

## ***Algoritmos e inteligencia artificial en el sistema de justicia penal***

-

Hace tiempo que desde la criminología y el derecho penal se advierte del uso cada vez más extendido de sistemas de predicción o estimación de riesgo basados en cálculos probabilísticos, realizados mediante algoritmos matemáticos y/o inteligencia artificial.

Las preocupaciones que criminólogos y penalistas, entre otros, manifiestan frente a este tipo de instrumentos son variadas y han sido objeto recientemente de una jornada monográfica que ha tenido lugar en la Facultad de Derecho de Valencia el 1 de octubre. De hecho, la jornada sirvió para presentar oficialmente una auditoría que Amnistía Internacional encargó a un grupo de investigadores españoles (liderados por la profesora Lucía Martínez Garay), sobre tres sistemas predictivos utilizados en la justicia penal de nuestro país: *VioGén*, *RisCanvi* y *VeriPol*<sup>1</sup>. Se trata de tres instrumentos con objetivos bien diferentes, que sin embargo tienen en común que estamos ante herramientas algorítmicas que realizan análisis predictivos sobre la base de una ingente recopilación y tratamiento de datos<sup>2</sup>.

En concreto, *VioGén* es un instrumento policial de valoración del riesgo, que realiza una estimación del riesgo que presenta una víctima de violencia de género de volver a ser victimizada<sup>3</sup>. *VeriPol* es también un instrumento utilizado por la policía que, a partir de técnicas de procesamiento del lenguaje natural y aprendizaje automático, promete predecir con una eficacia muy superior a la del cualquier policía experto la interposición de una denuncia falsa (en delitos contra el patrimonio)<sup>4</sup>.

---

<sup>1</sup> Vid. UNIVERSITAT DE VALÈNCIA, «Presenting: Three predictive policing approaches in Spain: Viogén, RisCanvi and Veripol (an assessment from a human rights perspective, by Martínez Garay et al.», 2024. <https://regulation.blogs.uv.es/presenting-three-predictive-policing-approaches-in-spain-viogen-riscanvi-and-veripol-an-assessment-from-a-human-rights-perspective-by-martinez-garay-et-al/>

<sup>2</sup> *Stricto sensu* ni *VioGén* ni *RisCanvi* pueden ser consideradas herramientas que utilizan Inteligencia Artificial (IA) ya que su algoritmo no tiene capacidad autónoma de aprendizaje. Por el contrario, *VeriPol* sí entraría dentro de la definición más asumida de IA.

<sup>3</sup> Vid. MINISTERIO DEL INTERIOR, «Sistema VioGén», 2024. <https://www.interior.gob.es/opencms/ca/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen/>

<sup>4</sup> POLICÍA NACIONAL, «La Policía Nacional pone en funcionamiento la aplicación informática VeriPol para detectar denuncias falsas», 27 de octubre de 2018. [https://www.policia.es/\\_es/comunicacion\\_prensa\\_detalle.php?ID=4433&idiomaActual=es#](https://www.policia.es/_es/comunicacion_prensa_detalle.php?ID=4433&idiomaActual=es#)

Por su parte, *RisCanvi* es una herramienta de predicción de riesgo utilizada por la Administración Penitenciaria catalana para valorar el riesgo que presenta todo interno en prisión respecto de una serie de escalas (reincidencia general, reincidencia violenta, quebrantamiento de condena, violencia auto-dirigida y violencia intra-institucional)<sup>5</sup>.

Los temores frente a estos instrumentos no se han hecho esperar y hacen referencia a diversas cuestiones, pero creo que, en la discusión relativa a sus pros y contras, ayudaría clarificar algunas cuestiones.

En primer lugar, debe tenerse claro que la estimación o predicción del riesgo, esto es, la probabilidad de que un evento (indeseado) futuro ocurra, es una cuestión de la que no puede prescindirse en ocasiones en el ámbito del derecho penal. Es más, la ley penal obliga en varios contextos a realizar tal predicción. Por ejemplo, el art. 544 ter de la LECrim, a la hora de decidir sobre la concesión o denegación de una orden de protección a una mujer que denuncia ser víctima de violencia de género, obliga al juez a realizar una predicción de este tipo, pues junto a la presencia de indicios de delito, exige la existencia de una «*situación objetiva de riesgo para la víctima*». Si la ley se contentara con una situación «*subjetiva de riesgo*» bastaría con preguntarle a la denunciante si teme volver a ser victimizada, pero como lo que se exige es que tal riesgo sea «*objetivo*», la predicción deberá tomar como base otros muchos elementos. La cuestión es, pues, si tal juicio predictivo se deja en manos del saber experto del profesional (sea policía, juez u otro agente del sistema penal), método que se conoce como «*juicio clínico*», en manos de una función matemática (como por ejemplo es *VioGén*), o de un sistema mixto que tenga en cuenta ambas posibilidades.

Planteada así la cuestión, algunas críticas recurrentes contra el uso de algoritmos o inteligencia artificial en el sistema de justicia penal, deberían reconducirse. En este sentido, cuando se denuncian los sesgos algorítmicos (un tema recurrente en EE. UU. con referencia a los sesgos raciales que se han descubierto en algunos algoritmos predictivos), debería recordarse que policías y jueces no están libres de tales sesgos. Y cuando se achaca, con razón, a algunos algoritmos su falta de transparencia, no deberíamos olvidar que muchas prácticas policiales (cacheos y registros, por ejemplo), actuaciones judiciales y de otros agentes de la justicia penal, distan mucho de ser modelos de transparencia. A la IA no debiera pedírsele cosas que no pedimos a la «Inteligencia Natural». Es más, las discusiones sobre los sesgos algorítmicos deberían hacernos reflexionar sobre los sesgos humanos.

Una segunda cuestión capital que debería presidir la discusión sobre los sistemas automatizados de predicción es la relativa a qué ventaja o avance suponen estas herramientas, respecto de la decisión experta de un individuo. Sin duda, lo primero que se arguye es su mayor efectividad en la predicción del riesgo. Ello porque la máquina, al manejar una cantidad de información mucho mayor que una persona, es capaz de realizar una mejor predicción que cualquier humano. Esta es una afirmación que debería ser validada científicamente, y no ser aceptada acríticamente fruto de una fascinación naif por la tecnología.

En tercer lugar y por último, no deberíamos olvidar una serie de peligros que me temo quedan enmascarados bajo la pátina de la tecnología. Primero: toda predicción de comisión futura de

---

<sup>5</sup> DEPARTAMENT DE JUSTÍCIA I QUALITAT DEMOCRÀTICA (GENERALITAT DE CATALUNYA), *Protocol de valoració del risc (RisCanvi)*, 2024. [https://justicia.gencat.cat/ca/ambits/reinsercio\\_i\\_serveis\\_penitenciaris/serveis\\_penitenciaris/RisCanvi/](https://justicia.gencat.cat/ca/ambits/reinsercio_i_serveis_penitenciaris/serveis_penitenciaris/RisCanvi/)

delitos se apoya en una teoría o comprensión concreta de las causas (factores de riesgo) de la delincuencia. Primero va la teoría y después la elaboración de un algoritmo (los datos sólo son datos en el contexto de una teoría). Quizás en un futuro la recopilación masiva de datos individuales permita a la IA elaborar algoritmos propios sin intervención previa humana y sin necesidad así de una teoría sobre los mismos. Pero de momento, los algoritmos que conocemos (*VioGén* y *RisCanvi*) son fruto de una reflexión humana previa sobre los factores de riesgo individuales conectados con la delincuencia. Los datos concretos que apoyan esa conexión deberían ser sometidos a escrutinio público. Esta exigencia de transparencia, no solo del algoritmo o de la función matemática, sino también de la teoría y datos que lo explican, es inexcusable si las herramientas tienen un uso público. Y hasta el momento ni *VioGén*, ni *RisCanvi* (a pesar de la gran mejoría reciente en este aspecto), ni mucho menos *VeriPol*, son herramientas totalmente transparentes.

Segundo (peligro de estas herramientas). No se pueden poner en manos de los agentes del sistema penal tales herramientas sin una formación especializada previa sobre su uso y sus limitaciones. La experiencia de *RisCanvi* en Cataluña es ilustrativa al respecto. Cuando las autoridades penitenciarias empezaron a utilizar esta herramienta, ni Jueces de Vigilancia Penitenciaria, ni Fiscalía, ni los abogados, recibieron información detallada sobre el potencial de esta herramienta. La desnuda información sobre si un preso presenta un determinado nivel de riesgo (bajo, medio o alto) sirve de bien poco si el operador no conoce las limitaciones del algoritmo. Así, normalmente bastaba tener un riesgo medio o alto en *RisCanvi* para que el fiscal impugnase cualquier permiso de salida concedido por la Junta de Tratamiento y el Juez de Vigilancia Penitenciaria. Sin embargo, solo cuando uno es consciente de la alta tasa de falsos positivos que presenta esta herramienta, puede cuestionar su utilización para negar un permiso o una progresión en grado. A pesar de ello, la información sobre el grado de la fiabilidad de la herramienta no acompaña a la propia valoración. Sin una buena formación en las limitaciones de la herramienta, el riesgo de des-responsabilización («si la máquina dice riesgo alto será por algo») es relevante, máxime en un ámbito en el que están en juego derechos fundamentales.

Tercer (peligro de estas herramientas). Reconociendo que, según hemos visto, en determinados ámbitos de la justicia penal es inevitable trabajar con estimaciones de riesgo, no deberíamos caer en la trampa de pensar que en la imposición de la pena tal predicción debe jugar el papel esencial. La prevención especial (sea positiva o negativa) tiene sin duda un papel a desempeñar en la elección del castigo, pero está lejos de ser una cuestión pacífica cuán importante ha de ser este papel. No deberíamos olvidar que una herramienta algorítmica de predicción de delincuencia futura puede hacer «responsable» al individuo por la posesión de unos rasgos (edad, género, educación recibida, amistades, etc.) que estadísticamente están asociados con la comisión de delitos, pero que sin embargo nunca estuvo en sus manos controlar<sup>6</sup>. El principio de culpabilidad individual y proporcionalidad todavía son considerados ejes esenciales del castigo y por ello, en la pena a imponer por la comisión de un delito, deberíamos ser más cautelosos todavía frente a la utilización de herramientas predictivas.

Cuarto (peligro de estas herramientas). Al final lo que sí ha conseguido el desarrollo de las herramientas algorítmicas de predicción de riesgos y delincuencia es situar el énfasis en el

---

<sup>6</sup> En este sentido, por ejemplo, *RisCanvi* cuenta entre las variables de riesgo factores como los antecedentes penales de los familiares, tener amistades antisociales, ser miembro de un «grupo social vulnerable», la falta de apoyo familiar, el nivel educativo, etc.

análisis y prevención de la delincuencia en el plano individual. Con ellas se pretende, en suma, evitar que un individuo concreto cometa delitos en el futuro analizando toda una serie de rasgos que ese individuo posee. Pero ¿y si resulta que es una variable relevante para explicar la delincuencia futura el abandono temprano de la escuela, el nacimiento en un barrio degradado de la ciudad, la falta de empleo o la realización de un trabajo mal remunerado? En estos casos, ¿la herramienta señala a un individuo o a una sociedad que distribuye de forma desigualitaria las oportunidades individuales?

En definitiva, la visión de la delincuencia que en últimas transmite la utilización de estos instrumentos es que estamos ante un problema individual que se podría resolver «científicamente», identificando toda una serie de factores de riesgo. Como criminólogos creo que no debiéramos dejar que estos instrumentos nos hagan olvidar el resto de «variables» e ignorar que estamos ante un problema social complejo que no se va a solucionar buscando la mejor fórmula matemática.

*Daniel Varona*