamos un arresto, le sucedió con más frecuencia a las personas negras que

45. ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L., "Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks", *op. cit.*, pp.1-2.

46. MAYSON, S. G., "Bias In, Bias Out", en *The Yale Law Journal*, 2019, pp. 2218-2300.

a las personas blancas en los datos anteriores, un análisis predictivo proyectará que sucederá con más frecuencia a las personas negras que a las personas blancas en el futuro. El evento predicho, llamado variable objetiva, es la clave de la disparidad racial en la predicción.⁴⁷

2. A propósito de la sentencia de la AN de 30 de septiembre de 2020: sistemas predictivos y violencia de género

Hace tiempo que se viene apuntando las dificultades inherentes a determinados conflictos que, aunque constitutivos de un delito, tienen su origen en un problema de naturaleza social o psicológica.⁴⁸ De manera incomprensible se pide al derecho penal (sustantivo y procesal) que dé solución a una situación de violencia en la (ex) pareja en la que una parte somete a la otra a todo tipo de vejaciones, dolor y miedo. Pero ¿puede el sistema judicial español resolver estos conflictos? Ni puede, ni es su cometido. La justicia puede y debe proteger a la víctima y castigar al victimario. También debe poder rehabilitarlo, al ser éste un mandato constitucional.⁴⁹ No es esto lo que sucedió en el caso objeto de estudio de este trabajo.

Con fecha 30 de septiembre la Sala de lo Contencioso-administrativo de la Audiencia Nacional ha condenado al Ministerio del Interior por la deficiente protección que la Guardia Civil otorgó a una mujer que solicitó una orden de protección. Un cuestionario de cribado le otorgó “riesgo bajo”. Sin realizar más averiguaciones los agentes calificaron el riesgo como “no apreciado”. Y esta misma valoración fue determinante para que también el juzgado denegase la medida de protección a la fallecida. Considera la sala que “la actuación de los agentes ante situaciones de violencia de género no debería quedar limitada a aspectos formales de atención a la denunciante, asistencia, información de derechos y citación a juicio, sino que su actuación exige una atención preferente de asistencia y protección de las mujeres que han sido objeto de comportamientos violentos en el ámbito familiar, a los efectos de prevenir y evitar, en la medida de lo posible, las consecuencias del maltrato”.

En este caso, el cuestionario de cribado de IA no apreció un riesgo que, con una entrevista personal y aplicando la perspectiva de género, se habría considerado. Fue la falta de concienciación social e institucional sobre la violencia de género unida a la confianza ciega que muchas veces depositamos en los sistemas de IA la que conllevó el fallecimiento de la solicitante de protección. La sentencia aprecia un funcionamiento

47. Ibidem, p. 2224.

48. MARTÍNEZ GARCÍA, E., “Capítulo 14: Posibilidades reales que ofrece la mediación penal en los procesos por violencia de género: violencia leve, primaria, perfiles de las partes y modelo procesal recomendable”, en: BARONA VILAR, Silvia (ed.), Justicia civil y penal en la era global, Tirant Lo Blanch, Valencia, 2017, p. 415.

49. De acuerdo con el apartado segundo del artículo 25 CE “Las penas privativas de libertad y las medidas de seguridad estarán orientadas hacia la reeducación y reinserción social y no podrán consistir en trabajos forzados.”

erróneo tanto en el servicio de la Guardia Civil como de la policía judicial. La respuesta se limitó a la recogida de datos automatizados, pero no previno la violencia ni reevaluó el riesgo por medio de agentes especializados en su tratamiento y sensibilizados con la lacra de la violencia de género. La sala reconoce el quebrantamiento de la obligación estatal positiva de proteger. Una víctima pidió ayuda y un sistema ajeno al género se la denegó. Existían indicios: su marido tenía antecedentes por maltrato en el país de origen, pero no se comprobaron; la violencia se ejercía delante de los menores e incluso delante de la madre del agresor, pero no se les tomó declaración. Las trabajadoras sociales describen una víctima totalmente sometida, con pánico a su agresor. Pero nada de esto fue suficiente para reevaluar el riesgo que un sistema de IA otorgó de manera errónea. Precisamente aquí es donde surge la responsabilidad estatal: no se puede depositar en la máquina la responsabilidad de la toma de decisiones que debe hacer el humano. Porque si alguien con la suficiente concienciación hubiese entrevistado a esta mujer, habría apreciado un riesgo muy alto y ahora estaría viva.

El derecho no es neutro y su aplicación igualitaria en casos donde una de las partes se encuentra en una posición de subordinación puede traer como consecuencia una nueva discriminación, esta vez desde el sistema judicial.⁵⁰ Juzgar con perspectiva de género implica hacer realidad el derecho a la igualdad. Responde a una obligación constitucional (artículo 9.2 CE) y supranacional (artículo 10.2 CE) de combatir la discriminación mediante el derecho procesal para garantizar el acceso en igualdad de condiciones a la justicia y remediar, en el caso concreto, situaciones asimétricas de poder.⁵¹ Pero la constitución lleva más de cuarenta años en vigor, y no se ha combatido esta discriminación. En este sentido, AVILÉS apunta que el principal obstáculo para alcanzar la igualdad real lo constituye la –paradójica– convicción social de que hombres y mujeres somos iguales.⁵²

Juzgar con perspectiva de género permite transformar las prácticas⁵³ de aplicación e interpretación del derecho y actuar de una manera global sobre el conflicto jurídico.

50. “Han sido tantos siglos de unidimensionalidad masculina que aun hoy en la actualidad es posible percibir que lo que se presenta como sexualmente neutral esconde en ciertos casos mecanismos de dominación masculina”, en: BARONA VILAR, S., “La necesaria deconstrucción del modelo patriarcal de justicia”, en: AAVV, *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018, p. 60

51. MARTÍNEZ GARCÍA, E., “Análisis de la justicia procesal desde la perspectiva de género”, Proyecto de investigación, Segundo ejercicio de Cátedra inédito defendido en la Universidad de Valencia, 2018, p. 44

52. AVILÉS PALACIOS, L., “La perspectiva de género como técnica jurídica e instrumento necesario para una justicia igualitaria”, en: AAVV, *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018, p. 282

53. Si no aplicamos la perspectiva de género “la diferencia femenina no sufre discriminación en el plano jurídico, puesto que en este plano resulta desconocida, oculta y enmascarada: las mujeres tienen los mismos derechos que los varones en cuanto son considerados o se finge que son (como los del) varón y se asimilan a ellos en los estilos de vida y en los modelos de comportamiento. Pero, precisamente porque desconocida de derecho, aquélla resulta penalizada de hecho –lo mismo para las mujeres que se asimilan que para las que no se asimilan– por los amplios márgenes de ineffectividad de la proclamada igualdad” FERRAJOLI, L., *La ley del más débil*, Editorial Trotta, 7 edición, 2010, p. 75

Aplicar esta perspectiva nos permite actuar sobre las personas, sobre los hechos y sobre la norma jurídica, aplicando una visión crítica de la realidad. Nos permite “ver” y nos impulsa a ser curiosos, testarudos y garantes de derechos, para reparar y dignificar a quién parte de una situación vital y social de desventaja frente a la desigualdad.⁵⁴ Es por ello que se hace necesario preguntarnos si la justicia tiene, o no, perspectiva de género.⁵⁵ Perspectiva que deberemos demandar también al sistema de IA que servirá de auxilio a las diversas decisiones que se tomen a lo largo del proceso.

III. PREDECIR EL FUTURO MIRANDO AL PASADO: GARBAGE IN, GARBAGE OUT

Las herramientas predictivas sistematizan lo que ya se hacía y se hace de manera tradicional y manual, probablemente con los mismos sesgos o incluso más.⁵⁶ La decisión de dónde se patrulla, dónde se ponen los radares o si se aplica, o no, la suspensión de la pena se tiene que seguir tomando. Todo aquello que aporte información científica y que supere nuestros prejuicios subjetivos, en principio, debería verse como algo positivo. Pero que debe ser evaluado con cautelas. En primer lugar, porque la IA imparcial es una utopía. Lo único que sabemos hasta el momento es que los algoritmos reflejan con precisión nuestro mundo, un mundo que está lleno de prejuicios. Por tanto, el sesgo humano puede perpetuarse e incluso incrementarse con este tipo de técnicas, siendo especialmente grave en las IA aplicadas al sistema de justicia penal.⁵⁷

El uso de la IA para decidir sobre la gravedad de la pena a imponer hace que los jueces puedan corroborar su intuición con un algoritmo de valoración del riesgo y, así, poner la sentencia con la conciencia más tranquila.⁵⁸ Deberemos hilar muy fino indicando qué puede y qué no puede hacer una máquina. A día de hoy, disponemos de los conocimientos y de la tecnología para crear máquinas que valoren la probabilidad de reincidencia del condenado por sentencia firme y, tras ello, opten por la libertad vigilada o

54. GUTIERREZ-SOLANA JOURNOUD, A., “La invisible perspectiva de género del TJUE en asuntos que afectan exclusivamente a las mujeres: generalidad frente a especificidad”, en: ETXEBARRÍA ESTANKONA, K.; ORDEÑANA GEZURAGA, I.; OTAUZA ZABALA, G. (Dirs.) *Justicia con ojos de mujer. Cuestiones procesales controvertidas*, Tirant Lo Blanch, Valencia, 2018, pp. 837-838

55. Por lo que respecta a la violencia de género, haremos referencia a una visión amplia omnicompreensiva tanto de la perspectiva que propugna NNUU como de la utilizada en el ordenamiento jurídico español, esto es, LOVG y Pacto de Estado.

56. “It is useful to remind ourselves that profiling is not new and can be done without modern automated and statistical techniques (‘human profiling’). Practices of ethnic profiling by police during stops and searches on the Street are an example of human profiling, albeit a problematic one”. LAMMERANT, H., DE HERT, P. “Predictive profiling and its legal limits: Effectiveness gone forever” en: VAN DER SLOOT, B.; BROEDERS, D.; SCHRIJVERS, E. (eds.), *Exploring the boundaries of big data*. Amsterdam University Press/WRR, vol. 32, 2016, p. 146

57. MIRÓ LLINARES, F., “Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots”, *op. cit.*, p. 126.

58. MARTÍNEZ GARAY, L., “Peligrosidad, Algoritmos y Due Process: el Caso State v Loomis”, *Revista de Derecho Penal Y Criminología*, núm. 20, 2018, p. 496.

la pena de prisión. Cuestión distinta es si deberíamos dejar que sea una máquina la que decida cuestiones tan sensibles como es la libertad de las personas.⁵⁹

Probablemente estemos frente a uno de los debates éticos más complejos e importantes a los que se va a enfrentar la justicia penal en las próximas décadas. En este debate es fácil avistar dos posiciones enfrentadas. En un lado, aquellos que defienden que los algoritmos no deben renunciar a determinadas variables como la edad, la etnia, el sexo, el grupo social y la demografía porque excluir variables altamente correlacionadas, aunque éticamente cuestionables, afectaría a la capacidad predictiva. En el otro lado, están quienes defienden que las cargas legales deben tener relación con la responsabilidad individual más que con otros factores problemáticos desde la perspectiva ético-legal y que, además, no dependen del individuo (por ejemplo, ser varón, de menos de treinta años y latino). Tal vez el equilibrio debería encontrarse en un punto medio entre ambas visiones. Así, deberíamos diferenciar dos tipos principales de sesgos: sesgos en datos de entrenamiento⁶⁰ y sesgos por una distribución desigual real⁶¹ de las variables.⁶²

Predecir el futuro mirando al pasado conlleva el riesgo de que en el futuro se sigan repitiendo los mismos estereotipos de género que vamos superando como sociedad. Debemos ser prudentes no solo en fase de aplicación de la IA, sino ir un paso más allá y preguntarnos quién hay detrás de la máquina, cómo se ha hecho el algoritmo, a quién beneficia y a quién perjudica –consciente o inconscientemente– porque puede acabar por controlar el pensamiento jurídico global.⁶³

Si queremos hacer uso de sistemas de IA debe ser siempre como apoyo y nunca sustituyendo la función del juez, pues esto atentaría directamente contra el artículo 117 de nuestra Constitución. Solo los jueces pueden juzgar, pero para juzgar pueden ayudarse

59. BARONA VILAR, S., "Inteligencia artificial o la algoritmización de la vida y de la justicia: ¿solución o problema?", *op. cit.*, p. 46.

60. Debido a que las IA aprenden con nuevos ejemplos, los datos de entrenamiento se perpetúan en el autoaprendizaje del algoritmo. Algunos de los errores que podrá arrastrar consigo este entrenamiento son: a) que la máquina sea alimentada con datos que han sido erróneamente calificados por el programador, viciando así el aprendizaje desde el comienzo; b) que los datos iniciales sean fruto de un muestreo no representativo. Consecuentemente se reduce su validez y aumenta el error de la interpretación de los resultados; c) que los parámetros del aprendizaje se realicen en intervalos temporales acotados. La inobservancia de este tipo de limitaciones perpetuaría la discriminación algorítmica hacia grupos infrarrepresentados, pudiendo lesionar o poner en riesgo sus derechos fundamentales.

61. En este supuesto se trata de variables que verdaderamente se distribuyen de manera desigual en la realidad. Son factores aparentemente neutrales que han sido recogidos por la IA pero que no se refieren al actuar de una persona en concreto, se refieren al actual pasado de aquellos con los que comparte género, etnia, raza, edad, etc. Estos son caracteres casi inmutables de los que el sujeto no se puede desprender.

62. MIRÓ LLINARES, F., "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *op. cit.*, pp. 122-123; MIRÓ LLINARES, F., "El modelo policial que viene: mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen" *op. cit.*, p. 111.

63. BARONA VILAR, S., "Cuarta revolución industrial (4.0.) o ciberindustria en el proceso penal: revolución digital, inteligencia artificial y el camino hacia la robotización de la justicia", *Revista Jurídica Digital UANDES*, vol. 3, núm. 1, 2019, p. 16.

de sistemas de IA. Estos sistemas deben ser tomados con cautelas y no depositar en ellos esperanzas irrealizables. Los algoritmos han sido creados por humanos que viven en esta, nuestra sociedad. Por tanto, estos algoritmos reproducen los sesgos de género existentes en nuestra sociedad. Considerar que los números son objetivos es una falacia. Los números han sido alimentados con datos que distan mucho de ser objetivos. Esta misma perspectiva de género que debe aplicar el juzgador, también hemos de exigirla a los programadores. Porque sin ella estamos condenados a seguir repitiendo los mismos estereotipos del pasado.

El alcance predictivo de estos algoritmos no es total, y debemos prestar especial atención en las decisiones jurídico-penales complejas. Ni el juicio de la IA es imparcial (reproduce nuestros sesgos) ni su fiabilidad es absoluta (su código de programación todavía no es capaz de interpretar con precisión y fiabilidad contextos confusos). Por ello, comparto con MIRÓ LLINARES que la IA del futuro deberá ser creada por equipos interdisciplinarios con “científicos sociales y juristas capaces de establecer tanto los criterios jurídicos y criminológicos de clasificación como de interpretación de resultados”. Porque si solamente programan los *data scientists* su objetivo será obtener el mejor rendimiento de la herramienta en términos estrictamente matemáticos,⁶⁴ olvidando factores como la equidad o la igualdad real.

BIBLIOGRAFÍA

- ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L., “Machine Bias. There’s software used across the country to predict future criminals. And it’s biased against blacks”, en *ProPublica*, 2016.
- AVILÉS PALACIOS, L., “La perspectiva de género como técnica jurídica e instrumento necesario para una justicia igualitaria”, en: AAVV, *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018.
- BAMBAUER, J.; ROGERS, J. E., “The Algorithm Game”, *Notre Dame Law Review*, núm. 1, 2018.
- BARONA VILAR, S., “La necesaria deconstrucción del modelo patriarcal de justicia”, en: AAVV, *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018.
- BARONA VILAR, S., “Inteligencia artificial o la algoritmización de la vida y de la justicia: ¿solución o problema?”, *Rev. Boliv. de Derecho*, núm. 28, 2019.
- BERK, R.; HEIDARIC, H.; JABBARIC, S.; KEARNSC, M.; ROTH; A., “Fairness in Criminal Justice Risk Assessments: The State of the Art”, *Berk*, 2017.
- COCA VILA, I., “Self-driving Cars in Dilemmatic Situations: An Approach Base don the Theory of Justification in Criminal Law”, *Criminal Law and Philosophy*, vol. 12, núm., 1, 2018.
- COUNCIL OF EUROPE COMMISSIONER FOR HUMAN RIGHTS, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, 2019.
- DE LA OLIVA SANTOS, A., “Justicia predictiva, interpretación matemática de las normas, sentencias robóticas y la vieja historia del ‘juztizklavier’”, *El Cronista del Estado Social y Democrático de Derecho*, núm. 50.

64. MIRÓ LLINARES, F., “El modelo policial que viene: mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen”, *op. cit.*, pp. 111-112.

- EUROPEAN COMMISSION, Brussels, 25.4.2018 COM (2018) 237 final, *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe*, {SWD(2018) 137 final}, 2018.
- EUROPEAN PARLIAMENT, *Understanding algorithmic decision-making: Opportunities and challenges*, 2019.
- FERRAJOLI, L., *La ley del más débil*, Editorial Trotta, 7 edición, 2010.
- FRA, *BigData: Discrimination in data-supported decision making*, 2018, p. 2. BARONA VILAR, Silvia, "Cuarta revolución industrial (4.0.) o ciberindustria en el proceso penal: revolución digital, inteligencia artificial y el camino hacia la robotización de la justicia", *Revista Jurídica Digital UANDES*, vol. 3, núm. 1, 2019.
- GUTIERREZ-SOLANA JOURNOUD, A., "La invisible perspectiva de género del TJUE en asuntos que afectan exclusivamente a las mujeres: generalidad frente a especificidad", en: ETXEBARRÍA ESTANKONA, K.; ORDEÑANA GEZURAGA, I.; OTAUZA ZABALA, G. (Dir.) *Justicia con ojos de mujer. Cuestiones procesales controvertidas*, Tirant Lo Blanch, Valencia, 2018.
- LAMMERANT, H., DE HERT, P. "Predictive profiling and its legal limits: Effectiveness gone forever" en: VAN DER SLOOT, B.; BROEDERS, D.; SCHRIJVERS, E. (eds.), *Exploring the boundaries of big data*. Amsterdam University Press/WRR, vol. 32, 2016.
- MARTÍNEZ GARAY, L., "Peligrosidad, Algoritmos y Due Process: el Caso State v Loomis", en *Revista de Derecho Penal Y Criminología*, núm. 20, 2018.
- MARTÍNEZ GARCÍA, E., "Análisis de la justicia procesal desde la perspectiva de género", Proyecto de investigación, Segundo ejercicio de Cátedra inédito defendido en la Universidad de Valencia, 2018.
- MARTÍNEZ GARCÍA, E., "Capítulo 14: Posibilidades reales que ofrece la mediación penal en los procesos por violencia de género: violencia leve, primaria, perfiles de las partes y modelo procesal recomendable", en: BARONA VILAR, Silvia (ed.), *Justicia civil y penal en la era global*, Tirant Lo Blanch, Valencia, 2017.
- MAYSON, S. G., "Bias In, Bias Out", en *The Yale Law Journal*, 2019.
- MIRÓ LLINARES, F., "El modelo policial que viene: mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen" en: *Libro Blanco de la Prevención y Seguridad Local Valenciana*, 2019.
- MIRÓ LLINARES, F., "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *Revista de Derecho Penal y Criminología*, núm. 20, 2018.
- O'DONNELL, R. M., "Challenging racist predictive policing algorithms under the equal protection clause", 2019, *New York University Law Review*, vol. 94, núm. 3.
- OSWALD, M.; GRACE, J.; URWIN S.; BARNES, G. C., "Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality", *Information & Communications Technology Law*, vol. 27, núm. 2, 2018.
- PALMA ARTIGOSA, A., "La inteligencia artificial en la toma de decisiones automatizadas. Análisis técnico de las fases que comprenden el desarrollo e implantación de los sistemas inteligentes", en prensa.
- PÉREZ ESTRADA, M. J., "Capítulo XI. El uso de algoritmos en el proceso penal y el derecho a un proceso con todas las garantías", en: BARONA VILAR, Silvia, *Claves de la Justicia Penal. Feminización, Inteligencia Artificial, Supranacionalidad y Seguridad*, Tirant Lo Blanch, 2019.