

¿DERECHO A UNA DECISIÓN HUMANA? EL USO DE LA INTELIGENCIA ARTIFICIAL EN LA PREDICCIÓN JUDICIAL DE LA REINCIDENCIA

Sección 41. ROBÓTICA, INTELIGENCIA ARTIFICIAL Y REALIDAD VIRTUAL Y AUMENTADA

RESUMEN

§. El presente trabajo analiza los presupuestos técnicos y jurídicos que serían necesarios para aceptar la introducción del software predictivo de la reincidencia en el proceso penal español.

§. Partiendo de las enseñanzas que proporciona la experiencia estadounidense, se analizan los principales peligros -y los potenciales beneficios- que supondría la incorporación de esta tecnología predictiva en la toma de decisiones judiciales.

§. Finalmente, el trabajo analiza también la pertinencia de esta utilización de la *inteligencia artificial* (IA) en función del tipo de resolución judicial concreta de que se trate -cautelar, para la fijación de la extensión de la pena, medidas de seguridad, etc.-, identificándose los principios de Derecho Penal que habrían de ser respetados en todo caso.

1. Introducción

§. La utilización de algoritmos y, en un sentido más amplio, de la inteligencia artificial (IA) -entendida como «sistemas para el tratamiento y análisis automático de la información¹», en la administración de la Justicia está lejos de ser un ejercicio de ciencia ficción. Se trata de una realidad, desde hace años, en países como EE. UU.² De hecho, la IA se ha introducido ya en el proceso penal español en un caso muy puntual: el de la violencia de género³.

§. No es casual que la aplicación de algoritmos para la toma de decisiones relativas a la prisión provisional/libertad bajo fianza (*parole*) o la extensión de la condena de privación de libertad (*evidence-based sentencing*) haya tomado impulso precisamente allí. Desde hace lustros, los Estados Unidos de Norteamérica se enfrentan a un gravísimo problema de política penitenciaria. El porcentaje de población reclusa por habitante ha crecido de manera constante desde hace años, lo que unido al alto coste que supone para las arcas públicas el mantenimiento de una persona en prisión (al margen de los costes sociales que la inocuización inevitablemente trae consigo) ha llevado tanto a las asociaciones de abogados, como a las autoridades de diversos Estados, a replantearse las bases de su sistema penal⁴.

§. De entre esos nuevos planteamientos, uno de los que ha tenido mayor eco es la introducción de mecanismos automatizados que permitan al juez obtener una información actuarial objetiva sobre el riesgo de reincidencia de un acusado (a la hora de acordar o no la prisión provisional) o de un condenado (para fijar la extensión de la pena o de una medida de seguridad). En algunos Estados se establece la obligatoriedad

¹ MIRÓ MIRALLES, “Inteligencia Artificial y Justicia Penal: más allá de los resultados lesivos causados por robots”, *UNED, Revista de Derecho Penal y Criminología*, nº 20, 2018, p. 90.

² Cf. STARR, “Evidence-Based Sentencing and the Scientific Rationalization of Discrimination.” *Stanford Law Review*, 66, no. 4, 2014, pp. 803-72.

³ Se trata de la herramienta VDG o VIOGÉN, de seguimiento integral en los casos de violencia de género, que encuentra su (discutible) base normativa en el art. 282 LECrim y en el Protocolo para la valoración policial del nivel de riesgo de violencia de género (LO 1/2004) y gestión de seguridad de las víctimas, aprobado mediante la Instrucción 7/2016, de la Secretaría de Estado de Seguridad. Se trata de un software que se alimenta de datos procedentes de registros oficiales, declaraciones en sede judicial, dispositivos GPS, etc., y que, mediante la aplicación de algoritmos, establece una valoración concreta del riesgo de reincidencia que clasifica en cinco niveles y que se comunica al Juzgado de Violencia sobre la Mujer y a la Fiscalía. Para una primera aproximación, SÁNCHEZ LÓPEZ, “La diligencia policial de valoración del riesgo de violencia de género en el sistema Viogén”, *Foro*, vol. 22, 2019, pp. 119-130.

⁴ KLINGELE, “The Promises and Perils of Evidence-Based Corrections”, nº 91, *Notre Dame Law Review*, 2016, pp. 544-547.

de que el juez consulte este tipo de aplicaciones informáticas antes de adoptar una decisión de esa naturaleza. En otros, se acoge como una posibilidad⁵.

§. Originariamente, este tipo de aplicaciones informáticas tenían como finalidad reducir el elevado número de prisiones preventivas acordadas, ante la creencia -fundada- de que había una sobreabundancia de “falsos positivos” (es decir, que se acordaban en exceso privaciones preventivas de libertad deambulatoria frente a individuos que no reincidían ni se sustraían de la acción de la Justicia). Muy pronto, este tipo de programas informáticos fueron aplicados también a la hora de fijar la extensión concreta de la sentencia a imponer, utilizándose este tipo de instrumentos algorítmicos para medir el riesgo de reincidencia futura (uno de los factores clave para los tribunales de EE. UU. a la hora de concretar la extensión de la condena dentro de la horquilla legal permitida).

§. El análisis de lo que está ocurriendo en el sistema judicial norteamericano no debe verse como ejercicio comparatista exótico. España presenta, al igual que Estados Unidos, una abultada tasa de población reclusa -en particular, si la comparamos con otros países de la UE-, a pesar de padecer, paradójicamente, uno de los índices de criminalidad más bajos de las sociedades occidentales⁶. El coste económico de mantener a esa población reclusa es también significativo, como lo son los costes sociales anudados a este tipo de medidas punitivas⁷.

§. La discusión no parece ser tanto si tales modelos predictivos deben o no ser empleados por los jueces españoles, sino *cuándo y cómo lo serán*. A este respecto, también debe reseñarse que la precisión de este tipo de aplicaciones informáticas sigue mejorando de forma muy rápida, por lo que más pronto que tarde no tendrá mucho sentido volver la espalda por completo a las aportaciones que pueden venir de la mano de la aplicación de la IA en la toma de decisiones judiciales. Porque, bien mirado, buena parte de la labor que realiza este tipo de *software* viene siendo ejecutada hasta el

⁵ Vid., “Using Offender Risk and Needs Assessment Information at Sentencing: Guidance for Courts from a National Working Group”, 2011, documento accesible en: <https://ncsc.contentdm.oclc.org/digital/collection/criminal/id/19>

⁶ RUIZ MORALES, “La evolución de la población reclusa española en los últimos treinta años. Una explicación integral”, *Anuario de derecho penal y ciencias penales*, Tomo 71, 2018, pp. 403-490.

⁷ Según detalló recientemente la Secretaría General de Instituciones Penitenciarias, el coste por preso asciende en España a 64,82 euros/día, siendo también relevante tomar en consideración que, a finales de 2019, el número de presos en las cárceles que dependen de la Administración General del Estado era de 50.129. Esta cifra supone un coste por cada interno cercano a los 2.000 euros al mes y más de 23.500 euros al año. En el documento se detalla que la cantidad diaria por recluso ha crecido de modo continuado desde los 52,53 euros de 2010 hasta la cifra actual, casi un 25% más. *Diario El País*, 20 de febrero de 2020.

momento, *informalmente*, por los jueces -pero sin contar una métrica precisa a su disposición- con el consiguiente riesgo de dejarse influenciar por estereotipos de género, de raza, de clase social, etc.

§. No es casual, por tanto, que el Anteproyecto de Ley de Enjuiciamiento Criminal aprobado el 24 noviembre de 2020 por el Consejo de Ministros contenga un precepto específicamente dirigido a regular esta materia⁸. En el presente trabajo se estudian las principales objeciones teóricas que plantea la utilización de este tipo de herramientas en el ámbito jurisdiccional y los presupuestos mínimos que deberían concurrir para aceptar su incorporación al proceso penal.

2. Los problemas que plantea su introducción en el proceso penal

§. Nos centraremos en identificar los principales obstáculos que presenta la utilización práctica, *en la actualidad*, de este tipo de herramientas automatizadas. Precisamos que ese diagnóstico es necesariamente coyuntural, a la vista del incesante progreso de la IA y los crecientes desafíos éticos y jurídicos que plantea⁹.

a) La (falta de) transparencia

§. En EE. UU., al estar usualmente protegidas por derechos de propiedad industrial y haber sido diseñados por corporaciones privadas las características técnicas de este tipo de *software* informático no son, en principio, de acceso público. Esta falta de transparencia supone un obstáculo no menor para su admisión en el seno de un proceso judicial. La imposibilidad actual de acceder a la denominada “caja negra” (*black box*)¹⁰ o código fuente de estas aplicaciones impide a las partes (y, en particular, al

⁸ Se trata del art. 485, bajo la rúbrica de “Instrumentos de valoración del riesgo de violencia o reincidencia”:

“1. Los instrumentos de valoración del riesgo de violencia o reincidencia deberán incluir todos los parámetros estadísticos que permitan evaluar tanto su fiabilidad como su capacidad predictiva.

2. Dichos instrumentos especificarán el tamaño de la población con la que se han construido, las variables utilizadas como factores de riesgo, los criterios de medición empleados para ponderar dichos factores asignando puntuaciones, y el tiempo de validez de la predicción. También habrán de identificar los estudios de validación realizados”.

⁹ Un buen punto de partida es la “Carta Europea para el uso ético de la Inteligencia Artificial en los sistemas judiciales y su entorno”, elaborado por el CEPEJ en 2018, en el se proclaman varios principios básicos: de respeto a los derechos humanos, de no discriminación, de calidad y seguridad o de transparencia, imparcialidad y equidad.

¹⁰ Yendo un paso más allá, puede distinguirse, en algunas aplicaciones, entre la “*legal black box*” y la “*technical black box*”. La primera contiene el código fuente, mientras que la segunda se refiere a los supuestos de aprendizaje automático del propio *software* (*machine learning* o *deep learning*), en los que las reglas decisionales se generan de forma automática, no siendo posible, ni para sus programadores, explicar por qué y cómo surgen.

investigado/condenado) cuestionar la corrección de los parámetros tenidos en cuenta, del peso otorgado a cada uno de ellos, etc. Esta cuestión, no por casualidad, fue la discusión central en el caso *State v. Loomis*¹¹, resuelto por el Tribunal Supremo de Wisconsin mediante su sentencia de 13 de julio de 2016¹². A falta de un pronunciamiento del Tribunal Supremo federal de EE. UU., esa resolución se considera en la actualidad como el precedente judicial más autorizado en la materia.

§. En la sentencia *Loomis*, el Tribunal concluyó que no se vulneró el derecho a un proceso debido (el *due process of law*, equivalente aproximado de nuestro derecho a la tutela judicial efectiva) por el hecho de que el juez tomara en cuenta a la hora de fijar la extensión de la sentencia condenatoria, entre otra serie de consideraciones, el riesgo de reincidencia previsible del condenado extraído de uno de esos programas informáticos (COMPAS, acrónimo de *Correctional Offender Management Profiling for Alternative Sanctions*). Y ello a pesar de que ni el órgano judicial ni el condenado tuvieran acceso a la metodología empleada por esa aplicación, al estar protegida por el derecho de propiedad industrial de su dueña. El TS de Wisconsin, no obstante, enfatiza que los jueces deben aproximarse con suma cautela a este tipo de modelos predictivos automatizados.

§. La doctrina *Loomis* aboga por no invalidarlos como un elemento más en la formación de la convicción judicial, pero no permite que la decisión final se adopte en virtud *exclusivamente* del resultado predictivo de este *software*. Y ello porque: (i) la propiedad privada del citado *software* impide desvelar cómo se calcula exactamente el riesgo efectivo de reincidencia; (ii) el sistema COMPAS es incapaz de predecir el riesgo individualizado de reincidencia, ya que lo que proporciona son rangos de probabilidad actuarial aplicables a grupos de individuos con circunstancias personales parecidas; (iii) el programa se basa en una muestra de datos a escala federal y no del Estado de Wisconsin; (iv) varios estudios científicos apuntan a que el programa COMPAS clasifica de forma desproporcionadamente alta (en el rango de riesgo de reincidencia) a las minorías raciales; y (v) COMPAS fue originariamente diseñado para asistir al Departamento de Justicia estatal en la fijación de los tiempos de reclusión más

¹¹ 881 N.W.2d 749 (Wis. 2016).

¹² La opacidad en el uso de los metadatos ha sido igualmente señalada por la Comisión Europea como uno de los grandes problemas que plantean las decisiones automatizadas. *Libro blanco sobre la inteligencia artificial. Un enfoque europeo para la excelencia y la confianza*, COM (2020) 65 final, Bruselas, 2020.

apropiados (con lo que sería razonable sospechar que su diseño podría estar sesgado en contra de la probabilidad de rehabilitación del reo y favor de una postura en exceso punitiva).

§. La sentencia *Loomis* -aun cuando muchos de sus planteamientos sean discutibles- tiene la virtud de poner sobre la mesa una parte de la extraordinaria constelación de problemas prácticos, éticos y jurídicos que plantea el uso de este tipo de herramientas en el ejercicio de la función jurisdiccional. Si algo llama la atención en una primera lectura es que, a pesar de los importantísimos *caveats* que la recorren, el Tribunal acabe concluyendo que su utilización por parte de los jueces de Wisconsin -de forma no vinculante y siempre acompañada de otras consideraciones adicionales- es admisible desde la perspectiva del *due process of law*. Porque lo cierto es que parece muy ingenuo pensar que los jueces serán capaces de descontar mentalmente -y en su justa medida, además- las limitaciones inherentes al programa COMPAS. En primer lugar, porque en ningún pasaje de *Loomis* se indica a los jueces cómo deberían hacerlo. En segundo lugar, porque las limitaciones son de tal entidad -recordemos que el código fuente del *software* y sus parámetros no son públicos, ni siquiera para los jueces- que cualquier prevención se tornaría fútil.

§. Dicho con otras palabras, los jueces no podrían ejercer de efectivos *gatekeepers* y comprobar si el *software* presenta o no sesgos raciales, de género, proacusación, etc., si no tienen acceso a sus datos de programación más elementales¹³.

§. También desde una perspectiva en exceso naíf, el TS de Wisconsin parece sostener que, dado que la última palabra la tendrá siempre un ser humano (el juez) y este estará obligado en todo caso a motivar las razones de su decisión (tomando en consideración de forma obligada otro tipo de factores adicionales) se garantizaría así una suerte de “inmunidad decisoria” del juez frente a una eventual influencia excesiva del resultado predictivo sobre la reincidencia. Sin embargo, una serie robusta de estudios empíricos evidencian que, por el efecto psicológico del denominado sesgo cognitivo de

¹³ En el voto particular concurrente formulado por la magistrada Abrahamson que acompaña a la sentencia del caso *Loomis* se alerta sobre este problema, al señalarse que «la falta de comprensión por parte de este tribunal de [el programa] Compas plantea un problema significativo. En la vista, el tribunal preguntó repetidamente tanto a los abogados de la defensa como a los del Estado cómo funciona Compas. Y obtuvo pocas respuestas».

“anclaje¹⁴”, existe un riesgo muy elevado de que esta información probabilística influya de forma desmedida -e imperceptible, por cierto, para el propio juez- en su decisión final. Esta disfunción cognitiva (*anchoring*) surge tras la realización de una estimación, por parte del sujeto, a partir de un valor inicial que progresivamente ajusta a medida que obtiene información adicional. Múltiples experimentos realizados en los últimos años acreditan cómo las personas sucumben a esta distorsión cognitiva de forma recurrente. Esto es, al magnetismo que, cuando hacemos estimaciones, ejercen sobre nosotros las cifras que nos rondan por la mente. Lo más llamativo es que esa atracción numérica se produce aunque tales cifras no guarden relación conceptual alguna con el valor que se tiene que estimar¹⁵.

§. Pero los riesgos de un mal uso de la información proveniente de estas herramientas no acaban aquí. En particular, parece fuera de discusión que los seres humanos (incluidos los jueces) *tendemos a dotar de mayor “peso mental” a la información de naturaleza empírica que a la que carece de tal condición*, de forma que es fácil colegir que el juez tenderá a otorgar mayor probabilidad de acierto al resultado extraído de este tipo de programas que a la postura meramente *narrativa* del encartado¹⁶.

§. En cualquier caso, y ya desde la perspectiva del ordenamiento jurídico español, no creo que sea compatible con el derecho a la tutela judicial efectiva (art. 24.1 CE) el más mínimo grado de opacidad con respecto a los datos de programación, forma de cálculo, elementos tomados en consideración, etc. Cualquier reforma futura de las normas procesales para la introducción de este tipo de aplicaciones en nuestro sistema de administración de Justicia -en línea con lo previsto en el Anteproyecto de LECrim de 2020- exigiría en todo caso la garantía de que, tanto el tribunal, como las partes, tuvieran pleno acceso al código fuente del *software*, de forma que pudieran cuestionar sus parámetros y testar científicamente su fiabilidad.

b) Los eventuales sesgos en el diseño de los parámetros

§. En estrecha relación con el requisito de la transparencia, es preciso hacer notar que no serían admisibles programas automatizados que discriminaran a los encartados en

¹⁴ MUÑOZ ARANGUREN, “La influencia de los sesgos cognitivos en las decisiones jurisdiccionales: el factor humano. Una aproximación”, *InDret*, Vol. 2, 2011, pp. 5-8.

¹⁵ FARIÑA, ARCE y NOVO, “Heurístico de anclaje en las decisiones judiciales”, *Psicothema*, Vol. 14, núm. 1, 2002, pp. 39-46.

¹⁶ *Vid.*, FENNELL y HALL, “Due Process at Sentencing: An Empirical and Legal Analysis of the Disclosure of Presentence Reports in Federal Courts”, *Harvard Law Review*, 1613, 1980, pp. 1628-30.

función de su género, raza u otra característica grupal. Aunque sea estadísticamente cierto que, por ejemplo, el hombre presenta una mayor tasa de reincidencia delictiva que la mujer¹⁷ (o que, en determinadas comunidades multirraciales, por razones fundamentalmente socioeconómicas, los miembros de determinada etnia puedan ser, en media, más propensos a delinquir de nuevo¹⁸), la toma en consideración de estas variables (tanto a la hora de acordar una medida cautelar personal, fijar la extensión de una condena o imponer una medida de seguridad) supondría penalizar (o beneficiar) al reo como consecuencia de su sexo, raza o estatus socioeconómico. Bien mirado, sería una suerte de aplicación tecnocrática del justamente denostado «derecho penal de autor¹⁹»: se castigaría al condenado -al menos, en parte- *en función de quién es*, no estrictamente *por lo que ha hecho*.

§. De lo que se sigue que el requisito de lo que podríamos denominar *neutralidad algorítmica* sea otra exigencia ineludible. No es posible aceptar que a través de este tipo de programas se otorgue un trato punitivo discriminatorio, contrario al art. 14 CE, en función de la raza o el sexo del acusado. No cabe duda de que la exclusión en los cálculos actuariales de estos factores reducirá la precisión de este tipo de *software* a la hora de predecir eventuales futuros comportamientos reincidentes. Pero se trata de un principio -el de la *personalidad de la pena*- irrenunciable desde el prisma de la dogmática penal: «El principio de personalidad de las penas, que forma parte del de legalidad penal y se encuentra, por tanto, incluido en el art. 25.1 CE, implica que *sólo se puede responder penalmente por los actos propios y no por los ajenos*» [STC 125/2001, de 4 de junio, FJ 6²⁰].

¹⁷ FREEMAN y SANDLER, “Female and Male Sex Offenders: A Comparison of Recidivism Patterns and Risk Factors”, *Journal of Interpersonal Violence*, 23(10), 2008, pp. 1394-1413.

¹⁸ WEHRMAN, “Race, concentrated disadvantage, and recidivism: A test of interaction effects”, *Journal of Criminal Justice*, 38(4), 2020, pp. 538-544. Aunque esa diferencia en la tasa de reincidencia no es completamente homogénea, ya que presenta variaciones en función del tipo delictivo de que se trate.

¹⁹ Para MUÑOZ CONDE, el *derecho penal de autor* se basa en determinadas cualidades de la persona de las que esta, en la mayoría de las ocasiones, no es responsable en absoluto y que, en todo caso, no pueden precisarse o formularse con toda nitidez en los tipos penales. “El nuevo Derecho Penal autoritario”, *Nuevo Foro Penal*, 2003, pp. 13-41. Dicho en otros términos, esta doctrina vincula la pena a la personalidad del autor, “siendo su asocialidad y el grado de la misma lo que decide la sanción. Lo que hace culpable aquí al autor no es ya que haya cometido un hecho, sino que solo el que el autor sea «tal» se convierte en objeto de la censura legal”, ROXIN, *Derecho Penal. Parte General*, tomo I, 2ª ed., Civitas, Madrid, 1997, p. 177.

²⁰ La STC 59/2008, de 14 de mayo, FJ 11 B), en cambio, vincula esta idea al principio de culpabilidad: «Ciertamente es que «la Constitución española consagra sin duda el principio de culpabilidad como principio estructural básico del Derecho penal» [...] como derivación de la dignidad de la persona [...], y que ello comporta que la responsabilidad es personal, por los hechos y subjetiva: que sólo cabe imponer una pena

§. Ahora bien, conviene no olvidar que los algoritmos aplican los datos brutos introducidos por seres humanos, de forma que si esa información incorpora a su vez sesgos -entendidos, en este caso, en sentido estadístico y no psicológico: un error sistemático en el que se puede incurrir cuando al hacer muestreos o ensayos se seleccionan o favorecen unas respuestas frente a otras- el algoritmo reproducirá y exacerbará estas disfunciones²¹.

§. De otro lado, el diseño del algoritmo -y no solo los datos que se introducen en él- es una creación humana, por lo que tampoco cabe descartar la existencia de sesgos sistémicos indebidos en su programación (raciales, de género, de clase social, proacusación o proacusado, etc.) que conduzcan a resultados inasumibles desde la perspectiva del derecho a la igualdad en la aplicación de la ley. Es preciso, consecuentemente, estar en guardia ante una potencial discriminación algorítmica derivada de la utilización de la IA para la perfilación de individuos.

c) La fiabilidad científica

§. Otro aspecto de indudable relevancia consiste en determinar si este tipo de aplicaciones informáticas tienen o no un margen de error técnico aceptable. Aunque la sentencia *Loomis* orilla la cuestión al indicar que las exigencias fijadas por el Tribunal Supremo federal de EE. UU. en *Daubert v. Merrell Dow Pharm., Inc.*²² solo serían aplicables durante la fase probatoria del proceso -pero no en momento de determinación de la pena-, es preciso detenerse en este punto²³.

§. En una primera aproximación, la naturaleza procesal de este tipo de programas informáticos parece revestir la condición de genuina prueba pericial²⁴. Por ello, su admisibilidad exige el cumplimiento de los *estándares mínimos de fiabilidad y científicidad*. No tendría sentido facilitar al juez un dato para que lo tomara en consideración a la hora de adoptar una decisión tan trascendente, si no estamos en condiciones de garantizar un *grado de acierto razonable*. En el presente momento, el

al autor del delito por la comisión del mismo en el uso de su autonomía personal. La pena sólo puede «imponerse al sujeto responsable del ilícito penal».

²¹ En el mismo sentido, *vid.* MARTÍNEZ GARAY, “Peligrosidad, algoritmos y *due process*: el caso *State v Loomis*”, *UNED, Revista de Derecho Penal y Criminología*, 3ª época, nº 20, 2018, p. 486.

²² 509 U.S. 579, 113 S.Ct. 2786, 125 L.Ed.2d 469 (1993).

²³ Para calibrar su influencia, haremos notar que el prelegislador español ha hecho suyas las reglas *Daubert* en el art. 483 del Anteproyecto de Ley de Enjuiciamiento Criminal de 2020.

²⁴ Con arreglo al art. 456 LECrim, “el juez acordará el informe pericial cuando, para conocer o apreciar algún hecho o circunstancia importante en el sumario fuesen necesarios o convenientes conocimientos científicos o artísticos”.

margen de error de este tipo de programas sigue siendo elevado, por lo que parece aconsejable guardar una cautela extrema²⁵.

§. Los criterios que, en su conjunto, conforman el denominado estándar o test *Daubert*, son los siguientes: a) controlabilidad y falsabilidad de la teoría científica sobre la que se funda la prueba; b) determinación del porcentaje o margen de error inherente a la técnica empleada; c) control por otros expertos en la materia a través de la denominada *peer review* (revisión por pares); d) existencia de consenso en la comunidad científica de referencia sobre su fiabilidad; e) finalmente, es exigible que la prueba científica sea directamente relevante respecto a los hechos que son enjuiciados.

§. Se trata de unas reglas que vienen siendo aceptadas de forma casi universal como parámetro para aceptar medios probatorios novedosos, ya que tienen la virtud de ser fácilmente exportables a otras culturas procesales distintas de la norteamericana²⁶. Si proyectamos estos criterios sobre los programas actualmente en el mercado que se utilizan en los Estados Unidos de Norteamérica para la toma semiautomatizada de decisiones judiciales (los denominados instrumentos de valoración del riesgo de tercera generación²⁷), cabe extraer las siguientes conclusiones:

§. 1. Es indudable que este tipo de *software* es científicamente falsable: puede contrastarse empíricamente su fiabilidad. De forma que este requisito se cumpliría sobradamente. Siempre y cuando, lógicamente, se permitiera el acceso al código fuente y demás características del programa utilizado. También se cumpliría otros de los requisitos de la doctrina *Daubert*, ya que se trataría de una prueba relevante con relación

²⁵ MARTÍNEZ GARAY y MONTES SUAY, “El uso de valoraciones de riesgo de violencia en Derecho Penal: algunas cautelas necesarias”, *InDret*, Vol.2, 2018, *passim*.

²⁶ VÁZQUEZ ROJAS, *De la prueba científica a la prueba pericial*, Madrid, Marcial Pons, 2015, pp. 103-127. Los tribunales españoles acuden cada vez con mayor frecuencia a las reglas *Daubert* para aceptar o no medios probatorios novedosos. Así, en el Auto núm. 87/2017, de 19 enero (ECLI: ECLI:ES:APSE:2017:130^a), de la Audiencia Provincial de Sevilla (Sección 4^a) se lee lo que sigue: “... a fecha actual los análisis de credibilidad del testimonio, que resultan imprescindibles en casos de víctimas de corta edad, por razones de dificultad de comunicación con los operadores jurídicos y de posible contaminación o manipulación por adultos, y aun así son discutidos en ese ámbito, carecen absolutamente de aceptación en la comunidad científico cuando tratan de discriminar la veracidad del testimonio adulto, de modo que la diligencia pericial que se propone no supera, por esta y por otras razones, el llamado canon o regla *Daubert*, establecido por el Tribunal Supremo de EE.UU. (*Daubert v. Merrell Dow Pharmaceuticals Inc*, 509 U.S. 579 [1993])”.

²⁷ Junto al ya citado COMPAS, el otro programa de tercera generación -que incorpora datos criminógenos muy precisos- más utilizado en EE. UU. es el LSI-R (*Level of Service Inventory-Revised*), cuyo uso ha sido validado -entre otras resoluciones- por la sentencia de 9 de junio de 2010 del Tribunal Supremo del Estado de Indiana, *Malenchik v. State*, 928 N.E.2d 564 (Ind. 2010).

a lo debatido en el proceso: la probabilidad de reincidencia del sujeto pasivo del proceso.

§. 2. El empleo de métodos actuariales en el cálculo de la probabilidad de reincidencia de una persona también permite la determinación del margen de error asociado a este tipo de programas. Porcentaje de error que, a buen seguro, se irá reduciendo con el paso del tiempo de una manera exponencial, pero que en la actualidad sigue siendo notable²⁸.

§. Una cuestión distinta es que el margen de error actual de estos programas informáticos sea aceptable. Conviene no olvidar que el estándar denominado de la “prueba prevalente” (una probabilidad de acierto mayor del 50%), admisible para resolver contiendas civiles, no podría serlo para entender enervado el principio de presunción de inocencia²⁹. En este último caso, se viene exigiendo por la doctrina un grado de probabilidad mucho más cualificado, de forma que la culpabilidad del acusado quede acreditada, con arreglo al adagio anglosajón, “más allá de cualquier duda razonable³⁰”. Aunque es altamente previsible que el desarrollo de la inteligencia IA y del *big data* [metadatos] perfeccionará en el futuro su tasa de acierto, en la actualidad su utilización presenta un margen de error (en el rango del 50%, aproximadamente) por encima de lo tolerable en el proceso penal. En efecto, si bien algunos estudios empíricos apuntan a que estos programas informáticos tendrían ya un grado de acierto bastante elevado, otros trabajos -de forma mayoritaria- sostienen exactamente lo contrario³¹.

²⁸ Vid., FÉREZ-MANGAS y ANDRÉS-PUEYO, “Eficacia predictiva en la valoración del riesgo del quebrantamiento de permisos penitenciarios”, *La Ley penal*, nº 134, 2018.

²⁹ El Auto de la Sala Especial del art. 61 LOPJ del Tribunal Supremo, «caso Sortu», de 1 de abril de 2011 -si bien como *obiter dictum*- afirma que «doctrinalmente el porcentaje de probabilidad necesario para una condena penal o civil se sitúa, en el primer caso, en torno al 90%, mientras que, para la segunda, con arreglo a diversas corrientes doctrinales, se daría entre el 51% y el 80%²⁹». De lo que se sigue que datos con una fiabilidad inferior al 90% no serían aceptables para fijar una pena o una medida de seguridad. Más dudoso sería el estándar teóricamente aplicable a una decisión sobre medidas cautelares personales, como la prisión provisional. Sobre esta cuestión, vid. VALENZUELA SALDÍAS, “Hacia un estándar de prueba cautelar en materia penal: algunos apuntes para el caso de la prisión preventiva”, *Política criminal*, vol.13 nº 26, 2018, pp. 836-857.

³⁰ Cfr. GASCÓN ABELLÁN, “Sobre la posibilidad de formular estándares de prueba objetivos”, *DOXA, Cuadernos de Filosofía del Derecho*, nº 28, 2005, pp. 127-139. También, FERNÁNDEZ LÓPEZ, “La valoración de las pruebas personales y el estándar de duda razonable”, *Cuadernos electrónicos de Filosofía del Derecho*, nº 15, 2007.

³¹ STARR, ob. cit., p. 853. En palabras de MARTÍNEZ GARAY y MONTES SUAY: “los instrumentos de valoración del riesgo de violencia evidencian valores positivos mucho más bajos, casi nunca superiores al 50%”, ob. cit., p. 25.

§. 3. En cuanto a la revisión por pares de este tipo de *software* -requisito formalmente cumplido en este caso-, lo cierto es que su grado de acierto y precisión se encuentra actualmente en discusión por la comunidad científica. El *peer review* implica la validación de las conclusiones científicas por expertos en esa misma disciplina, no su simple revisión acrítica.

§. 4. Como consecuencia de lo anterior, no puede decirse que exista un consenso mayoritario en la comunidad científica sobre la fiabilidad, en la actualidad, de este tipo de programas informáticos. En rigor, la doctrina *Daubert* no exige la “unanimidad” de la comunidad científica con relación a la fiabilidad de un medio de prueba concreto (algo, por definición, muy difícil de conseguir en cualquier discusión científica), sino que exista un acuerdo mayoritario en el seno de esta disciplina sobre la validez del medio probatorio en cuestión.

§. La inmensa mayoría de los estudios publicados sobre la fiabilidad de este tipo de programas proceden de Estados Unidos de Norteamérica, donde esta cuestión ha suscitado especial atención en la medida en que, como ya hemos visto, su utilización por parte de los jueces es habitual y, en algunos casos, obligada. En la actualidad, es mayoritaria la postura dentro de la comunidad jurídica norteamericana³² que niega que este tipo de programas informáticos reúnan los requisitos exigidos por el Tribunal Supremo federal de EE.UU. para admitirlo como medio de prueba en un procedimiento judicial³³. Si bien es cierto que, como ya hemos apuntado, varias resoluciones dictadas por órganos judiciales distintos del Tribunal Supremo han indicado que no serían de aplicación las reglas *Daubert* al no tratarse de un medio de prueba en sentido estricto -sino de un elemento complementario para la determinación de la extensión de la pena- se trata, a mi juicio, de un argumento débil. Al menos, desde la perspectiva del Derecho Procesal penal español.

³² Y de las comunidades estadística y psicológica. Cf. MASSIE, “Orange is the New Equal Protection Violation: How Evidence-Based Sentencing Harms Male Offenders”, 24 *Wm. & Mary Bill Rts. J.* 521, 2015, pp. 521-550; y REDDING, “Evidence-Based Sentencing: The Science of Sentencing Policy and Practice”, *Chapman Journal of Criminal Justice*, Vol. 1, nº. 1, 2009, pp. 1-19.

³³ Cfr. KLINGELE, ob. cit., p. 577 y la bibliografía allí citada. En cuanto a la comunidad jurídica española, *vid.* el exhaustivo trabajo de MARTÍNEZ GARAY y MONTES SUAY, ob. cit, *passim*.

§. Porque con independencia de que la utilización de este tipo de *software* se emplee, bien para acordar una medida de privación provisional de la libertad o bien para la fijación de la duración de una condena, su admisibilidad como instrumento procesal ha de reunir las mismas garantías de fiabilidad que si se tratara de un medio de prueba introducido en el plenario. Tanto desde un plano axiológico, como epistémico, *debe existir un umbral mínimo de la calidad de los datos* que se introducen en el proceso (es lo que conduce rechazar las aportaciones de la llamada *junk science*³⁴), cualquiera que sea su finalidad.

§. Podría discutirse si ese estándar mínimo debe ser superior o no al 90%, pero en ningún caso son aceptables magnitudes cercanas, en el mejor de los casos, al 50%³⁵. A resultas de lo cual, *en el estado actual de la ciencia*, no parece admisible la toma en consideración, aunque fuera con carácter complementario u orientativo, de este tipo de programas informáticos para la adopción de decisiones judiciales. Es indudable que más pronto que tarde el grado de precisión, acierto y fiabilidad de este tipo de programas mejorará de forma sustancial, de suerte que se genere un cierto consenso en la comunidad científica sobre su validez. Pero en tanto en cuanto no sea así, no puede aceptarse su introducción en el proceso penal. Si bien pudiera pensarse, a primera vista, que se trata de una información adicional inocua de la que dispondría el juez -que podría hacer uso o no con libertad de ella-, es preciso alertar de que el suministro de esos datos, con el aura de objetividad y acierto inherente a este tipo de aplicaciones informáticas, es susceptible de provocar una influencia indebida en la toma de decisiones judiciales³⁶. Como ya se ha apuntado, sería muy difícil para un juez desprenderse mentalmente (aunque quisiera) de esta información a la hora de tomar una decisión³⁷. Pretender que el juez pueda aislarse mentalmente de esta influencia y

³⁴ La llamada *junk science* o, literalmente, “ciencia basura”, comprende todo análisis seudocientífico cuyos resultados no se hayan testado y validado empíricamente. Sería el caso, por ejemplo, de la denominada “prueba del polígrafo”.

³⁵ Sobre el estándar aplicable, cfr. el excelente estudio de ALCÁCER GUIRAO, “Algunas dudas sobre la duda razonable. Prueba de descargo, estándares de prueba e *in dubio pro reo*”, *RECPC*, nº 23-09, 2021, pp. 1-47.

³⁶ En ese mismo sentido, GASCÓN INCHAUSTI, “Desafíos para el proceso penal en la era digital: externalización, sumisión pericial e inteligencia artificial”, en *La justicia digital en España y en la Unión Europea: situación actual y perspectivas de futuro*, VVAA, Atelier, Barcelona, 2019, pp. 202-205.

³⁷ Y ello, aunque el juez haga un esfuerzo mental *ad hoc*. Según numerosos experimentos llevados a cabo por científicos cognitivos, en virtud del denominado “proceso irónico” los esfuerzos por controlar nuestra mente e intentar obviar la influencia indebida de un dato suelen ser contraproducentes, especialmente en situaciones de tensión o carga mental (*mental load*). Como señala WEGNER (“Ironic processes of mental control”, *Psychological Review*, 1994, pp. 34-52), la explicación está en que nuestro intento de controlar

abstraerse de los sesgos cognitivos que podrían distorsionar su juicio es un desiderátum que se ha demostrado empíricamente incierto³⁸.

§. Finalmente, debe subrayarse que para que esa información predictiva se incorporara legítimamente al proceso penal sería necesario que se garantizara su previa *valoración crítica por parte del juez*, estableciéndose los mecanismos que fueren precisos para conjurar el riesgo de una previsible tendencia a un cierto automatismo humano en la asimilación de esos datos. Es fácil prever, de igual forma, que cuanto mayor sea su carga de trabajo, los jueces tenderán en mayor medida a confiar en este tipo de *software*, relajando correlativamente su evaluación personalizada.

3. ¿Es razonable distinguir entre las decisiones cautelares y las de imposición de una pena o de una medida de seguridad?

§. En el supuesto de que las anteriores objeciones fueran salvadas, es pertinente preguntarse si la utilización de este tipo de datos predictivos de la reincidencia sería igualmente admisible para la adopción de toda clase de decisiones en el proceso penal o si es necesario establecer algún tipo de distinción. Dicho de otra forma, en una primera aproximación tentativa es legítimo plantearse si la utilización de la inteligencia artificial sería igualmente admisible, tanto para acordar la prisión provisional del investigado (algo que, no lo olvidemos, ya se admite en España ante delitos de violencia de género), como para determinar la extensión de la pena a imponer al acusado, una vez enervado el principio de presunción de inocencia (o para la determinación del alcance y duración de eventuales medidas de seguridad³⁹). A mi modo de ver, se impone una respuesta matizada. Si bien en abstracto -siempre y cuando se cumpliera lo anteriormente expuesto- sería posible incorporar el resultado predictivo de este *software* al juicio de prognosis que realiza el juez al adoptar una medida cautelar personal, no me parece

la mente desencadena un doble proceso: uno consciente, que intenta evitar deliberadamente, con distracciones, el resultado no querido; y otro inconsciente, que verifica si surge o no el pensamiento prohibido (*ironic monitor*); cuando el primer proceso queda debilitado por un factor externo, aflora el segundo y surge el efecto paradójico.

³⁸ Una prevención similar puede encontrarse en el comentario de sentencia «Criminal Law–Sentencing Guidelines– State v. Loomis. Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing», *Harvard Law Review*, nº 130, 2017, pp.1531-1534.

³⁹ También merecerían una reflexión diferenciada las decisiones sobre la suspensión de la ejecución de la pena (art. 80.1 CP) o sobre el tratamiento penitenciario (art. 67 LOPG).

evidente que esa misma admisibilidad teórica pueda ser proyectada al ámbito de la imposición de la pena o de medidas de seguridad⁴⁰.

§. La razón fundamental es que en este último caso corresponde al legislador aquilatar el alcance de las penas guiado, al menos idealmente, por el principio de proporcionalidad. La circunstancia agravante de reincidencia, actualmente contemplada en el artículo 22. 8ª del Código Penal, ya cumple la función de exasperar la pena en caso de que un delincuente reitere su comportamiento delictivo. Sería contrario a los principios de culpabilidad y personalidad de la pena que a través de este *software* se incentivara el alargamiento del tiempo de privación de libertad. Y ello -nótese- como consecuencia del comportamiento pasado o presente del grupo o grupos al que pertenezca el condenado (por razón de su sexo, de su etnia, de su clase social, etc.). La evidencia empírica demuestra que los jueces, una vez facilitada la predicción de reincidencia, se separan de manera muy infrecuente de sus resultados a la hora de tomar una decisión⁴¹.

§. Corresponde al principio de proporcionalidad acotar el cuánto de la pena, mientras que el principio de culpabilidad preserva que el castigo no suponga una instrumentalización del penado -por decirlo en términos *kantianos*-, lo que se produciría si se le penara por conductas ajenas⁴². Aun cuando esta cuestión presenta una complejidad que excede de los márgenes del presente trabajo, es posible señalar, de manera sumaria, que al no basarse la decisión sobre la prisión provisional en verdaderos *actos de prueba* sino en un mero juicio de probabilidad de reincidencia (entre otros factores, como el riesgo de fuga o de destrucción de pruebas *ex art.* 503 LECrim), ni el principio de culpabilidad -ni el de personalidad de la pena- parecen verse directamente afectados. Conclusión que se refuerza si advertimos que la resolución cautelar no implica -al menos, en teoría- un pronunciamiento anticipado sobre la culpabilidad o

⁴⁰ Quizá las medidas de seguridad, en atención a su particular naturaleza, requerirían una respuesta todavía más matizada. Tradicionalmente se las ha definido como “aquellos medios penales que suponen una restricción o privación de bienes jurídicos por la realización previa de una conducta jurídicamente desaprobada considerada materialmente como delito”, RODRIGUEZ HORJADO, “Las medidas de seguridad” en *Manual de Introducción al Derecho Penal* (coord. LASCURAÍN SÁNCHEZ), Madrid, BOE, 2019, p. 268. Precisamente, a diferencia de las penas, “se fundamentan en la peligrosidad criminal del sujeto al que se impongan, exteriorizada en la comisión de un hecho previsto como delito” (art. 6.1 CP).

⁴¹ CHRISTIN *et al.*, “Courts and predictive algorithms”, accesible en la siguiente dirección: http://www.datacivilrights.org/pubs/20151027/Courts_and_Predictive_Algorithms.pdf [<https://perma.cc/5MBA-QPMQ>].

⁴² LASCURAÍN SÁNCHEZ, “El principio de culpabilidad: seis retos”, *Almacén de Derecho*, <https://almacenederecho.org/el-principio-de-culpabilidad-seis-retos/>

inocencia del investigado. Un pronunciamiento de esa naturaleza tan solo puede extraerse -con el grado de seguridad suficiente- tras la práctica de la prueba en el juicio oral⁴³.

4. Conclusiones

§. El impresionante progreso de la IA nos obligará a plantearnos, como sociedad, si estamos dispuestos a aceptar la introducción en el proceso judicial de este tipo de herramientas predictivas. A mi modo de ver, es preciso huir de planteamientos maximalistas. Tan poco deseable es aceptar la introducción en el proceso penal de este tipo de *software* sin meditar sobre las consecuencias que ello tendría -y sin tomar en cuenta su cuestionada fiabilidad actual-, como negarse de forma rotunda a que, en el futuro, cuando estos programas se hayan perfeccionado, puedan integrarse en el proceso de toma de decisiones judiciales con las debidas garantías.

§. Lo anterior supondría despreciar el conocimiento científico aplicado a la prevención de la reincidencia delictiva. No parece razonable rechazar instrumentos analíticos que permitan unas decisiones con un mayor respaldo empírico y, por consiguiente, de acierto. Porque, conviene no olvidarlo, hasta el día de hoy este tipo de valoraciones se realizan por el juez de forma *esencialmente intuitiva*. No se trata de sustituir en el futuro la decisión humana por la procedente de la IA, sino de abrir la posibilidad de que el juez, cuando el estado de la ciencia lo permita, pueda valerse de datos empíricos útiles, que en ningún caso pueden sustituir su criterio personal decisorio⁴⁴. Pero ello siempre bajo la premisa *de no abdicar de ninguna de las garantías penales que tanto esfuerzo histórico ha costado conquistar*.

⁴³ No debe olvidarse que, en la actualidad, la utilización de programas informáticos de predicción de reincidencia delictiva para la toma de decisiones judiciales relacionadas con la adopción de medidas de protección a la víctima de delitos de violencia de género (que incluyen la prisión preventiva) es una realidad en España. *Vid.*, GASCÓN INCHAUSTI, ob. cit., p. 204.

⁴⁴ Algo, por lo demás, vedado en la actualidad por el art. 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: “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”. Cfr. también el art. 11. 2 de Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

§. En los últimos años de su vida, Max WEBER advirtió agudamente que el progresivo proceso de racionalización de los ordenamientos jurídicos occidentales constituiría una destacada nota distintiva de la modernidad⁴⁵. Es fácil perder la perspectiva histórica y no darse cuenta de que ese proceso de racionalización jurídica no ha concluido. Precisamente, en esa evolución racionalista debe enmarcarse la creciente interacción entre el Derecho y las aportaciones de la *inteligencia artificial*.

⁴⁵ *Economía y Sociedad*, Madrid, Fondo de Cultura Económica, 1993, pp. 603-621.