IUS ET SCIENTIA
2025 • Vol. 11 • Nº 1 • ISSN 2444-8478
https://editorial.us.es/es/revistas/ius-et-scientia
https://dx.doi.org/10.12795/IETSCIENTIA • © Editorial Universidad de Sevilla 2025

Herramientas de predicción criminal y su impacto en nuestro proceso penal*

CRIMINAL PREDICTION TOOLS AND THEIR IMPACT ON THE SPANISH CRIMINAL PROCESS

Elisabet Cueto Santa Eugenia
Universidad Pontificia Comillas (ICADE)
ecueto@comillas.edu 0000-0001-5336-2003

Recibido: 22 de abril de 2025 | Aceptado: 16 de junio de 2025

RESUMEN

El presente trabajo analiza el uso de herramientas de predicción criminal, prestando especial atención a sus implicaciones éticas y jurídicas. En una primera parte, se expone el origen y desarrollo de estas tecnologías en el contexto anglosajón, especialmente el uso del *profiling* y el *predictive policing*, y se examinan los riesgos inherentes a su aplicación, como la posible reproducción de sesgos discriminatorios y la afectación desproporcionada a colectivos vulnerables. Asimismo, se aborda la cuestión de la transparencia en el funcionamiento de los algoritmos y el tratamiento de datos personales. En una segunda parte, el estudio se centra en el contexto español, analizando el uso de dos sistemas predictivos concretos: el sistema VioGén, orientado a la evaluación del riesgo en casos de violencia de género, y RisCanvi, utilizado en el ámbito penitenciario catalán para valorar la peligrosidad de las personas privadas de libertad. A través del examen de estos ejemplos, se reflexiona sobre la tensión entre la utilidad preventiva de estas herramientas y la necesidad de salvaguardar los derechos fundamentales en el marco del proceso penal.

PALABRAS CLAVE

Herramientas predictivas
Proceso Penal
Garantías
Inteligencia Artificial

ABSTRACT

This paper analyzes the use of crimiminal prediction tools, focusing on their ethical and legal implications. The first part tackles origin and development of these technologies in the anglo-saxon context, highlighting the use of profiling and predictive policing, and examines the risks inherent in their use, such as the potential reproduction of discriminatory biases or the disproportionate impact on vulnerable

KEYWORDS

Predictive Tools
Criminal Proceedings
Guarantees
Artificial Intelligence

* El presente trabajo forma parte del proyecto de I+D+i "Neuro-Derechos Humanos y Derecho Penal" (PID2023-149978NB-I00), financiado por MICIU/AEI/10.13039/501100011033 y por FEDER/ UE.

groups. It also addresses the issue of transparency in the way the algorithm operates and the processing of personal data. In the second part, the paper focuses on the Spanish context, analyzing the use of two specific predictive systems: the VioGén system, aimed at risk assessment in cases of gender-based violence, and RisCanvi, used in the Catalan penitentiary system to assess the dangerousness of inmates. Through the examination of these examples, the paper reflects on the tension between the preventive usefulness of these tools and the need to safeguard fundamental rights within the framework of criminal proceedings.

1. Herramientas de predicción criminal

El empleo de herramientas algorítmicas para predecir el riesgo penal ha ido cobrando relevancia y volviéndose especialmente popular en los últimos años. Estas herramientas, basadas en modelos matemáticos y técnicas de aprendizaje automático, buscan asistir a los distintos operadores jurídicos a la hora de tomar decisiones relativas a medidas cautelares, sentencias condenatorias o permisos de las personas que ya están privadas de libertad.

Estos sistemas a priori presentan una apariencia de objetividad, debido a que son capaces de procesar un gran volumen de datos y cabría pensar que podrían llegar a superar las limitaciones cognitivas del ser humano. Sin embargo, su implementación también ha suscitado debates sobre la transparencia y la fiabilidad de sus predicciones. De hecho, hay quien señala que, lejos de eliminar la discriminación, los algoritmos pueden perpetuar desigualdades existentes si se entrenan con datos históricos sesgados o si sus criterios de evaluación no son adecuadamente supervisados.

A lo largo del presente trabajo, nos dedicaremos a analizar el origen del empleo de estas herramientas, su modus operandi, las ventajas y las limitaciones que presentan, así como el uso que ya ha comenzado a hacerse de las mismas en el sistema penal español.

1.1. Origen y ejemplo anglosajón de uso de *profiling* y *predictive policing*

De cara a establecer políticas criminales tales como la distribución de patrullas de cuerpos del estado o la actuación de la policía en algunas circunstancias, en algunos lugares como Estados Unidos resulta habitual llevar a cabo una serie de análisis de las zonas geográficas y tipos de sujetos que las habitan[1], así como un estudio pormenorizado de las características de los sujetos con más probabilidad de cometer hechos tipificados como delito. Esto, muy a menudo conlleva terminar llevando a cabo lo que ellos denominan *"criminal profiling",* que no es otra cosa que elaborar listados de perfiles de personas que se considera que reúnen las condiciones como para llevar a cabo conductas antisociales o delictivas. El objetivo de este tipo de prácticas es optimizar los recursos y esfuerzos y prevenir posibles delitos, pero lo cierto es que conllevan el

1. Ejemplos notorios de esto son las teorías de la desorganización social de la escuela de Chicago. Al respecto, *vid.* Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology, 94*(4), pp. 774 y ss.

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

riesgo innegable de criminalizar colectivos que en principio deberían ser considerados vulnerables.

En este sentido cabe mencionar un fenómeno denominado *widening the net* (ensanchar la red), que consiste en una crítica a ciertas políticas que, si bien fueron diseñadas para para prevenir o combatir la delincuencia terminan derivando en que el número de personas bajo el radar del sistema penal crezca o se ensanche con sujetos que no necesariamente hayan realizado conductas tipificadas, sino que sencillamente hayan actuado de forma socialmente reprobable. Históricamente, la práctica de ampliar la red era un problema exclusivo de las autoridades de justicia juvenil, que anotaban y vigilaban perfiles de menores que estaban en lugares públicos en horario escolar o que merodeaban zonas consideradas focos de delincuencia –esto evidentemente granjeó muchas críticas al sistema porque implicaba la estigmatización de sujetos jóvenes y habitualmente vulnerables–. Con el paso del tiempo, lamentablemente, en vez de suprimirse estas prácticas de raíz, lo que sucedió es que se extendieron también a los adultos y el *criminal profiling* se da en muchos casos con fines preventivos respecto de sujetos que aún no han cometido delito alguno (Levinson, 2002).

En general, este tipo de herramientas responde a lo que los anglosajones denominan *preventive justice* (justicia preventiva*)*, que consiste en anticiparse a la comisión del ilícito para imponer medidas o establecer políticas. A pesar de que este tipo de políticas acostumbran a probarse útiles, lo cierto es que en ocasiones las políticas preventivas resultan demasiado intrusivas, cayendo en el riesgo de criminalizar una conducta antes de que ésta cause daño (Zedner, 2007).

De las prácticas del derecho anglosajón que consiste en crear perfiles y catalogar sujetos, una de las más habituales es lo que ellos denominan *"predictive policing".* Esta estrategia se basa en un análisis de datos que tiene el propósito de anticipar y prevenir la comisión de delitos Para ello se tienen en cuenta grandes cantidades de datos históricos, tales como informes policiales, patrones de comportamiento criminal, datos demográficos y geográficos, etc., con el fin de identificar zonas y personas con mayor probabilidad de verse involucradas en actividades delictivas (Susser, 2021).

Esta forma de identificar zonas y sujetos con apariencia de ir a delinquir presenta conflictos éticos que se ven a simple vista. En general, el origen del *"predictive policing"* buscaba ubicar *"hot spots"* o *puntos calientes* donde habitualmente se cometían delitos y ejercer un control de esas zonas enviando activos policiales para prevenir el crimen y disuadir a la ciudadanía de llevar a cabo conductas tipificadas. El problema –y las críticas sociales comprensibles y correspondientes– surgieron cuando se comenzaron a extrapolar criterios a "zonas vecinas que podrían contagiarse" y se empezó a difuminar la frontera entre prevenir y criminalizar a colectivos y barrios que presentaban características de vulnerabilidad (The Law Society of England and Wales, 2019).

De cara a analizar cómo funciona el *predictive policing* es imprescindible tener en cuenta que los patrones predictivos reflejan eventos que necesariamente tienen que seguir reglas predecibles. Por definición, las conductas que no sigan un patrón no podrán ser detectadas por medio de un algoritmo basado en patrones (Pérez Salazar, 2024). Esto es relevante a los efectos del presente estudio porque pone de manifiesto por un lado la necesidad de que las herramientas de las que posteriormente hablaremos dispongan de gran cantidad de información relativa a eventos pasados de cara

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

a afinar sus predicciones, pero también subraya la relevancia de revisión de circunstancias y la posibilidad de errar en las predicciones cuando hay un evento que se sale de los patrones establecidos y que puede resultar en un brote de violencia no predecible por una herramienta algorítmica.

En estrecha relación con el *predictive policing* –en realidad podríamos incluso afirmar que respondiendo a este– se han desarrollado teorías y estudios bajo el paraguas de lo que la doctrina anglosajona llama *crime prevention through enviromental design* (prevención de delitos mediante diseño ambiental), que no es otra cosa que un intento de intentar prevenir la comisión de delitos por medio de diseño arquitectónico de barrios y ciudades. En esta línea lo que se busca es diseñar espacios públicos en las ciudades en los que resulte difícil cometer delitos, ya sea porque son lugares en los que uno puede ser observado desde muchos prismas distintos, porque lugares en los que se logra un sentido de pertenencia de quienes están próximos –por ejemplo una plaza en el medio de una serie de edificios, que, por más que sea pública, los habitantes de dichos edificios perciben como propia– o por el modo en que se plantean actividades en los espacios, para promover su cuidado y buen mantenimiento (Reynald, 2011).

En general podemos afirmar que el *predictive policing* en sí mismo no es nocivo y de hecho sirve para dar respuestas a las necesidades comunitarias y promover la seguridad en los espacios públicos. No obstante, en algunas ocasiones puede resultar peligroso. Ejemplo de esto es que en ocasiones empleando técnicas de *predictive policing,* algunos entes públicos como la policía pueden disponer de listas de sujetos cuyo perfil entra dentro de los parámetros de peligrosidad suficiente como para cometer un delito. Esta manera de proceder, analizando a personas sin tener un fundamento empírico basado en algo distinto de su perfil, es evidentemente criticable desde el punto de vista de las garantías contra la discriminación, tal como analizaremos en el apartado siguiente. Cuestiones tan delicadas como presentar características que a todas luces demuestran cierta vulnerabilidad –provenir de una posición socioeconómica precaria, estar racializado o pertenecer a alguna minoría, padecer trastornos mentales, etc.– no deberían de bastar para asentar criterios de peligrosidad que definan políticas públicas o criterios de justicia (Daviera *et. al*., 2023).

El objetivo principal del *predictive policing* es ayudar a las fuerzas de seguridad a optimizar sus recursos y tomar decisiones informadas sobre la distribución de personal, la asignación de recursos, y la implementación de estrategias preventivas en áreas específicas. No obstante, existe el riesgo de convertir una herramienta útil en una imposición sesgada de restricciones a colectivos que precisarían cierta protección debido a sus características especialmente complejas. En este sentido, hay autores que consideran que la única manera de imponer medidas y soluciones justas y proporcionadas pasa por la necesidad de abolir las técnicas de *predictive policing* o, como mínimo, mitigarlas (Tonry, 2019).

En los Estados Unidos las técnicas de *predictive policing* se emplean a menudo debido a que muchas de sus políticas se articulan en base a la prevención, partiendo de la base de que hay sujetos que pueden resultar peligrosos y es necesario tenerlos ubicados (Robinson, 2001). En este sentido, es necesario manifestar que esta manera de diseñar herramientas policiales y jurídicas entraña peligro. Esto queda ilustrado, por ejemplo, en *Rummel v. Estelle*, en el que el acusado cometió un fraude por valor de 129,75 dólares

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

y, dado que era su tercera condena del mismo tipo –todas ellas delitos que podrían ser considerados de bagatela–, se le aplicó la cadena perpetua en virtud de lo que en Estados Unidos se conoce como la regla de los "tres strikes" –que básicamente implica que si cometes el mismo delito tres veces se te considera un sujeto peligroso y esta circunstancia se emplea como agravante–. El presente caso, si bien no se corresponde con el uso de una herramienta algorítmica, sirve para poner de manifiesto la necesidad de revisión de los protocolos establecidos. En concreto, a pesar de responder a un ilícito cometido y probado, puede mostrarse como claramente desproporcional, ya que ninguno de los tres delitos cometidos por el acusado tenía carácter grave o violento. Así, se pone en relieve que la respuesta legal establecida para la repetición de un delito puede resultar notoriamente desproporcional y quizá debería ser revisada y supervisada.

Un ejemplo claro de que las herramientas algorítmicas deben supervisarse es el caso *State v. Loomis.* En ese supuesto concreto, se llevó a cabo una previsión algorítmica del riesgo de que un sujeto reincidiese, y la discusión versó acerca de si emplear este mecanismo contravenía derechos fundamentales del sujeto o no. El caso es relevante, porque saca a la luz cuestiones que hay que tener en cuenta siempre que trabajemos con mecanismos predictivos. De este modo, si bien se consideró que el hecho de emplear un algoritmo para la predicción del riesgo no contravenía los derechos de Loomis, también subrayó que el hecho de que el procedimiento algorítmico resultase opaco –es decir, que no se publicase ni pudiese ver el modo en que el algoritmo funciona y llega hasta la respuesta que arroja– plantea importantes dudas sobre su precisión y fiabilidad. Estos problemas relacionados con la transparencia del sistema dificultan la capacidad para que los jueces analicen de forma crítica si la evaluación del riesgo es certera, y dejan patente la necesidad de que un ser humano revise el resultado de la máquina y tome la decisión final –pudiendo esta diferir de la que arroja el algoritmo–.

En resumen, quedan patentes los riesgos inherentes a la aplicación de herramientas de predicción en el ámbito penal, especialmente cuando su diseño y uso no consideran de manera suficiente los principios de proporcionalidad y no discriminación. Aunque la intención de prevenir delitos y optimizar recursos es razonable, lo cierto es que estas herramientas en ocasiones pueden derivar en prácticas que refuerzan desigualdades estructurales y perpetúan estereotipos sobre determinados colectivos. La experiencia estadounidense sugiere que el uso indiscriminado de este tipo de técnicas puede dar lugar a problemas notorios. El apartado siguiente pretende ocuparse de los riesgos de que las herramientas de predicción caigan en la discriminación, valorando hasta qué punto pueden ajustarse a un modelo de justicia que garantice una equidad efectiva.

1.2. Riesgos de caer en la discriminación

Antes de analizar los riesgos que pueden llegar a plantear las herramientas predictivas en el ámbito penal, es importante subrayar que el juicio humano tampoco está exento de sesgos. Es decir, a pesar de que la función jurisdiccional pretende realizarse de manera imparcial y justa, lo cierto es que los operadores que intervienen a diario en nuestro sistema de justicia penal –jueces, fiscales, policías, funcionarios penitenciarios, etc. –pueden tener prejuicios inconscientes relacionados con el género, el origen étnico,

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

la edad, la condición socioeconómica o incluso la apariencia física que lleguen a condicionar sus decisiones. Esto es importante tenerlo en cuenta porque uno de los argumentos esgrimidos a favor de la utilización de algoritmos en estos ámbitos es precisamente la concepción de que las máquinas no sienten. No obstante, lo cierto es que el uso de herramientas predictivas en el ámbito penal plantea un peligro significativo de discriminación, ya que su funcionamiento se basa en datos históricos que pueden reflejar sesgos sistémicos preexistentes (Mckay, 2020). Si las fuerzas de seguridad y los sistemas judiciales han operado históricamente con prejuicios raciales, socioeconómicos o de otro tipo, los algoritmos que alimentan estas herramientas tienden a replicar y amplificar esas desigualdades. Esto puede traducirse en una vigilancia desproporcionada sobre determinados grupos, reforzando estereotipos y perpetuando la criminalización de sectores vulnerables, como minorías étnicas, personas de bajos recursos o individuos con trastornos mentales. Es decir, que en ocasiones el algoritmo puede alimentarse de datos históricos que están sesgados y reflejan desigualdades sociales preexistentes, y perpetuar dichas desigualdades (Castagnedi Ramírez, 2024).

En este sentido, hay autores que reflexionan acerca de quién toma la decisión final cuando se emplean este tipo de herramientas y analizan desde todos los sujetos desde los cuales pueden provenir los sesgos: el programador, el legislador que diseña la política y el modo de emplear la herramienta algorítmica o la máquina en sí al nutrirse de datos que en sí mismos pueden estar sesgados (Hogan Doran, 2017).

En relación los posibles sesgos preexistentes y los riesgos que estos entrañan, resulta necesario mencionar que el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (en adelante Reglamento de Inteligencia Artificial) aborda de forma expresa en su párrafo 42 la prohibición de juzgar a las personas a partir de comportamientos predichos con IA que tengan base en la creación de perfiles. Esto es relevante porque se relaciona de forma directa con la presunción de inocencia y es un intento evidente de salvaguardar el derecho a la no discriminación de los sujetos. Se establece, de forma directa la prohibición de evaluar el resigo basada en características y rasgos de la personalidad.

Teniendo en cuenta la peligrosidad que pueden llegar a entrañar estas herramientas, queda clara la necesidad de establecer una serie de reglas o parámetros para procurar mantener tantas garantías como sea posible (Suárez Xavier, 2024). Así, es importante que quede establecido que los sistemas algorítmicos predictivos no tomen decisiones que tengan un componente de discrecionalidad –es decir, que si le damos a una máquina el poder de tomar decisiones estas sean de carácter automatizable–, en caso de hallarnos frente a decisiones que puedan colisionar con derechos fundamentales de las personas implicadas –como es el caso que nos ocupa, al estar tratando cuestiones relativas al proceso penal– lo ideal sería que los algoritmos no tomen decisiones, sino que sirvan como herramienta que apoye la función jurisdiccional, ejercida por un ser humano que supervise el buen hacer de la herramienta. En este sentido, también sería deseable que pudiera verse el modo en que el algoritmo llega al resultado, para poder seguir su línea de funcionamiento lógica y, en caso de hallarnos ante algún sesgo o error, poder corregirlo (Hogan Doran, 2017).

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

En general, cuando se emplean herramientas predictivas en el sistema penal una de las principales cuestiones a tener en cuenta es la búsqueda de transparencia. Esto es así porque el hecho de comprender por qué un algoritmo ofrece un resultado determinado, de un modo que se pueda someter a escrutinio la totalidad del proceso, puede permitir detectar posibles sesgos y corregir inexactitudes para evitar que perjudiquen de forma injusta a grupos especialmente vulnerables. Cuando el algoritmo es opaco o de *"caja negra"* –es decir, aquellos casos en los que el sistema o modelo en el que el funcionamiento interno no es accesible o comprensible para los usuarios o los involucrados en su aplicación–, los sesgos pueden perpetuarse sin posibilidad de corrección ni defensa adecuada por parte de quienes se ven afectados, cuestión innegablemente perjudicial para el sistema de justicia (Carlson, 2017).

La falta de transparencia no solo afecta a la posibilidad de que haya sesgos en el programa, sino que también implica una dificultad notable a la hora de evaluar la exactitud y eficacia de los resultados generados por el algoritmo. Al fin y al cabo, cuando no se tiene constancia del modo en que funciona la herramienta, es imposible seguir los pasos realizados por la misma de cara a comprobar que el sistema no haya cometido ningún error (Wang *et al*, 2022).

En resumen, el empleo de herramientas predictivas en el ámbito penal implica un riesgo claro en relación con la posibilidad de caer en la discriminación. Para evitar esto, resulta imprescindible que su utilización se realice bajo estrictos principios de legalidad y transparencia. Además, es importante que se realice una supervisión por parte de un humano en aquellos casos en los que las decisiones a tomar sean de naturaleza sensible. Las herramientas tecnológicas de esta índole, si bien se han mostrado útiles dada la cantidad de datos que pueden manejar y tener en cuenta, deben sin lugar a duda estar sometidas a un control que garantice su correcto empleo.

1.3. Transparencia y tratamiento de datos personales

Tal como ha quedado patente en los apartados anteriores, la utilización de herramientas algorítmicas para prevenir el riesgo puede conllevar una serie de riesgos tales como caer en sesgos preexistentes y perpetuarlos. Además, el modo en que este tipo de herramientas funciona no siempre es comprensible y esa falta de transparencia puede derivar en que la sociedad no confíe en la IA –es difícil percibir como justa una decisión arrojada por una máquina cuyo proceso no entendemos–. A todas estas cuestiones hay que añadir una que también es clave: los datos que alimentan al algoritmo, debido a la naturaleza de los mismos, a menudo son personales y sensibles. Es por eso que también es necesario abordar el modo en que los datos son tratados.

Resulta necesario detenerse en este punto y recordar alguna regulación existente al respecto. En concreto, el Reglamento (UE) 2016/679, también conocido como el Reglamento General de Protección de Datos (en adelante RGPD), establece en su artículo 15 que la protección de datos de las personas físicas debe ser tecnológicamente neutra y no debe depender de las técnicas utilizadas, haciendo hincapié que en aquellos casos en los que se empleen herramientas algorítmicas automatizadas también habrá que proteger los datos personales.

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

En general, los usuarios de algoritmos que afecten derechos de personas físicas deben informarlas acerca de la lógica del algoritmo de cara a que sus datos se vean protegidos (Castagnedi Ramírez, 2024). Esto es relevante ya que, tal como ha sido previamente abordado, en los casos en los que la herramienta algorítmica funciona con *caja negra* esta circunstancia de información no se da, y eso presenta notorios problemas éticos.

Además, cabe mencionar que el párrafo 71 del RGDP establece el derecho a no ser objeto de una decisión que se base únicamente en el tratamiento automatizado de sus datos y produzca efectos jurídicos en él o le afecte significativamente. Esto tiene especial relevancia en el contexto abordado por el presente trabajo, ya que las herramientas de predicción del riesgo pueden llegar a incidir en cuestiones tan delicadas como la concesión de permisos, grados penitenciarios o la libertad condicional. A tenor de lo establecido por el RGDP, el resultado que arroja el algoritmo en principio no debería ser una decisión, sino que las garantías deberían reforzarse por medio de la intervención humana que supervise el buen hacer de la herramienta y sea quien tenga la última palabra.

Queda patente pues, la necesidad de buscar transparencia en el modo en que lo datos son tratados de cara a que las herramientas respondan a criterios éticos básicos y puedan ser percibidos como justos. En este sentido sería necesario otorgar derechos de transparencia a quienes emplean y supervisan las herramientas, abriéndoles la posibilidad de encontrar la fuente de posibles resultados sesgados (Castagnedi Ramírez, 2024). Este asunto resulta problemático porque en muchas ocasiones quienes desarrollan las herramientas predictivas son empresas tecnológicas privadas que consideran que ofrecer la apertura de la *caja negra* colisiona de forma directa con sus secretos empresariales o *know how.* Es decir, que la opacidad puede llegar a provenir del propio diseño del algoritmo, ya que los desarrolladores técnicos privados quieren proteger su propiedad intelectual sobre el software y el diseño técnico de las herramientas (Zouave y Marquenie, 2017). No obstante, la doctrina coincide en que, dado que este tipo de herramientas ofrece un servicio para el sector público, es inaceptable que falte la transparencia (Borges Blázquez, 2024).

La doctrina ha discutido ampliamente sobre este problema relativo a la *caja negra*, y ha quedado patente que de cara a combatir la falta de transparencia resulta necesario que la *caja* se abra al menos para aquellos que deben tomar la decisión –quizá en este sentido podría promulgarse una regulación de protección de la propiedad intelectual de quienes diseñan el algoritmo pero que permita al órgano jurisdiccional entender el proceso lógico seguido por el sistema y revisarlo de algún modo–. Otra forma de abordar la problemática es combinar la transparencia con auditorías del buen funcionamiento de la herramienta concreta (Kroll, 2015)

El reglamento de Inteligencia Artificial de la UE aborda la cuestión de la transparencia, y en concreto en su párrafo 59 expone que en los casos en los que el sistema de IA no esté entrenado con datos de buena calidad o no sea suficientemente transparente o explicable se corre el riesgo de señale a personas de manera discriminatoria, incorrecta o injusta, pudiendo llegar incluso a impedir el ejercicio de importantes derechos procesales fundamentales, como el derecho a la tutela judicial efectiva y a un juez imparcial, así como el derecho a la defensa y a la presunción de inocencia.

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

En general, puede apreciarse que el uso de herramientas algorítmicas en el sistema penal exige especial atención en relación tanto con la necesidad de proteger datos personales como a la transparencia del modo en que se tratan y emplean esos datos. En este sentido, cabe destacar que la transparencia no debe entenderse como una mera exigencia técnica, sino como un principio esencial para garantizar el control democrático, el escrutinio público y la legitimidad de las decisiones tomadas con apoyo de herramientas algorítmicas.

2. Empleo de herramientas de este tipo en España

Tal como ha quedado patente en el apartado anterior, en los últimos tiempos los sistemas penales se han ido transformando y adaptando a las nuevas tecnologías de un modo que avanza hacia modelos de gestión orientados al riesgo o peligrosidad de los sujetos. Este cambio de enfoque genera un prisma de prevención en el que las herramientas de predicción y control de la peligrosidad cobran relevancia. España no ha resultado ajena a esta tendencia, y aunque la incorporación de este tipo de herramientas no está tan generalizada como en el sistema anglosajón, parece que poco a poco se dispone a seguir ese ejemplo (Pereira Puigvert, 2024).

Puede apreciarse esta tendencia en los casos que analizaremos en el presente apartado. En concreto, cabe destacar el sistema VioGén, que es el sistema de seguimiento integral en los casos de violencia de género, implantado por el ministerio del interior en 2007 y que implica la clasificación de víctimas de violencia de género en distintos niveles de riesgo por medio de un algoritmo, de cara a diseñar políticas policiales y judiciales que faciliten la protección de las víctimas. Además de esto, también existe la incorporación de herramientas de evaluación estructurada del riesgo de reincidencia y de comportamiento violento en el ámbito penitenciario. Si bien es cierto que la administración general del estado no ha implantado sistemas predictivos a nivel nacional, la administración penitenciaria catalana ha desarrollado e implementado el protocolo RisCanvi, que posteriormente analizaremos.

Tanto RisCanvi como VioGén constituyen ejemplos representativos del avance e implementación de herramientas de gestión del riesgo al sistema penal español, aunque sean herramientas con características y fundamentos distintos. Revisar ambas herramientas de forma conjunta nos permite tener una visión global de la actual transformación de nuestro sistema de justicia penal y nos ayuda a imaginar las posibles proyecciones de futuro para el mismo.

2.1. En materia de violencia de género: VioGén

El desarrollo de herramientas basadas en inteligencia artificial y aprendizaje automático ha abierto nuevas posibilidades en el abordaje de fenómenos tan complejos como la violencia de género. En ese contexto la prevención es absolutamente fundamental, y estas tecnologías presentan potencial para que se diseñen políticas públicas que resulten eficaces a la hora de prevenir (Pinto Muñoz *et al*, 2023).

El hecho de emplear herramientas de este tipo para asuntos tan complejos como la violencia de género presenta la ventaja de que poder analizar e incorporar al algo-

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

ritmo predictivo gran cantidad de datos que sirvan para identificar posibles patrones y evitar incidentes violentos (González Álvarez *et al.*, 2020). Esto tiene una relevancia notoria, ya que la cantidad de mujeres que padecen violencia de género ha ido aumentando porcentualmente en los últimos años, y este es un asunto que preocupa especialmente a la sociedad dado que habitualmente son delitos especialmente violentos, que producen lesiones, agresiones sexuales o incluso la muerte de la víctima. Para combatir estas formas de violencia que lamentablemente están en auge, las herramientas predictivas se han demostrado especialmente útiles ya que pueden identificar factores de riesgo específicos que se usan para diseñar medidas de protección más personalizadas que cubran las necesidades de las potenciales víctimas de forma eficaz (Pinto Muñoz *et al*, 2023).

En este sentido, el sistema español estableció el protocolo VioGén, que es una herramienta de evaluación del riesgo diseñada para mejorar la identificación y gestión de casos de violencia de género. Tiene como objetivo ofrecer a los diversos operadores jurídicos y policiales un marco para evaluar el nivel de riesgo de posibles víctimas y tomar decisiones relativas a su protección teniendo ese riesgo en cuesta. Entre sus principales características destacan el acceso multisectorial –que permite a juzgados, servicios sociales y organismos de igualdad consultar y aportar información–, la capacidad de emitir predicciones sobre el riesgo de reincidencia basadas en evaluaciones policiales, y una estrategia proactiva que ha permitido analizar más de 700.000 casos y realizar millones de valoraciones desde su puesta en marcha (Ministerio del Interior, 2022). Además, VioGén revisa de forma continuada posibles modificaciones en el riesgo, actualizándose de forma periódica o cuando suceden episodios nuevos de violencia. Los casos se monitorean y se ajustan a posibles nuevas circunstancias, cuestión que resulta innegablemente útil (Gordo y Rubio-Martín, 2024).

VioGén, a diferencia de otras herramientas, incorpora la posibilidad de que los profesionales policiales ajusten al alza el resultado generado automáticamente por el sistema, basándose en su experiencia, percepción del caso o información adicional que consideren relevante (Gordo y Rubio-Martín, 2024). Esto, que podría percibirse como subjetivo y ser criticado en relación con una posible arbitrariedad, en realidad puede resultar algo deseable, en tanto en cuanto implica una supervisión humana al resultado que arroja la máquina, que no siempre es certero.

Esto queda puesto de manifiesto en diversos estudios que indican que el poder de predicción de las herramientas algorítmicas en contextos de procesos penales es realmente limitado. Un ejemplo claro de esto es el estudio realizado en Reino Unido en relación con un intento de predecir homicidios domésticos por medio del protocolo DASH, cuyos resultados arrojaron que menos del 50 % de los casos contaban con antecedentes o intervenciones policiales previas, lo que limitaba severamente la capacidad predictiva basada únicamente en dichos registros (Thorton, 2017) y que hubo un 67% de casos erróneamente clasificados como de bajo o nulo riesgo, que finalmente resultaron letales para la víctima a pesar de que el algoritmo arrojase un resultado predictivo en contrario (Chalkley y Strang, 2017). Una revisión por parte de profesionales humanos que pueda contradecir lo que propone el algoritmo es, pues, una ventaja que puede combatir posibles errores en los resultados de la herramienta.

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

El sistema VioGén no es una excepción en relación con sus errores de predicción, y en esa dirección ha habido críticas, que abarcan diversas posibilidades, entre las que cabe destacar la alta incidencia de falsos negativos, los falsos positivos y los posibles niveles de riesgo no detectados. Esto evidencia la necesidad de introducir mejoras y complementar el sistema con otras formas de actuar (Gordo y Rubio-Martín, 2024). Un ejemplo que ilustra esto es la Sentencia de la Audiencia Nacional 2350/2020, que es un caso lamentable en el cual el resultado arrojado por el sistema VioGén fue la clasificación de riesgo como "no apreciado", cuestión que supuso que la protección policial proporcionada fuera mínima y no se adoptaran medidas específicas. Un mes más tarde su marido la asesinó de manera brutal. La Audiencia Nacional consideró que en ese caso la guardia civil incumplió su deber de seguimiento del caso, y así lo estableció en su fundamento jurídico segundo, en el que afirma *"La respuesta policial en violencia contra la mujer exige que el sistema pueda prevenirla violencia y reevaluar el riesgo, esto es, más allá de la recogida de datos automatizados, la predicción y la prevención son la finalidad primordial del sistema de evaluación que exige agentes especializados en su tratamiento y sensibilización en su seguimiento"* (Audiencia Nacional, 2020). Esa sentencia pone de manifiesto la necesidad de que la predicción arrojada por el sistema no sea la decisión final tomada por los operadores, así como la importancia de continuar revisando y siguiendo los casos periódicamente, ya que las circunstancias pueden cambiar drásticamente y tener consecuencias fatales. Además de eso, cabe destacar que a lo largo de la sentencia se analiza que el motivo por el cual el sistema arrojó un resultado de inexistencia de riesgo puede haber sido por la falta de prolijidad y exhaustividad a la hora de incluir datos y parámetros en el sistema –no se buscaron posibles testigos, no se tuvieron en cuenta las relaciones sexuales no consentidas mantenidas entre la víctima y el agresor, los antecedentes del agresor (que existían fuera de la jurisdicción española, en República Dominicana, país de origen de la mujer), ni tampoco se analizaron las circunstancias familiares, sociales, económicas y laborales de la víctima y su agresor– (Estévez Mendoza, 2021).

Esto pone de manifiesto la necesidad de que, de cara a que el algoritmo funcione, es necesario que se introduzcan la mayor cantidad posible de datos y estos han de ajustarse a la realidad tanto como sea posible. Para que los datos que alimentan al algoritmo sean de calidad y suficientes, es imprescindible que a la hora de rellenar el formulario el profesional revise exhaustivamente toda la información disponible, incluyendo el atestado policial, informes de servicios sociales, declaraciones de familiares y la declaración de la víctima. Uno de los problemas de base que pueden influir en los errores de la herramienta es que se ha evidenciado que en la práctica la revisión integral que debería llevarse a cabo no siempre se realiza, ya sea por falta de tiempo, recursos o sobrecarga profesional (Olaciregui, 2021).

En estrecha relación con lo anterior, cabe añadir que para que las herramientas de este tipo funcionen bien, no solo es necesario que los datos sean de calidad, sino que habrán de estar bien introducidos en el sistema. Esto pone de manifiesto que uno de los retos de la implementación de este tipo de sistemas es la necesidad imperiosa de formar a los usuarios policiales en su modo de empleo. Está claro que quienes vayan a trabajar con el sistema –policía, autoridad jurisdiccional correspondiente, otros operadores jurídicos implicados– tiene que estar debidamente familiarizado con la

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

herramienta, haber sido capacitado y contar con cierta cultura analítica relativa al funcionamiento del algoritmo. Esto sin duda facilitará su labor y contribuirá al buen uso de este tipo de herramientas (González Álvarez *et al*, 2020).

En lo que respecta al análisis e interpretación de la información, el sistema VioGén ha sido objeto de críticas en relación con los datos que se emplean para realizar el pronóstico. En este sentido, cabe destacar que se analizan factores de riesgo estáticos, que están relacionados con el pasado de la víctima o del agresor y que resultan poco modificables y no reflejan adecuadamente la evolución dinámica del riesgo a lo largo del tiempo ni su vinculación con contextos específicos. Esto no es ideal, sino que sería más deseable que el algoritmo conjugase factores individuales (que suelen ser estáticos) con otros factores ambientales o contextuales, ya que estos últimos por lo general acostumbran a ser más volátiles y pueden tener incidencia directa en una explosión de comportamiento violento (Gordo y Rubio-Martín, 2024).

Otra de las cuestiones que se critica es la relativa a la necesidad de transparencia de este tipo de herramientas. En este sentido, no solo es necesario que los datos de entrada al sistema sean exactos, sino que también es necesario que se revise y audite el modo en que el algoritmo procesa la información. Esto es importante de cara a reducir posibles sesgos y aumentar la fiabilidad del sistema. De momento, el sistema VioGén carece de transparencia ya que ni los evaluadores independientes ni las organizaciones de mujeres tienen acceso a los datos, lo cual es especialmente grave tratándose de un sistema financiado con fondos públicos y con un impacto social significativo (Borges Blázquez, 2024). Esta opacidad es problemática y sin duda debería ser revisada.

Una última cuestión a tener en cuenta en relación con los peligros que implica emplear una herramienta de este tipo se presenta en relación con el tratamiento de datos personales. En este sentido, y especialmente en el caso de VioGén, que gestiona datos de carácter personal que son innegablemente sensibles por razón de su naturaleza, resulta imprescindible regular el modo en que estos datos son tratados, de cara a evitar su indebida publicación, y tratar de proteger los derechos fundamentales de los sujetos implicados en cada caso. Al respecto, cabe mencionar que es resulta de aplicación el RGDP, sirviendo este para que los sujetos no sean objeto de decisiones basadas únicamente en el tratamiento automatizado de sus datos y teniendo derecho a que se les informe debidamente de cómo sus datos son utilizados y protegidos (Martín Ríos, 2024).

En definitiva, VioGén representa un avance significativo en el uso de tecnologías predictivas para la protección de víctimas de violencia de género, ofreciendo una herramienta valiosa para la toma de decisiones en el ámbito policial y judicial. No obstante, su implementación también revela limitaciones importantes que deben ser atendidas, especialmente en lo que respecta a su poder predictivo, la calidad y exhaustividad de los datos, y la transparencia del sistema. La incorporación de supervisión humana y el compromiso con auditorías externas, así como la revisión crítica de los factores de riesgo empleados, son pasos fundamentales para garantizar que este tipo de herramientas no solo refuercen la capacidad de respuesta institucional, sino que también respeten los derechos de todas las personas implicadas y cumplan con los estándares éticos que un sistema de esta naturaleza exige.

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

2.2. Para analizar la reincidencia: RisCanvi

Entre las herramientas de predicción del riesgo que se utilizan en España, destaca sin duda el protocolo "RisCanvi". Este se trata de un instrumento que comenzó a implementarse alrededor del año 2009 con el objetivo de realizar evaluaciones periódicas de diversos riesgos a todas las personas privadas de libertad en Cataluña. Su aplicación, por tanto, se limita a los centros penitenciarios gestionados por la Administración penitenciaria catalana. Sin embargo, su relevancia a nivel nacional es notoria, ya que a menudo es mencionado como posible ejemplo para el resto del estado, cuestión que hace que su análisis resulte clave para comprender la presencia e impacto de los instrumentos de evaluación y gestión del riesgo en el sistema penitenciario español (Alemán Aróstegui, 2023).

RisCanvi –que proviene de la conjunción de dos palabras en catalán, "risc" (riesgo) y "canvi" (cambio)–, es un protocolo que sirve para valorar el riesgo que representan las personas privadas de libertad, tanto en términos de reincidencia delictiva como de comportamiento dentro del centro penitenciario. En este sentido, el objetivo del protocolo es claro: lograr una gestión individualizada que ayude a definir si existen riesgos de que un preso concreto reincida. La idea es lograr prevenir una posible reincidencia y promover la reinserción de aquellos sujetos que no presentan riesgo.

En general, *riscanvi* establece un programa o tratamiento individualizado con base en un algoritmo que establece las posibilidades de reincidencia o peligrosidad de los sujetos. Así, el programa comienza realizando una valoración inicial del preso concreto, continúa con revisiones periódicas de las circunstancias y rasgos que podrían cambiar a lo largo del internamiento, y con base en esos análisis se obtiene un resultado de peligrosidad que ayuda a la toma de decisiones relativas a los permisos y demás posibilidades presentes en la gestión penitenciaria (Alemán Aróstegui, 2023).

El algoritmo funciona por medio de dos escalas de valoración del riesgo que revisan factores de carácter tanto estático como dinámico y ofrecen unos resultados que se divide en cinco tipos de comportamiento riesgoso: violencia autodirigida, violencia intrainstitucional, reincidencia general, reincidencia violenta y quebrantamiento de condena (Andrés Pueyo, 2016).

En este sentido, cabe destacar que los datos que se tienen en cuenta de cara a realizar el análisis tienen que ver con el tipo de delito, el comportamiento mostrado por el preso durante el internamiento, el número previo de ingresos en prisión si los hubiera, etc. (Dribia Data & Direcció General d'Afers Penitenciaris, 2024). Acerca de esto, cabe mencionar que en otros sistemas algorítmicos de predicción de riesgo de reincidencia se tienen en cuenta también datos de otra naturaleza, tales como el nivel educativo del interno, la calidad de sus lazos fuera de prisión basado en las visitas que este recibe –cuestión que puede resultar crucial de cara a establecer la voluntad del sujeto de actuar de manera prosocial si llega a obtener permisos de salida–, y un largo etc. de características que no necesariamente tienen que ver con el desempeño dentro del centro penitenciario (Zeng *et. al.,* 2017). En el sistema de *riscanvi*, este tipo de cuestiones no son analizadas, pero sí que son tenidas en cuenta de cara a la toma de decisiones, ya que existen equipos que visitan al preso y analizan su situación por medio de informes, que eventualmente son tenidos en cuenta por la autoridad jurisdiccional de forma conjunta con el resultado arrojado por *riscanvi*.

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

Una vez que el algoritmo arroja un resultado en relación con el riesgo de reincidencia o posibles comportamientos violentos dentro de la prisión, esto se emplea por parte de los operadores jurídicos para tomar decisiones. Estas decisiones son muy significativas y relevantes para la vida del interno, ya que muy a menudo versan acerca de permisos de salida, una posible concesión de tercer grado o la libertad condicional. En este sentido cabe mencionar que los resultados arrojados por la herramienta no son la decisión definitiva, sino que son una de las cuestiones a tener en cuenta por el juez de vigilancia penitenciaria, que también contará con otra información relevante –proveniente de los funcionarios de prisiones y otros profesionales que estén en contacto con el día a día del preso concreto–. No obstante, en los casos en los que el resultado que ofrece *riscanvi* y los equipos que han visitado al interno difieren en opiniones presentan complejidad, ya que el algoritmo es opaco en su funcionamiento. No entender bien el origen del disenso dificulta en grado sumo que la autoridad jurisdiccional correspondiente tome una decisión final debidamente informada (Alemán Aróstegui, 2023).

De cara a analizar y contrarrestar posibles sesgos del algoritmo, en enero de 2024 se realizó una auditoría a su funcionamiento por parte de la Generalitat de Cataluña. El informe pone de relieve varias cuestiones críticas en relación con el funcionamiento del algoritmo utilizado en el protocolo. En primer lugar, subraya la necesidad de que los resultados que generan puedan explicarse, ya que la opacidad del algoritmo hace que no quede claro el proceso mediante el cual se llega al resultado que arroja. Dado que el algoritmo es opaco, se dificulta la interpretación y el control sobre sus decisiones por parte de los profesionales que tienen que tomar la decisión. Esto puede resultar especialmente problemático en los casos en los que la decisión concreta afecte de forma directa a los derechos fundamentales de los internos (Dribia Data & Direcció General d'Afers Penitenciaris, 2024).

Además, el informe también alerta sobre la presencia de posibles sesgos discriminatorios en relación con características que no deberían ser tenidas en cuenta en aras de respetar la igualdad, tales como el sexo, el origen o la edad. Los análisis realizados indican que estos sesgos podrían incidir negativamente en la precisión de las predicciones de reincidencia, comprometiendo la equidad del sistema y afectando las decisiones de los profesionales penitenciarios (Dribia Data & Direcció General d'Afers Penitenciaris, 2024). Esto reviste interés porque uno de los argumentos más empleados a favor de las herramientas de valoración del riesgo que emplean algoritmos es que se les atribuye neutralidad, ya que se establecen por medio de criterios técnicos, y en ocasiones se obvia el carácter valorativo que puede venir predeterminado en la configuración de base (Alemán Aróstegui, 2023).

Otro de los aspectos destacados es la capacidad predictiva de acierto del algoritmo. Queda reflejado que *riscanvi* funciona bien para clasificar personas de bajo riesgo, pero su eficacia disminuye notablemente en la detección de casos de alto riesgo. Esta limitación puede generar una falsa sensación de seguridad respecto a algunos perfiles, al tiempo que subestima riesgos relevantes que podrían requerir intervenciones específicas (Dribia Data & Direcció General d'Afers Penitenciaris, 2024).

En estrecha relación con eso, se pone de manifiesto –al igual que sucedía con otras herramientas, como en el caso de VioGén recientemente revisado– la absoluta necesidad de supervisión humana del resultado arrojado por la máquina. En este sentido

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

se manifiestan las audiencias provinciales catalanas, y así la audiencia de Barcelona establece que *"riscanvi es un algoritmo al que no se puede dar valor absoluto debiendo valorarse el caso particular"* (Audiencia Provincial de Barcelona, 2024) y la de Tarragona apunta que *"no debe olvidarse que esta herramienta, útil para valorar con carácter objetivo estos riesgos, no puede desligarse de la valoración de los profesionales que tratan al interno, por lo que estimamos que debe prevalecer el informe de los especialistas frente a los ítems resultantes de los algoritmos utilizados por la herramienta informática"* (Audiencia Provincial de Tarragona, 2023).

También es necesario destacar que el informe advierte sobre las limitaciones en la recogida e integración de datos, señalando que la herramienta no incorpora de forma sistemática todas las fuentes de información disponibles. Esta deficiencia, unida a la falta de validación rigurosa y de transparencia sobre los factores que influyen en las clasificaciones, puede introducir sesgos adicionales y comprometer la fiabilidad del sistema de evaluación del riesgo (Dribia Data & Direcció General d'Afers Penitenciaris, 2024).

En conclusión, el protocolo *riscanvi* es un ejemplo paradigmático del uso de herramientas algorítmicas y su diseño pretende ayudar a que la gestión de los internos en centros penitenciarios catalanes pueda darse de forma individualizada. No obstante, es importante resaltar que es una herramienta de apoyo, que debido a las limitaciones que presenta –la opacidad del algoritmo, la existencia de sesgos, etc.– no puede emplearse como mecanismo que tome las decisiones, sino que precisa necesariamente de supervisión humana de los resultados que arroja.

3. Conclusiones

En el marco del proceso penal, el uso de herramientas algorítmicas de predicción del riesgo plantea importantes desafíos éticos, jurídicos y sociales. Está claro que su utilidad es considerable, ya que poseen la capacidad de procesar gran cantidad de datos y elaborar predicciones con base en ellos, pero lo cierto es que su implementación también genera tensiones con derechos fundamentales tales como la presunción de inocencia. Este riesgo se ve agravado por la posibilidad de que estas herramientas reproduzcan y perpetúen sesgos históricos y estructurales, especialmente cuando se nutren de datos previamente marcados por prácticas discriminatorias. En este sentido, resulta imprescindible prestar atención a posibles sesgos y combatirlos de cara a no discriminar a colectivos vulnerables.

Para poder actuar contra esos posibles sesgos, es necesaria la transparencia de las herramientas, que posibilita que se lleve a cabo una supervisión humana del correcto funcionamiento del algoritmo. No obstante, la mayor parte de los sistemas son opacos, y esto dificulta en grado sumo la comprensión de su funcionamiento y la posibilidad de impugnar los resultados que arrojan. Esto es socialmente inadmisible y resulta imprescindible que se regule de forma clara y restrictiva el uso de estas tecnologías para que se garantice, por medio de auditorías u otros tipos de revisión, que se salvaguardan las garantías y estándares éticos necesarios.

El equilibrio entre la prevención del delito y la protección de derechos no debe perder de vista el papel fundamental de la intervención humana. Las herramientas predictivas no pueden sustituir el juicio de los operadores jurídicos, sino asistirlo, y para ello habrán

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

de emplear criterios que permitan la supervisión por parte de las autoridades jurisdiccionales que serán quien tomarán las decisiones correspondientes.

El uso de herramientas algorítmicas en España no es aislado y los sistemas como VioGén y RisCanvi representan ejemplos claros de ello. En este sentido, VioGén ha demostrado ser una herramienta valiosa en la protección de víctimas de violencia de género al facilitar la clasificación del riesgo y permitir respuestas adaptadas. No obstante, su eficacia depende de la calidad de los datos introducidos y de una interpretación adecuada por parte de los profesionales, por lo que la transparencia y las auditorías son también aquí esenciales.

Lo mismo sucede en el caso de RisCanvi, que también debe ser utilizada como una herramienta de apoyo sin perder de vista las limitaciones existentes en relación con la opacidad y la necesidad de supervisión humana.

En última instancia, tanto VioGén como RisCanvi evidencian la necesidad de emplear este tipo de herramientas de una forma responsable, teniendo en cuenta que son modelos algorítmicos que presentan limitaciones y por tanto deben utilizarse de forma que se puedan equilibrar la intención de prevenir delitos y proteger víctimas con el respeto a las garantías inherentes al proceso penal.

Bibliografía

Alemán Aróstegui, L. (2023). El uso del RisCanvi en la toma de decisiones penitenciarias. *Estudios Penales y Criminológicos, 44*. https://doi.org/10.15304/epc.44.8884

Andrés Pueyo, A. (2016). ¿Es técnicamente posible anticipar la reincidencia delictiva? El protocolo RisCanvi en las prisiones de Cataluña. En *IX Jornadas de ATIP Almagro* (pp. 55–78).

Borges Blázquez, R. (2024). Algoritmización de la concesión de medidas cautelares en el proceso penal para la protección de víctimas de violencia de género. ¿Es capaz VIOGÉN de interpretar el *periculum in mora*? *Actualidad Jurídica Iberoamericana*, (21), 384–407.

Carlson, A. (2017). The Need for Transparency in the Age of Predictive Sentencing Algorithms. *Iowa Law Review,103,* 303-329.

Castagnedi Ramírez, A. E. (2024). La construcción de un algoritmo «ético». *Ius et Scientia, 10*(2), 123–151. https://doi.org/10.12795/IESTSCIENTIA.2024.i02.06

Chalkley, R. y Strang, H. (2017). Predicting domestic homicides and serious violence in Dorset: A replication of Thornton's Thames Valley analysis. *Cambridge Journal of Evidence-Based Policing, 1*(2), 81-92. https://doi.org/10.1007/s41887-017-0010-2

Daviera, A. L., Uriostegui, M., Gottlieb, A., & Onyeka, O. A. (2023). Risk, race, and predictive policing: A critical race theory analysis of the strategic subject list. *American Journal of Community Psychology, 71*, 1-13. https://doi.org/10.1002/ajcp.12671

Dribia Data & Direcció General d'Afers Penitenciaris. (2024). *Informe Tiresias: Auditoria de l'algorisme RisCanvi* (Versión del 9 de enero de 2024). Generalitat de Catalunya, Departament de Justícia, Drets i Memòria.

Estévez Mendoza, L. (2021). Inteligencia artificial y violencia contra las mujeres: ¿Funcionan los sistemas automatizados de evaluación del riesgo? *Perspectivas. Revista de Ciencias Jurídicas y Políticas, 3*, 127–141.

González-Álvarez, J. L., Santos-Hermoso, J. & Camacho-Collados (2020). Policía predictiva en España. Aplicación y retos de futuro. *Behavior & Law Journal, 6(1),* 26-41. https://doi.org/10.47442/blj.v6.i1.75

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

Gordo, Á., y Rubio-Martín, M. J. (2024). Incertidumbres algorítmicas en torno a las violencias de género. El caso del sistema VioGén y otros sistemas de predicción del riesgo. *Revista Española de Sociología, 33*(2), https://doi.org/10.22325/fes/res.2024.225

Hogan Doran, D (2017). Computer says 'no': Automation, algorithms and artificial intelligence in Government decision-making. *The Judicial Review, 13*.

Kroll, J (2015) Accountable Algorithms. *PhD Thesis, Princeton University*. https://dataspace.princeton.edu/jspui/handle/88435/dsp014b29b837r

Levinson, D. (2002). Net widening. *Encyclopedia of crime and punishment,* 4, 1088. https://doi.org/10.4135/9781412950664.n286

Martín Ríos, P. (2024). Predictive algorithms and criminal justice: expectations, challenges and a particular view of the Spanish VioGén system. *Rivista Italiana di Informatica e Diritto, 2,* 547-562.

McKay, C. (2020). Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making. *Current Issues in Criminal Justice, Sydney Law School Research Paper No. 19/67*. http://dx.doi.org/10.2139/ssrn.3494076

Ministerio del Interior (2022). VioGén cumple 15 años con más 700.000 casos analizados y 5,4 millones de valoraciones de riesgo realizadas. https://n9.cl/wm98b

Olaciregui, M P. (2021). *Prevenir la violencia contra las mujeres: análisis de las herramientas de evaluación y gestión del riesgo desde una perspectiva de género* [Tesis doctoral]. Universidad de Zaragoza.

Pereira Puigvert, S. (2024). La justicia y el proceso penal en clave algorítmica: Nuevos enfoques, nuevos riesgos. *Ius et Scientia*, *10*(2), 66–79. https://doi.org/10.12795/IESTSCIENTIA.2024.i02.03

Pérez Salazar, B. (2024). La criminología predictiva: ¿un futuro próximo o una ficción en lontananza? *Novum Jus, 18* (3), 343-369. https://doi.org/10.14718/NovumJus.2024.18.3.13

Pinto Muñoz, C. C., Zuñiga Samboni, J. A., & Ordoñez Erazo, H. A. (2023). *Machine learning applied to gender violence: A systematic mapping study*. Revista Facultad de Ingeniería, 32(64). https://doi.org/10.19053/01211129.v32.n64.2023.15944

Reynald, D. M. (2011). Translating CPTED into crime preventive action: A critical examination of CPTED as a tool for active guardianship. *European Journal on Criminal Policy and Research, 17*(1), 69–81. https://doi.org/10.1007/s10610-010-9135-6

Robinson, P. H. (2001). Punishing dangerousness: Cloaking preventive detention as criminal justice. *Harvard Law Review, 114*(5), 1429-1493. https://doi.org/10.2139/ssrn.183288

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology, 94*(4), 774-802. https://doi.org/10.1086/229068

Suárez Xavier, P. R. (2024). Algoritmización de la justicia penal: entre fe, confianza y garantías procesales. *Revista de derecho y proceso penal, 74*, 149-166.

Susser, D. (2021). Predictive policing and the ethics of preemption. En B. McCartney & J. Fagan (Eds.), *The ethics of policing* (pp. 268-292). New York University Press. https://doi.org/10.18574/nyu/9781479803729.003.0013

The Law Society of England and Wales (2019). Algorithms in the criminal justice system. *Commission on the Use of Algorithms in the Justice System* https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-inthe-criminal-justice-system-report/

Thornton, S. (2017). Police attempts to predict domestic murder and serious assaults: Is early warning possible yet? *Cambridge Journal of Evidence-Based Policing, 1*(2), 64-80. https://doi.org/10.1007/s41887-017-0011-1

Tonry, M. (2019). Predictions of dangerousness in sentencing: Déjà vu all over again. En *American sentencing: What happens and why?* (pp. 439-482). The University of Chicago Press. https://doi.org/10.1086/701895

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01

Wang, C., Han, B., Patel, B., y Rudin, C. (2022). In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction. *Journal of Quantitative Criminology*, 39 (4), https://doi.org/10.1007/s10940-022-09545-w

Zedner, L. (2007). Preventive justice or pre-punishment? The case of control orders. *Current Legal Problems, 60*(1), 174–203. https://doi.org/10.1093/clp/60.1.174

Zeng, J., Ustun, B., y Rudin, C. (2017). Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 180(3), 689-722. https://doi.org/10.1111/rssa.12227

Zouave, E. T., & Marquenie, T. (2017). An inconvenient truth: Algorithmic transparency & accountability in criminal intelligence profiling. *2017 European Intelligence and Security Informatics Conference (EISIC)*, 17–23. https://doi.org/10.1109/EISIC.2017.12

Jurisprudencia

Auto de la Audiencia Provincial de Barcelona 343/2024, de 15 de febrero de 2024. https://doi.org/10.69592/5-6673-N1-SEGUNDO-SEMESTRE-2024-ART-5

Auto de la Audiencia Provincial de Tarragona 90/2023, de 3 de febrero de 2023

Rummel v. Estelle, United States Supreme Court (1980)

Sentencia de la Audiencia Nacional 2350/2020, Sala de lo Contencioso, de 30 de septiembre de 2020

State v. Loomis (Wis. 2016)

Fuentes legales

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento General de Protección de Datos)

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

IUS ET SCIENTIA • 2025
Vol. 11 • Nº 1 • pp. 14-31
ISSN 2444-8478 • https://dx.doi.org/10.12795/IESTSCIENTIA.2025.i01.01