

EL USO DE ALGORITMOS PREDICTIVOS EN EL DERECHO PENAL. A
PROPÓSITO DE LA SENTENCIA DE LA CORTE DE JUSTICIA DEL DISTRITO
DE LA HAYA (PAÍSES BAJOS) SOBRE SyRI, DE 5 DE FEBRERO DE 2020*

*THE USE OF ALGORITHMS IN CRIMINAL LAW. ON THE JUDGMENT OF THE DISTRICT
COURT OF JUSTICE OF THE HAGUE (THE NETHERLANDS) ON SyRI, OF FEBRUARY 5, 2020*

María Sánchez Vilanova

*Profesora Ayudante Doctora de Derecho penal
Universitat de València*

RESUMEN

El presente estudio analiza los riesgos del uso de algoritmos predictivos en el Derecho penal, a propósito de la reciente resolución de la Corte de Justicia del Distrito de la Haya (Países Bajos), de 5 de febrero de 2020, que se pronunció sobre el uso de estos algoritmos por parte del Gobierno holandés (concretamente, de la herramienta SyRI) para rastrear posibles fraudes al Estado cometidos por la ciudadanía, estimando que el funcionamiento oculto de este instrumento vulneraba el derecho a la privacidad consagrado por el TEDH. Esta decisión contrasta con la respuesta que los tribunales estadounidenses han ofrecido respecto del uso de herramientas actuariales para la determinación del riesgo de reincidencia en los procesos penales, que se han mostrado proclives a su introducción al entender que no se vulneraría con ellas el derecho al debido proceso. No obstante, como se concluirá, se entiende que, en España, el uso en el proceso penal de herramientas que se sustenten en algoritmos secretos es contrario al derecho a la tutela judicial efectiva.

PALABRAS CLAVE

Algoritmos, privacidad, discriminación, factores de riesgo, sentencia.

ABSTRACT

This paper reflects the risks of using predictive algorithms in criminal law, based on the recent resolution of the District Court of Justice of The Hague (Netherlands), of February 5, 2020, which stated on its use by the Dutch government (specifically, SyRI tool) to trace possible frauds to the State committed by its citizens, considering that the hidden operation of this instrument violates the right to privacy enshrined by the ECHR. This decision contrasts with the response that the US courts have offered regarding the use of these actuarial tools to determine recidivism risk in criminal proceedings, showing a proclivity for their introduction on the understanding that the right to due process would not be violated with them. However, as will be concluded, it is understood that, in Spain, and with the support of the aforementioned resolution, the use in the criminal process of tools that are based on secret algorithms is contrary to the right to effective judicial protection.

KEYWORDS

Algorithms, privacy, discrimination, risk factors, sentence.

DOI: <https://doi.org/10.36151/TD.2022.059>

* Este artículo ha sido elaborado en el marco de los Proyectos de investigación «Derecho penal y comportamiento humano» (MICINN-RTI2018-097838-B-100) y «Ciberacoso sexual a menores: perfiles lingüísticos para el desarrollo de herramientas digitales forenses para prevención, detección y priorización en España» (PID2020-117964RB-I00), concedidos por el Ministerio de Ciencia, Innovación y Universidades de España, y «Sistema penal y exclusión social: un obstáculo para la reinserción de personas privadas de libertad» (GV/2021/047), concedido por la Conselleria de Innovación, Universidades, Ciencia y Sociedad Digital de la Generalitat Valenciana.

EL USO DE ALGORITMOS PREDICTIVOS EN EL DERECHO PENAL. A PROPÓSITO DE LA SENTENCIA DE LA CORTE DE JUSTICIA DEL DISTRITO DE LA HAYA (PAÍSES BAJOS) SOBRE SyRI, DE 5 DE FEBRERO DE 2020

María Sánchez Vilanova

Profesora Ayudante Doctora de Derecho penal
Universitat de València

Sumario: 1. Planteamiento: determinando el futuro. 2. Alcance del uso de algoritmos predictivos: análisis de la sentencia de la Corte de Justicia del Distrito de la Haya de 5 de febrero de 2020. 2.1. Supuesto de hecho. 2.2. Resolución de la Corte. 2.2.1. Caracterización de SyRI: *deep learning*, *datamining* y *big data*. 2.2.2. Controversias sobre el modelo de riesgo: algoritmos secretos. 2.2.3. ¿Criminalización de la pobreza? 2.2.4. Valoración: caso *S. & Marper vs. United Kingdom*. 2.2.4.1. Necesidad general en una sociedad democrática. 2.2.4.1. Proporcionalidad y subsidiariedad. 3. Algoritmos predictivos y Derecho penal. 3.1. Apuntes previos: los instrumentos de valoración del riesgo en el proceso penal. 3.2. Constitucionalidad del uso de instrumentos de valoración del riesgo en el proceso penal estadounidense: caso *State vs. Loomis*. 4. Reflexiones críticas: Algoritmos secretos y discriminación en el proceso penal. 5. Valoración final. Notas. Bibliografía.

1. PLANTEAMIENTO: DETERMINANDO EL FUTURO

En los últimos años, el escenario de ciencia ficción sobre la predicción de la delincuencia que muchos auguraban es ya una realidad. Casi 20 años después del estreno del largometraje *Minority Report*¹, dirigido por Steven Spielberg, se va abriendo el camino a un mundo en el que la tasa de criminalidad se reduzca al máximo mediante la utilización de sistemas que predicen la ocurrencia de los delitos poco antes de su comisión. No hay que desconocer, en este sentido, y especialmente en Estados Unidos (en adelante, EE. UU.), que algunos departamentos policiales —por ejemplo, el de la ciudad de Chicago— han implantado este tipo de técnicas, utilizando un algoritmo que, con apoyo en la tecnología del *big data*, facilita la gestión de grandes volúmenes de datos², permite confeccionar una lista de «posibles delincuentes» y actuar en la génesis del problema criminal³.

En realidad, son muchos los usos de estos algoritmos⁴ en nuestra vida diaria: desde los resultados de búsqueda o los anuncios que nos aparecen en línea hasta la determinación de las posibilidades de recibir un préstamo hipotecario o el precio de un seguro de vida. Excede de los propósitos del presente trabajo realizar un análisis de las vicisitudes y características de esta sociedad «algorítmica» ni abordar sus múltiples derivaciones prácticas, que abarcan desde sus usos más comerciales hasta el Proyecto del Genoma Humano; el objetivo de este estudio es más modesto, a saber, reflexionar sobre la relevancia (en cualquier caso, indirecta) de la decisión de un tribunal neerlandés que estimó, en sentencia de 5 de febrero de 2020⁵, que el sistema de análisis utilizado por el gobierno holandés para rastrear posibles fraudes al Estado (conocido por sus siglas SyRI, acrónimo de *Systeem Risicoindicatie/System Risk Indication*), no respetaba la privacidad de la ciudadanía y vulneraba el artículo 8 del Convenio Europeo de Derechos Humanos (en adelante, CEDH).

Esta sentencia, que se configura como la primera resolución europea limitadora del uso por un gobierno estatal de algoritmos predictivos, es especialmente importante en relación con una de las aplicaciones más controvertidas de los algoritmos: las herramientas actuariales para la determinación del riesgo en el Derecho penal. Si bien somos conscientes de la diferencia de ámbito y ordenamiento, estimamos que esta resolución sobre el uso de las administraciones públicas de algoritmos para las inspecciones tributarias incide en un aspecto esencial ante cualquiera de los potenciales usos de estos: sus posibles efectos discriminatorios, específicamente dentro de un modelo actuarial de penalidad que en España ha sido adoptado de forma relativamente tardía⁶.

Lo cierto es que, en EE. UU., tras la generalización a finales de la década de los años 80 del riesgo como la tecnología estrella en la *New Penology*⁷, desde la reciente promulgación de la *First Step Act*⁸ (legislación que promueve la reforma y modernización del proceso penal), los algoritmos son usados cada vez para adoptar un mayor número de decisiones por las jurisdicciones de todo el país, incluido el Gobierno federal⁹; desde decretar la libertad condicional (*parole*)¹⁰ o la suspensión de la pena (*probation*), hasta la imposición de la pena de muerte en los delitos capitales¹¹ o —cuestión particularmente interesante en el presente estudio— contribuir a la determinación del riesgo de reincidencia de los delincuentes. Sin

llegar a tales extremos, en España, siguiendo la tendencia del sistema jurídico-penal anglosajón, su uso es cada vez más frecuente.

Ahora bien, antes de empezar con el estudio de la sentencia objeto del presente trabajo, conviene responder a una pregunta previa: ¿en qué consisten los nuevos métodos de estimación del riesgo? En puridad, estos métodos se sirven de diferentes herramientas, pues desde su primera configuración estadística ha ido evolucionando a métodos actuariales más puros¹². A grandes rasgos, las herramientas de valoración del riesgo son listas que incorporan factores de riesgo, los cuales, cuando se administran a una persona en concreto, ofrecen como resultado un porcentaje estimativo del riesgo de esta persona en un intervalo temporal determinado conforme a la aplicación de modelos estadísticos y matemáticos¹³. En concreto, los métodos actuariales obtienen esta información de forma automática, mediante un algoritmo que combina de determinada manera los valores numéricos que estas herramientas asignan a cada factor de riesgo. En cualquier caso, estos últimos desarrollos, como precisa Romeo Casabona¹⁴, más que un método distinto, tan solo conforman una nueva herramienta tecnológica que trata «de forma inteligente los algoritmos que sirven de base al procedimiento actuarial».

Por consiguiente, ante la previsible tecnologización progresiva del proceso penal, estas nuevas formas actuariales en la medición del riesgo, que en fases más desarrolladas recurren al uso de la inteligencia artificial¹⁵, podrían llegar a comprometer importantes derechos fundamentales, requiriendo ello la adopción de las garantías penales necesarias al respecto. De hecho, como advierte Miró Llinars¹⁶, este moderno enfoque actuarial aportaría «un sustento metodológico inmejorable» para las vertientes partidarias del endurecimiento del Derecho penal proclives a la subordinación de los principios limitadores del sistema punitivo al control de la delincuencia, y para el consiguiente expansionismo penal¹⁷.

Así pues, se parte del análisis de la sentencia holandesa sobre SyRI de febrero del año 2020¹⁸ para efectuar una aproximación a las limitaciones, específicamente en atención a sus posibles efectos discriminatorios, de la aplicación de herramientas actuariales en el terreno penal, si bien en términos generales y sin profundizar específicamente en una concreta aplicabilidad, pues se reitera que el interés respecto de este pronunciamiento se circunscribe en el presente trabajo a su contraposición con la respuesta que, si bien diferente ámbito, utilidad y territorio, los tribunales estadounidenses han ofrecido en aquellos casos en los que estos instrumentos han sido introducidos propiamente en el terreno judicial para la valoración del riesgo de reincidencia, como, por ejemplo, en el famoso caso *State vs. Loomis*, en el que se permitió el uso de algoritmos secretos en el proceso penal; una respuesta cuestionable a la luz del citado tratamiento de la Corte de Justicia del Distrito de la Haya y, asimismo, de un gran número de especialistas en la materia, los cuales se muestran críticos con estos nuevos métodos, especialmente respecto de su uso en exclusiva, tanto con base en la escasa utilidad predictiva de estos pronósticos¹⁹ como en las preocupaciones éticas que el uso de estos algoritmos suscita²⁰.

En síntesis, como se concluirá, si bien las valoraciones del riesgo de reincidencia son necesarias en muchas etapas del proceso penal²¹, y, especialmente en EE. UU., cada vez

más autores abogan por el uso de herramientas actuariales, pues, según diferentes metaanálisis²², demostrarían una mayor capacidad predictiva que los juicios realizados por especialistas²³, los resultados de estos estudios no son concluyentes, sin olvidar que el marco ético de estos instrumentos no se ha abordado todavía. Por lo tanto, aunque se refuta el rechazo sistemático de estas herramientas, tampoco se apuesta por su adopción acrítica, especialmente porque el uso de estos algoritmos no es neutral. Por el contrario, se defiende un estudio previo y en profundidad de estos nuevos instrumentos del que se deriven las garantías necesarias para facilitar su uso por parte de los tribunales de justicia. Al final, el exponencial desarrollo de las herramientas computacionales²⁴ comporta no meramente cambios cuantitativos, sino también cambios estructurales cuyas implicaciones requieren un riguroso análisis²⁵, desarrollando la regulación adecuada para evitar extralimitaciones en su uso²⁶.

2. ALCANCE DEL USO SEW ALGORITMOS PREDICTIVOS: ANÁLISIS DE LA SENTENCIA DE LA CORTE DE JUSTICIA DEL DISTRITO DE LA HAYA (PAÍSES BAJOS), DE 5 DE FEBRERO DE 2020

2.1. SUPUESTO DE HECHO

El Ministerio de Asuntos Sociales y Empleo del Gobierno neerlandés utilizaba, desde el año 2014, la herramienta de valoración del riesgo de fraude fiscal SyRI, que recopilaba datos sobre ingresos, pensiones, deudas o el tipo de domicilio de los contribuyentes, entre muchas otras variables, para calcular, con base en algoritmos, las probabilidades que tenía cada ciudadano de defraudar a la Administración. El *modus operandi* era el siguiente: si alguna agencia gubernamental sospechaba de fraude sobre impuestos o beneficios en un determinado vecindario, podía recurrir a SyRI y así obtener un «informe de riesgos», en el que se indicaban las personas físicas o jurídicas que eran consideradas sospechosas y, por tanto, objeto de investigación²⁷.

Ante esta situación, varios grupos de interés de la sociedad civil, incluida la Sección Holandesa de la Comisión Internacional de Juristas (Nederlands Juristen Comité voor de Mensenrechten —NJCM—), o la Confederación de Sindicatos de los Países Bajos (Federatie Nederlandse Vakbeweging —FNV—), además de dos particulares, iniciaron el proceso contra el Estado de los Países Bajos alegando que la aplicación de SyRI por parte del Gobierno suponía una flagrante violación de los derechos humanos. Como expresamente destacaban, el derecho a la privacidad es un derecho fundamental protegido por el derecho internacional —arts. 8 del CEDH y 17 del Pacto Internacional de Derechos Civiles y Políticos (en adelante, PIDCP)—, vinculantes de conformidad con los arts. 93 y 94 de la Constitución holandesa (6.20), pues las intromisiones gubernamentales en el ejercicio de este derecho solo están permitidas cuando sea necesario en interés de la seguridad nacional, la seguridad pública o el bienestar económico del país (art. 8.2 CEDH).

Concretamente, el derecho a la vida privada, como afirma el TEDH, engloba «el derecho a la autonomía personal, el desarrollo personal y la autodeterminación y el derecho a establecer relaciones con otros seres humanos y el mundo exterior», constituyendo los principios de dignidad y libertad humana «la esencia misma de la Convención» (6.23)²⁸. Y, como se desprende de su misma jurisprudencia, el derecho a la protección de datos personales deviene fundamental para el respeto de la vida privada²⁹, que en el contexto del procesamiento de datos se articula como el «derecho a la igualdad de trato en casos iguales y al derecho a la protección contra la discriminación, los estereotipos y la estigmatización» (6.24).

Con base en esta doctrina consolidada, se puso en entredicho el cumplimiento de la herramienta SyRI de una serie de principios que tanto la Carta y el Tratado de Funcionamiento de la Unión Europea (TFUE) como el Reglamento Europeo de Protección de Datos (General Data Protection Regulation —GDPR-EU—) regulaban: el principio de transparencia³⁰, el principio de limitación de propósito³¹, el principio de minimización de datos³², el principio de precisión³³ y el principio de integridad y confidencialidad³⁴ y el principio de responsabilidad³⁵. Y, como seguidamente se abordará, la Corte interpretó el art. 8, apartado 2, del CEDH sobre la base de estos principios (6.41), analizando si la legislación SyRI cumplía con los requisitos establecidos en el articulado citado.

2.2. RESOLUCIÓN DE LA CORTE

2.2.1. Caracterización del SyRI: *deep learning*, *datamining* y *big data*

Uno de los primeros aspectos que la Corte de Justicia del Distrito de la Haya se detiene a analizar es la caracterización de la herramienta SyRI como un sistema de recopilación de datos personales para fines de investigación. En concreto, se trata de un sistema de seguimiento digital con el que los/as ciudadanos/as se clasifican en perfiles de riesgo y en cuyo contexto el Estado utiliza el «aprendizaje profundo» (*deep learning*: de forma simplificada, en relación con la temática de presente estudio, constituye una forma de automatización del análisis predictivo³⁶) y la minería de datos (*data mining*: proceso automatizado que detecta información de conjuntos grandes de datos, hallando patrones y facilitando la toma de decisiones³⁷) mediante la tecnología *big data*. No obstante, la Corte se hace eco de la controversia existente en torno a la definición exacta del término *big data*, y considera irrelevante para su evaluación adicional si el procesamiento de datos en SyRI debería ser calificado como tal (6.52).

En cualquier caso, en atención al objeto de la investigación que se presenta, lo verdaderamente relevante en este punto es la «Opinión consultiva independiente sobre los efectos de la digitalización en las relaciones constitucionales» que presentó la División Asesora del Consejo de Estado al gabinete³⁸, a la que el Tribunal se refiere en numerosas ocasiones en la caracterización de SyRI, donde se advirtió que el término «autoaprendizaje» que subyace al uso de estos instrumentos es confuso y engañoso, pues un algoritmo, literalmente, «no conoce ni comprende la realidad (6.46)»; una frase, sin lugar a duda, verdaderamente esclarecedora que, ante su uso en los procesos penales, debería tenerse más presente.

2.2.2. Controversias sobre el modelo de riesgo: algoritmos secretos

Al margen de disputas marginales, lo cierto es que el aspecto principal de debate en el presente caso fue la pertinencia del uso de algoritmos secretos. Como la Corte puso de relieve, el Estado no reveló el modelo de riesgo y los indicadores que componían el modelo de riesgo de la herramienta utilizada, pues no proporcionó información objetivamente verificable que permitiera a la ciudadanía evaluar la naturaleza de SyRI; esto es, no mostraba cómo funcionaba el modelo de decisión de este instrumento y qué indicadores podían usarse (6.49). En realidad, el modelo de riesgo que se está utilizando actualmente en esta herramienta y los indicadores de riesgo que lo componen son «secretos», pues el sistema no proporciona información sobre el funcionamiento del mismo, de modo que no es posible conocer el tipo de algoritmos utilizados en SyRI. En igual sentido, tampoco proporcionó información sobre el método de análisis de riesgos aplicado por la Inspección de Asuntos Sociales y Empleo.

Con carácter general, el Estado neerlandés únicamente explicó que este modelo de riesgo se basa en «*i*) indicadores de riesgo, *ii*) enlaces y *iii*) el denominado punto de corte». Y, en función del objetivo de la investigación, según se detalla, a cada indicador de riesgo se le otorgan determinados puntos, con base, entre muchos otros aspectos, en la probabilidad de ocurrencia del indicador de riesgo en cuestión, de modo que «cuanto más improbable sea que ocurra el indicador de riesgo específico, mayor será el puntaje», constituyendo el punto de corte predeterminado un valor umbral (6.89). Asimismo, aunque el Estado defiende la validación de estos modelos de riesgo por la Inspección de Asuntos Sociales y Empleo, lo cierto es que SyRI no permite conocer la validación del modelo de riesgo y la verificación de los indicadores de riesgo (6.89).

En el mismo sentido, se destaca otro aspecto importante, como es que «el modelo de riesgo, los indicadores y los datos que se han procesado concretamente no son públicos ni son conocidos por los interesados», ya que la legislación de SyRI no establece la obligación de informar a las personas de que sus datos han sido procesados por su sistema informático ni existe tampoco la obligación legal de informar a los interesados de que se ha presentado un informe de riesgos al respecto, cuando, como la Corte entiende, el mismo «tiene un efecto significativo en la vida privada de la persona a quien pertenece» (6.65). De hecho, aunque el inicio de un proyecto SyRI se publica en la *Gaceta del Gobierno*, puede conservarse un informe de riesgos en el registro durante dos años sin que el interesado lo sepa (6.60), con los problemas que de ello se derivan.

Por ende, no podría ser verificado el método de generación del conocido popularmente como «árbol de decisión» del sistema —esto es, el modelo de predicción— ni los pasos que lo componen, generando así indefensión respecto de las personas sobre las cuales se presenta el citado informe de riesgos. Igualmente, el requisito de transparencia del procesamiento tampoco se cumple cuando se procesan datos de un sujeto que no dieron como resultado un informe de riesgos, aunque los mismos deban destruirse, en todo caso, cuatro semanas máximo después del análisis (6.90).

2.2.3. ¿Criminalización de la pobreza?

Dicho esto, el uso de SyRI en este contexto comporta otro importante problema: los posibles efectos discriminatorios «no intencionales o de otro tipo», como denunció el relator especial de la ONU sobre pobreza extrema y derechos humanos, que afirmó, literalmente, que el uso de SyRI tiene «un efecto discriminatorio y estigmatizador». Concretamente, Philip Alston, remitió un *amicus curiae*³⁹ a la Corte, objetando que SyRI señalaba a los/as ciudadanos/as con menores rentas, así como a grupos de población de origen inmigrante, concluyendo que esta herramienta representaba una amenaza potencial para los derechos humanos, en particular, «para los más pobres de la sociedad»⁴⁰.

No en vano, como las partes demandantes alegaron y la Corte declaró probado, SyRI se utiliza, especialmente, para investigar a los vecindarios que se conocen como «áreas problemáticas». Y, obviamente, con ello aumentan las posibilidades de descubrir irregularidades en estas concretas áreas en comparación con otros vecindarios, contribuyendo a reforzar los estereotipos y la imagen negativa de estos. Aunque lo cierto es que, como posteriormente se detallará, no se estimó que ello implicase, *per se*, el incumplimiento del art. 8.2 CEDH, reconoció, considerando los datos utilizados, que podía crear enlaces basados en sesgos, como un menor nivel socioeconómico. De hecho, no hay que desconocer que este sistema contempla datos exclusivamente en disposición del Estado para generar el llamado «árbol de decisión».

2.2.4. Valoración: caso *S. & Marper vs. United Kingdom*

En atención a lo expuesto en líneas anteriores, no entiende la Corte que pueda ser posible fundamentar la ilegalidad del uso de esta herramienta, como así lo entendieron las partes demandantes⁴¹, con base en la jurisprudencia del TEDH en asuntos relacionados con la interceptación masiva no dirigida (vigilancia masiva) o la interceptación selectiva de datos en un contexto penal o de seguridad nacional, como, por ejemplo, sucedió en el caso *Sunday Times vs. United Kingdom*⁴², pues esta sentencia, como expresamente afirmó, «no puede considerarse como una guía individual para la evaluación del tribunal» (6.67).

Por el contrario, apuesta, con razón, por otro asunto: el caso *S. & Marper vs. United Kingdom*⁴³, que analizó la legalidad de la Ley Británica de Protección de Datos (1998), y la Base de datos de ADN nacional de Reino Unido, en relación con la retención de huellas digitales, muestras celulares y perfiles de ADN. Pese a reconocer que el contexto fáctico de este caso no es comparable al procedimiento enjuiciado, las consideraciones realizadas por el TEDH en esta controversia sobre protección de datos fueron de carácter más general que las vertidas en el proceso que las partes demandantes alegaron, lo que permite su aplicación aplicadas a la evaluación de la legalidad de la legislación SyRI (6.68).

Concretamente, en el citado caso, el TEDH se pronunció sobre la demanda de dos ciudadanos ingleses (que fueron detenidos, respectivamente, por intento de robo y acoso a su expareja), porque el Estado conservó en su poder sus huellas dactilares y muestras de ADN, una vez fueron absueltos o sus cargos retirados, como ocurrió precisamente con el

segundo de ellos. Siguiendo a González Fuster⁴⁴, este pronunciamiento subrayó especialmente la necesidad de limitar el tratamiento de estos datos, en aras de garantizar «un justo equilibrio entre los intereses públicos y privados en juego», trazando las directrices respecto de la conciliación de las relaciones entre el derecho a la vida privada y los principios de protección de datos personales. El TEDH remarca, en esta línea, que «la legislación nacional debe proporcionar una protección adecuada contra la arbitrariedad e indicar con suficiente claridad el alcance de la discreción conferida a las autoridades competentes y la forma de su ejercicio para cumplir los requisitos de accesibilidad y previsibilidad», especificando, asimismo, que «el nivel de precisión requerido de la legislación interna depende en gran medida del contenido del instrumento en cuestión, el campo para el que está diseñado y el número y el estado de las personas a las que se dirige» (6.69).

Con base en lo anterior, la valoración de la adecuación legal de la interferencia está relacionada, según la Corte, con su necesidad en una sociedad democrática (6.71), por lo que concluye que la legislación SyRI «contiene salvaguardas insuficientes para la conclusión de que es necesaria en una sociedad democrática a la luz de los propósitos de la legislación, como también requiere el art. 8 párrafo 2 CEDH. Como resultado, en su forma actual, esta legislación no pasa la prueba del art. 8 párrafo 2 CEDH y, por lo tanto, es ilegal» (6.72). Asimismo, destacó otro aspecto relevante, como es que, aunque el citado procesamiento de datos consta de dos fases, puesto que, tras la detección por este algoritmo de las personas en concreto, se lleva a cabo la «investigación humana» respecto de cada uno de estos, acerca de si cometió o no un caso de fraude, el Tribunal lo considera insuficiente, precisamente porque «la forma en que se lleva a cabo la selección definitiva del riesgo no es pública» (6.94).

Pues bien, al análisis detallado de estos presupuestos se dedican los epígrafes que siguen.

2.2.4.1. Necesidad general en una sociedad democrática

Como fundamento de su decisión, la Corte de Justicia del Distrito de la Haya parte del análisis de la necesidad de esta herramienta, en interés del bienestar económico del país, en una sociedad democrática. Entiende, en primer lugar, que SyRI tiene un propósito legítimo (6.74), si bien lo que se trata de determinar, en cualquier caso, es la existencia, según la Corte, de una «necesidad social apremiante» que legitime el uso de dicho sistema.

Por otro lado, las partes denunciadas, como líneas antes se destacó, pusieron de relieve la grave interferencia en la vida privada de la ciudadanía que este sistema comportaba, destacando en igual sentido que el Estado no había demostrado la necesidad de la implementación de un instrumento tan invasivo para mantener el sistema de seguridad social, no siendo tampoco efectivos los proyectos en los que se ha implantado el sistema para combatir el fraude (6.75). No obstante, el Tribunal rechaza estos argumentos, al entender que «la legislación SyRI en sí misma busca cumplir un propósito suficientemente convincente para justificar una interferencia con la vida privada», aludiendo a la importancia del fraude en el área de la seguridad social y el bienestar que tiene incluso efectos indirectos en la integridad del sistema económico y la confianza en las instituciones financieras.

Así pues, argumenta que «el daño directo e indirecto del fraude en esta área justifica la conclusión del legislador de que existe una necesidad social apremiante de tomar las medidas previstas por la legislación SyRI en interés del bienestar económico de los Países Bajos» (6.76). En definitiva, concluye la Corte que, en atención a sus fines, se cumple con el requisito de necesidad general del art. 8 del CEDH (6.78). Ahora bien, como seguidamente se analizará, esto no implica que el funcionamiento del instrumento de elección y los procedimientos y garantías asociados creados para su aplicación por el legislador respeten suficientemente la privacidad a la luz de la regulación europea (6.79).

2.2.4.2. Proporcionalidad y subsidiariedad

Tras abordar el requisito de su necesidad, la Corte evaluó, seguidamente, si el instrumento elegido por el Gobierno neerlandés cumplía con los requisitos de proporcionalidad y subsidiariedad de conformidad con el art. 8.2 del CEDH, a la luz de los objetivos perseguidos, valorando el «equilibrio justo» entre los propósitos de la legislación SyRI, esto es, evitar fraudes, y la invasión de la vida privada que causa este sistema (6.80), como se puso de manifiesto asimismo en el *amicus curiae* del Relator Especial de Naciones Unidas anteriormente visto⁴⁵. Y, como concluyó, efectivamente la aplicación de SyRI no alcanzaba el citado equilibrio justo (6.83).

Para justificar su decisión, recuerda la especial responsabilidad del Estado con el uso de toda nueva tecnología, sobre la que advirtió el TEDH en el asunto *S. & Marper vs. United Kingdom* (6.84), pues la recopilación y análisis de datos puede «interferir ampliamente con la vida privada de aquellos a quienes pertenecen los datos», por lo que el legislador tiene una «responsabilidad especial cuando aplica un instrumento como SyRI» (6.85). En atención al principio de transparencia, limitación de propósito y de minimización de datos, sostiene que la legislación SyRI no es lo suficientemente transparente y verificable (6.86). Concretamente, respecto del principio de transparencia (principio fundamental para la protección de datos), se arguye que este sistema no proporciona información sobre los datos objetivos que permitan fundamentar que existe un mayor riesgo, pues tan solo se aporta, a modo ilustrativo, una serie de indicadores que pueden revelar un mayor riesgo, sin explicar, de todos modos, en qué información objetivamente verificable se basan estos ejemplos (6.87-6.88).

Tras el análisis efectuado, se concluye que SyRI «contiene garantías insuficientes para proteger el derecho al respeto de la vida privada en relación con los indicadores de riesgo y el modelo de riesgo utilizados», y que por ello no respeta las exigencias del art. 8.2 del CEDH (6.95), pues no proporciona información sobre el funcionamiento y la validación de los indicadores de riesgo y el modelo de riesgo, con lo que los sujetos no tienen «la certeza suficiente de que su privacidad está protegida cuando se usa SyRI» (6.100).

3. ALGORITMOS PREDICTIVOS Y DERECHO PENAL

3.1. APUNTES PREVIOS: LOS INSTRUMENTOS DE VALORACIÓN DEL RIESGO EN EL PROCESO PENAL

Desde la mitad de la década de los años ochenta del pasado siglo XX, los procedimientos actuariales de valoración del riesgo en el ámbito penitenciario⁴⁶ se han multiplicado, generalizándose su uso en la práctica forense⁴⁷. Como en la introducción se puso de relieve, concretamente las herramientas estructuradas de valoración del riesgo, en contraste con los instrumentos clásicos, estructuran todo el proceso de estimación del riesgo, haciendo depender la probabilidad de que un evento suceda en el futuro de la combinación de diferentes factores que en el pasado ocasionaron su aparición. En concreto, seleccionan muestras de sujetos y observan sus circunstancias y comportamientos para determinar los factores que han estado asociados, en el concreto grupo, a la ocurrencia del fenómeno a predecir, de forma que, cuando se valora el riesgo de reincidencia de un determinado sujeto, se repasa la lista cerrada de factores de riesgo y se asigna a cada uno de ellos un valor numérico en función de si se presenta o no en el individuo en cuestión⁴⁸. Estas cifras se combinan posteriormente a través de un algoritmo estadístico que arroja una determinada puntuación del sujeto, la cual se compara con la escala de evaluación que dicho instrumento contiene, de tal modo que, en función del nivel de la escala en la que se sitúe la calificación del mismo, se le otorga un porcentaje de probabilidad de reincidencia. No obstante, como claramente se advierte, no se está ante una predicción individual para los individuos en concreto, sino que lo que se interpreta son unos valores obtenidos respecto a conjuntos de personas que presentan una serie de factores similares a los de este individuo⁴⁹; y si se tiene en cuenta la variabilidad interindividual inherente a cualquier grupo, los déficits de estas predicciones son más que evidentes⁵⁰, más aún cuando únicamente se toman en consideración factores estáticos.

Estos nuevos métodos, en los que se ha depositado una enorme esperanza, especialmente en el ámbito anglosajón, están empezando a utilizarse no solo en la fase del *pre-trial* para la adopción de medidas cautelares⁵¹, sino también en la fase de determinación de la pena⁵². Conocidas como *evidence-based sentencing* o *smart sentencing*, las valoraciones de riesgo son incorporadas como una de las variables a considerar para establecer aspectos como el tipo o duración de la pena, con el fin de reducir la reincidencia, fin legítimo del Derecho penal con base al cual se justifican. Concretamente, se destaca que estas herramientas permiten superar la «subjetividad» de los juicios intuitivos y personales que realizan los jueces, pues permiten una aproximación rigurosa al riesgo. Ahora bien, aunque estos instrumentos están adquiriendo cada vez mayor popularidad, y han sido incluso incorporados en una de las últimas revisiones del Model Penal Code⁵³, su uso en el proceso ha generado un importante debate doctrinal.

En cualquier caso, lo cierto es que, especialmente en EE. UU., cada vez más autores apuestan por su introducción⁵⁴ debido principalmente a su mayor capacidad predictiva, conforme se ha concluido en diferentes metaanálisis que agrupan los resultados de múltiples estudios que han comparado la precisión de los métodos clínicos y los actuariales en

la predicción de resultados que, no obstante, van desde el diagnóstico de una enfermedad hasta el riesgo de violencia y otros comportamientos delictivos⁵⁵. No en vano, como en el previamente citado metaanálisis se advierte, la precisión de estas predicciones varía según el tipo de predicción, el entorno en el que se recopilan los datos del predictor, el mismo tipo de fórmula estadística utilizada o la cantidad de información disponible. Y, aunque esta investigación concluyó que los métodos actuariales eran «claramente superiores al enfoque clínico», reconoció que esa superioridad era modesta⁵⁶.

De todos modos, cada vez más estudios efectúan comparativas específicas en el terreno judicial entre los métodos clínicos y los actuariales con base en la constatada débil relación entre las predicciones de los jueces y el riesgo real de los casos. A modo meramente ilustrativo, pueden destacarse dos investigaciones publicadas en el año 2017: la del grupo de investigación encabezado por Kleinberg⁵⁷, en la que se concluyó que reemplazar las decisiones judiciales por decisiones algorítmicas podría reducir los delitos previos al juicio en un 25 %, y la de Jung⁵⁸, que, en la misma línea, demostró la superioridad de las predicciones algorítmicas complejas sobre las decisiones de los jueces.

Al margen de lo apuntado, como incluso los partidarios de estas herramientas actuariales reconocen⁵⁹, lo decisivo sería la constatación de la superioridad de los instrumentos puramente actuariales sobre los instrumentos mixtos. Como Skeem y Monahan⁶⁰ ponen de relieve, entre las evaluaciones clínicas no estructuradas y las evaluaciones completamente actuariales encontramos formas de evaluación parcialmente estructuradas que, en el ámbito penal, pueden ser las más idóneas, ya que en la mayoría de metaanálisis específicos realizados no se han encontrados diferencias significativas en cuanto a la predicción entre los juicios profesionales estructurados y los actuariales⁶¹.

Como se infiere de lo expuesto, la discusión sobre cuál es el método de predicción más fiable y con menores tasas de error continúa, y si bien existe acuerdo en que actualmente el método intuitivo está desacreditado (tanto por la falta de fiabilidad y transparencia⁶² como su poca validez predictiva⁶³ o su escaso soporte empírico⁶⁴)⁶⁵ en relación con los demás sistemas, la cuestión es controvertida⁶⁶.

Por razones obvias, no es posible abordar un estudio en profundidad de todas las limitaciones técnicas que se han señalado sobre estos instrumentos de predicción del riesgo —desde la sensibilidad y especificidad al valor predictivo—, si bien, con base en una investigación que aborda de forma minuciosa las limitaciones de estos instrumentos en su aplicación específica en el terreno penal y, particularmente, respecto de su uso en las medidas de seguridad⁶⁷, puede concluirse que todos los métodos de predicción del riesgo de reincidencia, también los más desarrollados, tienen una capacidad predictiva verdaderamente baja cuando son utilizados para predecir la probabilidad de comisión de futuros delitos por parte de individuos concretos, sobre todo porque el valor predictivo positivo⁶⁸, que es el que se necesita para imponer medidas de seguridad, en muchas ocasiones es inferior al mero azar (esto es, inferior al 0,5), sin descontar los amplios márgenes de error. En este sentido, no debería olvidarse que, en la sentencia de la Haya anteriormente analizada, el ministro de Protección Legal de los Países Bajos destacó como aspecto a considerar res-

pecto del uso de SyRI la elevada cifra tanto de los falsos positivos como de falsos negativos del instrumento⁶⁹.

Tal vez por ello, en los últimos años se recurre con mayor frecuencia a estos métodos de juicio clínico estructurado⁷⁰, pues confieren mayor libertad al evaluador para valorar el grado determinado de peligrosidad del individuo, aunque con un grado de fiabilidad y transparencia superior al método clínico puro. Esto es así porque, de forma similar a los instrumentos actuariales, contienen una lista de factores de riesgo a valorar, junto con la regulación de la forma de recogida de la información, sin olvidar que estos instrumentos no registran únicamente información sobre datos estáticos e inmodificables, sino también ítems dinámicos. De hecho, numerosos estudios ponen de relieve que los factores de riesgo dinámicos predicen mejor la reincidencia delictiva de adultos que los estáticos⁷¹, superando a los instrumentos estructurados de evaluación del riesgo y al juicio clínico.

3.2. CONSTITUCIONALIDAD DEL USO DE INSTRUMENTOS DE VALORACIÓN DEL RIESGO EN EL PROCESO PENAL ESTADOUNIDENSE: CASO *STATE VS. LOOMIS*

Todas estas cuestiones pueden ser analizadas, precisamente, al hilo de un caso resuelto por el Tribunal Supremo de Wisconsin (en adelante, TSW) en el año 2016, en el que se examinó la constitucionalidad del uso de valoraciones del riesgo de violencia para determinar la clase y la extensión de la pena en atención al derecho al debido proceso (*due process of law*): el caso *State vs. Loomis*⁷².

El TSW se pronunció respecto del caso de Eric Loomis, acusado de cinco delitos al ser arrestado mientras conducía un vehículo utilizado en un tiroteo previo: dos cargos por posesión de armas de fuego, intento de evasión frente a un agente de tráfico, puesta en peligro de la seguridad pública, y conducción de un vehículo ajeno sin autorización. Loomis, que negó su participación en el tiroteo, reconoció haber conducido el vehículo con posterioridad a los hechos, llegando por ello a un acuerdo con la Fiscalía para aceptar los dos últimos cargos, si bien rechazó el resto de acusaciones.

Ahora bien, lo importante en relación con el objeto del presente estudio es que, al concretar la pena, el tribunal de instancia consideró, entre otros factores, el resultado de la evaluación del acusado mediante una herramienta estructurada de valoración del riesgo de reincidencia (COMPAS), que determinaba que Loomis presentaba un alto riesgo de reincidencia en general, así como de reincidencia violenta, para condenarlo a seis años de prisión, más otros cinco de supervisión postpenitenciaria. Asimismo, excluyó expresamente la suspensión de la condena conforme con tales predicciones. Pero, aunque Loomis solicitó la revisión de la cuantía de la pena basándose, entre otras razones, en la vulneración de su derecho a un debido proceso por la fundamentación de esta condena en el resultado ofrecido por COMPAS, el tribunal desatendió su solicitud, al entender que el resultado proporcionado por esta herramienta únicamente fue considerado por el tribunal de instancia para corroborar las conclusiones a las que llegó con base en el resto de información disponible.

Romeo Casabona⁷³ o Martínez Garay⁷⁴ son algunos de los autores que, sin desconocer la profusa literatura estadounidense⁷⁵, han analizado detenidamente este pronunciamiento en España, primero de una corte suprema estatal que permite el uso de cálculos matemáticos con un algoritmo secreto para determinar una condena criminal⁷⁶ y que, como esta última autora señala, demuestra el dilema que pone en evidencia la utilización de estos métodos entre la incorporación de información rigurosa, según el tribunal, para fijar la condena, y el correcto funcionamiento del sistema judicial, que, para una correcta resolución del caso, debe partir del respeto a los derechos fundamentales del acusado, de modo que, ante su lesión, debería prescindir de las posibles ventajas de estos métodos. No obstante, lo cierto es que la citada sentencia no establece ni tal rechazo ni fija los criterios necesarios para que estas valoraciones puedan ser introducidas con respeto a los derechos fundamentales de los acusados⁷⁷.

Con todo, aunque el TSW rechazó todos los motivos del recurso y se mostró partidario de la introducción de estos instrumentos, resulta especialmente significativo el análisis que el citado tribunal efectúa sobre los problemas que la literatura científica ha destacado en relación con el uso de estas herramientas de valoración del riesgo, afirmando que estos instrumentos no pueden ser utilizados, en exclusiva, para determinar la severidad de la sentencia o para decidir el encarcelamiento de las personas. Concretamente, reconoció que la herramienta estructurada que valoró el riesgo de reincidencia y las necesidades criminógenas del sujeto, COMPAS, que estimaba que Loomis presentaba un riesgo alto tanto de reincidencia en general como de reincidencia violenta, no desvelaba cómo funcionaba el algoritmo que calcula tales niveles de riesgo. Sin embargo, el TSW rechazó el recurso al aportar la empresa propietaria de COMPAS información respecto de la herramienta, considerando que extremos tales como el funcionamiento del algoritmo, es decir, la forma de medición, ponderación y combinación de las diferentes variables contempladas en el instrumento, quedaba dentro del secreto de empresa, imprescindible en el negocio. Así pues, se acogió al precedente establecido en el caso *Gardner vs. Florida*⁷⁸ para afirmar que los acusados no tienen un derecho al conocimiento detallado de la información contenida en su informe de investigación («derecho a la explicación»), sino tan solo un «derecho a la información»⁷⁹.

Por ello, aunque Loomis no pudo conocer el preciso funcionamiento del algoritmo, sí tuvo acceso tanto a los resultados como a la información con la que estos se calcularon, que refiere a los datos objetivos sobre su historial, de modo que el TSW entendió que no se había vulnerado su derecho, como acusado, a ser condenado sobre la base de información fiable y exacta solamente por no tuvo acceso a la explicación concreta sobre el tratamiento informático de los citados algoritmos⁸⁰. Desafortunadamente, como apunta Romeo Casabona⁸¹, aunque se solicitó la revisión del caso al Tribunal Supremo de EE. UU., este denegó su admisión⁸², por lo que se impidió la valoración por parte de la Corte Suprema del país de la legitimidad del uso de estas herramientas actuariales como soporte de una condena penal y la posible violación del derecho al debido proceso que el uso de un instrumento de evaluación de riesgos cuyo algoritmo se desconoce podría comportar.

4. REFLEXIONES CRÍTICAS: ALGORITMOS SECRETOS Y DISCRIMINACIÓN EN EL PROCESO PENAL

La decisión del TSW en el caso *State vs. Loomis* antes vista ha sido fuertemente cuestionada⁸³, pues supone la aceptación de que las sentencias puedan basarse, en parte, en valoraciones de riesgo realizadas mediante el uso de herramienta desarrolladas por empresas privadas que no revelan el funcionamiento de los algoritmos en los que basan sus resultados. En esta línea, Martínez Garay⁸⁴ entiende que los tribunales no deberían admitir valoraciones de riesgo con base en algoritmos secretos, pues, conforme con la Constitución española (en adelante, CE), el alcance de la tutela judicial efectiva (art. 24.2 CE) se extiende al poder de cuestionar no solo los resultados, sino también el diseño y funcionamiento de estas herramientas. De hecho, en una investigación encabezada por Eckhouse⁸⁵ se objetó precisamente este extremo, al entender que tanto jueces como acusados tendrían el derecho de examinar estos algoritmos y evaluar su equidad, como también opina Carlson⁸⁶, autora que sostiene que el uso de este tipo de herramientas en la justicia penal debería sujetarse «a los mismos requisitos de transparencia que otras agencias gubernamentales» y no proteger en exclusiva los intereses comerciales de las empresas privadas, pues estos deberían ser objeto de verificación y discusión⁸⁷. Con respecto a este último punto, conviene recordar que, al margen de las herramientas comercializadas por empresas privadas, también existen instrumentos públicos o de acceso gratuito, y que no existe en ningún caso evidencia científica de que el rendimiento de las herramientas privadas sea mejor que el de las públicas⁸⁸.

Pues bien, precisamente con la reciente sentencia de la Corte de Justicia del Distrito de la Haya de 5 de febrero de 2020, la negativa a la introducción de los algoritmos secretos en la justicia se refuerza, ya que, sin desconocer nuevamente su diferente ámbito administrativo y territorial, a diferencia de la resolución del TSW, el tribunal neerlandés entiende que la herramienta de predicción actuarial en concreto utilizada (SyRI) contenía garantías insuficientes para proteger el derecho a la privacidad, dado que no proporcionaba información sobre el funcionamiento y la validación de los indicadores y el modelo de riesgo. En este sentido, no hay que desconocer que, con carácter general, y en atención a la insuficiente protección existente en la actualidad, en el terreno administrativo y, específicamente, en materia de protección de datos, encontramos propuestas como la de Boix Palop⁸⁹, que aboga por la consideración por parte de nuestro Derecho como reglamentos de los algoritmos y programas utilizados por la Administración para adoptar decisiones administrativas (o que sean apoyo de estas), derivándose de dicha consideración todas las garantías tradicionales que nuestro ordenamiento establece para las normas reglamentarias, previsión que comporta, entre otros aspectos, la necesidad de que estos algoritmos estén debidamente publicados⁹⁰, pues cualquier ciudadano/a tiene que poder revisar los códigos fuentes de estos instrumentos en su integridad⁹¹. Y esta exigencia, como se concluirá, es de capital importancia para detectar posibles errores o discriminaciones, entre muchos otros aspectos técnicos que estos programas pueden comportar. Una exigencia que, cuando se trata de su uso para fundamentar resoluciones en el proceso penal, resulta si cabe más pertinente.

Efectivamente, el reconocimiento de que los sistemas automatizados no están libres de sesgos ha sido ampliamente analizado, especialmente en atención a la discriminación

racista y sexista⁹². No obstante, en la sentencia del TSW en el caso *State vs. Loomis*, pese a reconocer y citarse en la misma sentencia diferentes estudios que cuestionaban los efectos discriminatorios del instrumento utilizado (al que no se tiene acceso) respecto a las minorías, se aceptó su introducción. Se trata de una decisión incomprensible, dado que, como previamente se ha apuntado, existen también herramientas de valoración del riesgo públicas, por las que han optado otras jurisdicciones, que permiten el acceso a la citada información. No hay que desconocer que las empresas privadas suelen entorpecer los esfuerzos para examinar los algoritmos que emplean, evitando con ello la reducción de los sesgos que, más que manipulados intencionalmente, en la mayoría de ocasiones reproducen los prejuicios inconscientes de la sociedad actual. Como algunos autores destacan, los resultados discriminatorios son, casi siempre, «una propiedad emergente no intencional del uso del algoritmo en lugar de una elección consciente por parte de sus programadores»⁹³. Y, precisamente en atención al caso resuelto por el TSW, se ha destacado que la herramienta COMPAS tendría como falsos positivos el doble de acusados negros que de blancos⁹⁴.

Más preocupante aún es que, como se concluirá, los algoritmos, incluso los neutros, pueden incrementar la discriminación. Como sostienen Hannah-Moffat y Struthers Montford⁹⁵, se ha demostrado que, en la «confección» de la «población criminal» la discriminación racial, junto con la socioeconómica, es determinante, y que estos instrumentos pueden potenciar estos sesgos⁹⁶. De forma meramente ilustrativa, para ser conscientes del alcance de lo enunciado, conviene recordar un anuncio de Naciones Unidas⁹⁷ que ponía de relieve que, mediante el uso de la función de autocompletar de Google, que utiliza búsquedas previas de otros usuarios para ayudar a predecir lo que estás buscando, se reflejaba el propio sesgo sexista de la sociedad; se trata de un claro ejemplo de cómo un algoritmo neutral produce resultados sexistas, simplemente porque «responde a los aportes de personas sexistas»⁹⁸.

No en balde, como en el punto central del presente estudio hemos tenido ocasión de examinar al abordar la respuesta de la Corte de Justicia del Distrito de la Haya en el caso SyRI, el relator especial de la ONU sobre pobreza extrema y derechos humanos no tuvo reparos en afirmar que el uso de SyRI tenía «un efecto discriminatorio y estigmatizador». El problema se acentúa si pensamos, como advierte Martínez Garay⁹⁹, que con carácter general se afirma que la información que proporcionan las herramientas estructuradas de valoración del riesgo es objetiva —y, por lo tanto, más fiable que las valoraciones personales que efectúan los jueces— presenta múltiples prejuicios inconscientes. Ello omite que, en cualquier caso, los algoritmos, al final, solamente reflejan los datos que incorporan estas técnicas, de modo que, si estas reúnen sesgos, el instrumento los reproducirá y, como se acaba de apuntar, hasta puede aumentarlos. De hecho, en EE. UU. se denuncia, de forma constante, que la valoración estructurada de este riesgo incrementa la sobrerrepresentación de las minorías en el sistema penitenciario.

Para terminar, conviene apuntar, con carácter general, que cada vez más autores reclaman la necesidad de una regulación pública que evite tales desigualdades¹⁰⁰. Además, como destaca Boix Palop¹⁰¹, las exigencias a los poderes públicos deberían ser más estrictas que las requeridas en las relaciones privadas, dado que los/as ciudadanos/as están sujetos al

cumplimiento de sus decisiones, por lo que la normativa reguladora de los concretos usos de estos algoritmos en la adopción de decisiones debería ser más garantista que la derivada de las normas de protección de datos (utilizadas, como vimos, por la Corte de Justicia del Distrito de la Haya en el caso SyRI) que, como sostiene el autor antecitado, únicamente se refiere a una de las cuestiones implicadas en el uso de la inteligencia artificial.

5. VALORACIÓN FINAL

El recurso por parte del sistema penal a las nuevas herramientas de valoración del riesgo se inserta en una dinámica en la que, en línea con la corriente expansiva del Derecho penal y la creciente cientifización de la vigilancia, estos nuevos medios tecnológicos, que se basan en algoritmos predictivos, son implementados en cada vez más sectores de la vida social, con los problemas, especialmente éticos, que de su uso se derivan, pues ni su funcionamiento ni la responsabilidad en su empleo se ha analizado de forma detenida por parte de los operados jurídicos.

Como hemos tenido ocasión de comprobar, en febrero del año 2020 la Corte de Justicia del Distrito de la Haya se pronunció por primera vez sobre el uso por parte del Gobierno neerlandés de una de estas herramientas, que usaban algoritmos predictivos para rastrear posibles fraudes al Estado, entendiendo que su apuesta por un algoritmo oculto no respetaba la privacidad de la ciudadanía y que, por tanto, vulneraba el art. 8.2 del CEDH¹⁰². Esta resolución se contrapone con la cuestionable respuesta que, desde EE. UU., se ofreció respecto de su uso concretamente en el proceso penal¹⁰³ en la resolución del TSW en el caso *State vs. Loomis*, pues se aceptó la introducción de una herramienta actuarial pese a no disponer información exacta el ciudadano en cuestión de su funcionamiento.

Sin lugar a dudas, uno de los usos más controvertidos de los algoritmos son las herramientas actuariales para la determinación del riesgo de reincidencia en el Derecho penal, especialmente en atención al respeto de los principios y garantías que deben informar el proceso penal. La polémica que envuelve a este concepto es incesante, posiblemente porque la predicción del comportamiento futuro, al igual que el mismo concepto de ciencia, no responde a certezas, sino a grados de probabilidad. Con respecto a su uso concreto en la adopción de medidas de seguridad, autores como Romeo Casabona¹⁰⁴ destacan que, a pesar de las limitaciones de los métodos clásicos¹⁰⁵, estas nuevas herramientas se configuran como una fuente «despersonalizada» e incompatible con los principios que imperan en nuestro ordenamiento jurídico. Por el contrario, sus defensores argumentan que, finalmente, estos instrumentos solamente muestran los sesgos grupales en los que también los jueces se basan al tomar sus decisiones, pues se sustentan en suposiciones sobre el tipo de persona que es el delincuente¹⁰⁶. En cualquier caso, si bien los defensores de estos métodos señalan sus ventajas cuando se trata de grandes casos, reconocen las limitaciones de los mismos respecto de su uso en el proceso penal, pues ofrecen muy poca información sobre los individuos reales evaluados¹⁰⁷. Sin olvidar, en igual sentido, que estas estimaciones de pro-

babilidad específicas lo son dentro de un corto período de tiempo (por ejemplo, uno o dos años), con las dificultades que ello comporta para su implementación en el proceso penal.

Esto no significa que deba renunciarse a toda pretensión de objetividad; es más, se apuesta por la introducción de los procedimientos actuariales como uno de los criterios a valorar de forma complementaria por los jueces, junto con los informes individuales realizados por profesionales especializados, como psiquiatras o criminólogos, que evalúan la personalidad y el entorno social del delincuente¹⁰⁸. No obstante, en la actualidad no parece conveniente, en línea con lo que sostienen Martínez Garay o Romeo Casabona¹⁰⁹, la introducción con carácter general de procedimientos de evaluación del riesgo en exclusiva para determinar estos pronósticos, rechazando especialmente aquellos que se sirven de algoritmos actuariales o automatizados ocultos¹¹⁰. En contraste, debería reflexionarse sobre la equidad y riesgos en la introducción de estas nuevas herramientas en el sistema de justicia penal, confeccionando el marco legal adecuado para la introducción de unos modelos que, de todos modos, requieren del juicio humano en su interpretación¹¹¹. El problema que quedaría por resolver, como ya ha podido comprobarse en EE. UU., es la influencia que estas herramientas estructuradas de valoración del riesgo, pese a ser complementarias, podrían tener en los jueces; especialmente cuando determinan puntajes altos de reincidencia y centran su la atención en estos instrumentos más que en otros factores¹¹².

De hecho, al margen de las múltiples objeciones técnicas de estos instrumentos, inclusive aquellos autores que en EE. UU. defienden su uso en el proceso penal con base en la peligrosa idea de su rentabilidad para con la protección de la sociedad¹¹³, entienden que estas herramientas deberían someterse a aquellos controles necesarios para garantizar el respeto de los derechos fundamentales de la ciudadanía. En este sentido, sería conveniente, como Romeo Casabona¹¹⁴ pone de relieve, que para su uso por parte del sistema judicial estas herramientas fueran sometidas a la supervisión y homologación por agencias independientes colaboradoras con la Administración de Justicia. A tal efecto, autores como Carlson¹¹⁵ proponen que, en lugar de confiar en el sector comercial privado, los gobiernos deberían desarrollar sus propios instrumentos actuariales y algorítmicos¹¹⁶.

Al final, la elección del algoritmo más adecuado para una concreta aplicación depende de muchos factores: desde el número de elementos que se ordenan hasta la medida en que estos están clasificados. Y, para ello, como en la sentencia neerlandesa analizada en este estudio, se afirma, se requiere el conocimiento de su concreto funcionamiento por parte de la ciudadanía junto con la validación de los indicadores y el modelo de riesgo. Esto es más importante, si cabe, en el proceso penal, de modo que, en España, los tribunales no deberían admitir valoraciones de riesgo hechas con algoritmos secretos si quieren respetarse las garantías del proceso penal vinculadas a nuestro Estado de Derecho, pues, como previamente se ha destacado, con su introducción se atentaría al derecho a la tutela judicial efectiva (art. 24.2 CE). Asimismo, no cabe desconocer la colisión con el derecho a la intimidad, consagrado en el art. 18 de la CE, que autores como Miró Llinares¹¹⁷ analizan en profundidad respecto del tratamiento automatizado de datos personales por los sistemas judiciales.

En definitiva, en la pretensión de la consecución de un conocimiento neutral y objetivo, libre de los sesgos subjetivos del especialista, se desconoce que los algoritmos predictivos nunca son neutrales, pues incorporan, como se ha tenido ocasión de comprobar, una importante carga de valor, con el riesgo de perpetuar tratamientos discriminatorios. Estos aspectos deberán ser tenidos en cuenta a la hora de elaborar el marco normativo que garantice la compatibilidad del uso de estos nuevos instrumentos con los derechos y libertades reconocidos a la ciudadanía. Ello requiere, asimismo, que estas herramientas partan de una regulación ética, pues, como declaró la Corte de Justicia del Distrito de la Haya respecto de la herramienta SyRI, su recurso puede estar justificado en el Derecho penal en atención al fin legítimo pretendido, como es en este caso la reducción de la reincidencia.

La sociedad no puede poner obstáculos al avance tecnológico, pues estas herramientas están demostrando su utilidad cada vez en más sectores para dar respuesta a problemas complejos en los que intervienen múltiples variables, entre ellos la Administración de Justicia¹⁸. Ahora bien, deviene imprescindible reflexionar sobre los diferentes usos de estas herramientas junto con sus potenciales problemas derivados, lo que exige un estándar normativo que tenga en cuenta no únicamente los intereses de la comunidad, sino también los derechos y libertades de los particulares afectados, vinculados al mantenimiento de un Estado social y democrático de Derecho.

NOTAS

1. Para más información, especialmente jurídica, consúltese Martínez Garay (2014: 579-606).
2. Sagioglu y Sinanc (2013: 42-47).
3. Cabe destacar, en este sentido, un interesante documental que analiza, de forma cercana, estos sistemas de vigilancia predictiva, los cuales no solamente han llegado a implementarse en ciudades estadounidenses, sino también en metrópolis europeas, entre ellas Londres o Munich. Al respecto, Hielscher y Heeder (2017).
4. Se estima imprescindible aclarar el significado de un concepto que, pese a su frecuente uso, en ocasiones se desconoce de forma precisa. Los algoritmos pueden ser definidos, de forma sencilla, como una secuencia de pasos de cálculo que transforman entradas (*input*) en salidas (*output*). Este procedimiento computacional toma un valor o conjunto de valores como entrada, y produce un valor o conjunto de valores como salida. Así pues, los algoritmos no serían más que «una herramienta para resolver un problema computacional bien especificado»; el conjunto de sencillas instrucciones a seguir para resolver un problema concreto, analizando conjuntos de datos para identificar tendencias y crear predicciones. Cormen, Leiserson, Rivest y Stein (2009: 10).
5. C/09/550982/HA ZA 18-388. Disponible en: <<https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878>>.
6. Brandariz García (2016: 27).
7. Castro Liñares (2019: 79). Para un estudio en profundidad, *vid.* Feeley y Simon (1992: 449).
8. H.R.5682 - FIRST STEP Act 115th Congress (2017-2018). Disponible en: <<https://www.congress.gov/bill/115th-congress/house-bill/5682>>.

9. Christin, Rosenblat y Boyd (2015); y Goel, Shroff, Skeem y Slobogin (2018).
10. No en vano, como indica Castro Liñares, los instrumentos de valoración y gestión del riesgo surgieron con la implementación de esta institución. Castro Liñares (2019: 63).
11. Martínez Garay (2018: 488)
12. Romeo Casabona (2018: 42-54)
13. Esto se explica porque parten de la base de que la conducta humana es comprensible a través de la causalidad, de cuyo curso pueden extraerse predicciones sobre la conducta humana en el futuro.
14. Romeo Casabona (2018: 47).
15. Miró Llinares (2018). En este trabajo se analizan, de forma minuciosa, las implicaciones del uso de sistemas de inteligencia artificial en la justicia penal, abordando en qué consisten exactamente los mismos y sus usos en relación con la justicia criminal. Asimismo, es de particular interés en el presente estudio su examen de COMPAS, herramienta automatizada de valoración del riesgo que, como habrá ocasión de comprobar en el tercer epígrafe, está siendo utilizada por los tribunales estadounidenses. Efectúa, asimismo, un interesante análisis de los problemas generados por el progresivo uso de estas herramientas automatizadas Nieva Fenoll (2018: 23).
16. Miró Llinares (2018: 41 ss.).
17. Simon (1998: 455).
18. Para una interesante lectura de esta sentencia, que pone al descubierto las deficiencias que conforme a la misma se estarían produciendo en el sector público valenciano, consúltese Cotino Hueso (2020). Por otro lado, para un análisis focalizado en sus implicaciones en relación con la privacidad, y los principios de seguridad jurídica e interdicción de la arbitrariedad, Lazcoz Moratinos y Castillo Parrilla (2020).
19. Así lo estima, por ejemplo, Andrés-Pueyo (2013: 483 ss.).
20. Angwin, Larson, Mattu y Kirchner (2016); Barabas, Dinakar, Ito, Virza y Zittrain (2018); y Martin (2019: 835-850).
21. Un repaso a las notas que caracterizan la introducción de estas herramientas en el proceso penal, en Brandariz García (2016: 93-110).
22. Ægisdóttir, White, Spengler, Maugherman, Anderson, Cook, Nichols, Lampro-poulos, Walker, Cohen y Rush (2006: 341-382).
23. Goel, Shroff, Skeem y Slobogin (2019: 2); y Ægisdóttir, White, Spengler, Maugherman, Anderson, Cook, Nichols, Lampro-poulos, Walker, Cohen y Rush (2006: 341-382).
24. Tegmark (2017: 61-71).
25. No obstante, algunos autores, como Castro Liñares (2019: 91-92), específicamente respecto de su aplicación en el terreno penal, entienden que comportarían una «revolución formal» más que un cambio en sentido material.
26. Un resumen de los avances en el marco regulatorio actual de los usos de la inteligencia artificial en: Boix Palop (2020: 226 ss.). Interesa destacar, en relación con las iniciativas estatales, la prohibición en el Derecho público alemán —si bien en su Ley de procedimiento administrativo (§35VwVfG)— del

recurso a algoritmos para la adopción de decisiones que afecten a los derechos de los/as ciudadanos/as con contenido discrecional.

27. Concretamente, fueron cinco las ciudades que solicitaron el uso de esta tecnología: Róterdam, Eindhoven, Haarlem y Capelle aan den IJssel. Se puede ver de forma ilustrativa el funcionamiento de este sistema, con base en el caso de Haarlem. Disponible en: <<https://algorithmwatch.org/en/story/high-risk-citizens/>>.

28. Al respecto, consultar, entre otras: ECtHR 27 octubre 1995, no. 20190/92 (*C.R. vs. the United Kingdom*), para. 42 y ECtHR 29 abril 2002, no. 2346/02 (*Pretty vs. the United Kingdom*), para. 65.

29. ECtHR 4 diciembre 2008, nos. 30562/04 y 30566/04 (*S. and Marper vs. the United Kingdom*), para. 66.

30. El principio de transparencia exige «la accesibilidad de la información y el fácil entendimiento de esta, comunicación y lenguaje claro, y provisión de información al interesado sobre la identidad del controlador y los propósitos del procesamiento de datos. Además, requiere el suministro de información adicional de manera activa para garantizar un procesamiento de datos sólido y transparente, siendo las personas físicas conscientes de los riesgos, reglas, salvaguardas y derechos en relación con el procesamiento de datos personales y también de cómo pueden ejercer sus derechos con respecto al procesamiento» (6.31).

31. El principio de limitación de propósito alude a que «los datos personales deben ser recopilados para fines específicos, explícitos y legítimos y no procesarse de forma posterior de un modo incompatible con dichos fines» (6.32).

32. El principio de minimización de datos requiere que «los datos personales sean adecuados, relevantes y limitados a lo necesario en relación con los fines para los que se procesan». Además, el tiempo de conservación de estos datos personales no debe exceder del tiempo necesario para el propósito para el que se procesan los mismos (6.33).

33. De acuerdo con el principio de precisión «han de tomarse todas las medidas razonables para garantizar que los datos personales que sean inexactos, con base en los fines para los que se procesan, sean borrados o modificados sin demora» (6.34).

34. El principio de integridad y confidencialidad exige el «procesamiento de los datos personales mediante técnicas que garanticen su seguridad» (6.34).

35. Refiere a la obligación de cumplimiento con los principios anteriores (6.34).

36. Skansi (2018).

37. Han, Pei y Kamber (2011).

38. *Parliamentary Papers II* 2017/18, 26643, 557.

39. Alston (2019). Disponible en: <<https://www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf>>.

40. *Ibidem*: 12.

41. Las partes demandantes argumentaron su postura conforme con las declaraciones del TEDH en el citado caso, entendiendo que la base legal de este sistema debería ser lo suficientemente clara como para permitir que un individuo regule su conducta en consecuencia.

42. STEDH, n.º 6538/74, 26.04.1979. En este famoso proceso, conocido como «caso de la talidomida», el TEDH confirmó que la prohibición de la publicación de un artículo del periódico *Sunday Times*, a petición de la sociedad farmacéutica británica Dsitillers, violaba el art. 10 del CEDH, pues la interferencia no correspondía a intereses públicos. Concretamente, en este artículo se cuestionaban los acuerdos que se estaban negociando entre los padres y la farmacéutica respecto de las malformaciones neonatales sufridas por la ingesta durante el embarazo de un medicamento con efectos sedantes que contenía talidomida. Esto es, el Tribunal estimó que dicha interferencia no cumplía con el estándar de necesidad en una sociedad democrática, dado que la misma «no correspondía con una necesidad social lo suficientemente urgente para pesar más que el interés público de libertad de expresión dentro del significado de la Convención».

43. STEDH, n.º 30562/04 y 30566/04, 04.12.2008.

44. González Fuster (2009: 620).

45. Alston (2019: 10).

46. Para un conciso análisis de estas herramientas, Castro Liñares (2019: 105-253).

47. Martínez Garay (2018: 488).

48. Dix (1975: 327-344).

49. Los problemas de estos instrumentos en su aplicación al terreno judicial han sido puestos de manifiesto desde hace décadas por Vives Antón (1974: 410 ss.).

50. De hecho, como subraya Martínez Garay (2014:16-17), en Alemania la doctrina mayoritaria y la jurisprudencia del BGH consideran que este tipo de predicciones no pueden ser nunca suficientes para decidir sobre la peligrosidad del sujeto, siendo necesario un estudio completo e individualizado del individuo en cuestión. No en balde, como destacó el Parlamento Europeo en el año 2017, si bien respecto de la utilización de modelos de actuación policial predictiva es radicalmente distinto, la realización de predicciones probabilísticas sobre lugares o acontecimientos, de la perfilación individual o predicción que se efectúa sobre individuos particulares (Miró Llinares, 2018: 128).

51. Solar Cayón (2020a:130).

52. Martínez Garay (2018:488).

53. Tras su última revisión, se establece que las *Sentencing Commissions* «deben» desarrollar instrumentos de valoración de riesgos y de necesidades criminógenas, y que cuando estos demuestren ser suficientemente fiables, «pueden» incorporarlos en las *sentencing guidelines* (Martínez Garay, 2018: 489).

54. Goel, Shroff, Skeem y Slobogin (2018: 2); y Hogan-Doran (2017: 1-39).

55. Ægisdóttir, White, Spengler, Maugherman, Anderson, Cook, Nichols, Lampro-poulos, Walker, Cohen y Rush (2006: 341-382).

56. Goel, Shroff, Skeem y Slobogin (2018: 2 ss.).

57. Kleinberg, Lakkaraju, Leskovec, Ludwig y Mullainathan (2017).

58. Jung, Concannon, Shroff, Goel y Goldstein (2017).

59. Goel, Shroff, Skeem y Slobogin (2018: 3).

60. Skeem y Monahan (2011: 38-42).

61. Campbell, French y Gendreau (2009: 329-353).
62. Andrés Pueyo y Redondo Illescas (2007: 167).
63. Singh, Grann y Fazel (2011: 501).
64. Por ello, Slobogin (2006: 296) entiende que solamente deberían admitirse los métodos actuariales o de juicio clínico estructurado con base en los protocolos que se siguen. De hecho, según este autor, solamente si el propio acusado consiente en presentar un juicio clínico para acreditar una baja peligrosidad se podría permitir a la acusación su uso para probar lo contrario. Objetan también las limitaciones de los métodos no estructurados Harcourt (2005); y Skeem y Monahan (2011: 39).
65. Interesa destacar en este punto la crítica de Monahan (1995: 32) a los cuatro puntos ciegos de estos instrumentos clínicos: la falta de especificidad al definir el criterio utilizado, la falta de base estadística, la dependencia de correlaciones falsas o una inadecuada valoración de los factores situacionales ambientales.
66. Skeem y Monahan (2011: 39 ss.). De hecho, Martínez Garay (2014: 18) se hace eco de cómo Fazel, Sing, Doll y Grann (2012: 5) concluyeron en su metaanálisis que los instrumentos actuariales no predecían mejor este riesgo que los métodos de juicio clínico estructurado, en contraste con lo que habían afirmado algunos estudios anteriores.
67. Martínez Garay (2014: 1 ss.).
68. El cual nos dice con qué probabilidad, cuando consideramos peligroso a un sujeto, su comportamiento posterior confirmará la predicción.
69. Documentos parlamentarios II, 2019/20, 26643, 641.
70. Asimismo, conviene apuntar, siguiendo a Martínez Garay (2014: 17), que métodos que, en principio, son usados para realizar juicios clínicos estructurados, como es el PCL-R, pueden ser utilizados de forma actuarial, asignando a los ítems puntuaciones numéricas (0/1/2) y usando la puntuación global obtenida por el sujeto para estimar directamente su peligrosidad. Concretamente, el PCL-R proporciona una puntuación general de psicopatía y puntuaciones en dos factores específicos (Interpersonal/Afectivo y Desviación social), los cuales se dividen a la vez en cuatro facetas (Interpersonal, Afectiva, Estilo de vida, Antisocial).
71. Andrews y Bonta (2006); Gendreau, Little y Goggin (1996: 575-607); Grove, Zald, Lebow, Snitz y Nelson (2000: 19-30); Hanson y Morton-Bourgon (2004); y McNeil, Sandberg y Binder (1998: 655-669).
72. Caso *Loomis vs. Wisconsin*, Decisión de 26 de junio de 2017. 881 N.W.2d 749 (Wis. 2016).
73. Romeo Casabona (2018).
74. Martínez Garay (2018: 489 ss.).
75. Dietrich, Mendoza y Brennan (2016). Disponible en: <https://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf>.
76. La primera fue el caso *Malenchik vs. State of Indiana*, 928 N.E.2d 564 (Ind. 2010).
77. Martínez Garay (2018: 494).
78. 430 U.S. 349 (1977).
79. Un análisis detallado de estas cuestiones en Romeo Casabona (2018: 49-51).

80. De hecho, como se explica expresamente en la sentencia, la Guía del Usuario del COMPAS de 2015 de Northpointe, Inc. (*Practitioner's Guide to COMPAS Core*) contiene una explicación de la base de la que se extraen las puntuaciones de los sujetos; concretamente, se basan tanto en información estática (historial criminal) como en diferentes variables dinámicas, entre ellas, por ejemplo, el abuso de sustancias.

81. Romeo Casabona (2018: 51).

82. Caso *Loomis vs. Wisconsin*, Decisión de 26 de junio de 2017. 881 N.W.2d 749 (Wis. 2016).

83. Además de los autores antes citados, se muestran igualmente críticos con esta resolución, cuestionando su compatibilidad con el derecho al debido proceso o la presunción de inocencia, entre muchos otros: Freeman (2016: 106); De Miguel Beriaín (2018: 48); y Lightbourne (2017: 334-337).

84. Martínez Garay (2018: 499).

85. Eckhouse, Lum, Conti-Cook y Ciccolini (2019: 185).

86. Carlson (2017: 329).

87. Hildebrandt (2018: 34).

88. Martínez Garay (2018: 499).

89. Boix Palop (2020: 223 ss.). No obstante, como este autor destaca, desafortunadamente no encontramos ningún tipo de normativa europea que exija la publicación del código de estos algoritmos, de modo que la Administración española está empezando a utilizar herramientas basadas en algoritmos predictivos cuyo funcionamiento no es accesible por la ciudadanía, como ocurre, por ejemplo, con el Sistema de Seguimiento Integral de los casos de Violencia de Género (el famoso algoritmo VioGén).

90. Concretamente, Boix Palop se hace eco de trabajos, como el de Valero Torrijo, que exige, en palabras del mismo «un derecho por parte de los ciudadanos a obtener toda aquella información que permita la identificación de los medios y aplicaciones utilizadas, del órgano bajo cuyo control permanezca el funcionamiento de la aplicación o el sistema de información; debiendo incluir, asimismo, en su objeto no solo el conocimiento del resultado de la aplicación o sistema informático que le afecte específicamente a su círculo de intereses sino, además y sobre todo, el origen de los datos empleados y la naturaleza y el alcance del tratamiento realizado, es decir, cómo el funcionamiento de aquellos puede dar lugar a un determinado resultado». Y, este último aspecto, esto es, el origen de los datos y su alcance, resulta especialmente importante en el terreno forense, debido a las implicaciones jurídico penales que estos pronósticos conllevan. Boix Palop (2020: 260).

91. Yeung y Lodge (2019: 1-17).

92. Hannah-Moffat (2018: 453 ss); Ferguson (2017); Pasquale (2015: 23 ss); Barocas y Selbst (2016); y Sweeney (2013: 44-54).

93. Barocas y Selbst (2016).

94. Angwin, Larson, Mattu y Kirchner (2016).

95. Hannah-Moffat y Struthers Montford (2019).

96. Especialmente interesante, específicamente respecto de la justicia juvenil, *vid.* Goddard y Myers (2017: 151-167).

97. UN WOMEN, *UN Women Ad Series Reveals Widespread Sexism*, (Oct. 21, 2013). Disponible en: <<https://www.unwomen.org/en/news/stories/2013/10/women-should-ads>>.
98. Chander (2016: 1036).
99. Martínez Garay (2018: 496).
100. Coglianese y Lehr (2017).
101. Boix Palop (2020: 248).
102. Con carácter general, Ranchordás y Schuurmans (2020) ponen al descubierto los riesgos del creciente fenómeno de la privatización en las administraciones públicas, específicamente respecto de la subcontratación y el uso de estas herramientas automatizadas. Asimismo, en nuestro país destaca el estudio en el ámbito laboral de Todolí Signes (2020), donde reflexiona sobre los retos legales que plantea el uso de algoritmos en la selección de sujetos a investigar en las inspecciones de trabajo.
103. Se reitera que no se desconoce, en cualquier caso, el carácter disímil de estas, tanto respecto del actor de uso (Administración Tributaria *vs.* Administración de Justicia), como su procedencia (Países Bajos *vs.* EE. UU.), que determina distintos principios informadores.
104. Romeo Casabona (2018: 54).
105. Respecto de las diferencias entre los métodos clínicos y actuariales, entre los que, como previamente se ha destacado, no puede efectuarse una división tajante, *vid.* Brandariz (2016: 95-98).
106. Goel, Shroff, Skeem y Slobogin (2018: 15).
107. Darjee, Russell, Forrest, Milton, Savoie, Baron, Kirkland y Stobie (2016).
108. Andrés-Pueyo y Redondo (2007: 157-173).
109. Romeo Casabona (2018: 55).
110. En este punto, se estima interesante señalar la advertencia de McKay (2020: 35), que, pese a mostrarse partidaria del uso complementario de estos instrumentos, advierte de los riesgos existentes si estas herramientas se convirtiesen en árbitro único, pues, como señala la autora: «At least an imperfect decision by a judge may be tested on appeal, whereas an imperfect algorithm may be forever concealed».
111. Eckhouse, Lum, Conti-Cook y Ciccolini (2019).
112. Carlson (2017) analiza un caso en el que el puntaje que ofreció COMPAS fue tan alto que el juez de sentencia anuló el acuerdo de culpabilidad y sentenció al delincuente a dos años, reconociendo el mismo que, sin esta evaluación de riesgos, solo le habría impuesto una sentencia de un año.
113. Slobogin (2018: 6).
114. Romeo Casabona (2018: 55).
115. Carlson (2017).
116. Para una aproximación a estos algoritmos públicos o de acceso abierto, *vid.* Wexler (2017). Disponible en: <<https://washingtonmonthly.com/magazine/junejulyaugust-2017/code-of-silence/>>; y Martin (2019: 847).
117. Miró Llinares (2018: 114-118).

118. De todos modos, como Solar Cayón (2020b: 26) reconoce respecto del uso de la inteligencia artificial en la Administración de Justicia, es la aplicación de estas herramientas para la toma de decisiones judiciales la que más controversias presenta, pues, pese a los importantes avances: «nos hallamos lejos aún del punto en el que los sistemas inteligentes puedan identificar y, mucho menos aún, realizar, argumentos jurídicos de cierta complejidad».

BIBLIOGRAFÍA

- ÆGISDÓTTIR, Stefania, Michael WHITE, Paul SPENGLER, Alan MAUGHERMAN, Linda ANDERSON, Robert COOK, Cassandra NICHOLS, Georgios LAMPRO-POULOS, Blain WALKER, Genna COHEN y Jeffrey RUSH (2006): «The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction», *The Counseling Psychologist*, 34(3), 341-382.
- ALSTON, Philip (2019): «Amicus brief in the case of NJCM c.s./De Staat der Nederlanden (SyRI): Implications of the use of digital technologies in welfare states» [en línea], <<https://www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf>>. [Consulta: 29/04/2022.]
- AMERICAN LAW INSTITUTE (2007): «Model Penal Code: Sentencing. Proposed Final Draft. §6B.09: Evidence-Based Sentencing; Offender Treatment Needs and Risk of Reoffending» [en línea], <https://robinainstitute.umn.edu/sites/robinainstitute.umn.edu/files/2022-02/mpcs_proposed_final_draft.pdf>. [Consulta: 27/10/2022.]
- ANDRÉS PUEYO, Antonio (2013): «Peligrosidad criminal: análisis crítico de un concepto polisémico», en E. Demetrio Crespo y M. Maroto Calatayud (coords.), *Neurociencias y Derecho Penal. Nuevas perspectivas en el ámbito de la culpabilidad y tratamiento jurídico-penal de la peligrosidad*, Madrid: Edisofer, 483-503.
- ANDRÉS PUEYO, Antonio y Santiago REDONDO ILLESCAS (2007): «Predicción de la violencia. Entre la peligrosidad y la valoración del riesgo de violencia», *Papeles del Psicólogo*, 28(3), 157-173.
- ANDREWS, Donald A. y James BONTA (2006): *The psychology of criminal conduct*, Cincinnati: Anderson.
- ANGWIN, Julia, James LARSON, Surya MATTU y Lauren KIRCHNER (2016): «Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks» [en línea], <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>. [Consulta: 29/04/2022.]
- BARABAS, Chelsea, Karthik DINAKAR, Joichi ITO, Madars VIRZA, y Jonathan ZITTRAIN (2020): «Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment», *Proceedings of Machine Learning Research*, 81, 62-76.
- BAROCAS, Solon y Andrew SELBST (2016): «Big Data's Disparate Impact», *California Law Review*, 104, 671-372.
- BOIX PALOP, Andrés (2020): «Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones», *Revista de Derecho Público: Teoría y Método*, 1, 223-269.
- BRANDARIZ GARCÍA, José Ángel (2016): *El Modelo Gerencial-Actuarial de Penalidad. Eficiencia, riesgo y sistema penal*, Madrid: Dykinson.
- CAMPBELL, Mary Ann, Sheila FRENCH y Paul GENDREAU (2009): «The prediction of violence in adult offenders: a meta-analytic comparison of instruments and methods of assessment», *Criminal Justice & Behavior*, 36(6), 567-590.
- CARLSON, Alyssa (2017): «The Need for Transparency in the Age of Predictive Sentencing Algorithms», *Iowa Law Review*, 103, 303-329.
- CASTRO LIÑARES, David (2019): *Los instrumentos de valoración y gestión de riesgos en el modelo de penalidad español*, Madrid: Reus.

- CHANDER, Anupam (2016): «The Racist Algorithm?», *UC Davis Legal Studies Research*, 498, 1023-1044.
- CHRISTIN, Angèle, Alex ROSENBLAT y Danah BOYD (2015): «Courts and Predictive Algorithms» [en línea], <https://www.law.nyu.edu/sites/default/files/upload_documents/Angele%20Christin.pdf>. [Consulta: 27/10/2022.]
- COGLIANESE, Cary y David LEHR (2017): «Regulating by Robot: Administrative Decision Making in the Machine Learning Era», *Faculty Scholarship at Penn Law*, 1734, 1147-1223.
- CORMEN, Thomas, Charles LEISERSON, Ronald RIVEST y Clifford STEIN, (2009): *Introduction to Algorithms*, Cambridge (Massachusetts): The MIT Press.
- COTINO HUESO, Lorenzo (2020): «Holanda “SyRI, ¿a quién sanciono?” Garantías frente al uso de inteligencia artificial y decisiones automatizadas en el sector público y la sentencia holandesa de febrero de 2020», *La Ley privacidad*, 4, 1-20.
- DARJEE, Rajan; Katharine RUSSELL, Lauren FORREST, Erica MILTON, Valerie SAVOIE, Emily BARON, Jamie KIRKLAND y Stewart STOBIE (2016): «Risk of Sexual Violence Protocol (RSVP): A real world study of the reliability, validity and utility of a structured professional judgement instrument in the assessment and management of sexual offenders in South East Scotland» [en línea], <https://www.researchgate.net/publication/294718597_Risk_of_Sexual_Violence_Protocol_RSVP_A_real_world_study_of_the_reliability_validity_and_utility_of_a_structured_professional_judgement_instrument_in_the_assessment_and_management_of_sexual_offenders_>. [Consulta: 27/10/2022.]
- DE MIGUEL BERIAÍN, Iñigo (2018): «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*, 17(1), 45-53.
- DIX, George (1975): «Determining the continued dangerousness of psychologically abnormal sex offenders», *Journal of Psychiatry and the Law*, 3, 327-344.
- ECKHOUSE, Laurel, Kristian LUM, Cynthia CONTI-COOK y Julie CICCOLINI (2019): «Layers of Bias: A Unified Approach for Understanding Problems with Risk Assessment», *Criminal Justice and Behavior*, 46(2), 185-209.
- FAZEL, Seena, Jay SINGH, Helen DOLL y Martin GRANN (2012): «Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis», *BMJ*, 345-469.
- FERGUSON, Andrew Guthrie (2017): *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York: New York University Press.
- FREEMAN, Katherine (2016): «Algorithmic Injustice: how the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis», *North Carolina Journal of Law & Technology*, 18, 75-106.
- GARCÍA RIVAS, Nicolás (2013): «La libertad vigilada y el Derecho Penal de la seguridad», en E. Demetrio Crespo y M. Maroto Calatayud (coords.), *Neurociencias y Derecho Penal. Nuevas perspectivas en el ámbito de la culpabilidad y tratamiento jurídico-penal de la peligrosidad*, Madrid: Edisofer, 601-628.
- GENDREAU, Paul, Tracy LITTLE y Claire GOGGIN (1996): «A metaanalysis of the predictors of adult offender recidivism: What works!», *Criminology*, 34, 575-607.
- GODDARD, Tim y Randolph MYERS (2017): «Against evidence-based oppression: Marginalized youth and the politics of risk-based assessment and intervention», *Theoretical Criminology*, 21(2), 151-167.
- GOEL, Sharad, Ravi SHROFF, Jennifer SKEEM y Christopher SLOBOGIN (2018): «The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment» [en línea], <<https://ssrn.com/abstract=3306723>>. [Consulta: 29/04/2022.]
- GONZÁLEZ FUSTER, Gloria (2009): «TEDH – Sentencia de 04.12.2008, S. y Marper c. Reino Unido, 30562/04 y 30566/04 – Artículo 8 CEDH – vida privada – injerencia en una sociedad democrática – los límites del tratamiento de datos biométricos de personas no condenadas», *Revista de Derecho Comunitario Europeo*, 33, 619-633.

- GROVE, William, David ZALD, Boyd LEBOW, Beth SNITZ y Chad NELSON (2000): «Clinical versus mechanical prediction: A meta-analysis», *Psychological Assessment*, 12, 19-30.
- HAN, Jiawei, Micheline PEI y Jian KAMBER (2011): *Data mining: concepts and techniques*, Massachusetts: Elsevier.
- HANNAH-MOFFAT, Kelly (2018): «Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates», *Theoretical Criminology*, 23(4), 453-470.
- HANNAH-MOFFAT, Kelly y Kelly STRUTHERS MONTFORD (2019): «Unpacking Sentencing Algorithms Risk, Racial Accountability and Data Harms», en J. W. de Keijser, J. V. Roberts y J. Ryberg, *Predictive Sentencing Normative and Empirical Perspectives*, Londres: Hart Publishing, 87-103.
- HANSON, Karl y Kelly MORTON-BOURGON (2004): «Predictors of sexual reidivism: An updated meta-analysis» [en línea], < <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/2004-02-prdctrs-sxl-rcdvsm-pdtd/index-en.aspx>>. [Consulta 27/10/2022.]
- HARCOURT, Bernard (2005): «Against Prediction: Sentencing, Policing, and Punishing in an Actuarial Age», *University of Chicago Public Law & Legal Theory Working Paper*, 94, 1-43.
- HILDEBRANDT, Mireille (2018): «Law as computation in the era of artificial legal intelligence: Speaking law to the power of statistics», *University of Toronto Law Journal*, 68(1), 12-35.
- HOGAN-DORAN, Dominique (2017): «Computer says “no”: Automation, algorithms and artificial intelligence in Government decision-making», *Judicial Review*, 13, 1-39.
- JUNG, Jongbin, Connor CONCANNON, Ravi SHROFF, Sharad GOEL y Daniel GOLDSTEIN (2017): «Simple rules for complex decisions», *Stanford University Working Paper*, Feb-17, 1-9.
- KEINBERG, Jon, Himabindu LAKKARAJU, Jure LESKOVEC, Jens LUDWIG y Sendhil MULLAI-NATHAN (2017): «Human Decisions and Machine Predictions», *NBER Working Paper*, 23180, 1-76.
- LAZCOZ MORATINOS, Guillermo y José Antonio CASTILLO PARRILLA (2020): «Valoración algorítmica ante los derechos humanos y el Reglamento General de Protección de Datos: El caso SyRI», *Revista Chilena de Derecho y Tecnología*, 1(1), 207-225.
- LIGHTBOURNE, John (2017): «Damned lies & criminal sentencing using evidence-based tools», *Duke Law & Technology Review*, 15, 327-343.
- LOINAZ, Ismael (2017): *Manual de evaluación del riesgo de violencia. Metodología y ámbitos de aplicación*, Madrid: Pirámide.
- FEELEY, Malcolm y Jonathan SIMON (1992): «The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications», *Criminology*, 30(4), 449-474.
- MARTIN, Kristen (2019): «Ethical Implications and Accountability of Algorithms», *Journal of Business Ethics*, 160, 835-850.
- MARTÍNEZ GARAY, Lucía (2014a): «La incertidumbre de los pronósticos de peligrosidad: consecuencias para la dogmática de las medidas de seguridad», *InDret*, 2, 1-77.
- (2014b): «Minority Report: pre-crimen y pre-castigo, prevención y predicción», en T. Vives Antón, J. C. Carbonell Mateu, J. L. González Cussac, A. Alonso Rimo y M. Roig Torres (coords.), *Crímenes y Castigos. Miradas al Derecho Penal a través del arte y la cultura*, Valencia: Tirant lo Blanch, 579-606.
- (2018): «Peligrosidad, algoritmos y due process: el caso *State vs. Loomis*», *Revista de Derecho Penal y Criminología*, 20, 485-502.
- MCKAY, Carolyn (2020): «Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making», *Current Issues in Criminal Justice*, 32(1), 22-39.
- MCNEIL, Dale, David SANDBERG y Renée BINDER (1998): «The relationship between confidence and accuracy in clinical assessment of psychiatric patients' potential for violence», *Law and Human Behavior*, 22, 655-669.
- MIRÓ LLINARES, Fernando (2018): «Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots», *RDPC*, 20, 87-130.
- MONAHAN, John (1995): *The clinical prediction of violent behavior*, Northvale: Jason Aronson Inc.
- NIEVA FENOLL, Jordi (2018): *Inteligencia artificial y proceso judicial*, Madrid: Marcial Pons.

- OLVER, Mark, Keira STOCKDALE y Stephen WORMITH (2009): «Risk assessment with young offenders: A meta-analysis of three assessment measures», *Criminal Justice and Behavior*, 36(4), 329-353.
- PASQUALE, Frank (2015): *The Black Box Society: The Secret Algorithms That Control Money and Information*, Massachusetts: Harvard University Press.
- RANCHORDÁS, Sofia y Imre SCHUURMANS (2020): «Outsourcing the Welfare State: The Role of Private Actors in Welfare Fraud Investigations», *European Journal of Comparative Law and Governance*, 7, 5-42.
- ROMEO CASABONA, Carlos María (2018): «Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad», *REDS*, 13, 39-55.
- SAGIROGLU, Seref y Duygu SINANC (2013): «Big data: A review» [en línea], <<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6567202>>. [Consulta 01/11/2022.]
- SINGH, Jay, Martin GRANN y Seena FAZEL (2011): «A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25.980 participants», *Clinical Psychology Review*, 31, 499-513.
- SKANSI, Sandro (2018): *Introduction to Deep Learning from Logical Calculus to Artificial Intelligence*, Londres: Springer.
- SKEEM, Jennifer y John MONAHAN (2011): «Current directions in violence risk assessment», *Current Directions in Psychological Science*, 20(1), 38-42.
- SLOBOGIN, Christopher (2006): «Dangerousness and Expertise Redux», *Emory Law Journal*, 275, 1-51.
- (2018): «A Defense of Modern Risk-Based Sentencing», *Vanderbilt Law Research*, 18(52), 1-23.
- SOLAR CAYÓN, José Ignacio (2020a): *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, Madrid: Servicio de Publicaciones de la Universidad de Alcalá.
- (2020b): «La inteligencia artificial jurídica: nuevas herramientas y perspectivas metodológicas para el jurista», *Revus: Journal for Constitutional Theory & Philosophy of Law*, 41, 1-27.
- SWEENEY, Latania (2013): «Discrimination in Online Ad Delivery», *Communications of the ACM*, 56(5), 44-54.
- TEGMARK, Max (2017): *Life 3.0. Being human in the age of Artificial Intelligence*, Londres: Penguin Books.
- TODOLÍ SIGNES, Adrián (2020): «Retos legais do uso do big data na selección de suxeitos a investigar pola Inspección de Traballo e da Seguridade Social», *REGAP: Revista galega de administración pública*, 1(59), 9-102.
- UN Women (2013): «UN Women Ad Series Reveals Widespread Sexism» [en línea], <<https://www.unwomen.org/en/news/stories/2013/10/women-should-ads>>. [Consulta: 29/04/2022.]
- U.S. Congress (2018) «H.R.5682 - FIRST STEP Act 115th Congress (2017-2018)» [en línea], <<https://www.congress.gov/bill/115th-congress/house-bill/5682>>. [Consulta: 29/04/2022.]
- URRUELA MORA, Asier (2015): «¿Hacia un cambio de paradigma? la configuración de un derecho penal de la peligrosidad mediante la progresiva introducción de medidas de seguridad aplicables a sujetos imputables en las recientes reformas penales españolas», *CDPC*, 115, 119-160.
- (2018): «Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad», *Revista Penal*, 42, 165-179.
- VIVES ANTÓN, Tomás S. (1974): «Métodos de determinación de la peligrosidad», en J. M. Canivell (dir.), *Peligrosidad social y medidas de seguridad*, Valencia: Universidad de Valencia, 389-417.
- WEXLER, Rebecca (2017): «Code of Silence. How private companies hide flaws in the software that governments use to decide who goes to prison and who gets out» [en línea], <<https://washington-monthly.com/magazine/junejulyaugust-2017/code-of-silence/>>. [Consulta: 29/04/2022.]
- YEUNG, Karen y Martin LODGE (2019): «Algorithmic regulation. An Introduction», en K. Yeung y M. Lodge, *Algorithmic Regulation*, Oxford: Oxford University Press, 1-17.

JURISPRUDENCIA

Caso *Loomis vs. Wisconsin*, Decisión de 26 de junio de 2017. 881 N.W.2d 749 (Wis. 2016).
 Caso *Malenchik v. State of Indiana*, 928 N.E.2d 564 (Ind. 2010).
 STEDH, caso *S. y Marper vs. Reino Unido*, n.º 30562/04 y 30566/04, 4 de diciembre de 2008.
 STEDH, caso *Sunday Times*, n.º 6538/74, 26 de abril de 1979.
 STEDH, caso *C.R. vs. Reino Unido*, n.º 20190/92, 27 de octubre de 1995.
 STEDH, caso *Pretty vs. Reino Unido*, n.º 2346/02, 29 de abril de 2002.
 C/09/550982 / HA ZA 18-388, [en línea], <<https://uitspraken.rechtspraak.nl/inziendocument?id=EC LI:NL:RBDHA:2020:1878>>[Consulta: 29/04/2022.]

VIDEOGRAFÍA

Hielscher, Matthias y Monika Heeder (2017): *Pre-Crime* [Documental], Alemania: Gravitas Ventures.

LEGISLACIÓN

Ley Orgánica 2/2015, de 30 de marzo, por la que se modifica la L.O. 10/1995, de 23 de noviembre, del Código Penal [en línea], <BOE núm. 77, de 31-03-2015>.
 Ley Orgánica 5/2010, de 22 de junio, por la que se modifica la Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal [en línea], <BOE núm. 152, de 23-06-2010>.

Fecha de recepción: 28 de abril de 2022.

Fecha de aceptación: 28 de septiembre de 2022.