

Artículos / Articles

Incertidumbres algorítmicas en torno a las violencias de género. El caso del sistema VioGén y otros sistemas de predicción del riesgo

Algorithmic concerns surrounding gender-based violence. The case of the VioGén system and other risk prediction systems

Ángel Gordo * 

Departamento de Sociología: Metodología y Teoría, Instituto Complutense de Sociología para el Estudio de las Transformaciones Sociales Contemporáneas (TRANSOC), Universidad Complutense de Madrid, España.
ajgordol@ucm.es

María José Rubio-Martín 

Departamento de Sociología: Metodología y Teoría, Universidad Complutense de Madrid, España.
majrubio@ucm.es

Recibido / Received: 01/03/2023
Aceptado / Accepted: 12/02/2024



RESUMEN

En la sociedad actual los algoritmos juegan un papel destacado en la toma de decisiones, desde la búsqueda de información hasta la estimación de conductas delictivas. Los sistemas algorítmicos de evaluación del riesgo se han convertido en una herramienta clave para la prevención y la protección de las víctimas de violencia de género. Este artículo analiza las principales bondades y críticas de los sistemas algorítmicos de evaluación del riesgo, con especial énfasis en el caso de la violencia de género. En la primera parte se identifican las principales ventajas y limitaciones de estos sistemas, según sus defensores y detractores. En la segunda parte se revisan los principales modelos y programas vigentes en numerosos países occidentales, así como los distintos sistemas de evaluación del riesgo de violencia de género vigentes en España, con especial detenimiento en el sistema VioGén. En la tercera parte, se discuten propuestas éticas y sociales que surgen en torno a estos sistemas.

Palabras clave: Algoritmos, discriminación, evaluación del riesgo, femicidios de pareja, sistemas de ayuda a la toma de decisiones, falsos negativos.

ABSTRACT

In today's society, algorithms play a central role in decision-making, from information retrieval to the estimation of recidivism. In the context of gender-based violence, algorithmic risk assessment systems have become a critical tool for prevention and victim protection. This article analyzes the main strengths and criticisms of algorithmic risk assessment systems, with a special focus on the case of gender-based violence. In the first part, the main advantages and limitations of these systems are identified, according to their proponents

*Autor para correspondencia / Corresponding author: Ángel Gordo, ajgordol@ucm.es

Sugerencia de cita / Suggested citation: 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), a225. <https://doi.org/10.22325/fes/res.2024.225>

and opponents. The second part reviews the main models and programs implemented in numerous Western countries, along with the various gender-based violence risk assessment systems operational in Spain, with a particular focus on the VioGén system. The third part discusses ethical and social proposals that have emerged concerning these systems.

Keywords: Algorithms, Decision Supported Systems (DSS), discrimination, false negative, intimate partner femicide, risk assessment.

INTRODUCCIÓN

En el contexto de una sociedad dominada por una incertidumbre desatada y radical como principales rasgos de la “semántica del riesgo” (Ramos-Torre y Callejo, 2018), los algoritmos, inteligencias artificiales (IAs) y sistemas de decisión automatizados, juegan un papel destacado en un sinfín de actividades: desde las recomendaciones de búsquedas de todo tipo, las conversaciones con los interfaces conversacionales (*chatbots*) u otras tomas de decisiones en ámbitos muy diversos entre las que se incluyen los drones utilizados en los campos de batalla o la estimación de conductas delictivas reincidentes. La evaluación algorítmica del riesgo, destinada a facilitar la toma de decisiones, también ha experimentado un creciente protagonismo en muchos sistemas policiales.

El objetivo de este artículo es analizar las principales bondades y críticas de los sistemas algorítmicos de evaluación del riesgo, con especial énfasis en el caso de la violencia de género. Para ello, se hace una revisión temática de la literatura acerca de la evaluación de riesgo basada en IA y aprendizaje automático en el contexto judicial y en particular en el contexto de las violencias de género. Los bloques temáticos seleccionados inicialmente fueron: métodos actuariales y algorítmicos para la evaluación de riesgos: sesgos y beneficios; predicciones algorítmicas de la violencia de género: principales modelos; y propuestas y avances para tomas de decisiones más humanas y socialmente éticas en materia de violencia de género.

Hemos dividido los principales resultados en tres apartados. En el primero presentamos, de forma general, algunas de las principales correspondencias y continuidades entre los procedimientos actuariales previos y la actual lógica algorítmica puesta al servicio de la toma de decisiones policiales y judiciales. También identificamos las principales bondades y críticas esgrimidas por sus defensores y detractores respectivamente. El segundo apartado se detiene en el campo específico de la violencia de género desde el prisma de las lógicas algorítmicas actuariales. Un interés común a gran parte de la literatura analizada en torno a esta temática son los programas informáticos vigentes en numerosos países occidentales destinados a asistir en la toma de decisiones sobre niveles de riesgo o exposición a violencias de género. En este apartado también damos cuenta de los distintos modelos teóricos que subyacen a estos programas en el contexto europeo. El último apartado pone el acento en propuestas de más reciente cuño centradas en una perspectiva humana y en el elemento ético-social como pieza complementaria a las lógicas actuariales algorítmicas.

LAS LÓGICAS ALGORÍTMICAS ACTUARIALES: BONDADES Y SESGOS EN LA EVALUACIÓN DEL RIESGO

La toma de decisiones basadas en métodos actuariales para la evaluación de riesgos se remonta a principios del siglo XX (Harcourt, 2006; Wang et al., 2022). En el ámbito jurídico estos métodos recurren a la teoría de la probabilidad para decidir “a quién

buscar, qué zonas geográficas vigilar, la posibilidad de obtener la libertad bajo fianza o la libertad condicional, la duración de una condena penal y el tipo de prisión en la que debe ser encarcelado un delincuente condenado” (McNamara et al., 2019, p. 96). Estas lógicas experimentaron un notable desarrollo en la década de los 90 coincidiendo con los procesos crecientes de digitalización y uso de sofisticadas técnicas estadísticas (análisis de regresión) y computacionales (inteligencia artificial y aprendizaje automático – *machine learning*). Copas y Marshall (1998), años atrás, reparaban en la continuidad entre los por aquel entonces incipientes usos de algoritmos en el sistema de justicia criminal y los métodos basados en regresión estadística utilizados durante décadas para determinar la posibilidad de reincidencia. McNamara et al. (2019, p. 96), siguiendo las tesis de Berk (2012), proponen que las lógicas actuariales podrían considerarse como un antecedente de la actual evaluación algorítmica de riesgos, con la salvedad de que “utilizan métodos de aprendizaje automático sobre grandes volúmenes de datos y se comercializan bajo el paraguas de la inteligencia artificial”.

En un escenario marcado por el paradigma del procesamiento de la información y las ciencias cognitivas, uno de los principales debates que acompañaron a estos primeros sistemas de evaluación algorítmica del riesgo se ceñía a la disyuntiva central que planteaba en qué medida las evaluaciones producidas por personas expertas son más precisas que las emitidas por estos sistemas (Dawes, Faust y Meehl, 1989; Grove y Meehl, 1996). En la actualidad los programas actuariales de naturaleza algorítmica son de uso común en el ámbito judicial, mayormente en vistas previas y concesiones de libertad provisional o condicional. *The Public Safety Assessment, LSI-R* y *Ohio Risk Assessment System* son algunos de los sistemas propietarios utilizados en las jurisdicciones estadounidenses, con un código fuente que no está disponible públicamente, lo que significa que los/as usuarios/as no pueden ver cómo funciona el sistema operativo o modificarlo para satisfacer sus necesidades específicas. Algunas jurisdicciones han adoptado estas evaluaciones de riesgo en la fase previa al juicio para reemplazar la fianza en efectivo, por suponer un fuerte sesgo para las personas (Zweig, 2010; Desmarais, Garrett y Rudin, 2019).

Otros beneficios que comportan los sistemas de ayuda a la toma de decisiones (DSS - *Decision Support Systems*), además de una supuesta mayor precisión, consistencia y escalabilidad en las resoluciones judiciales (Kleinberg, Mullainathan y Raghavan, 2017), son la transparencia que brinda sus modelos algorítmicos cada vez más abiertos y accesibles (Zeng, Ustun y Rudin, 2017), así como la capacidad de ajuste y refinamiento de la metodología y el reajuste de dichos modelos. En una tónica parecida a Wang et al. (2022), McNamara et al. (2019) expresan su confianza en los enfoques cuantitativos para implementar mejoras y, con ellas, mayor justicia en los sistemas de evaluación de riesgo basados en la IA y aprendizaje automático en su caso en el contexto judicial australiano.

Sin menoscabar las bondades de este tipo de decisiones automatizadas, durante la última década han aparecido otros estudios que reparan en los efectos indeseados que acompañan a determinados usos de las lógicas actuariales algorítmicas. El algoritmo *Correctional Offender Management Profiling for Alternative Sanctions* [Gestión de perfiles de delincuentes correccionales para sanciones alternativas] (COMPAS), utilizado por los tribunales estadounidenses para evaluar el riesgo de reincidencia, se convirtió en un ejemplo paradigmático de los sesgos y prejuicios de la Inteligencia Artificial y el aprendizaje automático (Angwin et al., 2016). COMPAS evalúa el riesgo potencial de reincidencia a partir de tres escalas (reincidencia general, reincidencia violenta y liberación previa al juicio) (Larson et al., 2016). Las puntuaciones en la escala de riesgo de liberación previa al juicio, que estima la posibilidad de que una persona no comparezca y/o cometa nuevos delitos mientras está en libertad, como indica el manual de uso de COMPAS (Northpointe, 2015, p. 27), se calculan a partir de los siguientes indicadores: “cargos actuales, cargos pendientes, historial de arrestos previos, fallas previas al juicio, estabilidad residencial, situación laboral, vínculos con la comunidad y abuso de sustancias”. Angwin et al. (2016) han denunciado que COMPAS puntuaba casi dos veces más a las personas afroamericanas

que a las caucásicas, como de riesgo alto, cuando en realidad no llegaban a reincidir (véase también Flores, Bechtel y Lowenkamp, 2016; López-Flores, 2022).

El sistema equivalente a COMPAS en Australia, el *Suspect Targeting Management Plan* [Plan de Gestión de Objetivos Sospechosos] (STMP) también adolece de serios sesgos en el caso de los/as jóvenes y la población indígena australiana (Englezos, 2023). El STMP encuentra un máximo referente en el *Harm Assessment Risk Tool* [Instrumento para Determinar el Riesgo de Delito] (HART), un programa utilizado en el Reino Unido para evaluar el riesgo -alto, medio, bajo- de reincidencia. También Simón-Castellano (2021) ha denunciado las lógicas discriminatorias y deterministas de HART.

Ante semejantes sesgos raciales y “fallos”, algunos analistas abogan por la diversificación de los datos de entrada con el propósito de desarrollar sistemas más inclusivos (Eubanks, 2018; Noble, 2018; West, Whittaker y Crawford, 2019). Otros estudios, por el contrario, sostienen que, en el caso del reconocimiento facial, mejorarlos con sistemas más precisos puede exacerbar las tendencias discriminatorias de estas tecnologías, por ejemplo, cuando se utilizan de forma indebida para vigilar a las minorías étnicas (Taylor, 2019; Hassein, 2017). Los trabajos a favor de la diversificación de las entradas de datos, según Pasquale (2015), formarían parte de la primera ola de la investigación sobre la “responsabilidad algorítmica”, mientras que el segundo grupo, conocido como los “estudios algorítmicos críticos”, estaría más preocupado por las fuerzas estructurales, las cuestiones de gobernanza y las desigualdades de poder endógenas a las sociedades de control.

Este segundo grupo encuentra uno de sus principales referentes en el libro de Noble (2018) *Algorithms of Oppression*. El texto arroja evidencias sobre los sesgos contra las personas de color en los motores de búsqueda. Según Noble (2018) los algoritmos refuerzan las relaciones sociales opresivas y promulgan nuevas discriminaciones raciales. La autora denomina este proceso como *redlining* tecnológico (tachado o discriminación tecnológica). Buolamwini y Gebru (2018) dan buena cuenta de la tesis principal de Noble cuando evalúan los tres programas comerciales más utilizados de reconocimiento facial en los Estados Unidos. Los resultados muestran que las mujeres de piel más oscura experimentan mayores tasas de error: un 34,7% frente al 0,8% para los hombres de piel clara. El informe de Lo (2022) y el trabajo de Wang et al. (2022) ofrecen nuevos ejemplos de los posibles sesgos racistas de estos programas.

El informe coordinado por Lo (2022) propone el término “criminalización de los datos” para denunciar y desnaturalizar “las prácticas de clasificación predictivas impulsadas para justificar el castigo y la exclusión social” de las personas migrantes en el contexto estadounidense. Desde una aproximación afín, Wang et al. (2022) sostienen que el Servicio de Inmigración y Control de Aduanas (ICE) opera como una agencia de vigilancia nacional. En sus esfuerzos por detener y deportar a los registros digitales de los gobiernos estatales y locales, el ICE funciona como una infraestructura de vigilancia “sin ningún tipo de supervisión judicial, legislativa o pública” (Lo, 2022). En estos programas de evaluación del riesgo, según Harcourt (2015), los antecedentes penales son uno de los principales factores predictivos, además de un indicador indirecto de la raza. Završnik (2021, p. 632) sugiere que la asociación entre los antecedentes penales y la sentencia penal no es un indicador fiable de riesgo, sino que refleja, en cambio, las desigualdades raciales en el sistema de justicia penal.

Otra idea común a la mayoría de los estudios críticos algorítmicos es la necesidad de abogar por la equidad y la transparencia en los sistemas algorítmicos. Lin y Chen (2023) amplían los requisitos cuando proponen la certeza, la interpretabilidad y la confiabilidad como los tres principios éticos fundamentales para los sistemas algorítmicos. Barabas et al. (2017) insisten en la necesidad de analizar los algoritmos subyacentes a las decisiones automatizadas para identificar y mitigar los posibles sesgos.

Esta última oleada de estudios críticos algorítmicos, más afines a las propuestas de “responsabilidad algorítmica”, propone atender igualmente a nuevos conceptos y actuaciones policiales entre los que se incluyen: el “análisis de sentimientos”, la “minería de opiniones” y el “tratamiento computacional de la subjetividad”, los cuales, como apunta [Završnik \(2021, p. 627\)](#) “reorganizan y difuminan los límites en el ámbito de la seguridad y el control de la delincuencia”.

Otros trabajos recientes, además de reconocer los sesgos existentes, apuestan por soluciones técnicas para conseguir mayores dosis de justicia e igualdad. Estas posturas tan comprometidas como propositivas se hallan representadas en los trabajos de [Berk, Kuchibhotla y Tchetgen \(2023\)](#), [Skeem y Lowenkamp \(2020\)](#) y [Mitchell et al. \(2021\)](#). También dejan constancia de estas posturas [Ashford, Spivak y Shepherd \(2022\)](#) cuando realizan una revisión sistemática de los desafíos para optimizar la equidad racial en los instrumentos de evaluación de riesgo en convivencia con una mayor precisión. Por su parte, [Berk et al. \(2021\)](#) proporcionan un examen integrado de la equidad y precisión en las evaluaciones de riesgos de la justicia penal a partir de la revisión de la literatura existente en criminología, informática y estadística. Con el mismo propósito de diseñar instrumentos más específicos y eficaces para la evaluación del riesgo cabe mencionar los trabajos centrados en el refinamiento de las tipologías de las víctimas de violencia de género ([Pineda et al., 2023](#); [Vignola-Lévesque y Léveillé, 2022](#); [Santos et al., 2022](#)).

Pero, en materia específica de violencia de género ¿Cuáles son las oportunidades e incertidumbres derivadas del uso de algoritmos y de los sistemas de decisión automatizados?

OPORTUNIDADES E INCERTIDUMBRES ALGORÍTMICAS EN MATERIA DE VIOLENCIA DE GÉNERO

La violencia de género ha sido una de las prioridades desde sus inicios de las políticas sociales basadas en las evaluaciones de riesgo ([Grogger et al., 2021](#)). Los trabajos de [Dutton y Kropp \(2000\)](#), [Campbell, Webster y Glass \(2009\)](#), [Ericson \(1997\)](#), [Rodríguez-Rodríguez et al. \(2020\)](#) y [Zambrano-Guerrero y Rodríguez-Pabón \(2021\)](#) dan buena muestra del interés suscitado por la aplicación de sistemas de decisión automatizados en la predicción de la violencia de género. En este sentido, [Delgadillo-Alemán et al. \(2019\)](#) utilizaron los datos proporcionados por el Instituto Mexicano de las Mujeres y, con la ayuda de ecuaciones diferenciales, demostraron la capacidad de su modelo para diagnosticar el riesgo de violencia de género en una determinada pareja. [De la Poza, Jódar y Barreda \(2016\)](#) desarrollaron un modelo matemático para inferir la incidencia oculta de la violencia de género. El modelo incluye factores como la conciencia social de los hombres, la edad, el consumo de drogas y las estadísticas de mujeres asesinadas, todo ello para estimar la población oculta de agresores. Desde esta misma perspectiva, [Rodríguez-Rodríguez et al. \(2020\)](#) comparan cuatro técnicas de algoritmos para predecir el número de denuncias por violencia de género presentadas ante un juzgado en España, en un horizonte de predicción de seis meses.

La mayoría de los países europeos han incorporado sistemas de evaluación del riesgo como parte destacada de sus políticas contra la violencia de género. Los sistemas más utilizados son el *Spousal Assault Risk Assessment (SARA)* [Valoración del riesgo de violencia contra la pareja] (Dinamarca, Irlanda, Eslovaquia, Chequia, Suecia, Italia, Portugal y España), el *Domestic Abuse, Stalking, and Harassment and Honor-Based Violence (DASH)* [Maltrato doméstico, acoso, hostigamiento y violencia contra la integridad] (Estonia, Irlanda, Eslovaquia, Reino Unido), el *Brief Spousal Assault Form for Risk Assessment (B-SAFER)* [Formulario breve para la valoración del riesgo de violencia contra la pareja] (Italia y Suecia), además del *Ontario Domestic Assault Risk Assessment (ODARA)* [La valoración Ontario del riesgo de violencia doméstica] (Alemania y Eslovaquia)

y el *Danger Assessment (DA)* [Evaluación del peligro] (Alemania, Italia, Austria, Portugal y Suecia) ([Instituto Europeo de Igualdad de Género \[European Institute for Gender Equality\] EIGE, 2019](#); [Fernández Teruelo, Fernández-Rivera González y García Amez, 2022](#)).

La evaluación del riesgo de violencia de género se lleva a cabo de acuerdo con diferentes enfoques y, en algunos casos, se utiliza una combinación de estos. Los principales modelos de evaluación del riesgo de violencia de género que prevalecen en el contexto europeo son: (i) la toma de decisiones clínicas no estructuradas, (ii) el enfoque actuarial y (iii) el enfoque del juicio profesional estructurado ([EIGE, 2019](#)).

El modelo no estructurado recurre a técnicas más cualitativas y flexibles, menos estructuradas. El modelo de evaluación actuarial del riesgo persigue una estimación precisa de la probabilidad de reincidencia y supone el uso de una lista de control de factores de riesgo (como los antecedentes penales) que actúan a modo de indicadores de reincidencia/homicidio. Las herramientas de evaluación del riesgo actuarial más utilizadas son el *Ontario Domestic Assault Risk Assessment (ODARA)* y el *Danger Assessment (DA)*. El enfoque estructurado supone una combinación de los otros dos enfoques, el actuarial y el clínico no estructurado, con el propósito de superar sus respectivas limitaciones. En él se hace hincapié en el desarrollo de directrices y herramientas para la evaluación del riesgo (enfoque actuarial), que sean lo suficientemente flexibles como para tener en cuenta las especificidades de cada caso de violencia (enfoque clínico). Dos de las herramientas más utilizadas que siguen este enfoque son el *Spousal Assault Risk Assessment (SARA)* y el *Brief Spousal Assault Form for Risk Assessment (B-SAFER)*.

En el Estado Español el sistema más empleado es VioGén, un programa gubernamental a cargo del Ministerio del Interior dirigido a recopilar, analizar y proporcionar medidas preventivas relacionadas con los casos de violencia de género. VioGén se creó en cumplimiento de la Ley Orgánica 1/2004, de 28 de diciembre, de Medidas de Protección Integral contra la Violencia de Género. Se trata de un cuestionario que administra la Policía Nacional, la Guardia Civil y algunas Policías Autonómicas y Locales a las mujeres que han sido víctimas de violencia de género para evaluar el riesgo de exposición a este tipo de violencia. Las valoraciones del nivel de riesgo o pronósticos individuales del sistema VioGén se obtienen a partir de valores algorítmicamente calculados en base a cohortes de personas que han experimentado distintos niveles de riesgos.

Los responsables de la administración del sistema VioGén son las Fuerzas y Cuerpos de Seguridad del Estado. No obstante, las Instituciones Penitenciarias, Juzgados, Institutos de Medicina Legal y Ciencias Forenses, Oficinas de Asistencia a las Víctimas, Fiscalías, Servicios Sociales y organismos de Igualdad de las diferentes Comunidades Autónomas también tienen acceso al sistema VioGén ([Ministerio del Interior, 2019](#)) y “pueden consultar el sistema para obtener información sobre situaciones específicas o niveles de riesgo de las víctimas” ([EIGE, 2019](#), p. 53). El sistema admite el acceso simultáneo de múltiples usuarios. Actualmente, ofrece soporte para archivar y procesar en paralelo todos los informes de delitos de género que ocurren en todo el país. Según estos informes, el sistema emite una predicción del riesgo de reincidencia para cada caso.

En el País Vasco y Cataluña se utilizan herramientas de valoración del riesgo específicas para su territorio. En el País Vasco se utiliza la Escala de Predicción de Riesgo Grave de Violencia contra la pareja (EPV-R), desarrollada por el Instituto Vasco de Criminología ([Echeburúa et al., 2009](#)). La EPV-R es un instrumento actuarial utilizado por psicólogos forenses, los profesionales de la policía, de la justicia y de los servicios sociales en el País Vasco para evaluar el riesgo de violencia grave contra la pareja en los casos de violencia doméstica y permitir “la adopción de medidas de protección a las víctimas, cuando se produce la primera denuncia, adecuadas a sus necesidades específicas y basadas en criterios empíricos” ([Echeburúa et al., 2010](#), p. 1055).

La EPV-R evalúa los principales factores de riesgo, incluyendo antecedentes de violencia, consumo de sustancias y amenazas. En una evaluación posterior de la herramienta los/as autores/as identificaron que los ítems que resultaban más

discriminativos del grupo de riesgo alto de violencia, presentes en un 75% de los casos analizados, eran los “relacionados directamente con características propias del agresor, tales como las conductas controladoras o de celos, el consumo abusivo de alcohol/ drogas y la falta de empatía, así como el incremento de las amenazas y de la violencia física” (Echeburúa et al., 2010, p. 1056).

En Cataluña, los Mossos d'Esquadra, aplican el Qüestionari Policial de Valoració de Risc [Cuestionario policial de valoración del riesgo] (QPVR), que consta de 25 a 30 preguntas. Las preguntas no se hacen directamente, sino que los Mossos d'Esquadra las cumplimentan a partir de las propias declaraciones de las mujeres. En un estudio reciente sobre el uso de herramientas algorítmicas en la valoración de la situación de riesgo de la mujer, que incluye el análisis de expedientes judiciales de 2014 a 2019 sobre peticiones de órdenes de protección en Cataluña, Bodelón González et al. (2019, p. 70), se señalan que los Mossos d'Esquadra consideran al QPVR como un instrumento objetivo que, además, les proporciona una estimación o “adelanto” a la hora de valorar la situación de riesgo de la mujer.

Por otro lado, el Circuito Barcelona contra la Violencia hacia las Mujeres utiliza el Protocolo de valoración del riesgo de violencia contra la mujer por parte de su pareja o ex pareja (RVD-BCN), un instrumento semi-estructurado que tiene como objetivo identificar el riesgo de violencia grave contra la mujer por parte de su pareja o ex pareja. Este protocolo integra los principales ámbitos implicados en el abordaje de la violencia machista en Barcelona (ámbito sanitario, social, policial, educativo, justicia y judicial), e implica directamente tanto a la administración autonómica, como al Ayuntamiento de Barcelona, la Fiscalía, los consorcios Sanitario, de Educación y de Servicios Sociales de la ciudad de Barcelona (Álvarez Freijo et al., 2011). El RVD-BCN se basa en una evaluación multidimensional que incluye factores de riesgo individuales, relacionales y contextuales desde su puesta en marcha en 2011.

Un aspecto común a los programas actuariales es la administración de un cuestionario a las víctimas (Grogger et al., 2021). Algunos de estos cuestionarios incluyen rúbricas y puntuaciones preestablecidas, mientras que otras se dejan al arbitrio de las personas entrevistadoras expertas (Kropp, 2004; Messing y Thaller, 2013). Por ejemplo, el primer componente del DASH es un cuestionario que consta de 27 elementos, administrado a la víctima por un/a agente de policía. Su segundo componente es el grado de riesgo atribuido por el/la agente, que equivale a una predicción del riesgo futuro (Richards, 2009). La clasificación del riesgo que hace el/la agente está basada en las respuestas de la víctima al cuestionario, si bien en última instancia tiene que recurrir a su propio juicio profesional para emitir una evaluación de riesgo (Robinson et al., 2016). La Escala de Predicción del Riesgo de Violencia Grave contra la Pareja (EPV-R), desarrollada y validada en el País Vasco por Echeburúa, Fernández-Montalvo, de Corral y López-Goñi (2009), es un cuestionario breve, heteroaplicado, que incluye 20 ítems, para la evaluación de los 10 principales factores de riesgo de violencia de género.

La *Guía de procedimiento VPR5.0 y VPER4.1 Protocolo de valoración policial del riesgo y gestión de la seguridad de las víctimas de violencia de género* (Ministerio del Interior, 2019) detalla los distintos formularios que incluye el protocolo VioGén, además de la definición de los factores e indicadores. En el sistema VioGén la información inicial se recopila a través de un formulario estándar llamado VPR, que proviene de la 'Valoración Policial del Riesgo'. Los datos posteriores son registrados mediante un formulario diferente llamado VPER, que deriva de la 'Valoración Policial de la Evolución del Riesgo'. Como señalan González-Prieto et al. (2021), el formulario VPR (actualmente en la versión VPR5) permite recabar datos básicos para la evaluación, que serán posteriormente analizados por el programa VioGén. En esta fase de la evaluación, el programa VioGén se limita a estimar la probabilidad de reincidencia (en una escala de 0 a 4, probabilidad escasa-extrema) en lugar de su intensidad, gravedad o frecuencia de ocurrencia. En base a la estimación arrojada por VioGén, el/la agente de policía determinará diferentes medidas preventivas.

A diferencia de otras herramientas actuariales (Como SARA), VioGén, el Questionari Policial de Valoració de Risc (QPVR) y el EPV-R contemplan la posibilidad de que los profesionales policiales modifiquen, al alza, el resultado automático en función de la información o percepción que tengan del caso o su experiencia. Así lo muestran, por ejemplo, las siguientes declaraciones de Mossos d'Esquadra tomadas de [Bodelón González et al. \(2019, p. 72\)](#):

Porque a veces como, claro, que valoras una agresión física muy grave, claro, esto valora algo muy grave, pero quizás el que está recogiendo la denuncia percibe que hay un riesgo y no le gusta el nivel que ha dado. Tiene el poder de subirlo un nivel... (P5).

[...] hay veces que la experiencia evidentemente te da el plus que no tiene la máquina, para subir el nivel de riesgo si dices no, esta persona, por lo que me está explicando, necesita una protección policial, y por lo que sea, el cuestionario no la da. Entonces, tiene la potestad de subir el nivel de riesgo (P2).

El nivel de riesgo es reevaluado cada cierto tiempo o, en caso de producirse un nuevo episodio de violencia durante la fase de seguimiento (o periodo de vigilancia), por medio de un formulario de Valoración Policial de la Evolución del Riesgo (VPER). El empleo de dos formularios distintos es una particularidad de protocolo VioGén. Esta valoración permite calcular el número de reincidencias, siendo este uno de los principales índices utilizados para estimar la probabilidad de nueva ocurrencia.¹

Desde su creación en 2007, el sistema VioGén ha analizado 701.563 casos y ha realizado más de 5,4 millones de valoraciones policiales de riesgo y está considerado como uno de los principales sistemas de valoración de riesgo de violencia de género a escala mundial ([Ministerio del Interior, 2022](#)). Las sucesivas evaluaciones de la eficacia del sistema VioGén han estimado un alto valor predictivo de la VPR ([López-Ossorio, González-Álvarez y Andrés-Pueyo, 2016](#); [López-Ossorio et al., 2019](#); [López-Ossorio et al., 2021](#)).

A pesar del uso tan amplio y la centralidad otorgada a los programas algorítmicos actuariales como VioGén, estos programas están expuestos a un continuo cuestionamiento. Un primer grupo de críticas están dirigidas al bajo poder predictivo observado en contextos de justicia penal ([Farrington y Tarling, 1985](#); [Campbell, et al., 2005](#); [Éticas, 2022](#)). [Thornton \(2017\)](#) intentó predecir los homicidios domésticos y actos de violencia graves a partir de los registros disponibles en una base de datos en el condado de Dorset (Reino Unido) y de la evaluación obtenida a través de las evaluaciones de riesgo de la policía local, incluido el protocolo DASH. Los resultados del estudio indicaban que menos del 50% de estos casos contaba con una intervención o presencia policial previa, por lo que la predicción no podía basarse únicamente en los registros policiales. También concluyó que en los casos en los que se disponía de registros, el protocolo no evaluaba de 'alto riesgo' al 89% de los casos. Según [Chalkley y Strang \(2017\)](#), el 67% de los casos de violencia mortal fueron clasificados por el sistema como carente de riesgo o falsos negativos, incluso cuando los agresores habían tenido antecedentes o contacto previo con la policía.

Otras críticas apuntan a la inconsistencia en la toma de decisiones ([Grove y Meehl, 1996](#); [Gottfredson y Moriarty, 2006](#)). En el caso de DASH el estudio a cargo del colectivo [Her Majesty's Inspectorate of Constabulary \(2014\)](#) informó que, de un total de 28 unidades de policía, 10 unidades calificaron menos del 10% de los casos de maltrato doméstico como de "alto riesgo", mientras que tres unidades de policía designaron más del 80% como de "alto riesgo". Este nivel de discrepancia entre juicios de profesionales pone en duda la fiabilidad de las predicciones. Otros estudios ofrecen resultados parecidos sobre la fiabilidad de otros sistemas actuariales de evaluación de riesgo en materia de violencia de género en distintos entornos ([Gottfredson y Moriarty, 2006](#); [Kahneman et al., 2016](#)).

¹ Para un metaanálisis y una revisión sistemática de la validez y fiabilidad de los principales sistemas y herramientas de evaluación del riesgo de violencia de género en el contexto internacional, véase respectivamente [Van Der Put, Gubbels y Assink \(2019\)](#) y [Graham et al. \(2021\)](#).

En el estudio referido acerca de órdenes de petición de alejamiento en Cataluña, [Bodelón González et al. \(2019\)](#) concluyen que, contrariamente a las valoraciones positivas del QPVR por parte de los Mossos d'Esquadra, la abogacía y la judicatura, como las otras partes que participan en el proceso de evaluación y la toma de decisiones acerca de las medidas de seguridad a tomar, no consideran el QPVR una herramienta que dé una pauta de la situación real de riesgo de la mujer. Una gran parte de los miembros de las judicatura y abogacía entrevistados, según [Bodelón González et al. \(2019\)](#), no dan valor alguno o ignoran el resultado del cuestionario. Esto es así hasta el punto de que “algunos miembros de la judicatura, que con riesgos no apreciados o bajos dan órdenes de protección”, o que muchas veces solo tienen en cuenta las valoraciones de riesgo alto o muy alto debido a que, “si normalmente no aprecian el riesgo, aunque exista, cuando lo aprecian es que la situación puede ser muy grave” ([Bodelón González et al., 2019, p. 71](#)).

En lo que respecta al sistema VioGén, encontramos críticas referidas tanto al inicio del proceso y la recogida de información, como al análisis y la interpretación de los datos de acuerdo con el protocolo establecido. Los errores de predicción (altos porcentajes de riesgo no percibido, falsos negativos y falsos positivos) advierten de inconsistencias que deberían ser mejoradas y complementadas con otro tipo de actuaciones. Los datos y la información que se incluyen en el sistema son de vital importancia, por eso en la guía del formulario se aconseja al profesional que lo cumplimenta, que revise toda la información disponible (atestado policial, informes de servicios sociales, declaraciones de familiares), además de la declaración de la víctima. Sin embargo, diferentes estudios indican que ya sea por falta de tiempo o de recursos esto no siempre se hace así ([Olaciregui, 2021](#)).

Por otro lado, [Muñoz-Vicente y López-Ossorio \(2016\)](#) indican como una limitación que VioGén no incluya de una manera solida algunos indicadores que influyen en la conducta violenta relacionados con la valoración psicológica del agresor y la víctima, así como creencias, distorsiones cognitivas o patologías que tienden a justificar la violencia. Tampoco el conocimiento y la formación de los agentes policiales sobre este tipo de cuestiones permite suplir esta carencia. En este sentido, hay que advertir que además del peritaje policial, existe la posibilidad de hacer una valoración forense en sede judicial. Esto cobra particular interés en los casos clasificados de “especial relevancia” en los que, junto a la VPR, VioGén genera una diligencia automatizada que recomienda al juez la práctica de evaluaciones forenses complementarias en las Unidades de Valoración Forense Integral. Sin embargo, según la [Fiscalía General del Estado \(2019\)](#), el uso del *Protocolo médico-Forense de valoración urgente del riesgo de violencia de género* sigue siendo muy escaso, y reconoce que los informes policiales de valoración del riesgo son en la práctica el único instrumento disponible para adoptar o no medidas cautelares de protección de la víctima.

En cuanto al análisis y la interpretación de la información, VioGén recibe críticas centradas en sesgos de tipo estadístico, y en otros sesgos más sociales en línea con los estudios algorítmicos críticos. Entre las críticas de corte estadístico están aquellas que reparan en la falta de medición directa de la evaluación del riesgo, la cual sólo se estima a partir de una medida indirecta: las recaídas/reincidencia. Otras limitaciones estadísticas señaladas son la marcada dependencia del sistema respecto de los factores de riesgo estáticos (pertenecen al pasado de la persona -víctima/agresor- y son difícilmente modificables), que no captan cómo puede fluctuar el riesgo de violencia a lo largo del tiempo y en función de los diferentes contextos. Precisamente, uno de los rasgos diferenciales de las valoraciones del riesgo, frente a los tradicionales y criticados pronósticos de peligrosidad, consiste en contemplar tanto factores individuales (y, por ende, en buena medida estáticos) como factores ambientales y contextuales (cambiantes), que pueden influir en su comportamiento. Según fuentes e informes oficiales, el número de casos calificados con cada nivel de riesgo se mantiene constante a lo largo del tiempo, con variaciones muy pequeñas. Estas limitaciones a la

hora de capturar las fluctuaciones de la violencia de género suponen una perspectiva más orientada a la predicción del riesgo, que a la prevención de la violencia. Como señalan [González-Prieto et al. \(2021\)](#), la ausencia de reincidencia en violencia de género puede responder a las medidas de protección adoptadas o ser debida a los fallos de detección del sistema. A esta incertidumbre se suma el hecho que el algoritmo VioGén está aprendiendo a clasificar el riesgo de sufrir o ser víctima de la violencia de género, una variable en sí misma intangible ([González-Prieto et al., 2021](#)).

El segundo grupo de críticas recibidas por el sistema VioGén se centra en los sesgos sociales y en la necesidad de reconsiderar la falta de valores éticos y de presencia humana en el sistema. El informe basado en la auditoría del sistema VioGén a cargo de [Éticas Research and Consulting \(Éticas, 2022\)](#) indica que más del 80% de las víctimas consultadas tuvieron dificultades a la hora de responder a los formularios de 'Valoración Policial del Riesgo' (VPR) o de 'Valoración Policial de la Evolución del Riesgo' (VPER). Dependiendo de su nivel educativo y cultural, las mujeres entienden mejor o peor las preguntas. La auditoría informaba, además, que la cumplimentación del cuestionario coincide con el momento de interponer la denuncia policial por violencia de género, por lo que muchas mujeres se encuentran en tal estado de shock que les hace difícil contestar adecuadamente. [Éticas \(2022\)](#) también señala que VioGén tiene una supervisión humana mínima y que las fuerzas policiales a cargo de VioGén han modificado sólo el 5% de la puntuación de riesgo calculada por el algoritmo. Además, esta participación disminuye en momentos de mayor presión de trabajo, como ocurrió durante la pandemia del COVID-19.

ENTRE DEPENDENCIAS ALGORÍTMICAS Y DECISIONES ÉTICAMENTE VERIFICADAS

La comisaría de Policía Nacional María Jesús Cantos, responsable del Área contra la Violencia de Género de Interior, tras un significativo repunte de homicidios por violencia de género, en diciembre de 2022 en unas declaraciones a título personal dijo ser partidaria de que todas las mujeres a las que el sistema VioGén estimara que corren "riesgo de homicidio" se les asignara el dispositivo telemático de detección y prevención de la violencia ([Europa Press, 2022](#)). Este comentario acerca de la conveniencia de generalizar el dispositivo de detección de riesgo reparaba en la escasez de recursos, en lugar de los posibles fallos e insuficiencias de VioGén, como una de las principales razones del repunte de homicidios experimentado durante esos meses. Por otro lado, la propia comisaría indicaba que algunos de los homicidios recientes eran casos inactivos, pues el sistema desactiva aquellos que no han registrado ninguna nueva incidencia tras un periodo de diez años. En consecuencia, las autoridades consideraron entonces alargar la vida de los casos dentro del sistema. Como se puede ver, los esfuerzos acordados un año atrás apuntaban en la dirección de reforzar el algoritmo, sin ponerlo en tela de juicio ni pensar en actuaciones complementarias.

Pero, no solo se trata de una cuestión de falta de tiempo y de recursos. Apostar por la automatización en la toma de decisiones como parte de los dispositivos del sistema penal y de las políticas de seguridad confiere un aura de objetividad a las decisiones adoptadas ([Harcourt, 2015](#)), además de otorgar una falsa sensación de seguridad ([Hannah-Moffat, 2019](#)). Al mismo tiempo, las lógicas algorítmicas actuariales, como hemos señalado, pueden llegar a fomentar sistemas de vigilancia que redirigen las actividades de las fuerzas de seguridad y del sistema jurídico penal hacia actos no cometidos e individuos potencialmente sospechosos.

Un ejemplo de estas prácticas de vigilancia preventiva lo encontramos en la legislación antiterrorista alemana posterior al 11-S y su desarrollo del constructo "terrorista durmiente", basado en los supuestos estados psicológicos personales y

otras circunstancias conocidas de presuntos terroristas. Esta identificación algorítmica preventiva, permite criminalizar antecedentes delictivos menores y distantes en el tiempo además de alteraciones psicológicas transitorias. La idea de la detección preventiva supuso el arresto de ciudadanos alemanes musulmanes debido a la “situación general de amenaza” (Završnik, 2021, p. 627). Jiménez y Douhaibi (2023) han denunciado las asunciones de los sistemas actuariales preventivos de naturaleza algorítmica (PRODERAI/PRODERAEV) utilizados en España tras los atentados en las Ramblas de Barcelona y en Cambrils, en 2017, para prevenir la radicalización, que recuerdan el constructo “terrorista durmiente”.

Un año después de haber empezado a escribir este trabajo, reparamos en las nuevas cifras de violencia de género ofrecidas por el ministro del Interior, Fernando Grande Marlaska, en su reunión con la nueva ministra de Igualdad, Ana Redondo, a principios de diciembre 2023: 55 mujeres fueron asesinadas por su pareja o expareja, seis más que en 2022, de las cuales solo 5 de los 13 casos denunciados estaban activos en VioGén (García Martín y Rubio, 2023; Soler, 2023). La respuesta de las autoridades ante esta situación reitera la necesidad de ampliar los recursos destinados al programa VioGén. Según el ministro del Interior, desde 2018 se han duplicado las mujeres protegidas por VioGén y el número de funcionarios para su atención. Además, se anima encarecidamente al entorno social de las presuntas víctimas a que avise a las fuerzas y cuerpos de seguridad. De este modo, se podría activar el “protocolo cero” que permite a la policía investigar el supuesto caso de violencia de género sin necesidad de la denuncia de la víctima. Los avisos por parte del entorno de las víctimas representan actualmente el 30% de los casos de VioGén.

En estas declaraciones del ministro del Interior y de la ministra de Igualdad, la violencia de género, aludida en un primer momento como un “problema estructural”, pasa a ser calificada como un “problema de Estado” y un “compromiso político” respectivamente. A diferencia de las declaraciones de la comisaria de Policía Nacional María Jesús Cantos en diciembre de 2022, donde se planteaba la necesidad de destinar más recursos al sistema VioGén, ahora, un año después, las medidas de seguridad de las víctimas requieren a su vez “la confianza en las Fuerzas y Cuerpos de Seguridad”. Como se deriva de las declaraciones mencionadas, se requiere del “entorno” social una doble confianza: por un lado, en la capacidad de nuestras fuerzas del orden de protegernos y de hacer un uso preciso y veraz de los sistemas de evaluación y detección del riesgo de violencia de género; y, por otro, confianza, en el sentido de cercanía y familiaridad, para poder avisarles de las violencias que puedan ocurrir en nuestro entorno.

La denuncia ciudadana, y como paso previo re/construir la confianza en las fuerzas de orden, y sus procedimientos algorítmicos, aparecen ahora como el gran reto en las políticas contra la violencia de género. De ahí que el ministro del Interior haya insistido en la idea de que “La denuncia es la que nos permite abrir el paraguas protector del Estado español”. Ni las predicciones algorítmicas ni los sistemas policiales preventivos son suficientes por sí mismos para mitigar el riesgo de la violencia de género.

Más allá de la violencia de género, la necesidad de complementar los programas asistidos por IA con la dimensión más social, además de la verificación humana, también la podemos apreciar en el modo que las Fuerzas de Defensa de Israel (FDI) utilizan un sistema basado en IA, llamado Habsora (El Evangelio), para seleccionar de manera masiva los objetivos a bombardear en Gaza (Davies, McKernan y Sabbagh, 2023). El rendimiento de esta unidad secreta de selección de objetivos, potenciado por “las capacidades de la IA”, cuenta con la participación de “cientos de soldados y oficiales” con el propósito, según la FDI, de que “haya una coincidencia total entre la recomendación de la máquina y la identificación llevada a cabo por una persona” (Davies, McKernan y Sabbagh, 2023, párr. 12). Especialistas en inteligencia artificial y derecho internacional humano, como Marta Bo, afirman que, incluso cuando hay humanos en estos procesos de tomas de decisiones, “existe el peligro de desarrollar una ‘preferencia por lo automatizado’, así

como una ‘dependencia excesiva en sistemas que terminan ejerciendo una influencia demasiado grande sobre decisiones humanas complejas’ (citado en [Davies, McKernan y Sabbagh, 2023](#), párr. 34).

CONCLUSIONES

En una “sociedad de incertidumbres”, que busca certezas para amortiguar el malestar provocado por los crecientes problemas sociales, así como de los posibles daños no deseados que se deriven de la toma de decisiones institucionales, se recurre a sistemas de decisión y lógicas algorítmicas actuariales.

Bajo el ideal de objetividad que confiere la tecnología y la matematización de la información, las herramientas algorítmicas aparecen como supuestamente exentas de sesgos (subjetivos) e intereses espurios (humanos). Todo ello otorga a los sistemas de evaluación y gestión de riesgos (como VioGén) una aureola de certezas y seguridades. No obstante, la predicción sobre el riesgo de violencia que ofrecen los instrumentos actuariales no son pronósticos individuales efectuados en relación con sujetos concretos, sino valores calculados en base a cohortes de personas que presentan una serie de características y factores similares a los del individuo al que se aplica la estimación de riesgo de reincidencia. Usos indebidos de estos sistemas algorítmicos de toma de decisiones agravan habitualmente la situación de grupos y sectores sociales ya de por sí vulnerables o en el foco de la estigmatización social.

Los múltiples sesgos sociales y estadísticos intrínsecos a los programas de evaluación de riesgos de violencia de género no han supuesto hasta la fecha un impedimento para su institucionalización y expansión en las políticas y legislaciones de una buena parte de países del Norte Global. Por el contrario, como hemos visto en el caso de España y el uso del sistema VioGén, la respuesta de las autoridades competentes ante los posibles fallos del sistema consiste en reforzar, legitimar y ahora, en el momento que escribimos este texto, “amplificar” la precisión de sus valoraciones a través de incentivar las denuncias por parte del entorno de las presuntas víctimas de género. No obstante, las lógicas predictivas algorítmicas siguen utilizándose sin un debate suficiente sobre la ética de sus modelos. Esta dimensión ética, que va más allá de la transparencia y posibilidades de mejorar de los pronósticos, aparece como un valor añadido en lugar de una barrera u obstáculo para la ciencia y el desarrollo tecnológico. La propia incertidumbre de los pronósticos del riesgo inherentes a las valoraciones o tomas de decisiones en materia de violencia de género, así como sus limitaciones, han de ser social y éticamente considerados, verificados, para lo que el “factor” humano sigue siendo imprescindible.

FINANCIACIÓN

Este artículo ha sido escrito en el marco del proyecto “Visibilizar y Dimensionar las violencias sexuales en las universidades” (RTI2018-093627-B-I00) que está cofinanciado por el Ministerio de Ciencia, Innovación y Universidades (MCIU), Agencia Estatal de Investigación (AEI), Fondo Europeo de Desarrollo Regional (FEDER), dentro del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad 2018. Las publicaciones y comunicaciones que derivan de este trabajo reflejan únicamente las visiones de sus autoras/es, y el Ministerio no se hace responsable de cualquier uso derivado de las informaciones contenidas en las mismas.

REFERENCIAS

- Álvarez Freijo, M., Andrés-Pueyo, A., Augé Gomà, M., Choy Vilana, A., Fernández Velasco, R., Fernández Rodríguez, C., Foulon, H., López Ferré, S., Martínez-Izquierdo, M.T., Martínez Bueno, C., Saiz-Lloret, M., y Serratusell Salvadó, L. (2011). *Protocolo de valoración del riesgo de violencia contra la mujer por parte de su pareja o ex pareja (RVD-BCN)*. Circuito Barcelona contra la Violencia hacia las Mujeres.
<https://bcnroc.ajuntament.barcelona.cat/jspui/bitstream/11703/91573/1/10964.pdf>
- Angwin, J., Larson, J., Mattu, S. y Kirchner, L. (2016, 23 de mayo). Machine bias. *ProPublica*.
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Ashford, L. J., Spivak, B. L., y Shepherd, S. M. (2022). Racial fairness in violence risk instruments: a review of the literature. *Psychology, Crime & Law*, 28(9), 911-941.
<https://doi.org/10.1080/1068316X.2021.1972108>
- Barabas C., Dinakar K., Ito, J., Virza, M. y Zittrain J. (2017). *Interventions over predictions: Reframing the ethical debate for actuarial risk assessment*. Cornell University.
<https://doi.org/10.48550/arXiv.1712.08238>
- Berk, R. A., Kuchibhotla, A. K. y Tchetgen Tchetgen, E. (2023). Fair Risk Algorithms. *Annual Review of Statistics and Its Application*, 10, 165-187.
<https://doi.org/10.1146/annurev-statistics-033021-120649>
- Berk, R.A., Heidari, H., Jabbari, S., Kearns, M. y Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1), 3-44.
<https://doi.org/10.1177/0049124118782533>
- Berk, R.A. (2012). *Criminal justice forecasts of risk: A machine learning approach*. Springer Science & Business Media.
- Bodelón González, E., Barcons Campmajó, M., Ortiz Amaro, L., Murillo Blasco, E., Pisonero Francisco, A. y González Pardo, P. (2019, diciembre). *Órdenes de protección y derechos de las mujeres que han sufrido violencia de género: obstáculos para una efectiva protección*. El Institut Català de les Dones (ICD).
https://dones.gencat.cat/web/.content/03_ambits/docs/cdoc_publicacions_ordenes_proteccion.pdf
- Buolamwini, J. y Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1-15.
<http://proceedings.mlr.press/v81/buolamwini18a.html>
- Campbell, J. C., O'Sullivan, C., Rohel, J. y Webster, D. (2005). *Intimate Partner Violence Risk Assessment Validation Study. Final Report*. National Institute of Justice (NIJ).
<https://www.ojp.gov/ncjrs/virtual-library/abstracts/intimate-partner-violence-risk-assessment-validation-study-final>
- Campbell, J. C., Webster, D. W. y Glass, N. (2009). The danger assessment: Validation of a lethality risk assessment instrument for intimate partner femicide, *Journal of Interpersonal Violence* 24(4), 653-674. <https://doi.org/10.1177/0886260508317180>
- Europa Press (2022, febrero 5). *La comisaria de Viogén sobre las pulseras: "Las pondría a todas con riesgo de homicidio"* [Entrevista].
<https://www.europapress.es/videos/video-comisaria-viogen-pulseras-pondria-todas-riesgo-homicidio-20230205132136.html>

- 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>
- Copas, J. y Marshall, P. (1998). The offender group reconviction scale: A statistical reconviction score for use by probation officers, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 47(1), 159-171. <https://doi.org/10.1111/1467-9876.00104>.
- Davies, H., McKernan, B. y Sabbagh, D. (2023, 9 de diciembre). 'El Evangelio': cómo utiliza Israel la inteligencia artificial para seleccionar objetivos en Gaza. *The Guardian/El diario.es* [traducción de Francisco de Zárate]. https://www.eldiario.es/internacional/theguardian/evangelio-utiliza-israel-inteligencia-artificial-seleccionar-objetivos-gaza_1_10740704.html
- Dawes, R. M., Faust, D. y Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668-1674. <https://doi.org/10.1126/science.2648573>
- De la Poza, E., Jódar, L. y Barreda, S. (2016). Mathematical modeling of hidden intimate partner violence in Spain: A quantitative and qualitative approach. *Abstract and Applied Analysis*. <https://doi.org/10.1155/2016/8372493>
- Delgadillo-Alemán, S., Ku-Carrillo, R., Pérez-Amezcuca, B. y Chen-Charpentier, B. (2019). A mathematical model for intimate partner violence. *Mathematical and Computational Applications*, 24(1), 29. <https://doi.org/10.3390/mca24010029>
- Desmarais, S., Garrett, B. y Rudin, C. (2019). Risk assessment tools are not a failed 'minority report'. *Law360*. <https://www.law360.com/access-to-justice/articles/1180373/risk-assessment-tools-are-not-a-failed-minority-report->
- Dutton, D. G. y Kropp, P. R. (2000). A review of domestic violence risk instruments. *Trauma, Violence & Abuse*, 1(2), 171-181. <https://doi.org/10.1177/152483800001002004>
- Echeburúa, E., Amor, P. J., Loinaz, I. y De Corral, P. (2010). Escala de Predicción del Riesgo de Violencia Grave contra la pareja-Revisada-(EPV-R). *Psicothema*, 22(4), 1054-1060. <http://www.redalyc.org/articulo.oa?id=72715515077>
- Echeburúa, E., Fernández-Montalvo, J., Corral, P. y López-Goñi, J.J. (2009). Assessing risk markers in intimate partner femicide and severe violence. A new assessment instrument. *Journal of Interpersonal Violence*, 24(6), 925-939. <https://doi.org/10.1177/0886260508319370>
- Englezos, E. (2023). Policing by algorithm: NSW Police's Suspect Target Management Plan. *Alternative Law Journal*, 48(1), 17-24. <https://doi.org/10.1177/1037969X221147745>
- Ericson, R. V. (1997[1948]). *Policing the risk society*. University of Toronto Press. <http://search.ebscohost.com/login.aspx?direct=true&scope=sitedb=e000xnaAN=683064>
- Éticas (2022). *The external audit of the VioGén system*. Association Éticas Research and Innovation https://eticasfoundation.org/wp-content/uploads/2022/04/ETICAS_-_Auditori%CC%81a-Externa-del-sistema-VioGe%CC%81n_-_20220308.docx.pdf
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- European Institute for Gender Equality (EIGE) (2019). *Risk assessment and management of intimate partner violence in the EU*. Publications Office of the European Union. <https://doi.org/10.2839/37233>

- Farrington, D. P. y Tarling, R. (1985). Criminological prediction: An introduction. En D.P. Farrington y R. Tarling (Eds.), *Prediction in criminology* (pp. 2-33). SUNY Press.
- Fernández Teruelo, J. G., Fernández-Rivera González, P. y García Amez, Javier (Dir.) (2022). *Nuevas formas de prevención y respuesta jurídico-social frente a la violencia de género*. Thomson Reuters Aranzadi.
- Fiscalía General del Estado (2019). *Memoria de la Fiscalía General del Estado, 2018*. https://www.fiscal.es/memorias/memoria2019/FISCALIA_SITE/index.html
- Flores, A. W., Bechtel, K., y Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals and it's biased against blacks. *Federal Probation Journal*, 80(2), 38-46. https://www.uscourts.gov/sites/default/files/80_2_6_0.pdf
- García Martín, J., y Rubio, E. (2023, 7 de diciembre). *Repunte de asesinatos machistas - Se revisarán los criterios para dar de baja un caso en VióGén* [Reportaje]. <https://www.rtve.es/v/7030197/?t=55s>
- González-Prieto, Á., Brú, A., Nuño, J. C. y González-Álvarez, J. L. (2021). Machine learning for risk assessment in gender-based crime. *Computers and Society, arXiv preprint arXiv*, 2106.11847. <https://doi.org/10.48550/arXiv.2106.11847>
- Gottfredson, S. D. y Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency*, 52(1), 178-200. <https://doi.org/10.1177/0011128705281748>
- Graham, L. M., Sahay, K. M., Rizo, C. F., Messing, J. T. y Macy, R. J. (2021). The Validity and Reliability of Available Intimate Partner Homicide and Reassault Risk Assessment Tools: A Systematic Review. *Trauma, Violence, & Abuse*, 22(1), 18-40. <https://doi.org/10.1177/1524838018821952>
- Grogger, J., Gupta, S., Ivandic, R. y Kirchmaier, T. (2021). Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases. *Journal of Empirical Legal Studies*, 18(1), 90-130. https://www.nber.org/system/files/working_papers/w28293/w28293.pdf
- Grove, W. M. y Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293-323. <https://doi.org/10.1037/1076-8971.2.2.293>
- Hannah-Moffat, K. (2019). Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates. *Theoretical Criminology*, 23(4), 453-470. <https://doi.org/10.1177/1362480618763582>
- Harcourt, B.E. (2015). Risk as a proxy for race: The dangers of risk assessment. *Federal Sentencing Reporter*, 27(4), 237-243. <https://doi.org/10.1525/fsr.2015.27.4.237>
- Harcourt, B.E. (2006). *Against prediction: Profiling, policing and punishing in an actuarial age*. University of Chicago Press.
- Hassein, N. (2017, 15 de agosto). Against Black Inclusion in Facial Recognition. *Digital Talking Drum*. <https://digitaltalkingdrum.com/2017/08/15/against-black-inclusion-in-facial-recognition/>

- Her Majesty's Inspectorate of Constabulary (2014). *Everyone's business: Improving the police response to domestic abuse*. Home Secretary Inspectorate of Constabulary (HMIC). https://www.scie-socialcareonline.org.uk/everyones-business-improving-the-police-response-to-domestic-abuse/r/a11G0000_003FG_3FIAW
- Jiménez, A., y Douhaibi, A. (2023). The Islamophobic consensus: Datafying racism in Catalonia. En Z. Bednarz, y M. Zalnieriute (Eds.), *Money, Power, and AI: Automated Banks and Automated States* (pp. 152-170). Cambridge University Press. <https://doi.org/10.1017/9781009334297.012>
- Kahneman, D., Rosenfield, A. M., Gandhi, L. y Blaser, T. (2016). Noise: How to overcome the high, hidden cost of inconsistent decision making. *Harvard Business Review*, 94(10), 38-46. <https://hbr.org/2016/10/noise>
- Kleinberg, J., Mullainathan, S. y Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. En *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science*.
- Kropp, P R. (2004). Some questions regarding spousal assault risk assessment. *Violence against women*, 10(6), 676-697. <https://doi.org/10.1177/1077801204265019>
- Larson, J., Mattu, S., Kirchner, L. y Angwin, J. (2016, 23 de mayo). How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Lin, J., y Chen, C. (2023). The epistemic ethical concerns involving algorithms in intelligent communication. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 20(Special Issue), 27-36. <https://doi.org/10.5209/tekn.87552>
- Lo, P. (2022). *From data criminalization to prison abolition*. Community Justice Exchange Project. <https://abolishdatacrim.org/en>
- López-Flores, L. M. (2022). Colonialidad algorítmica: Racialización y sexualización mecanizada en el capitalismo digital: Una perspectiva de género. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 19(2), 113-122. <https://doi.org/10.5209/TEKN.78681>
- López-Ossorio, J. J., González-Álvarez, J. L., Loínaz, I., Martínez-Martínez, A. y Pineda, D. (2021). Intimate partner homicide risk assessment by police in Spain: The Dual Protocol VPR5.0-H. *Psychosocial Intervention*, 30(1), 47-55. <http://dx.doi.org/10.5093/pi2020a16>
- López-Ossorio, J. J., González-Álvarez, J. L., Muñoz Vicente, J. M., Urruela Cortés, C. y Andrés-Pueyo, A. (2019). Validation and calibration of the Spanish police intimate partner violence risk assessment system (VioGén). *Journal of police and criminal psychology*, 34, 439-449. <https://doi.org/10.1007/s11896-019-09322-9>
- López-Ossorio, J. J., González-Álvarez, J. L. y Andrés-Pueyo, A. (2016). Predictive effectiveness of the police risk assessment in intimate partner violence. *Psychosocial Intervention*, 25(1), 1-7. <https://doi.org/10.1016/j.psi.2015.10.002>
- McNamara, D., Graham, T., Broad, E. y Ong, C. S. (2019). Trade-offs in algorithmic risk assessment: An Australian domestic violence case study. En A. Daly, K. Devitt y M. Mann (Eds.) *Good data* (pp. 96-116). Institute of Network Cultures.
- Messing, J. Th. y Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence*, 28(7), 1537-1558. <https://doi.org/10.1177/0886260512468250>

- Ministerio del Interior (2022, 28 de julio). VioGén cumple 15 años con más 700.000 casos analizados y 5,4 millones de valoraciones de riesgo realizadas. <https://n9.cl/wm98b>
- Ministerio del Interior (2019). *Guía de procedimiento VPR5.0 y VPER4.1. Protocolo de valoración policial del riesgo y gestión de la seguridad de las víctimas de violencia de género*. <https://violenciadegenerotic.files.wordpress.com/2019/05/instruccion-4-2019.pdf>
- Mitchell, S., Potash E., Barocas S., D'Amour A, Lum. K. (2021). Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Applications*, 8, 141-163. <https://doi.org/10.1146/annurev-statistics-042720-125902>
- Muñoz-Vicente, J. M. M. y López-Ossorio, J. J. (2016). Valoración psicológica del riesgo de violencia: alcance y limitaciones para su uso en el contexto forense. *Anuario de Psicología Jurídica*, 26(1), 130-140. <https://doi.org/10.1016/j.apj.2016.04.005>
- Noble, S. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Northpointe (2015). *A Practitioner's Guide to COMPAS Core*. Northpointe. <https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>
- Olaciregui, M. de la 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. <https://zuan.unizar.es/record/109559>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Pineda, D., Rico-Bordera, P., Galán, M., Piqueras, J. A., and González-Álvarez, J. L. (2023). Women Victims of Intimate Partner Violence and Intimate Partner Homicide: A Typology Based on Victimization Variables. *Psychosocial Intervention*, 32(1), 43-53. <https://doi.org/10.5093/pi2023a3>
- Pueyo, A. A. y Redondo, S. (2007). Predicción de la violencia: Entre la peligrosidad y la valoración del riesgo de violencia. *Papeles del Psicólogo*, 28(3), 157-173. <https://www.papelesdelpsicologo.es/pdf/1500.pdf>
- Ramos-Torre, R., y Callejo-Gallego, J. (2018). "Semántica social del riesgo: una aproximación cualitativa", *Política y Sociedad*, 55(1), 235-256. <https://doi.org/10.5209/POSO.54062>
- Richards, L. (2009). *Domestic Abuse, Stalking and Harassment and Honour Based Violence (DASH, 2009) Risk Identification and Assessment and Management Model*. Association of Police Officers (ACPO). <https://reducingtherisk.org.uk/wp-content/uploads/2022/08/DASH-2009.pdf>
- Robinson, A. L., Myhill, A., Wire, J., Roberts, J. y Tilley, N. (2016). *Risk-led policing of domestic abuse and the DASH risk model*. Cardiff University, College of Policing y UCL Department of Security and Crime Science. <https://library.college.police.uk/docs/college-of-policing/Risk-led-policing-2-2016.pdf>
- Rodríguez-Rodríguez, I., Rodríguez, J. V., Pardo-Quiles, D. J., Heras-González, P. y Chatzigiannakis, I. (2020). Modeling and forecasting gender-based violence through machine learning techniques. *Applied Sciences*, 10(22), 8244. <https://www.mdpi.com/2076-3417/10/22/8244/htm>
- Santos-Hermoso, J., González-Álvarez, J.L., Alcázar-Córcoles, M.Á. y Carbonell-Vayá, E. J. (2022). Intimate Partner Homicide Against Women Typology: Risk Factor Interaction in Spain. *European Journal on Criminal Policy and Research*. <https://doi.org/10.1007/s10610-022-09517-7>

- Simón-Castellano, P. (2021). Inteligencia artificial y Administración de Justicia: ¿Quo vadis, justitia? *IDP. Revista de Internet, Derecho y Política*, 33. <https://doi.org/10.7238/idp.v0i33.373817>
- Skeem, J. y Lowenkamp, C. (2020). Using algorithms to address trade-offs inherent in predicting recidivism. *Behavioral Sciences & the Law*, 38(3), 259-278. <https://doi.org/10.1002/bsl.2465>
- Soler, S. (2023, 7 de diciembre). *Interior revisará los criterios que descartan a las mujeres del sistema de protección de violencia de género*. <https://www.rtve.es/noticias/20231207/interior-igualdad-reunion-violencia-genero/2465163.shtml>
- Taylor, K.-Y. (2019). Predatory inclusion. *N+1*, 35. <https://nplusonemag.com/issue-35/essays/predatory-inclusion/>
- 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>
- van Der Put, C. E., Gubbels, J., y Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*, 47, 100-116. <https://doi.org/10.1016/j.avb.2019.03.008>
- Vignola-Lévesque, C. y Léveillé, S. (2022). Intimate Partner Violence and Intimate Partner Homicide: Development of a Typology Based on Psychosocial Characteristics. *Journal of Interpersonal Violence*, 37, 17-18, NP15874-NP15898. <https://doi.org/10.1177/08862605211021989>
- 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*, <https://doi.org/10.1007/s10940-022-09545-w>
- West, S.M., Whittaker, M., y Crawford, K. (2019). *Discriminating systems: Gender, race, and power in AI* (Informe técnico). AI Now Institute. <https://ainowinstitute.org/discriminatingystems.pdf>
- Zambrano-Guerrero, C. A., y Rodríguez-Pabón, D. M. (2021). Design thinking como herramienta para prevenir la violencia basada en género en estudiantes universitarios. *Revista Guillermo de Ockham*, 19(2), 293-306. <https://doi.org/10.21500/22563202.5316>
- Završnik, A. (2021). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 18(5), 623-642. <https://doi.org/10.1177/1477370819876762>
- 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>
- Zweig, J. (2010). Extraordinary conditions of release under the bail reform act. *Harvard Journal on Legislation*, 47, 555-585. <https://n9.cl/phutj>