**JUSTICIA ALGORÍTMICA Y DERECHO PENAL.**

Asier Urruela Mora*[[1]](#footnote-1)\**

1. ***EVIDENCE BASED SENTENCING* COMO PARADIGMA DE INTRODUCCIÓN DE LOS INSTRUMENTOS DE EVALUACIÓN DEL RIESGO DE VIOLENCIA EN LA ESFERA DEL PROCESO PENAL**

En las últimas décadas se perciben nítidamente distintos intentos operacionales provenientes del ámbito de disciplinas científicas como la Psicología o la Criminología en países como EE.UU., Reino Unido o Canadá, entre otros, objeto de recepción en España y consistentes en sustituir el concepto de peligrosidad criminal (entendido como un referente dogmático en la esfera penal) por un supuestamente más objetivo “riesgo de violencia”(variable que, en países como Estados Unidos, entre otros, sirve para incidir en la modalidad y extensión de la consecuencia jurídico-penal aplicable a sujetos imputables)[[2]](#footnote-2).

Resulta fundamental, en este punto, analizar de manera pormenorizada el funcionamiento de los principales instrumentos actuariales de evaluación del riesgo de violencia empleados en la práctica forense (al objeto del presente trabajo nos centraremos en la praxis judicial estadounidense) con el fin de contrastar lo acertado de los postulados favorables a la incidencia de dichos instrumentos en la sentencia penal, tal como sostiene un sector importante de la doctrina en EE.UU. A efectos de contextualizar las actuales tendencias en materia de concreción del riesgo de violencia cabe diferenciar, siguiendo a Hamilton[[3]](#footnote-3), cuatro generaciones de instrumentos de evaluación del riesgo[[4]](#footnote-4). La primera se corresponde con la actuación de clínicos procediendo a llevar a cabo entrevistas no estructuradas o semi-estructuradas con el fin de obtener información relevante que, sobre la base de la experiencia y la competencia del profesional, determinaba los factores de riesgo de reincidencia. Las mejoras introducidas posteriormente en la validez predictiva de los instrumentos provocaron que dichas metodologías fueran superadas por instrumentos de generaciones ulteriores. La segunda generación de evaluaciones consistía en instrumentos de puntuación basados en la evidencia empírica integrados por aquellas variables que mostraban una correlación estadística mayor con la reincidencia (en esta generación se incluyen instrumentos tan conocidos como el *Violence Appraisal Guide*-VRAG, el Static-99 o el *Pre-Trial Risk Assessment tool*-PTRA). La tercera generación de instrumentos combinaba evaluación actuarial con juicio profesional dirigido, integrando asimismo factores estáticos y dinámicos (ejemplos de herramienta de esta naturaleza son el HCR-20 que integra 20 factores de riesgo en ámbitos históricos, clínicos y de gestión del riesgo o el *Level of Service Inventory Revised*; LSI-R[[5]](#footnote-5)). Finalmente, la cuarta generación de sistemas de evaluación procedió a añadir a la combinación de riesgo-necesidades criminógenas, los principios de capacidad de respuesta y una perspectiva más amplia del caso, gestión que abarca desde la admisión hasta el cierre del caso. Uno de los instrumentos de cuarta generación más extendidos en la práctica forense estadounidense es el *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS)[[6]](#footnote-6) empleado en numerosas jurisdicciones penales estadounidenses y que representa el eje de lo que ha constituido el caso de referencia en el análisis de la admisibilidad de las herramientas actuariales de predicción del riesgo en el sistema penal de dicho país (caso *State of Wisconsin v. Loomis[[7]](#footnote-7)*)[[8]](#footnote-8).

Es en el referido marco de evolución, mejora y perfeccionamiento de los instrumentos actuariales de evaluación del riesgo de violencia a los que ya se atribuye un alto grado de fiabilidad en la predicción del comportamiento humano (particularmente, en lo que aquí nos interesa, del delictivo), en el que se inscribe el actual movimiento de incorporación de los mencionados instrumentos de evaluación del riesgo no sólo a esferas de la intervención penal en las que ya se encontraban asentados (en la praxis estadounidense, entre otros, la *parole* o la asignación de los sujetos a los distintos programas de tratamiento) sino igualmente al ámbito de la determinación de la sentencia en relación con sujetos imputables. Se ha llegado a acuñar dicho proceso como Sentencia Basada en la Evidencia (EBS de sus siglas en inglés, *Evidence Based Sentencing*) entendiendo por evidencia la investigación sobre factores que predicen la reincidencia delictiva con base en la cual se proporcionan a los jueces puntuaciones de riesgo para cada acusado en función de variables que, además de los antecedentes penales, suelen incluir sexo, edad, estado civil, y factores socioeconómicos como el empleo y la educación[[9]](#footnote-9).

1. ***CONSIDERACIONES CRÍTICAS EN TORNO AL EMPLEO DE INSTRUMENTOS ACTUARIALES DE EVALUACIÓN DEL RIESGO PARA LA DETERMINACIÓN DE LA SENTENCIA PENAL***
2. **La capacidad de las herramientas actuariales de evaluación del riesgo para predecir el comportamiento criminal futuro a nivel individual**

Una primera cuestión que procede abordar, en relación con el empleo de instrumentos de evaluación del riesgo de violencia en el marco del proceso penal con la finalidad de incidir en la sentencia a imponer, es la relativa a la capacidad predictiva misma de la herramienta. En este sentido, un sector muy relevante y cualificado de la doctrina estadounidense ha puesto seriamente en entredicho la supuesta fiabilidad y exactitud de dicho instrumental al objeto de predecir el comportamiento criminal futuro del sujeto. Resulta fundamental evidenciar que los instrumentos de evaluación del riesgo de violencia adolecen de una capacidad predictiva de los futuros delitos de un concreto individuo relativamente baja por lo que su incorporación a la esfera penal exige importantes cautelas[[10]](#footnote-10).

Como destacan Monahan y Skeem[[11]](#footnote-11) el análisis de los factores de riesgo empleados en dichas herramientas sugiere que los instrumentos abarcan cuatro dimensiones fundamentales: antecedentes penales, estilo de vida irresponsable, psicopatía y actitudes delictivas, y problemas relacionados con el abuso de sustancias, cada una de las cuales sería igualmente predictiva de la reincidencia.

En este sentido, una cuestión central a la hora de analizar la fiabilidad de las herramientas de predicción de la peligrosidad de base actuarial viene constituida por la forma de configuración de dichos instrumentos y cómo lo anterior redunda en el número de falsos positivos y falsos negativos generados. Para ello, debemos partir del hecho de que algunos de los algoritmos empleados por el sistema penal en EE.UU. pertenecen a compañías privadas con ánimo de lucro (tal es el caso del software COMPAS) quedando protegidos por el secreto industrial[[12]](#footnote-12). En consecuencia, el éxito comercial de dichos algoritmos se vincula fundamentalmente a la consecución de un número estadísticamente bajo de falsos negativos (máxime si la predicción de reincidencia se orienta a la delincuencia violenta o sexual) dada la visibilidad social e impacto mediático que dichos nuevos crímenes pueden ostentar con el consiguiente perjuicio económico para el desarrollador de la herramienta de predicción. Por el contrario, la posibilidad de impacto reputacional para la compañía derivado de los falsos positivos es realmente bajo, ya que el encarcelamiento o la prolongación de la estancia en prisión de un individuo cuya peligrosidad criminal no resulta importante difícilmente será detectado. Por ello, no es descartable que se introduzcan por parte de los desarrolladores del algoritmo mecanismos autocorrectores que introduzcan el oportuno sesgo en el sentido indicado[[13]](#footnote-13).

1. **La interpretación de los resultados de los instrumentos actuariales de evaluación del riesgo por parte de los operadores jurídicos**

Una cuestión relevante en el ámbito analizado en este trabajo es la capacidad real del profano en técnicas actuariales para interpretar adecuadamente los resultados suministrados por las herramientas de evaluación del riesgo de reincidencia[[14]](#footnote-14). Diferentes estudios empíricos en EE.UU. han puesto de manifiesto que la forma de comunicación de los resultados obtenidos de las herramientas actuariales a los jueces y miembros de los jurados puede condicionar en gran parte su comprensión de los mismos.

Al dato anterior se añade que en numerosas ocasiones se aportan índices de fiabilidad de la herramienta actuarial en cuestión completamente extraños para el profano, tales como el Área Bajo la Curva (curva ROC, de su acrónimo en inglés *Receiver Operating Characteristic*). Por otro lado, los estudios de validación de los distintos instrumentos de evaluación del riesgo suelen aportar resultados de parámetros como la sensibilidad o la especificidad de la herramienta, siendo que el valor relevante a efectos de la estimación a efectuar en el ámbito del sistema de justicia penal es fundamentalmente el valor predictivo.

Resulta evidente, en este caso, que únicamente personas con un alto grado de formación en estadística podrán discriminar adecuadamente dichos conceptos y que, en relación con los profanos en dicha disciplina científica (como lo son la casi totalidad de los distintos operadores jurídicos: miembros de un jurado, partes e incluso la mayoría de la judicatura) la forma de presentación de los propios resultados y de los estudios de validación de la herramienta de predicción del riesgo pueden distorsionar completamente el sentido atribuido al resultado presentado en el caso concreto.

1. **Herramientas actuariales de evaluación del riesgo de violencia y sesgo racial**

Una de las cuestiones que han ocupado de manera más intensa a la doctrina estadounidense reside en la posibilidad de que dichos instrumentos estadísticos de predicción de la reincidencia acaben produciendo un importante sesgo racial, sobrevalorando la peligrosidad de los miembros de las minorías étnicas[[15]](#footnote-15).

En este sentido, si bien en las actuales herramientas actuariales de evaluación del riesgo de violencia la raza no se incluye directamente como ítem o factor predictivo, resulta probable que existan otras variables en dichas herramientas que funcionen como sustitutivos indirectos de aquella, tales como el área de residencia del sujeto[[16]](#footnote-16), el nivel educativo o de empleo o el número de arrestos previos (a la vista del distinto grado de vigilancia policial y de la tasa diferencial de arrestos producidos en relación con las minorías étnicas)[[17]](#footnote-17). Es decir, sin necesidad de incluir directamente la raza como factor de riesgo otros ítems en la herramienta acaban correlacionando con la raza de una manera estadísticamente significativa.

1. **Inferencia Grupo a Individuo (G2i)**

Uno de los grandes problemas que plantea el empleo de instrumentos actuariales de evaluación del riesgo de violencia en el ámbito de la sentencia viene determinado por lo que se conoce como Inferencia Grupo a Individuo (G2i de su acrónimo inglés *Group To Individual Inference*). Procede poner de manifiesto que el empleo de los distintos instrumentos estadísticos de evaluación del riesgo se funda en un proceso de inferencia, es decir, partiendo de datos extraídos a nivel grupal (diseño a nivel agregado) se obtienen conclusiones que se aplican al concreto individuo sobre el que se lleva a cabo el juicio de peligrosidad criminal sobre la base de la presencia en este último de una serie de factores de riesgo o ítems que en el plano grupal han demostrado correlacionar con altas tasas delictivas. Con ello se produce un salto argumental fundamental, pues pasamos de la causalidad como eje de la aproximación metodológica (concepción clásica en la esfera penal) a tratar de determinar una correlación, es decir, el grado de relación entre dos variables[[18]](#footnote-18).

1. **Sesgos cognitivos en la toma de decisiones en materia de reincidencia.**

Resulta procedente poner de manifiesto la presencia de diferentes sesgos cognitivos que pueden afectar a un operador jurídico (juez, responsables del tratamiento penitenciario, etc.) a la hora de adoptar decisiones sobre un individuo concreto con base en los resultados suministrados por un instrumento actuarial de evaluación del riesgo, particularmente cuando se incorpora la inteligencia artificial y el *machine learning* a dichos procesos (destacadamente, tal sería el caso del algoritmo COMPAS empleado en numerosas jurisdicciones penales en EE.UU.). De todos ellos, el que mayor impacto tiene en la decisión final es, probablemente el sesgo de automatización. Dicho sesgo consiste en la tendencia a atribuir mayor poder y autoridad a las ayudas automatizadas que a otras fuentes de asesoramiento. Como pone de manifiesto Wisser[[19]](#footnote-19) resulta probable que los jueces asuman que los métodos cuantitativos son superiores al razonamiento verbal ordinario[[20]](#footnote-20), lo que puede fácilmente mutar una sugerencia tecnológica en una decisión final dotada de autoridad.

No obstante, el sesgo de automatización, siendo el más relevante, no constituye el único sesgo susceptible de concurrir en la toma de decisiones por parte de operadores jurídicos que actúan con apoyo de algoritmos de predicción del riesgo de violencia. Otros sesgos cognitivos sirven para afianzar aún más el papel de dichos instrumentos estadísticos en la determinación del contenido de la sentencia. En este sentido, diferentes autores ponen de manifiesto el peso que en dicho ámbito puede ostentar el sesgo de confirmación -entendido como la tendencia a buscar información que valide una noción preconcebida mientras se rechaza otra información de signo contrario- que puede llevar a los jueces a ignorar otros factores en un Informe de Investigación Previo a la Sentencia que contrarrestan una puntuación de riesgo elevada. Asimismo, se menciona la posible incidencia del denominado sesgo de punto ciego o tendencia de las personas a verse a sí mismas como menos susceptibles al sesgo que los demás, lo que puede hacer que los jueces subestimen el grado en que sus decisiones de sentencia están siendo realmente influidas por la propia herramienta actuarial. Como resultado, no solo es probable que se violen los derechos al debido proceso de los acusados al dictar sentencia cuando dichos instrumentos estadísticos de predicción de la reincidencia se encuentran involucrados, sino que también es probable que los jueces que dictan sentencia actúen, sin darse cuenta, como facilitadores de prejuicios raciales como resultado de esta dependencia excesiva[[21]](#footnote-21).

1. ***HACIA UN MARCO NORMATIVO EUROPEO EN MATERIA DE INTELIGENCIA ARTIFICIAL (IA). IMPLICACIONES PARA EL DERECHO PENAL***

En los últimos años, las instituciones europeas han adquirido plena conciencia del impacto de las aplicaciones basadas en la IA en los derechos y libertades fundamentales de los ciudadanos. Dicha incidencia resulta particularmente importante cuando los citados algoritmos ostentan una proyección en la esfera del derecho penal y procesal penal, por lo que, en aras a hacer frente a los peligros que pueden derivarse de la proliferación de las mencionadas tecnologías en los referidos ámbitos, comienza a diseñarse un futuro marco normativo europeo que haga frente a los riesgos antedichos. Por otro lado, y con el fin de contextualizar las reflexiones contenidas en el presente apartado en el objeto de nuestro trabajo, procede poner de manifiesto que gran parte de los instrumentos de evaluación del riesgo de violencia-reincidencia existentes en la esfera penal en la actualidad utilizan tecnología fundada en la IA, lo que obliga a una toma en consideración específica de la problemática que ello genera.

En este sentido, particular relevancia ostentan una serie de textos sobre el particular surgidos de las instituciones de la UE. A nuestros efectos, especial consideración merecen la Resolución del Parlamento Europeo sobre la IA en el Derecho Penal y su utilización por las autoridades policiales y judiciales en asuntos penales (6 de octubre de 2021)[[22]](#footnote-22) y, sobre todo, dada su vocación de convertirse en el marco normativo de referencia en toda la UE en materia de IA, la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión cuyas primeras versiones fueron conocidas en abril de 2021.

Comenzando por la Resolución del Parlamento Europeo sobre la IA en el Derecho Penal, la misma contiene reflexiones de gran calado orientadas a encauzar el uso de la IA en la esfera que aborda. En primer lugar, dicha Resolución considera fundamental evitar que estas tecnologías produzcan efectos discriminatorios o generen sesgos o perjuicios ya constatados a lo largo de nuestro trabajo. Partiendo del dato de que dichos sesgos pueden resultar inherentes a los conjuntos de datos subyacentes se evidencia que los sesgos tienden a aumentar de forma gradual por lo que el potencial discriminatorio se incrementa de manera exponencial (en particular para personas pertenecientes a colectivos tradicionalmente racializados).

La referida Resolución del Parlamento Europeo aborda igualmente cuestiones de naturaleza procesal penal, que han emergido al hilo de la utilización de algoritmos fundados en la IA en la esfera norteamericana fundamentalmente. En este sentido, la Resolución evidencia el impacto cierto de las herramientas fundadas en la IA en el ejercicio del derecho de defensa como consecuencia básicamente de la dificultad para obtener información significativa sobre su funcionamiento y la consiguiente dificultad para impugnar sus resultados ante los tribunales (en particular, por parte de los investigados).

En segundo lugar, en la esfera de la UE fundamental resulta la Propuesta de Reglamento (Ley de Inteligencia Artificial[[23]](#footnote-23)) puesto que constituye el embrión de la norma llamada a sentar el marco normativo común en toda la Unión en relación con la IA. A pesar de que la Ley de Inteligencia Artificial ostenta una vocación amplia, pues se orienta a la regulación de los sistemas de IA con independencia de su ámbito de incidencia, algunas de sus disposiciones merecen una toma en consideración particular por su impacto cierto en la esfera penal.

A nivel aplicativo, le Ley de Inteligencia Artificial se estructura a partir de una clasificación de los sistemas de IA en función del riesgo que implican, distinguiéndose entre aquellos que a) implican un riesgo inaceptable y, en consecuencia, constituyen prácticas de IA prohibidas (art. 5 AIA), b) aquellos que conllevan un riesgo alto, que integran la base de la regulación contenida en la Ley de Inteligencia Artificial, y c) los de riesgo bajo o mínimo.

En relación con el particular el considerando 40 AIA estima que deben conceptuarse como de alto riesgo ciertos sistemas de IA destinados a la administración de justicia y los procesos democráticos, dado que pueden tener efectos potencialmente importantes para la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial. En particular, a fin de evitar el riesgo de posibles sesgos, errores y opacidades, se estima que deben considerarse de alto riesgo aquellos sistemas de IA cuyo objetivo es ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho y a aplicar la ley a unos hechos concretos.

En coherencia con lo anterior, el art. 6 AIA establece las reglas de clasificación para los sistemas de IA de alto riesgo y, entre otros, asigna a dicho grupo en su apartado 2º los que figuran en el anexo III de la Ley de Inteligencia Artificial. El apartado 6 del Anexo III se refiere a los asuntos relacionados con la aplicación de la ley, estableciendo un repertorio de sistemas de IA susceptibles de aplicación en la esfera judicial-penal.

La calificación de un sistema como de alto riesgo implica importantes consecuencias regulatorias, pues la mayor parte de las disposiciones de la propia Ley de Inteligencia Artificial se orientan hacia la determinación de las características que deben cumplir los referidos sistemas. En este sentido, como ponen de manifiesto LAZCOZ/URRUELA/DE MIGUEL[[24]](#footnote-24) para estos se fijan una serie de mecanismos de cumplimiento y aplicación *ex ante* y *ex post* que, por lo general, se traducen en obligaciones jurídicas para los proveedores de los sistemas en la fase de diseño y desarrollo. Como evidencian los citados autores, este sistema normativo se ha diseñado para asegurar que los sistemas de alto riesgo cumplan con una serie de requisitos definidos como obligatorios antes de su puesta en el mercado; a saber, gobernanza de datos, documentación y registro, la transparencia y la comunicación de información a los usuarios, la supervisión humana, o la solidez, la precisión y la ciberseguridad.

No obstante, la clasificación de los sistemas de IA con incidencia en la esfera penal como de alto riesgo no goza de unanimidad. En este sentido, el Supervisor Europeo de Protección de Datos y el Comité Europeo de Protección de Datos en su Dictamen conjunto 5/2021[[25]](#footnote-25) vienen a establecer que las formas intrusivas de IA, en especial las que pueden afectar a la dignidad humana, deberían considerarse sistemas de IA prohibidos con arreglo al art. 5 de la Ley de inteligencia artificial en lugar de clasificarse simplemente como de alto riesgo en el anexo III, como las del art. 6. Ahondando en dicho argumento, se considera que la determinación o clasificación por un ordenador de la conducta futura con independencia de la voluntad propia también afecta a la dignidad humana. Por otro lado, se argumenta que los sistemas de IA destinados a ser utilizados por las autoridades encargadas de la aplicación de la ley para llevar a cabo evaluaciones de riesgos individuales de personas físicas con el objetivo de determinar el riesgo de que cometan infracciones penales (anexo III, apartado 6, letra a), o para predecir la frecuencia o reiteración de una infracción penal real o potencial con base en la elaboración de perfiles de personas físicas o en la evaluación de rasgos y características de la personalidad o conductas delictivas pasadas (anexo III, apartado 6, letra e) utilizados según su finalidad prevista, conducirán a la dominación fundamental de la toma de decisiones policiales y judiciales, con la consiguiente cosificación de la persona afectada. A la luz de lo anterior se pone de manifiesto que dichos sistemas, que afectan a la esencia del derecho a la dignidad humana, deben prohibirse en virtud del artículo 5.

En definitiva, la incidencia de la IA en la esfera penal constituye un campo particularmente necesitado de una importante reflexión multidisciplinar (ya abordada -aunque no culminada-, tal como se ha puesto de manifiesto, a nivel de UE) con la finalidad de evitar que estos sistemas inteligentes acaben suplantando la decisión humana en un ámbito en el que se encuentran en juego los derechos y libertades fundamentales de la ciudadanía.

1. \* Dr. *iur.* Dr. *med.* Catedrático de Derecho Penal (Universidad de Zaragoza). Instituto Universitario de Investigación en Empleo, Sociedad Digital y Sostenibilidad (IEDIS) [↑](#footnote-ref-1)
2. Andrés-Pueyo, A., «Peligrosidad criminal: análisis crítico de un concepto polisémico», en *Neurociencias y Derecho Penal*, Madrid, Edisofer, 2013, pp. 483 y s. [↑](#footnote-ref-2)
3. Hamilton, M., «Risk-Needs Assessment: Constitutional and Ethical Challenges», *American Law Review*, vol. 52, núm. 2, 2015, pp. 236-239. En el mismo sentido, Brennan, T./ Dieterich, W./ Ehret, B., «Evaluating the predictive validity of the COMPAS Risk and Needs Assessment System», *Criminal Justice and Behavior*, vol. 36, núm. 1, 2009, pp. 1-2. [↑](#footnote-ref-3)
4. En la doctrina médico-legal española, uno de sus más ilustres representantes, Esbec Rodríguez, hace referencia a cinco generaciones en el diagnóstico de la peligrosidad/riesgo de violencia. (Esbec Rodríguez, E., «Valoración de la peligrosidad criminal (Riesgo-Violencia) en Psicología Forense. Aproximación conceptual e histórica», *Psicología Clínica Legal y Forense*, vol. 3, núm. 2, 2003, pp. 52 y ss.). [↑](#footnote-ref-4)
5. Para un análisis de las principales herramientas de evaluación del riesgo de violencia, véase Loinaz, I., *Manual de evaluación del riesgo de violencia*, Pirámide, Madrid, 2017, pp. 145 y ss.; Armaza Armaza, E.J., *El tratamiento del delincuente imputable peligroso*, Comares, Granada, 2013, pp. 98 y ss. [↑](#footnote-ref-5)
6. COMPAS es un algoritmo propiedad de la empresa *Equivant* (previamente *Northpointe, Inc.)* dirigido a predecir el riesgo de reincidencia individual sobre la base de un análisis que implica el empleo de información extraída de un cuestionario compuesto por 137 ítems divididos en diferentes secciones, así como de información derivada de los registros penales del sujeto. La evaluación de riesgo producida con base en COMPAS se funda en información obtenida del historial criminal, así como en una entrevista con el acusado (ROMEO, C.M., «Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad», *Revista Penal,* núm. 42, 2018, p. 174). [↑](#footnote-ref-6)
7. En dicho caso, la Corte Suprema de Wisconsin hubo de pronunciarse acerca de un supuesto en el que tribunal de circuito determinó en parte la sentencia con base en los resultados del algoritmo COMPAS que habían sido incluidos en el Informe de Investigación Previo a la Sentencia (PIR-*Presentence Investigation Report*). La Sentencia de la Corte Suprema de Wisconsin de 13 de julio de 2016 rechazó los argumentos de la defensa fundados en la supuesta vulneración del derecho al debido proceso en lo que constituye una resolución fundamental acerca de la constitucionalidad del uso de algoritmos en el marco de una sentencia penal (De Miguel Beriain, I., «Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling», *Law, Probability and Risk*, núm. 17, 2018, pp. 45-46). Véase *State v. Loomis*, 881, N.W.2d 749 (Wis, 2016); *Harvard Law Review*, 2017, pp. 1530-1537. [↑](#footnote-ref-7)
8. En definitiva, en la evolución de los instrumentos de evaluación del riesgo se ha llegado a la configuración de sistemas que, partiendo de una base actuarial, utilizan datos convertidos en algoritmos en los que la evaluación de aquellos se automatiza y objetiviza mediante su procesamiento por medio de procedimientos informáticos de sistemas de inteligencia artificial. (Romeo, C.M., 2018, p. 172). [↑](#footnote-ref-8)
9. Starr, S.B., «Evidence-based Sentencing and the Scientific Rationalization of Discrimination», *Stanford Law Review*, vol. 66, núm. 4, 2014, p. 805. [↑](#footnote-ref-9)
10. Martínez Garay, L., «La incertidumbre de los pronósticos de peligrosidad: consecuencias para la dogmática de las medidas de seguridad», *InDret*, 2/2014, p. 52. [↑](#footnote-ref-10)
11. Monahan, J./Skeem, J.L., «Risk Assessment in Criminal Sentencing», *Annual Review of Clinical Psychology*, vol. 12, 2016, pp. 500 y 501. [↑](#footnote-ref-11)
12. Ello genera indudables problemas desde el punto de vista procesal si los resultados de los referidos algoritmos protegidos por el secreto industrial son empleados para determinar (al menos parcialmente) la pena, cuestión que ya se planteó oportunamente en el caso *State of Wisconsin v. Loomis* al alegar la defensa que la utilización del referido instrumental en dichas condiciones viola el derecho del acusado a ser sentenciado con base en información precisa, en parte porque el secreto industrial aplicable a COMPAS le impide a aquel evaluar la exactitud de la puntuación de riesgo asignado (opacidad que convierte el algoritmo en una verdadera “caja negra” al impedirse al acusado acceder al código fuente, tal como evidencia Freeman, K., «Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis», *North Carolina Journal of Law&Tecnology*, vol. 18, 2016, p. 88). [↑](#footnote-ref-12)
13. De Miguel Beriain, I., 2018, p. 49. [↑](#footnote-ref-13)
14. Dicha cuestión ya se planteó en el caso *State v. Loomis* a raíz del voto concurrente de la juez Abrahamson quien expresó su acuerdo con el sentido de la resolución, pero se mostró preocupada por el hecho de que el tribunal tuviera dificultades para entender evaluaciones de riesgo algorítmicas. [↑](#footnote-ref-14)
15. En relación específicamente con el algoritmo COMPAS (uno de los más extendidos en el sistema penal de EE.UU.), ProPublica, un consorcio independiente dedicado al periodismo de investigación, llevó a cabo en 2016 un análisis del mismo que puso de manifiesto fuertes inconsistencias en los resultados obtenidos mediante dicho instrumental predictivo que afectaban no solo a su fiabilidad sino a la posibilidad de que produjera un fuerte sesgo racial. (Larson, J. et al., «How we analyzed the COMPAS Recidivism Algorithm», *ProPublica*, 2016). [↑](#footnote-ref-15)
16. Véase Starr, S.B., 2014, p. 838. [↑](#footnote-ref-16)
17. Tonry, M., «Legal and Ethical Issues in the Prediction of Recidism», *Fed. Sentencing R.,* vol. 26, núm. 3, 2014, p. 172; Lee Park, A., «Injustice ex machina: Predictive Algorithms in Criminal Sentencing», *UCLA Law Review*, 2019. [↑](#footnote-ref-17)
18. Gil González, E., *Big data, privacidad y protección de datos*, Madrid, Agencia española de protección de datos, 2016, pp. 28 y 33. [↑](#footnote-ref-18)
19. Wisser, L., «Pandora’s algorithmic black box: the challenges of using algorithmic risk assessments in sentencing», *American Criminal Law Review*, vol. 56, núm. 4, 2019, p. 1824. [↑](#footnote-ref-19)
20. Pasquale, F., «Secret Algorithms Threaten the Rule of Law», *MIT Tech. Review*, 2017. [↑](#footnote-ref-20)
21. Lee Park, A., 2019. [↑](#footnote-ref-21)
22. Previamente cabe citar, por su carácter fundamental en materia de IA -si bien no específicamente orientada a la esfera penal- la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2021(INL)). [↑](#footnote-ref-22)
23. *Artificial Intelligence Act* (en adelante, AIA). [↑](#footnote-ref-23)
24. Véase G. Lazcoz/A. Urruela/I. De Miguel, “La Propuesta de Reglamento “Ley de Inteligencia Artificial”: análisis de su posible impacto en la utilización de sistemas automatizados en el ámbito del Derecho Penal”, *en prensa*. [↑](#footnote-ref-24)
25. *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act),* 18 June 2021. [↑](#footnote-ref-25)