INTELIGENCIA ARTIFICIAL EN LA JUSTICIA CON PERSPECTIVA DE GÉNERO: AMENAZAS Y OPORTUNIDADES*

ARTIFICIAL INTELLIGENCE WITH GENDER PERSPECTIVE: THREATS AND OPPORTUNITIES

Actualidad Jurídica Iberoamericana Nº 21, agosto 2024, ISSN: 2386-4567, pp. 566-597

* Estudio redactado en el marco del Proyecto "Claves para una justicia digital y algorítmica con perspectiva de género", PID2021-123170OB-I00 financiado por MCIN/ AEI/10.13039/501100011033.

Ana
MONTESINOS
GARCÍA

ARTÍCULO RECIBIDO: 29 de mayo de 2023
ARTÍCULO APROBADO: 1 de julio de 2024

RESUMEN: Comienza a instaurarse, aunque todavía de manera incipiente, el uso de aplicaciones y herramientas basadas en inteligencia artificial en la Justicia. En este trabajo analizamos sus riesgos y beneficios desde una perspectiva de género. En primer lugar, estudiamos las amenazas que suponen los sesgos algorítmicos que refuerzan los estereotipos de género, así como la posible incorporación en el acervo probatorio de un proceso de los denominados deepfakes, nuevas formas de violencia ejercida a través de la IA cuyas principales víctimas son las mujeres. En segundo lugar, abordamos el empleo de herramientas algorítmicas en la lucha contra el mayor exponente de la desigualdad como es la violencia de género. En concreto, para atender a las víctimas, investigar determinados delitos y prevenir su ejecución mediante técnicas predictivas que valoran el riesgo de reincidencia.

PALABRAS CLAVE: Inteligencia artificial; justicia; perspectiva de género; sesgos; deepfakes y herramientas de valoración del riesgo.

ABSTRACT: *The use of applications and tools based on artificial intelligence in the justice system is beginning to be implemented, although it is still in its infancy. In this paper, we analyse their risks and benefits from a gender perspective. First, we examine the threats posed by algorithmic biases that reinforce gender stereotypes, as well as deepfakes as new forms of violence perpetrated by AI, whose main victims are women. Second, we look at the use of algorithmic tools in the fight against the greatest exponent of inequality, such as gender-based violence. Specifically, to support victims, to investigate certain crimes and to prevent their execution through predictive techniques that assess the risk of recidivism.*

KEY WORDS: *Artificial intelligence, justice, gender perspective, biais, deepfakes, risk assessment instruments.*

I. INTRODUCCIÓN: JUSTICIA ALGORÍTMICA CON PERSPECTIVA DE GÉNERO.

La inteligencia artificial (en adelante, IA[1]) ha experimentado un significativo progreso, convirtiéndose en una de las tecnologías estratégicas del siglo XXI. En concreto y en lo que a este trabajo se refiere, en los últimos años se ha constatado la oportunidad de su utilización en la Administración de Justicia. Posee la capacidad de generar notables beneficios en términos de eficiencia, eficacia y precisión, pero no se puede obviar que también conlleva riesgos sustanciales para los derechos fundamentales. Precisamente a sus bondades y peligros dedicamos este trabajo, en el que vamos a analizar ambas vertientes desde una perspectiva de género.

Para ello, partiremos de una doble premisa. Por un lado, la IA comienza, aunque todavía de manera incipiente, a instaurarse en la Administración de Justicia. Distintas herramientas computacionales, sistemas algorítmicos y softwares informáticos de última generación empiezan a utilizarse para cumplir múltiples objetivos. Desde simples tareas instrumentales de gestión procesal, burocráticas, organizativas y rutinarias (donde inicialmente se han implantado), hasta otras funcionales de mayor complejidad. Entre estas últimas, cabría distinguir aquellas herramientas asistenciales o colaboradoras, de aquellas otras que directamente ofrecen la solución (propositiva o imperativa)[2]. Consciente nuestro legislador del amplio abanico de posibilidades que pueden brindar a la Justicia, recientemente se ha regulado el empleo de nuevas herramientas algorítmicas para fines que sirvan de apoyo a la función jurisdiccional y a la tramitación de procedimientos judiciales en el Real Decreto-ley 6/2023, de 19 de diciembre, por el que se aprueban

1 Aunque somos conocedoras de que no todos los softwares son IA, vamos a utilizar el término IA en sentido amplio, sin ajustarnos a la definición del Reglamento de IA, que es mucho más estricta "Sistema de IA: un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales" (art. 3.1).

2 Barona Vilar, S.: "Datalización de la justicia (Algoritmos, Inteligencia Artificial y Justicia, ¿el comienzo de una gran amistad?)", *Revista Boliviana de Derecho*, 2023, núm. 36, p. 26.

• Ana Montesinos García
Profesora titular de Derecho Procesal, Universitat de València. Correo electrónico: ana.montesinos@uv.es.

medidas urgentes para la ejecución del Plan de Recuperación, Transformación y Resiliencia en materia de servicio público de justicia, función pública, régimen local y mecenazgo[3].

Por otro lado, nuestro legislador, incitado por compromisos internacionales y europeos[4], aboga por la incorporación de la perspectiva de género en la Justicia en aras de remover los obstáculos que dificultan la consecución de la igualdad efectiva entre hombres y mujeres[5]. En este sentido, la Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres califica expresamente la igualdad de trato y de oportunidades entre mujeres y hombres, como principio informador del ordenamiento jurídico y, como tal, se integrará y observará en la interpretación y aplicación de las normas jurídicas (art. 4)[6]. Se interpela así a los operadores jurídicos a aplicar la perspectiva de género. Además, su artículo 15 dispone que el principio de igualdad de trato y oportunidades entre mujeres y hombres informará, con carácter transversal, la actuación de todos los Poderes Públicos. Las Administraciones públicas lo integrarán, de forma activa, en la adopción y ejecución de sus disposiciones normativas, en la definición y presupuestación de políticas públicas en todos los ámbitos y en el desarrollo del conjunto de todas sus actividades. De ahí que, como defiende Barona Vilar, "una de las claves de desarrollo de la Justicia en el siglo XXI es indudablemente la feminización de la misma"[7].

Nos encaminamos, por consiguiente, hacia un modelo de Justicia cada vez más digital y algorítmico, cuya transformación debe necesariamente llevarse a cabo desde un enfoque de género, pues solo así podrá garantizarse la igualdad consagrada en el artículo 14 de nuestra Carta Magna[8]. Sin embargo, la intersección

3 BOE núm. 303, de 20.12.2023.

4 Tanto el Tratado de Ámsterdam como el Tratado de Lisboa establecen como objetivo de la Unión la eliminación de las desigualdades entre el hombre y la mujer y promueven su igualdad (art. 3.2 y art 8, respectivamente). A lo que debemos añadir que la Carta de Derechos Fundamentales de la UE proclama como valor fundamental la igualdad y reconoce la no discriminación por razón de sexo (art. 21) así como la igualdad entre mujeres y hombres (art. 23). Por su parte, el Convenio de Estambul, ratificado por España, en su art.4, condena "todas las formas de discriminación contra las mujeres" de manera que el Estado "tomará, sin demora, las medidas legislativas y de otro tipo para prevenirla, en particular: indicando en sus constituciones nacionales o en cualquier otro texto legislativo adecuado el principio de la igualdad entre mujeres y hombres, garantizando la aplicación efectiva del mencionado principio; prohibiendo la discriminación contra las mujeres, recurriendo incluso, en su caso, a sanciones; derogando todas las leyes y prácticas que discriminan a la mujer".

5 Puede definirse la perspectiva de género como un mecanismo o metodología que permite identificar, cuestionar y valorar la discriminación y la desigualdad en el trato entre hombres y mujeres en aras a implementar acciones positivas con el ánimo de avanzar y alcanzar la tan anhelada igualdad material. *Guía de actuación con perspectiva de género en la investigación y enjuiciamiento de los delitos de violencia de género*, Unidad de coordinación de violencia sobre la mujer de la FGE, diciembre 2020.

6 BOE núm. 71, de 23.03.2007.

7 Barona Vilar, S.: "La necesaria deconstrucción del modelo patriarcal de justicia", en AA.VV.: *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018, p. 32.

8 En esta línea, el Comité consultivo para la igualdad de oportunidades entre mujeres y hombres de la Comisión Europea ha elaborado un Dictamen sobre la IA que analiza, entre otras cuestiones, las repercusiones de esta última en la igualdad de género (*AI – opportunities and challenges for gender equality*,

entre la perspectiva de género, la inteligencia artificial y la Justicia, todavía está en ciernes[9], por lo que deviene imperativo adoptar medidas y precauciones para prevenir un impacto indeseado sobre los derechos fundamentales, especialmente sobre los derechos a la no discriminación e igualdad.

Al respecto se pronuncia la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, que contiene la primera regulación positiva en nuestro ordenamiento del empleo de la IA por las Administraciones Públicas[10]. Esta norma, que nace con el objetivo de garantizar y promover el derecho a la igualdad de trato y no discriminación, recoge medidas destinadas a prevenir, eliminar, y corregir toda forma de discriminación, directa o indirecta, en los sectores público y privado (art.1.1). Concretamente en lo que a la IA se refiere, su artículo 23 dispone que "en el marco de la Estrategia Nacional de Inteligencia Artificial, de la Carta de Derechos Digitales y de las iniciativas europeas en torno a la Inteligencia Artificial, las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio".

Hasta el momento la regulación de la IA, tanto nacional como europea, ha sido de *soft law*, es decir, se ha limitado a Comunicaciones, Informes y Cartas u otros documentos. Destáquese, en nuestro país, la Carta de Derechos digitales adoptada el 14 de julio de 2021, que contempla el derecho a la igualdad y a la no discriminación en el entorno digital. En particular, fomenta que los procesos de transformación digital apliquen la perspectiva de género y adopten, en su caso, medidas específicas para garantizar la ausencia de sesgos de género en los datos y algoritmos usados (VIII)[11].

2020). Por su parte, la Estrategia de la Unión Europea para la igualdad de género 2020-2024 también hace referencia al vínculo entre la IA y la igualdad de género (Comunicación de la comisión al Parlamento Europeo, al Consejo, al Comité Económico y social europeo y al Comité de las Regiones. Una Unión de la igualdad: Estrategia para la Igualdad de Género 2020-2025, Bruselas, 5.3.2020, COM (2020) 152 final) y la Red europea de organismos para la igualdad (Equinet) ha publicado el informe "Regulating for European AI that Protects and Advances Equality. An Equinet Position Paper", 22/06/2022.

9 En sentido similar, pero con referencia al Derecho privado, se pronuncia Navas Navarro, S.: "La perspectiva de género en la inteligencia artificial", *Diario La Ley*, núm. 48, sección Ciberderecho, 8 de marzo de 2021, p. 1.

10 BOE núm. 167, de 13.07.2022.

11 Un estudio detallado de la misma puede verse en Catalán Chamorro, M.J.: "La carta de derechos digitales y su implicación en el derecho procesal español", en AA. VV.: *Digitalización de la justicia: prevención, investigación y enjuiciamiento* (dir. por M. Llorente Sánchez-Arjona y S. Calaza López), Aranzadi, Cizur Menor, 2022, pp. 179 - 208. Véase también de esta autora: L*a justicia digital en España. Retos y desafíos*, Tirant Lo Blanch, Valencia, 2023.

En Europa, tras diversas normas de *soft law* (Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno de 2018[12], Directrices éticas para una IA fiable de 2019[13] y Libro Blanco de la Inteligencia Artificial de la Comisión de 2020[14], entre las más destacadas), finalmente ha llegado el Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión, cuyo objeto principal reside en impulsar la innovación a partir de normas que promuevan la confianza en las tecnologías[15]. En el mismo se clasifican los sistemas de IA en cuatro categorías atendiendo al riesgo potencial que implica su uso: prohibidos, alto riesgo, riesgo medio/bajo y resto de sistemas. En particular, a fin de evitar el riesgo de posibles sesgos, errores y opacidades, procede a considerar de alto riesgo aquellos sistemas de IA destinados a ser utilizados por una autoridad judicial o en su nombre para ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho y a aplicar la ley a unos hechos concretos (Considerando 61 y Anexo III punto 8.)[16]. A los sistemas de alto riesgo, a los que se dedica la mayor parte del articulado de la norma, se les exige el cumplimiento de toda una serie de requisitos y obligaciones específicas previstas en los capítulos 2 y 3.

Para finalizar la introducción de este trabajo, quisiera resaltar la labor llevada a cabo por la UNESCO en esta materia. Especialmente su Recomendación sobre la

12 Aprobada por la Comisión europea para la eficacia de la justicia (CEPEJ) el 4 de diciembre de 2018.

13 Elaboradas por un Grupo de expertos de alto nivel sobre IA, por encargo de la Comisión Europea, el 8 de abril de 2019. Entre los entre los siete requisitos que establece para una IA fiable, incluye la diversidad, la no discriminación y la equidad.

14 Libro Blanco sobre la IA -un enfoque europeo orientado a la excelencia y la confianza de 19 de febrero 2020 (COM (2020) 65 final).

15 Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (P9_TA(2024)0138). El Reglamento complementa el Derecho de la Unión vigente en materia de no discriminación al establecer requisitos específicos que tienen por objeto reducir al mínimo el riesgo de discriminación algorítmica, en particular en lo tocante al diseño y la calidad de los conjuntos de datos empleados para desarrollar sistemas de IA, los cuales van acompañados de obligaciones referentes a la realización de pruebas, la gestión de riesgos, la documentación y la vigilancia humana durante todo el ciclo de vida de tales sistemas (EM, 1.2)

16 No obstante, dicha clasificación no debe hacerse extensiva a los sistemas de IA destinados a actividades administrativas meramente accesorias que no afectan a la administración de justicia en casos concretos, como la anonimización o seudonimización de las resoluciones judiciales, documentos o datos; la comunicación entre los miembros del personal o las tareas administrativas. En tal sentido, el art. 6.3 manifiesta que, no obstante, un sistema de IA no se considerará de alto riesgo si no plantea un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, en particular al no influir sustancialmente en el resultado de la toma de decisiones. Así será cuando se cumplan una o varias de las condiciones siguientes: a) que el sistema de IA tenga por objeto llevar a cabo una tarea de procedimiento limitada; b) que el sistema de IA tenga por objeto mejorar el resultado de una actividad humana previamente realizada; c) que el sistema de IA tenga por objeto detectar patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores y no esté destinado a sustituir la evaluación humana previamente realizada sin una revisión humana adecuada, ni a influir en ella; o d) que el sistema de IA tenga por objeto llevar a cabo una tarea preparatoria para una evaluación pertinente a efectos de los casos de uso enumerados en el anexo III. Los sistemas de IA siempre se considerarán de alto riesgo cuando lleven a cabo la elaboración de perfiles de personas físicas.

ética de la inteligencia artificial de 23 de noviembre de 2021[17], en cuyo ámbito de actuación número 6, que versa sobre género, insta a los Estados para que velen por que se optimice plenamente el potencial de las tecnologías digitales y la IA para contribuir a lograr la igualdad de género, así como por que los estereotipos de género y los sesgos discriminatorios no se trasladen a los sistemas de IA, sino que se detecten y corrijan de manera proactiva (puntos 89 y 90)[18].

Siendo este el estado de la cuestión, pasamos a analizar en el siguiente apartado algunas de las amenazas que acechan los sistemas de IA.

II. AMENAZAS: REFUERZO DE ESTEREOTIPOS Y NUEVAS FORMAS DE VIOLENCIA CONTRA LAS MUJERES.

Entre las diversas amenazadas asociadas al uso de IA en el ámbito de la Justicia, queremos destacar dos. En primer lugar, advertimos del riesgo de que las herramientas algorítmicas empleadas en los Juzgados puedan contener sesgos que reproduzcan estereotipos que perjudiquen a las mujeres. En segundo lugar, nos referimos a las nuevas formas de violencia que pueden ejercerse a través de la IA, particularmente a los "deepfakes" y a su posible incorporación en el acervo probatorio de un proceso judicial.

1. Los sesgos algorítmicos de género. Ejemplos en el marco de un proceso judicial.

Uno de los principales desafíos vinculados al uso de sistemas de IA en la Justicia reside en la posibilidad de que se generen situaciones discriminatorias que puedan comprometer el principio de igualdad.

Es ya un lugar común admitir que los sesgos algorítmicos reproducen los sesgos humanos[19]. En este sentido, requieren especial atención los sesgos algorítmicos de género que encontramos cuando un sistema informático propone o adopta

17 Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa

18 Véase asimismo su informe "Artificial intelligence and gender equality" de 2020, en el que propone que la igualdad de género se constituya en un principio autónomo dentro del elenco de principios éticos de la IA (p. 16) y "I'd Blush if I Could: closing gender divides in digital skills through education" de 2019 ("Me sonrojaría si pudiera: cerrando brechas de género en la esfera digital a través de la educación"; título que proviene de la respuesta proporcionada por Siri al insulto "eres una puta"), en el que aborda el potencial de los sesgos algorítmicos de propagar y reforzar estereotipos de género. Ambos informes se encuentran disponibles en https://unesdoc.unesco.org/ark:/48223/pf0000374174 y https://unesdo c.unesco.org/ark:/48223/pf0000367416.page=1. Otra iniciativa a destacar, en este caso privada, ha sido la llevada a cabo por Amnistía Internacional, Access Now y otras organizaciones en 2018 reflejada en la Declaración de Toronto sobre la protección del derecho a la igualdad y la no discriminación en los sistemas de aprendizaje automático.

19 Un refrán muy conocido entre los informáticos "Garbage in, garbage out" (si entra basura, sale basura), transmite la idea de que cualquier resultado algorítmico discriminatorio procede de prejuicios inyectados en los algoritmos por seres humanos. En otras palabras, y tal y como lo replantea Mayson, "Bias in, bias out". Mayson, S. G.: "Bias In, Bias Out", *Yale Law Journal*, 2018, núm. 128, pp. 2218- 2300.

decisiones erradas que reproducen estereotipos de género[20]. Se entrelazan así, los algoritmos por un lado y, por otro, los estereotipos de género[21]. Estos últimos se definen como percepciones generalizadas o prejuicios acerca de los atributos o características que se supone que hombres y mujeres poseen, o deberían poseer, así como las funciones sociales que se espera que desempeñen[22].

Son tres, principalmente, las etapas clave en las que se pueden introducir los sesgos en los sistemas de IA: (i) cuando se decide el objetivo a alcanzar por el sistema; (ii) cuando se recopilan los datos (que son poco representativos[23] o reflejan prejuicios existentes en la realidad social) y (iii) cuando se seleccionan los atributos que se quiere que tenga en cuenta el algoritmo[24]. Aunque en ciertas situaciones los sesgos algorítmicos responden a un objetivo claramente discriminatorio (discriminación directa), en la mayoría de casos simplemente son provocados por el desinterés hacia su impacto colateral[25].

Si los sistemas de IA no se diseñan y desarrollan desde una perspectiva de género que sea capaz de detectar, analizar y corregir los sesgos, estos no solo serán susceptibles de ser replicados sino incluso acrecentados y reforzados. Sus efectos podrán ser infinitamente más rápidos y devastadores[26], lo que resulta especialmente preocupante si se augura que en un futuro cercano los jueces van a auxiliarse cada vez más de este tipo de herramientas.

Existen numerosos ejemplos de discriminación algorítmica hacia las mujeres, como el caso del chatbot "Tay.AI" de Microsoft, los créditos bancarios de Apple

20 Danesi, C.: "Sesgos algorítmicos de género con identidad iberoamericana: las técnicas de reconocimiento facial en la mira", *Revista Derecho de Familia*, 2021, núm.100, p. 161.

21 El Índice de Normas Sociales de Género (GSNI, por sus siglas en inglés) revela la falta de avances en la superación de los prejuicios contra las mujeres en la última década, ya que aproximadamente 9 de cada 10 hombres y mujeres en el mundo siguen manteniendo en la actualidad un sesgo contra las mujeres. "Una década de estancamiento: el PNUD presenta nuevos datos que muestran la persistencia de los sesgos de género", Comunicado de prensa, Programa de las naciones unidas para el desarrollo, Nueva York, 12 de mayo de 2023, disponible en file:///C:/Users/Admin/Documents/gsni_2023_pr_sp.pdf

22 Belloso Martin, N.: "La problemática de los sesgos algorítmicos (con especial referencia a los de género) ¿Hacia un derecho a la protección contra los sesgos?", en AA.VV.: *Inteligencia artificial y filosofía del derecho* (dir. por F. llano Alonso), Laborum, Murcia, 2022, p. 55.

23 Los algoritmos necesitan entrenarse con una gran cantidad de datos y, además, deben ser de calidad, esto es, representativos de toda la población. De lo contrario, se pueden producir situaciones en las que el sesgo de la muestra de entrenamiento se incorpora como un criterio que se ha de cumplir, lo que dificulta que se avance en la igualdad de oportunidades. Fernández, A., "Inteligencia artificial en los servicios financieros", *Boletín económico - Banco de España*, 2019, núm. 2, p. 5.

24 Ei, D., y Moser, G.: "Human arbitrators (the undisputed champion) v (the robots challenger)", *Hong Kong L.J*, 2020, vol. 50, p. 239. Hao, K., "This Is How AI Bias Really Happens - and Why It's So Hard to Fix", MIT *Technology* Review, 4 febrero 2019, disponible en: https://www.technologyreview.com/s/612876/this-ishow-ai-bias-really-happensand-why-its-so-hard-to-fix/

25 Rivas Vallejo, P.: "Sesgos de género en el uso de inteligencia artificial para la gestión de las relaciones laborales: análisis desde el derecho antidiscriminatorio", *e-Revista Internacional de la Protección Social*(*e-RIPS)*, 2022, vol. VII, núm. 1, p. 54.

26 O'Neil manifiesta que es crucial entender que, bajo la apariencia de neutralidad de los algoritmos, hay decisiones morales que perpetúan y aumentan las desigualdades sociales. Por eso los denomina "armas de destrucción matemática". *Armas de destrucción matemática*, Capitán Swing, Madrid, 2017.

Card, el reclutador inteligente de Amazon o los motores de búsqueda de Google, entre otros[27]. Excede de nuestro trabajo profundizar en ellos. No obstante, si vamos a mostrar, a modo de hipótesis, algunos ejemplos que ilustran las nocivas consecuencias del empleo de herramientas algorítmicas sesgadas en el marco de un proceso judicial. Para ello, partiremos de la clasificación, previamente mencionada, de las tres etapas clave en las que se pueden introducir los sesgos.

En primer lugar, los sesgos pueden incorporarse en un sistema de IA desde el momento en el que se establece el objetivo a alcanzar. Trasladado al marco de un proceso penal, imaginemos una herramienta que se diseña con el único fin de detectar y, consiguientemente, perseguir denuncias falsas presentadas por víctimas de violencia de género. La decisión de crear un instrumento específico para este propósito, asumiendo que estas denuncias son tan frecuentes como las denuncias falsas por robo (para las cuales ya se ha desarrollado una herramienta[28]), revela una percepción errónea de que las denuncias por violencia de genero son más propensas a ser falsas. Este prejuicio podría acarrear repercusiones devastadoras para las víctimas de tales delitos, perpetuar la desconfianza hacia ellas y contribuir a un entorno que desaliente la presentación de denuncias legítimas.

En segundo lugar, las herramientas de IA pueden introducir sesgos que estén presentes en las bases de datos de las que se nutren, bien porque se recopilan datos que son poco representativos bien porque reflejan prejuicios existentes en la realidad o tejido social. Probablemente este sea el supuesto más común en la práctica.

Los softwares de reconocimiento facial son un buen ejemplo de cómo este tipo de herramientas pueden tener consecuencias perjudiciales en la justicia penal. De hecho, se alerta constantemente acerca de su potencial discriminatorio, dado que muchos de estos sistemas se entrenan con conjuntos de datos que están sesgados en términos de representación. Fundamentalmente porque sobrerepresentan a hombres caucásicos e infrarrepresentan a mujeres. El resultado que ofrecen, por consiguiente, puede ser un reconocimiento facial menos preciso y más propenso a cometer errores cuando se aplica a mujeres de piel oscura. Este sesgo en los datos de entrenamiento puede llevar a disparidades en el trato, ya que las personas

27 Nos remitimos a los brillantes trabajos de Soriano Arnanz, A.: "Discriminación algorítmica: garantías y protección jurídica", en AA.VV.: *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas* (dir. por L. Cotino Hueso), Aranzadi, 2022, pp. 139-169 y "Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules", *Revista de Internet, Derecho y Política*, 2023, núm. 38.

28 Nos referimos a Veripol, herramienta utilizada por la policía en nuestro país que estima la probabilidad de que una denuncia por robo con violencia e intimidación sea falsa. El sistema identifica el delito basándose en el texto de la denuncia. Utiliza el procesamiento del lenguaje natural y la IA para analizar y calcular las combinaciones de palabras más comunes cuando se miente ante un policía.

pertenecientes a los grupos infrarrepresentados pueden enfrentarse a tasas más altas de falsas identificaciones[29].

Por otro lado, los datos contenidos en las bases que alimentan el sistema de IA también pueden reflejar prejuicios existentes en nuestra sociedad. Pensemos, de nuevo, un ejemplo que podría darse en un proceso judicial. El juez se auxilia de una herramienta de IA para resolver el caso que le proporciona argumentos a favor y en contra. La herramienta, que se nutre de un repertorio de sentencias previas que han resuelto casos de agresiones sexuales, podría aprender y replicar los prejuicios y estereotipos contenidos en las mismas sobre el comportamiento que se espera de una víctima (resistirse de determinada manera, denunciar inmediatamente o mantenerse alejada de la vida social). Esto podría llevar al dictado de una resolución que resuelva que la violencia no ha existido, además de reflejar juicios sesgados y discriminatorios que contribuyen a la perpetuación de desigualdades en el sistema judicial[30]. Ello inevitablemente podría comportar consecuencias perjudiciales para las mujeres que intervienen en los procesos judiciales, así como favorecer automáticamente la desconfianza hacia sus declaraciones.

En tercer y último lugar, los sesgos también pueden provenir del modo en que se entrena al algoritmo, a la hora de seleccionar los atributos que deben tenerse en cuenta.

Veamos el ejemplo ficticio planteado por MARTÍNEZ, BORGES y SIMÓ. Se crea un software para ayudar al juez a valorar la declaración de la víctima y ese algoritmo ha sido entrenado bajo el criterio de que la tardanza en denunciar es relevante para valorar su credibilidad, fruto de un estereotipo humano vigente en los tribunales. Ante un supuesto en el que la víctima denunciara los hechos inmediatamente después de que ocurrieran, su nivel de credibilidad sería alto. Sin embargo, si la víctima tardara meses en denunciar, el algoritmo le reportaría al juez un grado de credibilidad mínimo. Estaríamos rotundamente ante un supuesto

29 Vid. European Union Agency for fundamental rights, *Facial recognition technology: fundamental rights considerations in the context of law enforcement,* 2020 y European Parliamentary Research Service, *Regulating facial recognition in the EU*, septiembre 2021, Bruselas.

30 Como señala GIL, existe el pensamiento generalizado de que una víctima que no cuenta desde el principio que sufrió abuso o agresión sexual, y lo cuenta después, no está diciendo la verdad; o que la víctima se pierda o desordene fechas o momentos en que suceden los hechos denota que tampoco dice la verdad: o que recuerde con mayor nitidez unos u otros momentos implica un testimonio poco fiable; o que no haya precisado atención psiquiátrica indica que quizá lo que relata no ocurrió; o que una víctima que efectúa su relato con entereza y sin llorar o con buena apariencia física, ya parece que no es víctima. Estos prejuicios son los que planea sobre los testimonios de las víctimas. Son una muestra de los estereotipos que existen y persisten en nuestra sociedad, y en el ámbito concreto de la justicia; además denota las enormes carencias que hay en la formación de los operadores jurídicos. Es no entender como un acto violento de estas características produce bloqueo en las víctimas, sentimientos de vergüenza o de culpabilidad. "La perspectiva de la mujer víctima del sistema judicial ajeno al género", en AA.VV.: *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018, pp. 238 y 239.

de discriminación algorítmica[31]. Lo mismo ocurría si se entrena al algoritmo pautándole que si la víctima no recuerda los detalles, no quiere declarar o quiere retirar la denuncia contra su pareja, entonces debe entenderse que es falsa[32].

Un último ejemplo podríamos hallarlo en las herramientas algorítmicas de evaluación de riesgos que, al entrenarse con datos históricos, reflejan y perpetúan estereotipos de género existentes en los casos anteriores de los que se nutre. Supongamos que, en el pasado, las denuncias de ciertos tipos de violencia contra las mujeres fueron consideradas de menor gravedad. Si el algoritmo se entrena con esos datos, podría aprender patrones sesgados y asignar automáticamente menores niveles de riesgo a casos similares en el futuro, lo que podría afectar negativamente a la atención y recursos asignados a las víctimas.

Vistos estos ejemplos, no podemos sino concluir que hay que esforzarse por procurar que no se reproduzcan sesgos históricos de género con herramientas algorítmicas empleadas en el marco de la Justicia. La tecnología no puede convertirse en un vehículo que facilite la vulneración de los derechos que principian el proceso, concretamente, el de igualdad. El riesgo de que estos instrumentos puedan discriminar resulta inaceptable. No olvidemos que estamos ante sistemas de alto riesgo dado que se destinan a auxiliar a los jueces, por lo que pueden tener efectos importantes sobre los derechos fundamentales, entre otros, el derecho a la tutela judicial efectiva y a un juez imparcial. De manera que resulta crucial trabajar para mejorar la detección de los sesgos algorítmicos y poner en marcha las medidas necesarias para neutralizarlos. Esto implica la adopción de medidas tanto de índole política como jurídica y tecnológica[33].

En este contexto, resulta imprescindible que los algoritmos sean auditados y controlados tanto de manera previa a su puesta en marcha como de forma periódica. Para ello, los programas, en la medida de lo posible, tendrán que ser

31 Martínez García, E.; Borges Blázquez, R., y Simó Soler, E.: "Inteligencia artificial y perspectiva de género en la justicia penal", *Diario* La *Ley*, Sección Ciberderecho, 20 de enero de 2021, núm. 47, p. 6. No olvidemos que el propio Tribunal Supremo ha declarado en su sentencia 184/2019, de 2 de abril (ES:TS:2019:1071) que la credibilidad de la víctima no debe ser menoscabada por el hecho de retrasar la denuncia.

32 En lugar de pensar en las dudas que puedan tener las mujeres por las posibles consecuencias que afectarían al "padre de sus hijos", o los miedos por las presiones y amenazas de los entornos. Lorente Acosta, M.: "Justicia, género y estereotipos", en AA.VV.: *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018, p.154.

33 Belloso Martin, N.: "La problemática", cit., p. 69.
Se propone la incorporación de una mayor presencia femenina en los equipos que diseñan estas herramientas. Además, que se trate de equipos formados capaces de introducir la perspectiva de género desde el diseño (*gender-by-design*) del propio sistema de IA. "La perspectiva de género en la inteligencia artificial", cit., p. 10. Como señalan, Ortiz De Zárate y Guevara, la diversidad en los equipos puede ofrecer nuevas perspectivas y traer a colación experiencias que permitan reconocer los sesgos y trabajar para corregirlos. *Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España*, Fundación alternativas, 2021, núm. 101, p. 24. En esta línea, la nueva Agenda España Digital 2026 en su noveno eje, referido a las competencias digitales, prevé que el reto para 2026 sea reforzar las competencias digitales de la ciudadanía, reduciendo las brechas digitales, consiguiendo una paridad de género en los especialistas digitales.

trasparentes y explicables. Solo así se podrá mitigar el problema de los sesgos, en tanto en cuanto se permitirá, a *priori*, su corrección y perfeccionamiento[34].

2. *Deepfakes:* violencia basada en el género con empleo de IA.

La manipulación de imágenes, audios o vídeos no es un fenómeno novedoso. Sí lo es su ejecución mediante técnicas de IA que implican un mayor grado de sofisticación y producen resultados que se asemejan mucho más a la realidad, hasta el punto de dificultar considerablemente el discernimiento entre la autenticidad y la falsedad de los contenidos generados.

A través de redes generativas adversariales (*generative adversarial networks* o GAN por sus siglas en inglés) pueden crearse falsos contenidos audiovisuales hiperrealistas que dan lugar a los denominados *deepfakes*. Este término, en inglés, deriva de la combinación de las palabras "fake" (falsificación) y "deep learning" (aprendizaje profundo)[35]. Los *deepfakes* ganaron notoriedad a finales de 2017, cuando un usuario anónimo de la plataforma Reddit (conocido con el alias "deepfake") compartió videos pornográficos falsos que superponían los rostros de celebridades como Taylor Swift o Scarlett Johansson, en cuerpos de mujeres desnudas. A pesar de la pronta eliminación de estos videos, esta técnica de manipulación se ha propagado rápidamente por Internet, lo que se ha debido en gran medida a que la creación de este tipo de material falso está al alcance de cualquiera, dado que existen aplicaciones gratuitas cada vez más populares que facilitan la edición de contenidos de manera relativamente sencilla.

Los *deepfakes* tienen múltiples usos. Algunos son legítimos. Por ejemplo, en el ámbito de la publicidad se pueden crear campañas más impactantes y personalizadas; en el de la educación se puede generar material didáctico interactivo; en el cine se pueden realizar efectos especiales más realistas y en medicina se pueden simular escenarios clínicos para propósitos de formación. Sin embargo, al mismo tiempo los *deepfakes* ostentan un enorme potencial para ejercer una variedad de fines maliciosos e incluso delictivos, que incluyen, entre otros, difusión de noticias falsas, atribución a políticos (caso de Obama) o empresarios (caso de Zuckerberg) de

34 Macchiavelli, N.: "La violencia de género y el uso de algoritmos como herramienta efectiva para la protección de los derechos fundamentales", *AFD*, 2022 (XXXVIII), p. 64.

35 La tecnología utilizada para generar estos contenidos depende de distintos tipos de falsificaciones digitales que, en términos generales, pueden clasificarse de la siguiente manera: a) Sustitución de caras: intercambio de la cara de una persona, fusionándola con la de otra; b) Reinterpretación facial: manipulación de los rasgos faciales de un sujeto para parecer que está diciendo algo que en realidad no es así; d) Generación de rostros: creación de imágenes sintéticas convincentes y ficticias de personas; e) Síntesis de voz: generación de contenido de audio mediante el uso y entrenamiento de algoritmos para crear una voz falsa o un archivo de audio sintético y, f) Shallowfakes: falsificaciones audiovisuales menos sofisticadas creadas mediante técnicas de edición rudimentarias. T Trend Micro Research, United Nations Interregional Crime and Justice Research Institute (UNICRI) Europol's European Cybercrime Centre (EC3), "Malicious Uses and Abuses of Artificial Intelligence", 2020, pp. 53 y 54. Disponible en: file:///C:/Users/Admin/OneDrive%20-%20Universitat%20de%20Valencia/vid.%20pp.%2053%20y%2054.pdf.

declaraciones que nunca realizaron, destrucción de la imagen y credibilidad de una persona, acoso o humillación en línea, perpetración de extorsiones y fraudes, distribución de desinformación y manipulación de la opinión pública, chantajes, desbloqueo de dispositivos, suplantaciones de identidad para estafas, falsificación o manipulación de pruebas en procesos judiciales, etc[36]. De entre todos ellos, la producción de pornografía, especialmente en los casos de "pornovenganza" (*revenge porn*), se erige como su principal manifestación[37]. Recuérdese en este sentido, lo sucedido en Almendralejo (Badajoz), donde decenas de chicas adolescentes han sido víctimas de estos 'desnudos', que han circulado rápidamente entre los móviles de sus compañeros[38].

Consciente de estos peligros, el Reglamento de IA ha tomado – aunque muy prudentemente- cartas en el asunto. A pesar de no prohibir esta tecnología, sí que impone ciertos requisitos mínimos, especialmente en lo que se refiere a obligaciones de transparencia. De este modo, los creadores de *deepfakes* deben hacer público que el contenido se ha generado de forma artificial o ha sido manipulado. Esta obligación no se aplicará cuando su uso esté autorizado por la ley para detectar, prevenir, investigar o enjuiciar infracciones penales (art. 50.2)[39].

Dicho esto, queremos hacer hincapié en que el empleo de las tecnologías se está convirtiendo en un componente cada vez más habitual en la violencia contra las mujeres[40]. En este sentido, el surgimiento de la IA ha dado lugar a nuevos modos de violencia con el mismo propósito de control y dominación[41]. Como ocurre

36 Trend Micro Research, "Malicious Uses", cit., p. 52.

37 Excede del objeto de este trabajo analizar la calificación jurídico-penal que merecen estas conductas. Nos remitimos a las reflexiones de Bello San Juan, P.: "La inteligencia artificial al servicio del crimen: La revolución del deepfake desde una perspectiva criminológica", en AAV.: *La justicia en la sociedad 4.0: nuevos retos para el siglo XXI* (dir. por L. Fontestad Portalés), Colex, 2023, p. 239. Si queremos mencionar que la recientemente aprobada Directiva sobre la lucha contra la violencia contra las mujeres y la violencia doméstica de 14 de mayo de 2024 (DOUE: 2024/1385), expresamente se refiere a los *deepfakes* que representan actividades sexuales e insta a los Estados a penalizar la producción, manipulación o divulgación no consentida del material manipulado.

38 Borraz, M. y Pastor, A.: "'Deepfakes sexuales: el caso de las menores de Almendralejo consolida una nueva forma de violencia machista", *el Diario.*es, 19 de septiembre de 2023.

39 Véase el Informe que con carácter previo emitió el Parlamento Europeo titulado "Tackling deepfakes in European Policy" de julio 2021. Véase asimismo en nuestro país la Proposición de Ley Orgánica de regulación de las simulaciones de imágenes y voces de personas generadas por medio de la inteligencia artificial, presentada por el Grupo parlamentario Plurinacional SUMAR. BOE núm. 23-1 de 13 de octubre de 2023.

40 Entre los comportamientos específicos que utilizan tecnología como medio para ejercer violencia contras las mujeres (ciber violencia), se incluye el ciberacoso, ciberhostigamiento, el sexting, el stalking, la publicación de contenido sexual en línea sin consentimiento, etc. Estas prácticas, sin lugar a dudas, afectan principalmente a las mujeres. Vid. Lloria García, P.: *Violencia sobre la mujer en el siglo XXI. Violencia de control y nuevas tecnologías: habitualidad, sexting y stalking*, Iustel, Madrid, 2020.

41 La tecnología incrementa el riesgo de violencia, especialmente de la violencia psicológica porque permite a los agresores crear "una sensación de omnipresencia" que erosiona la sensación de seguridad. La mayoría de estudios sobre violencia de género facilitada por la tecnología se centra en ciberriesgos "convencionales", como el abuso a través de redes sociales. Pero hay un área más nueva de la tecnología que merece también atención: el "Internet de las Cosas" o el "IoT", término que describe la red de dispositivos autónomos conectados a Internet que las personas pueden supervisar o controlar desde una ubicación remota. Estos dispositivos abarcan toda una serie de tecnologías como electrodomésticos

con los deepfakes, "el fenómeno de la violencia contra las mujeres no es nuevo, lo es el procedimiento a través del cual se procura su ejercicio'"[42]. Precisamente su propagación está estrechamente ligada a la elaboración y distribución de contenido pornográfico falso protagonizado por mujeres. La generación de este tipo de imágenes no deja de ser una herramienta y un atentado más que afecta directamente contra su imagen, dignidad e integridad. Además, tengamos presente que no solo se han visto afectadas celebridades, sino también mujeres anónimas cuyos exnovios o amantes han utilizado esta tecnología para vengarse de ellas y humillarlas en línea. Es, por tanto, un arma que puede ser muy peligrosa contra ellas para acosarlas, intimidarlas y degradarlas[43].

En este sentido, como señala Bello San Juan, resulta insoslayable el componente de género subyacente tras estas conductas, al ser principalmente mujeres las perjudicadas por este tipo de acciones con independencia de su condición social, económica o la posición que represente en la sociedad[44]. La abrumadora mayoría de víctimas de estos vídeos son mujeres[45]. Como muestra, la aplicación DeepNude permite desnudar artificialmente incorporando una foto de un rostro a un cuerpo

inteligentes (altavoces, frigoríficos, televisores), dispositivos personales (relojes, dispositivos médicos, coches), sistemas domésticos (termostatos, cámaras de seguridad, iluminación), asistentes domésticos (Alexa), etc. La creciente prevalencia de estos dispositivos "inteligentes" proporciona a los agresores una nueva y poderosa herramienta para ampliar y magnificar los daños tradicionales de la violencia doméstica. Permiten superar los límites geográficos y espaciales que de otro modo les impediría vigilar, controlar, acosar, aislar y amenazar a las víctimas. Nos podemos hacer una idea de la gravedad que puede alcanzar este asunto con los ejemplos expuestos por MADISON LO, de conductas que un maltratador podría llevar a cabo. Entre otras, el apagado a distancia de los aparatos de aire acondicionado, el cambio diario de las contraseñas digitales de la puerta principal, el timbre de la puerta sonando incesantemente, cambiar la temperatura de una vivienda a kilómetros de distancia, hervir un hervidor de agua para recordar que el maltratador está mirando, utilización de sensores que controlan las cerraduras inteligentes para restringir la capacidad para salir de casa, control del historial de búsqueda de los asistentes virtuales por voz para asegurarse de que no se busca ayuda, etc. LO, M.: "A Domestic Violence Dystopia: Abuse via the Internet of Things and Remedies Under Current Law", *California Law Review*, 2021, vol. 109, pp. 277- 315.

42 Simó Soler, E.: "Retos jurídicos derivados de la Inteligencia Artificial Generativa Deepfakes y violencia contra las mujeres como supuesto de hecho", *InDret*, febrero 2023, núm.2, p. 498.

43 Soto Santana, M.: "Justice for Women: Deep fakes and Revenge Porn", 3rd Global Conference on woman´s studies, 25-27 septiembre 2022, p. 113. Disponible en file:///C:/Users/Admin/OneDrive%20-%20Universitat%20de%20Valencia/Inteligencia%20artificial/Deepfakes/Justice%20for%20woman.pdf

44 Bello San Juan, P.: "La inteligencia artificial", cit., p. 244. Son una manifestación más de la cosificación de la mujer. Ellas protagonizan falsas escenas eróticas y pornográficas; ellos, discursos y circunstancias relacionados con el humor o con la política, apareciendo normalmente vestidos. Ellas asoman en espacios privados; ellos, en espacios públicos ostentando el poder o un protagonismo sano. Ellas son cosificadas y sus rostros se pegan al cuerpo de una actriz despersonificada. Ellos tienen otro cuerpo, pero no pierden su esencia personal ni son tratados como objetos porque lo llamativo es lo que dicen o hacen. Ellas son sujetos pasivos; ellos protagonistas activos. Cerdán Martínez, V. y Padilla Castillo, G.: "Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso", *Historia y comunicación social*, 2019, núm. 24 (2), pp. 505-520.

45 En un informe elaborado en el 2023 por Home Security Heroes, empresa especializada en ciberseguridad, se alcanzaron las siguientes conclusiones: el número total de vídeos deepfake en línea en 2023 fue de 95.820, lo que representa un aumento del 550% con respecto a 2019; La pornografía deepfake representa el 98 % de todos los deepfake en línea; El 99% de las personas a las que se dirige esta pornografía son mujeres; Una de cada tres herramientas de deepfake permite a los usuarios crear pornografía; Se tarda menos de 25 minutos y cuesta 0 dólares crear un vídeo pornográfico de 60 segundos con tan solo utilizar una imagen de la cara. "2023 State of Deepfakes. Realities, threats, and impact", disponible en: https://www.homesecurityheroes.com/state-of-deepfakes/#appendix

desnudo obtenido de una base de datos que únicamente contiene imágenes de mujeres.

En definitiva, los *deepfakes* sexuales no consentidos plantean un riesgo significativo porque los agresores pueden utilizarlos para amenazar, controlar, intimidar, aislar, avergonzar, chantajear y abusar de las víctimas que son, en su mayoría, mujeres[46]. Esto, trasladado al marco de un proceso judicial podría tener unos efectos claramente perniciosos. Como afirma Simó Soler, no sería descabellado pensar que los maltratadores puedan generar *deepfakes* para poner en duda la versión de las futuras denunciantes («Tengo la prueba de que hubo consentimiento»), destruir su imagen y credibilidad y forzar el desistimiento («¿Quién te va a creer si eres una buscona?»)[47].

Hasta el momento la doctrina se ha centrado en analizar cómo prevenir, mitigar y sancionar el uso malicioso de esta tecnología. Pero hay otro aspecto que no debe olvidarse: los *deepfakes* también van a llegar, de manera inevitable, a los juzgados. En tal caso, los contenidos audiovisuales falsos creados mediante GAN podrían socavar seriamente la integridad de los procesos judiciales de varias maneras. Entre otras, podrían ser presentados como pruebas, bien con la intención de engañar al juzgador, bien sin ser la parte que los aporta consciente de su falsedad. Podrían también viciar las declaraciones de los testigos en la medida en que se hayan visualizado o escuchado grabaciones manipuladas por estos sistemas, creyéndolas reales[48]. Es más, incluso en situaciones donde los contenidos no fueran falsos, la mera existencia de los *deepfakes* podría complicar seriamente la tarea de demostrar la veracidad de las pruebas. Es decir, la parte contraria podría argumentar que se trata de un video falso e impugnar la prueba con el fin de descartarla como evidencia o, al menos, sembrar la duda acerca de su autenticidad[49]. Todo ello va a conllevar cargas adicionales para los diferentes operadores jurídicos (abogados, jueces, fiscales, peritos, etc.), al tener que dedicar tiempo, dinero y esfuerzo en la detección y comprobación de falsificaciones cada vez más sofisticadas.

Dada la complejidad técnica inherente, son precisos conocimientos específicos que el juez no dispone, por lo que la respuesta más directa e inmediata a los *deepfakes* pasaría por la prueba pericial. Se plantea así, como vía para despejar

46 Kweilin, L.T.: "Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology", *Victims & Offenders*, 2022, vol. 17, núm. 5, p. 648.

47 Simó Soler, E.: "Retos jurídicos", cit., p. 501.

48 Bello San Juan, P.: "La inteligencia artificial", cit., p. 238.

49 Este problema podría alcanzar proporciones especialmente graves cuando el sujeto afectado por el *deepfake* fuera una persona particularmente indefensa, o incapaz por sí mismo de reclamar justicia ante una prueba que no se corresponde con la realidad. Miguel Freita, P.: "Deepfakes, conteúdo gerado por inteligência artificial e verdade processual", en AA.VV.: *El proceso penal ante una nueva realidad tecnológica europea* (dir. por C. Arangüena Fanego, m. De Hoyos Sancho y E. Pillado González), Thomson Reuters Aranzadi, 2023, p. 204.

las dudas sobre la veracidad o falsedad de las pruebas que contienen imágenes, videos o audios, que se adjunte un informe pericial que acredite que, a la luz de los conocimientos tecnológicos actuales, no ha sido posible detectar indicios de que ese contenido sea parcial o totalmente artificial. Y ante la duda acerca de la parcialidad de dicho informe técnico, podría encomendarse dicha tarea a los peritos forenses oficiales designados por el juez, como ocurre con el resto de pruebas periciales[50].

En este sentido, se advierte acerca de "la necesidad de un peritaje judicial avanzado, sofisticado e hiperexperto, requiriendo incluso de la propia IA para detectar los vídeos falsos"[51]. Ahora bien, no consideramos que vaya a resultar imprescindible acudir siempre y en todo lugar a sofisticadas herramientas forenses para detectar las manipulaciones. Cuando los vídeos falsificados sean de mala calidad, la tarea no será tan difícil. Por otro lado, si demostrar la falsedad resulta demasiado complejo, será más fácil demostrar que el vídeo no ha sido manipulado, por ejemplo, adjuntando metadatos adicionales en el momento de grabar el vídeo, con el objetivo de dar fe de la autenticidad de la grabación del vídeo[52]. Además, no olvidemos que también podrá acudirse a otras pruebas, como, por ejemplo, a la testifical (llamar a quien tomó el video, a quien sale en él, a quien presenció la grabación, etc.) o al interrogatorio de la parte que ha presentado el video como prueba.

Teniendo en cuenta este escenario, y sin soslayar sus riesgos, no es, sin embargo, nuestra intención incurrir en un alarmismo desproporcionado. Es probable que nos enfrentemos a algunos de estos casos, pero no por ello vislumbramos la inminencia de una avalancha de falsificaciones profundas en los procesos judiciales (¡al menos no por el momento ¡). Y en el caso de que así fuera, aunque pudieran implicar un costo adicional, confiamos en que los tribunales afronten los desafíos que plantean, tal y como han hecho en el pasado con generaciones anteriores de falsificaciones, sin necesidad de modificar las reglas probatorias ni de imponer, con carácter general, normas más restrictivas para verificar la autenticidad de las pruebas[53].

50 Miguel Freita, P.: "Deepfakes, conteúdo", cit., p. 203

51 Simó Soler, E.: "Retos jurídicos", cit., p. 505.
Parece haber consenso en que el enfoque más eficaz para identificar estas representaciones sintéticas es emplear la misma tecnología utilizada para generar *deepfakes*, esto es, redes generativas adversarias.

52 Pfefferkorn, R.: "Deepfakes" in the Courtroom", *BU Pub. Int. LJ*, 2020, vol. 29, p. 268.

53 Pfefferkorn, R.: "Deepfakes" in the Courtroom", cit., p. 246. En sentido contrario se pronuncia Delfino, que considera que debe exigirse a los tribunales que adopten medidas adicionales para determinar la autenticidad de las imágenes antes de admitirlas como prueba. "Deepfakes on trial: a call to expand the trial judge's. Gatekeeping role to protect legal proceedings from technological fakery", *Hastings Law Journal*, 2023, vol. 74, núm. 2, p. 297.

III. IA AL SERVICIO DE LA LUCHA CONTRA LA VIOLENCIA HACIA LAS MUJERES.

A pesar de haber expuesto los riesgos de que se refuercen y perpetúen los sesgos, debe al mismo tiempo reconocerse que los sistemas de IA bien diseñados pueden ayudar a identificarlos y, por ende, a corregirlos[54]. En este sentido, la Carta ética europea sobre el uso de la IA en los sistemas judiciales y su entorno, hace referencia a su capacidad para revelar la discriminación existente. Por ello, no podemos desatender las virtudes y ventajas que la IA puede ofrecernos[55].

Como señalan XENIDIS y SENDEN, si bien los algoritmos aumentan los problemas de discriminación en algunos casos, también ofrecen la oportunidad de reducir la arbitrariedad mediante una mayor explicabilidad de los procedimientos de toma de decisiones. Mientras que las decisiones humanas podrían asimismo calificarse de "caja negra" por su naturaleza opaca y no reproducible, los algoritmos de aprendizaje automático ofrecen la posibilidad de una toma de decisiones más responsable, siempre que se cumplan ciertos requisitos de transparencia. Las decisiones humanas, a diferencia de las decisiones algorítmicas, no pueden reproducirse cambiando un factor para comprobar de dónde procede la discriminación. Por lo tanto, ciertos principios como la transparencia, la explicabilidad y la rendición de cuentas son fundamentales para desarrollar aplicaciones de IA si el objetivo es convertir los riesgos existentes de discriminación en una oportunidad para aumentar la igualdad[56].

Del mismo modo, aunque hemos visto que la IA ha facilitado el surgimiento de nuevos modos de violencia contra las mujeres, no podemos obviar la otra cara de la moneda. La IA también puede emplearse en la lucha contra el mayor exponente de la desigualdad, como es la violencia de género. Entre otras, con herramientas que ayuden a proteger a las víctimas, a investigar determinados delitos, así como a prevenir su ejecución mediante técnicas predictivas que estimen la probabilidad de reincidencia. Así las cosas, se entiende que la IA tiene un enorme valor para

54 Sunstein, C. R., "Algorithms, Correcting Biases", *Forthcoming, Social Research*, 12 diciembre 2018, disponible en SSRN: https://ssrn.com/abstract=3300171

55 Interesantes resultan al respecto las reflexiones vertidas por Simó Soler, que propone el uso de modelos de *machine learning* para la detección de estereotipos de género en las sentencias. *Estereotipos de género en procesos por violencia sexual,* Tirant lo Blanch, Valencia, 2023.

56 Senden Xenidis, R. y Senden, L.: "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination", en AA.VV.: *General Principles of EU law and the EU Digital Order* (ed. por U. Bernitz *et al.*), Kluwer Law International, Países Bajos, 2020, p. 30. Señala al respecto Sanchis Crespo, La diferencia con los sesgos robóticos es que éstos se exteriorizan claramente —en tanto en cuanto el algoritmo sea transparente y esté bien evaluado— y desde esa perspectiva es más fácil mitigarlos. Los sesgos humanos pueden, sin embargo, pasar desapercibidos a menos que se muestren a las claras. "Inteligencia artificial y decisiones judiciales: crónica de una transformación anunciada", Scire: Representación y organización del conocimiento, 2023, vol. 29, núm. 2, p. 80.

mitigar ciertos aspectos de este tipo de violencia[57]. Veamos a continuación algunas de las bondades que puede brindarnos.

1. Tecnologías para atender a las víctimas. De la teleasistencia a la IA.

En los últimos años se han creado diversas herramientas "inteligentes" que pueden ser utilizadas para atender o auxiliar a las mujeres víctimas de violencia.

Ya en el año 2004 se puso en marcha Atenpro (Servicio Telefónico de Atención y Protección para víctimas de violencia contra las mujeres), un servicio de teleasistencia complementario al 016. Mediante el mismo se ofrece a las víctimas un dispositivo móvil a través del cual pueden recibir asistencia inmediata durante las 24 horas del día los 365 días del año. Atenpro, que se basa en la utilización de tecnologías de comunicación telefónica móvil, permite que las mujeres puedan entrar en contacto en cualquier momento con un Centro atendido por personal específicamente preparado para responder a sus necesidades, incluso en situaciones de riesgo con carácter de urgencia[58]. Durante muchos años este servicio apenas se ha modernizado. Sin embargo, recientemente se ha informado que va a incorporar IA para mejorar la protección de las usuarias y crear un sistema integral de seguimiento. Por un lado, los nuevos dispositivos se compondrán no solo de dispositivos móviles, sino también de relojes inteligentes y pulsadores que permiten a la policía la geolocalización de la víctima por GPS en caso de emergencia. Por otro lado, para mejorar su potencialidad, se pretende crear una aplicación informática dotada con IA que mejore la prevención y atención. Entre otras, la aplicación permitirá clasificar a las víctimas en función del nivel de riesgo, de forma que el personal de Atenpro contacte más con aquellas mujeres con mayor riesgo de ser agredidas[59]. Además de la protección a las víctimas de violencia de género, la actualización y modernización de este servicio, permitirá la atención a las víctimas de agresión sexual, acoso sexual y violencia sexual cometida en el ámbito digital, así como otras formas de violencia ejercida contra la mujer reconocidas en el Convenio de Estambul.

57 Macchiavelli, N.: Perspectiva de género en las nuevas tecnologías. El problema de los sesgos, *Diario Suplemento Derecho y Tecnología*, 2021, núm. 84, p. 9. Vid. asimismo Llorente Sánchez-Arjona, M.: "La inteligencia artificial como nueva estrategia de prevención en los delitos de violencia sexual", en AA.VV.: *Uso de la información y de los datos personales en los procesos: los cambios en la era digital* (dir. por I. Colomer Hernández), Aranzadi, 2022. La autora analiza el empleo de la IA en la lucha contra otras formas de violencia como la trata, pornografía infantil, agresores sexuales en serie, etc.

58 https://violenciagenero.igualdad.gob.es/informacionUtil/recursos/servicioTecnico/home.htm

59 El Gobierno ha ampliado el presupuesto estatal destinado al programa. En concreto, los fondos europeos Next Generation han consignado 32 millones para la modernización de Atenpro. En la puesta en marcha de la aplicación creada con IA está colaborando la Cátedra en Inteligencia Artificial de la Universidad de Alcalá. Martin, P., "España usará la IA y el 'big data' para proteger mejor a las víctimas del machismo", *Diario el Periódico*, 23 de septiembre de 2023, disponible en: https://www.elperiodico.com/es/sociedad/20230923/violencia-genero-machista-inteligencia-artificial-proteccion-big-data-victimas-92338576.

Aunque no se trata propiamente de un servicio de asistencia a las víctimas, queremos hacer mención a un interesante proyecto que, hasta donde alcanza nuestro conocimiento, todavía no se ha implantado. Nos referimos al Proyecto de investigación que recibe el nombre de "Certeza de Voz" del Instituto Andaluz de la Mujer (IAM), en colaboración con la Empresa Pública de Emergencias Sanitarias "EPES 061", dependiente de la Consejería de Salud y Familias. Se trata de un *software* inteligente creado para la detección precoz de supuestas víctimas de violencia de género mediante la voz de la mujer que llama a los Centros de Coordinación de Urgencias y Emergencias Sanitarias de Andalucía[60]. A través del mismo, se permitirá saber si la entonación, expresiones, uso de palabras, pausas o suspiros de la mujer que llama, muestra un patrón en las personas que sufren esta violencia. En tal caso, se generará una alerta de sospecha de un caso de violencia de género[61].

Fuera de España, destacamos PROTOBADI, creada en Bangladesh con el objetivo de proporcionar seguridad a las mujeres. Se trata de una aplicación para teléfonos inteligentes que crea mapas "calientes" que determinan las zonas con mayor probabilidad de producirse acoso sexual hacia las mujeres. Cuenta con un botón en la pantalla, que al ser presionado enciende una alarma muy ruidosa y a continuación, envía mensajes de texto a los contactos de la mujer indicando su ubicación y solicitando ayuda. Además, como hemos mencionado, permite recopilar los datos para configurar un mapa que señale las áreas más peligrosas, así como una especie de blog donde las usuarias pueden compartir sus experiencias[62]. Al igual que PROTOBADI, existen otras muchas aplicaciones destinadas a proteger a las mujeres en diversas partes del mundo, tales como: Eyewatch SOS for Women, SpotnSave Feel secure, iGoSafely, bSafe, Chilla, etc.[63]

Otras tecnologías que están utilizando IA para atender y ayudar a las víctimas de violencia de género son los chatbots. Entre otros, My*Sis Bot*, desarrollado en Tailandia, proporciona información, ayuda de emergencia, asistencia jurídica y acceso a diversos servicios (centros de llamadas sin ánimo de lucro, policía o juzgados de familia) a las mujeres que han sufrido violencia. La aplicación se descarga en el propio teléfono y permite a las usuarias mantener una conversación

60 Proyecto de investigación financiado con Fondos Feder, dentro del Pacto de Estado contra la Violencia de Género.

61 Constanza Gamboa, N., "La inteligencia artificial como herramienta al servicio de la erradicación de la Violencia de Género", Observatorio violencia, septiembre 2020, disponible en: https://observatorioviolencia.org/la-inteligencia-artificial-como-herramienta-al-servicio-de-la-erradicacion-de-la-violencia-de-genero/

62 El término Protobadi significa "alguien que protesta" en bengalí. Marks, P., "Bangladesh: Sex harassment app helps women map abuse", *NewScientist,* mayo 2014.

63 Vid. Gopika, B., "10 Safety Apps For Women", 12 junio 2018, BW Business World, disponible en https://www.businessworld.in/article/10-Safety-Apps-For-Women/12-06-2018-151793/

en tiempo real durante la cual reciben la asistencia que necesitan[64]. O el chatbot holandés elaborado por la Universidad de Maastricht que, con el fin de atender a las víctimas de acoso y agresión sexual, les permite contar su historia con libertad y ofrece a continuación consejos del lugar al que se debe acudir en función de cada caso (comisaría, hospital, psicólogo o refugio)[65]. Por su parte, en América Central se ha desarrollado el chatbot o asistente virtual inteligente Sara[66], diseñado con IA, que orienta a las mujeres sobre el riesgo de sufrir o haber sufrido violencia y proporciona información acerca de donde denunciar, los derechos que le asisten, etc[67].

En lo que a nuestro país respecta, se están desarrollando algunos proyectos, como el programa Improve (*Improving Access to Services for Victims of Domestic Violence by Accelerating Change in Frontline Responder Organisations*), financiado por la Unión Europea (*Horizon Europe*). Este robot conversacional multilingüe con IA, ofrecerá a las víctimas, si prefieren no acudir a una comisaría o llamar a la policía, asesoramiento inmediato, evaluación de riesgos además de orientarles sobre los servicios y recursos disponibles.

Son otras muchas las tecnologías que pueden proporcionar ayuda a las víctimas de violencia de género y que se están desarrollando en todo el mundo. Desde herramientas que analizan imágenes de vídeo para detectar comportamientos agresivos o violentos hacia las mujeres hasta análisis de llamadas de emergencia realizadas por mujeres con base en el tono de su voz y el lenguaje utilizado o identificación de publicaciones en redes sociales que contengan contenido de acoso[68].

2. Sistemas de IA en el marco de la investigación.

La IA puede resultar extremadamente útil en la investigación criminal, especialmente por lo que respecta a la mejora de los métodos de trabajo de las autoridades policiales. Como bien sabemos, aunque la instrucción sea dirigida por los jueces, son ellas quienes se encargan en realidad de llevarla a cabo.

64 "Using AI in accessing justice for survivors of violence", 30 mayo 2019, disponible en: https://www.unwomen.org/en/news/stories/2019/5/feature-using-ai-in-accessing-justice-for-survivors-of-violence.

65 https://eldiariofeminista.info/2019/10/11/chatbot-para-victimas-de-acoso-sexual/

66 https://chatbotsara.org/

67 Desarrollado por el Proyecto Regional Infosegura, iniciativa del Programa de las Naciones Unidas para el Desarrollo en colaboración con la Agencia de los EEUU para el Desarrollo Internacional.

68 Ejemplo paradigmático es el de Suecia, que ostenta el índice de igualdad de género más alto de la Unión Europea, que ha adoptado diversas iniciativas para usar tecnologías disruptivas de una forma proactiva a favor de la igualdad de género. Véase el informe presentado por Ortiz De Zárate Alcarazo, l. y Guevara Gómez. A.: "Inteligencia artificial", cit., p. 49.

Es evidente que la investigación penal está experimentando una transformación[69]. Las diligencias de investigación tecnológicas, recogidas en nuestra LECrim desde la reforma operada por la Ley 13/2015, han incorporado, en mayor o menor medida, sistemas de IA para el esclarecimiento y descubrimiento de los delitos[70]. Así, la policía empieza a emplear tecnologías de reconocimiento facial[71] (por ejemplo, para buscar en bases de datos de sospechosos e identificar a víctimas de trata de seres humanos o abuso y explotación sexual infantil), de identificación por voz, reconocimiento del habla, análisis autónomos de bases de datos identificadas, técnicas predictivas (actuación policial predictiva y análisis de puntos críticos de delincuencia), herramientas avanzadas de autopsia virtual para ayudar a determinar la causa de la muerte, vigilancia de las redes sociales (rastreo [scraping] y recopilación de datos para detectar conexiones), etc[72].

La implementación de sistemas de IA puede optimizar considerablemente algunas diligencias de investigación ya preexistentes. Pensemos, entre otras, en la práctica del agente encubierto informático (282 bis 6 de la LEcrim). Podríamos recurrir a la IA generativa, es decir, a sistemas que generan imagen, audio y vídeo para simular la identidad del agente e incluso para crear el material necesario para intercambiar el archivo ilícito[73]. De este modo, podría ponerse la técnica de los *deepfak*es al servicio de la investigación de determinados delitos y, por ejemplo, crear material "ultrafalso" pornográfico con la intención de desarticular una red

69 Como señala Barona Vilar, las ciencias forenses, la criminalística, han mutado; los métodos empleados se sostienen sobre algoritmos, software, que han introducido técnicas idóneas para ubicar, analizar, e introducir evidencias y pruebas en el proceso penal. *Algoritmización del Derecho y de la Justicia. De la inteligencia artificial a la Smart e Justice*, Tirant Lo Blanch, Valencia, 2021, p. 502.

70 Recordemos que la Ley 13/2015, de 5 de octubre, de modificación de la LECrim para el fortalecimiento de las garantías procesales y la regulación de las medidas de investigación tecnológicas, introdujo numerosos preceptos (arts. 588 bis a hasta 588 octies) para regular nuevos medios de investigación tecnológicos. Entre otros, la interceptación de las comunicaciones telefónicas y telemáticas (arts.588 ter a) y ss.); grabación de las comunicaciones orales directas (art. 588 quater a); captación de imágenes en lugares o espacios públicos (art. 588 quinquies a); utilización de dispositivos de geolocalización (art. 588 quinquies b); registro de dispositivos de almacenamiento masivo (art. 588 sexies b); registro remoto sobre equipos informáticos (art. 588 septies y ss.), etc.

71 Téngase en cuenta que el Reglamento de IA, en su art. 5, cataloga como "Prácticas de IA prohibidas", el uso de sistemas de identificación biométrica a distancia «en tiempo real» en espacios de acceso público con fines e aplicación de la ley, salvo y en la medida en que dicho uso sea estrictamente necesario para alcanzar determinados objetivos, entre los que se encuentra: la búsqueda selectiva de víctimas concretas de secuestro, trata de seres humanos o explotación sexual de seres humanos, así como la búsqueda de personas desaparecidas; la prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas físicas o de una amenaza real y actual o real y previsible de un atentado terrorista; o la localización o identificación de una persona sospechosa de haber cometido una infracción penal a fin de llevar a cabo una investigación o un enjuiciamiento penales o de ejecutar una sanción penal por alguno de los delitos mencionados en el anexo II que en el Estado miembro de que se trate se castigue con una pena o una medida de seguridad privativas de libertad cuya duración máxima sea de al menos cuatro años..

72 Vid. en este sentido, el Considerando M. de la Resolución del Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI)).

73 González Pulido, I.: "El uso de la inteligencia artificial generativa en la investigación de la ciberdelincuencia de género: ante el auge de los deepfakes", *Ius et Scientia*, 2023, vol. 9, núm. 2, p. 176.

criminal, ganándose la confianza de sus miembros[74]. Algo similar es lo que ha hecho la holandesa *Sweetie*, de la ONG *Terre des Hommes*, que a través de un bot que se hace pasar por una niña, persigue la pornografía infantil en la red y ha conseguido localizar a miles de pederastas. A pesar de sus logros, la polémica está servida. Se plantean numerosas dudas acerca de su legalidad: presupuestos de legitimidad en la utilización de un agente encubierto robótico, los efectos jurídicos asociados a la provocación del delito, así como las dificultades para reconocer el derecho a la indemnidad sexual en un robot[75].

Como señala Llorente Sánchez-Arjona, el escenario en el que se desarrollan determinados delitos (abuso y explotación sexual infantil) resulta propicio a los sistemas de IA que pueden convertirse en una tecnología clave para hacer frente a los mismos. La autora cita algunos ejemplos, como el software denominado iCOP que permite identificar pederastia en la red; las herramientas creadas por Google y Apple (Neural Match) con el objetivo de luchar contra este tipo de delitos; C-SEX que analiza el comportamiento de los usuarios en el entorno de la pornografía infantil, etc.[76]. En nuestro país, la Secretaría de Estado de seguridad ha impulsado un proyecto para implementar una versión española de la herramienta *Chat Analysis Triage Tool* (CATT) para casos de *online child grooming*. Su objetivo es identificar, mediante el análisis del discurso y las tácticas utilizadas por los groomers en los chats, a aquellos abusadores que pretenden tener un encuentro físico con el menor, y diferenciarlos de aquellos que solo buscan satisfacer sus fantasías sin buscar un contacto real. De manera que se prioricen los recursos policiales en los primeros supuestos[77].

Sin duda alguna, la IA va a repercutir en el aumento de la eficacia de la lucha contra determinados delitos, entre los que destacamos, la explotación sexual en línea. Pero también debemos ser conscientes de los riesgos que implica. De ahí que nos cuestionemos, junto a Marchena Gómez, los límites que hay que imponer para que la investigación de esos delitos por el Estado no desborde los presupuestos que legitiman el ejercicio del *ius puniendi*. Como señala el magistrado, la necesidad de actualizar la metodología de la investigación penal no puede ser cuestionada.

74 Blázquez Moreno, R.: "Deepfakes en el procedimiento probatorio", *Revista vasca de derecho procesal y arbitraje*, 2023, vol. 35, núm. 3, p. 231.

75 Sea como fuere, la utilización de Sweetie ha sido asociada a la ventaja que proporciona, no ya como herramienta de investigación, sino para paliar los negativos efectos que los delitos de pornografía infantil producen en los agentes que los investigan. Los agentes infiltrados que operan en chats con el objetivo de localizar pedófilos tienen que soportar una fuerte carga psicológica por la exposición continuada a contenidos de esta pornografía, por lo que han de ser sustituidos cada cierto tiempo y pueden tener secuelas psicológicas, problema que se eliminaría si fuera un robot el que tratara con esos contenidos. Marchena Gómez, M.: "Inteligencia artificial y jurisdicción penal", Discurso con motivo de su ingreso como Académico de Número de la Real Academia de Doctores de España el 26 de octubre de 2022, separata de la Real Academia de Doctores de España, Madrid, p. 14.

76 Llorente Sánchez-Arjona, M.: "La inteligencia", cit., pp. 274 y 275.

77 En idéntico sentido, González-Álvarez, J.L., Santos-Hermoso, J. y Camacho-Collados, M.: "Policía predictiva en España. Aplicación y retos futuros", *Behavior & Law Journal*, 2020, núm. 6(1), p. 30.

Sin embargo, la constatación de ese hecho no debe llevarnos a legitimar, al amparo de las ventajas técnicas de la IA, una investigación en la que todo vale, sin reparar en la intensa injerencia estatal y consiguiente sacrificio del espacio de intimidad que cada ciudadano dibuja frente a los poderes públicos y a terceros. Por eso la importancia de que la regulación de estas diligencias ligadas a las nuevas tecnologías sea encabezada por una referencia a los principios rectores a los que se refiere el artículo 588 bis a) de la LECrim, es decir, a los principios de especialidad, idoneidad, necesidad y proporcionalidad[78].

En definitiva, no podemos olvidar que el empleo de sistemas de IA en el seno de una investigación judicial entraña numerosos riesgos que pueden afectar seriamente a los derechos y garantías constitucionales que deben presidir el proceso. La eficiencia no puede en modo alguno anteponerse o ir en detrimento de los mismos[79]. Deben, por tanto, necesariamente salvaguardarse los derechos de las personas investigadas. En este sentido, si bien no se prohíben en el Reglamento europeo de IA, se supedita su utilización al cumplimiento de determinados requisitos (previstos en los artículos 8 y ss) dado que, como ya hemos adelantado, se califican de alto riesgo. Entre otros, se exige garantizar un nivel de transparencia suficiente, permitir una efectiva supervisión humana y contar con un nivel adecuado de precisión, solidez y ciberseguridad.

3. Instrumentos de valoración del riesgo de reincidencia.

Tanto en el seno del proceso penal como en el ámbito penitenciario, asistimos al empleo de instrumentos de valoración del riesgo de reincidencia (*risk assessment instruments*, conocidos por sus siglas, RAIs)[80]. Estas herramientas estructuradas de valoración del riesgo han ido evolucionando hasta alcanzar su automatización y

78 Marchena Gómez, M.: "Inteligencia artificial", cit., p. 15. Como señala este autor, una puntualización es obligada. Los principios a los que hacemos referencia representan límites axiológicos que la LECrim contempla como presupuestos de legitimidad para validar diligencias intrusivas en el círculo de derechos definidos por el art. 18 de la CE. Sin embargo, son otros muchos los derechos afectados cuando el ciudadano se expone a las técnicas de investigación de IA que aspiran al esclarecimiento del hecho investigado. Algunos de estos derechos y los principios que han de condicionar su limitación por el Estado tienen hoy nombre propio en nuestro sistema constitucional -principio de contradicción, igualdad, derecho de defensa, imparcialidad del órgano judicial, protección de datos *ex* art. 24 de la CE. Otros son principios y derechos de nueva generación que discurren, hoy por hoy, en el terreno dogmático -principio de transparencia algorítmica, principio de trazabilidad, principio de imparcialidad del validador o derechos a la dignidad algorítmica y a la identidad algorítmica-, que, a buen seguro, adquirirán, antes o después, tratamiento normativo.

79 Alerta BARONA VILAR, además, sobre el peligro de emplear las medidas de investigación tecnológicas de alta fiabilidad investigadora para reducir o eliminar riesgos, empero no para investigar hechos cometidos. Obviamente, la confusión de funciones no es neutra, ni las consecuencias que se van a producir en el respeto a los derechos humanos tampoco lo es. Abandonamos el derecho penal *ex post*, para construir el derecho penal *ex ante*, que reacciona ante riesgos y amenazas, y lo hace con toda la carga en profundidad sobre las garantías y los derechos. *Algoritmización del,* cit., p. 503.

80 Aunque vamos a analizarlos únicamente en el marco de un proceso penal, estas herramientas también se emplean en el ámbito penitenciario. Así ocurre en nuestro país, en concreto en las prisiones catalanas, que utilizan la herramienta RISCANVI para evaluar la conducta de los internos, y en función de ello, asistir en la decisión acerca de la situación del privado de libertad, ya sea para concederle un permiso de salida, clasificarle en un grado o incluso, otorgarle la libertad condicional.

digitalización con un elevado grado de sofisticación. De hecho, las más novedosas se asisten de algoritmos para emitir el pronóstico, incluso recurriendo a sistemas inteligentes computarizados para su tratamiento[81], con el fin de conjurar el riesgo de reiteración delictiva basándose en la información que obra en los expedientes y en las informaciones estadísticas de casos previos[82].

Los instrumentos de valoración del riesgo pueden ser útiles en el asesoramiento al juez acerca del riesgo de reincidencia del presunto maltratador. Este riesgo, recordamos, es determinante en la adopción de las órdenes de protección del artículo 544 ter LEcrim u otras medidas cautelares para proteger a las víctimas de violencia de género (medidas de alejamiento, prohibición de comunicación, etc.) así como a sus hijos/as menores (suspensión cautelar del régimen de visitas del art. 544 ter.7 LEcrim)[83]. En este sentido, estas herramientas pueden auxiliar al juez a la hora de adoptar su decisión. Eso sí, solo como un elemento más, que deberá en todo caso ser corroborado por otros datos o circunstancias.

El carácter crónico y repetitivo de la violencia contra la pareja, así como la relación de afectividad existente entre la víctima y el agresor, conlleva un riesgo adicional para las víctimas de violencia de género que difiere del de otros delitos, y muestra una urgente necesidad de protección frente a una posible reincidencia. De ahí que, en los últimos tiempos, los instrumentos de valoración del riesgo de violencia contra la pareja se hayan multiplicado. Entre otros, destacan: SARA (*Spousal Assault Risk Assessment*), DASH (*Domestic abuse, stalking y harassment and honour-based violence*) y SVR-20 (*Sexual Violence Risk Assessment*[84]).

En España contamos desde el año 2007 con el sistema VioGén (Sistema de Seguimiento Integral en los casos de Violencia de Género). Esta herramienta informática permite el seguimiento y protección de las víctimas de violencia de género y de sus hijos/as. Entre sus objetivos, destaca la realización de valoraciones policiales del riesgo de las víctimas denunciantes de sufrir una nueva agresión, y en función del resultado, poder protegerlas. Para dicha tarea se sigue un Protocolo en el que se emplean dos instrumentos complementarios: la Valoración Policial del Riesgo (VPR4.0) para realizar una estimación inicial y la Valoración Policial de

81 Romeo Casabona, C. M.: "Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad", *Revista de Derecho, Empresa y Sociedad (REDS)*, julio-diciembre 2018, núm. 13, p. 43

82 Simón Castellano, P.: *Justicia Cautelar e inteligencia artificial. La alternativa a los atávicos heurísticos judiciales*, Bosch, Barcelona, 2021, p. 25.

83 Como refiere Magro Servet, nos encontramos ante un fenómeno claramente repetitivo y con un aspecto conductual que se reproduce en el tiempo, que tiene unos parámetros de actuación homogéneos en la mayoría de los casos y con un carácter predecible en cuanto a los hechos que han ocurrido y a la protección a las víctimas de lo que pueda ocurrir. Pocas materias existen en la actualidad en donde el mimetismo conductual se reproduce con tanta repetición como en la violencia de género. "La inteligencia artificial para mejorar la lucha contra la violencia de género", en AA.VV.: *Inteligencia artificial legal y Administración de Justicia* (dir. por S. Calaza López y M. Llorente Sánchez-Arjona), Cizur Menor: Aranzadi, 2022, pp. 398 y 405.

84 Esta herramienta valora el riesgo de violencia sexual.

la Evolución del Riesgo (VPER4.0) para el seguimiento del caso. Si bien VioGén se diseñó inicialmente para evaluar el riesgo de reincidencia de una agresión, posteriormente se creó otro algoritmo para valorar la posibilidad de sufrir violencia letal (formulario VPR5.0).

De manera que, cuando una víctima de violencia de género interpone una denuncia, deviene preceptiva la realización por parte de la policía de un análisis de la valoración del riesgo al que se encuentra sometida. Para ello, se le formulan una serie de preguntas y se cumplimenta por los agentes policiales el formulario VPR. Los datos introducidos se someten a un algoritmo que, tras valorar automáticamente cada ítem, establece uno de los cinco niveles de riesgo que presenta la víctima de sufrir una nueva agresión a corto plazo: "extremo", "alto", "medio", "bajo" y "no apreciado". En función del resultado asignado, se llevan a cabo de forma inmediata las medidas provisionales de protección policial aparejadas a cada nivel de riesgo[85]. El informe obtenido se incluirá en el atestado (junto al resto de diligencias policiales) y servirá para informar al juez a la hora de decidir las medidas de protección de la víctima que deben adoptarse en cada caso. Efectuada la valoración inicial, la estimación del riesgo debe mantenerse actualizada. Para ello, la policía cumplimenta el segundo formulario (VPER) que, mediante valoraciones periódicas, permite la monitorización de las víctimas[86].

Sin negar la utilidad de esta herramienta, no podemos obviar que puede fallar. Muestra de ello, es la Sentencia de 30 de septiembre de 2020 de la Audiencia Nacional, que ha condenado al Estado español por la deficiente protección que la Guardia Civil proporcionó a una mujer que solicitó una orden de protección[87]. El sistema VioGén asignó el nivel de "no apreciado" y sin mayores indagaciones, a pesar de la existencia de indicios suficientes de maltrato, las autoridades policiales calificaron el caso en este sentido, lo que determinó que el juez denegara la orden

85 Adviértase que el agente puede modificar (al alza) el riesgo apreciado si estima que existen razones para ello.

86 Conviene dejar claro que el sistema VioGén, al igual que otras herramientas actuariales de valoración del riesgo, no puede considerarse IA en sentido estricto. Miró Llinares, F.: "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *Revista de Derecho Penal y Criminología*, 2018, núm. 20, p. 103. Cierto es que la conexión con la IA es evidente, pero la IA va más allá. Como señala Simón Castellano, es una diferencia sutil, puesto que casi todas las herramientas actuariales tienen una parte automática y automatizable, o digital, como es el cálculo de la probabilidad, pero esto no en realidad motivo bastante para considerarlo por si solo IA. *Justicia Cautelar,* cit., p. 94

87 Recordamos la importancia de formar en perspectiva de género a quienes hacen uso de estas herramientas. Como señala Arruti Benito, este caso evidencia la ausencia de perspectiva de género en la implementación de los sistemas de IA. Por un lado, muestra la fe ciega a la objetividad de los sistemas de IA, olvidando su carácter complementario y asistencial –que no sustitutivo–depositando, en la valoración de un algoritmo, la toma de decisión que corresponde a la inteligencia humana y al sentido común humano. Por otro lado, pone de manifiesto la ausencia de conocimiento sobre el potencial discriminatorio que entrañan este tipo de sistemas debido a los sesgos de género que pueden incorporarse. "Justicia e inteligencia artificial en clave de género", en AA.VV.: *Investigación y género. Proyectos y resultados en estudios de las mujeres: VIII Congreso Universitario Internacional de Investigación y Género* (ed. por M. E. García y A. M. de la Torre Sierra), Universidad de Sevilla, 2022, p. 402.

solicitada y su consecuente trágico final por el que la mujer murió asesinada a manos de su marido[88].

Son muchas las voces que advierten acerca de las falencias de este sistema[89]. Entre otras, se critica que actualmente VioGén se rige por unos parámetros que han ha quedado arcaicos, no siempre se ajustan a la realidad, y en ocasiones arroja resultados poco adecuados[90]. Se ha evidenciado que el sistema necesita una mejora. De ahí que, tal y como se ha anunciado, el Área de Violencia de Género, Estudios y Formación de la Secretaría de Estado de Seguridad ha incorporado la plataforma analítica de la empresa de software SAS Iberia, que va a "facilitar actualizaciones mucho más rápidas y eficaces del Protocolo VPR del Sistema VioGén, ponderando mejor los actuales indicadores de riesgo de reincidencia e identificando nuevas variables que ayuden a afinar aún más esa valoración de riesgo mediante análisis automatizados y en tiempo real de grandes cantidades de datos"[91].

V. A MODO DE SÍNTESIS FINAL.

La transformación digital de la Justicia debe realizarse desde una perspectiva de género. Solo así podrá promoverse una justicia que garantice el respeto del derecho constitucional a la tutela judicial efectiva en condiciones de igualdad.

El empleo de herramientas y aplicaciones de IA puede tener un impacto negativo significativo en la justicia, si no se identifican, abordan y afrontan los posibles sesgos algorítmicos que refuerzan estereotipos de género. Al mismo tiempo, preocupa la posibilidad de que la IA pueda ser utilizada como un instrumento para ejercer nuevas formas de violencia de género. Entre otras, a través de los *deepfake*s, que, además, podrían ser incorporados en un juicio socavando la integridad de los

88 Adviértase, que no se condena por error judicial, sino por no recibir una adecuada información de la Guardia Civil acerca de cuál era la situación real del riesgo de la víctima, que fue determinante a la hora de orientar al juez en el (no) dictado de una orden de protección.

89 Como alerta el Informe de auditoría externa del Sistema VioGén llevado a cabo en el año 2022 por la Fundación Ana Bella (p. 26), el sistema se basa en el supuesto de que las mujeres entienden y responden con claridad a los 35 puntos del formulario VPR y los agentes de policía transforman objetivamente las declaraciones de las mujeres en respuestas binarias (presente/no presente). Pero en la realidad, el proceso rara vez funciona de este modo idealizado. Esto significa que la calidad de los datos introducidos podría verse comprometida durante la fase de generación de datos, lo que daría lugar a posibles fuentes de sesgo y tergiversación.

90 En este sentido, Llorente Sánchez-Arjona, M.: "La inteligencia artificial", cit., p. 268. De otro lado, las Fuerzas y Cuerpos de Seguridad del Estado son las que se ocupan de introducir todos los datos en el sistema, muchos de los cuales desconocen y deben contestar de forma intuitiva a partir de la toma de declaración de la víctima, cuando presenta la denuncia, y del agresor.

91 Con el *software* de SAS Iberia se incorpora tecnología de analítica avanzada e IA que automatizará el análisis de una mayor cantidad de datos de criminalidad, combinados incluso con datos de fuentes abiertas, lo que ayudará a ponderar mejor los algoritmos, identificando nuevos indicadores de riesgo, y en periodos de tiempo mucho más cortos. Además, son algoritmos más sensibles a la evolución de la criminalidad y mejoran con ello la predicción de aquellos casos en los que es previsible que se produzcan agresiones reincidentes. La Moncloa. Interior 15.12.2020, disponible en: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2020/151220-inteligencia.aspx

procesos judiciales. Debemos, por tanto, tener en cuenta estos peligros y tratar de establecer medidas de índole política, jurídica y tecnológica adecuadas para prevenirlos.

A pesar de los riesgos mencionados, no podemos dejar de reconocer las bondades que las nuevas herramientas algorítmicas pueden ofrecer a la justicia en términos de colaboración o auxilio. Especialmente y en lo que a este trabajo se refiere, cuando nos enfrentamos ante uno de los mayores exponentes de la desigualdad como es la violencia de género. Estas herramientas pueden utilizarse para apoyar a las víctimas, investigar delitos y prevenir la reincidencia mediante sistemas de evaluación del riesgo, siempre y cuando no se menoscaben los derechos y garantías que deben presidir un proceso.

En definitiva, si bien la IA tiene el potencial de mejorar la eficiencia, efectividad e incluso, la calidad de la justicia, deviene imprescindible abordar los riesgos asociados a las misma. Para ello, la integración de la perspectiva de género no solo en su diseño y desarrollo sino también en su aplicación por parte de nuestros Juzgados resulta fundamental en aras de garantizar su impacto positivo y su alineación con el principio de igualdad.

BIBLIOGRAFIA

Arruti Benito, S.: "Justicia e inteligencia artificial en clave de género", en AA.VV.: *Investigación y género. Proyectos y resultados en estudios de las mujeres: VIII Congreso Universitario Internacional de Investigación y Género* (ed. por M. E. García y A. M. de la Torre Sierra), Universidad de Sevilla, 2022, pp. 395-406.

Barona Vilar, S.: *Algoritmización del Derecho y de la Justicia. De la inteligencia artificial a la Smart Justice*, Tirant lo Blanch, Valencia, 2021.

Barona Vilar, S.: "La necesaria deconstrucción del modelo patriarcal de justicia", en AA.VV.: *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018.

Barona Vilar, S.: "Dataización de la justicia (Algoritmos, Inteligencia Artificial y Justicia, ¿el comienzo de una gran amistad?)", *Revista Boliviana de Derecho*, 2023, núm. 36, pp. 14-45.

Belloso Martin, N.: "La problemática de los sesgos algorítmicos (con especial referencia a los de género) ¿Hacia un derecho a la protección contra los sesgos?", en AA.VV.: *Inteligencia artificial y filosofía del derecho* (dir. por F. llano Alonso), Laborum, Murcia, 2022.

Bello San Juan, P.: "La inteligencia artificial al servicio del crimen: La revolución del deepfake desde una perspectiva criminológica", en AA.VV.: *La justicia en la sociedad 4.0: nuevos retos para el siglo XXI* (dir. por L. Fontestad Portalés), Colex, 2023.

Blanco García, A. I.: "Retos para una inteligencia artificial inclusiva de los colectivos vulnerables", *Revista Actualidad jurídico Iberoamericana*, 2024, núm. 21.

Blázquez Moreno, R.: "Deepfakes en el procedimiento probatorio", *Revista vasca de derecho procesal y arbitraje*, 2023, vol. 35, núm. 3.

Borraz, M. y Pastor, A.: "'Deepfakes' sexuales: el caso de las menores de Almendralejo consolida una nueva forma de violencia machista", e*l Diario*.es, 19 de septiembre de 2023.

Catalán Chamorro, M.J.: "La carta de derechos digitales y su implicación en el derecho procesal español", en AA. VV.: *Digitalización de la justicia: prevención, investigación y enjuiciamiento* (dir. por M. Llorente Sánchez-Arjona y S. Calaza López), Aranzadi, Cizuer Menor, 2022, pp. 179 - 208.

Catalán Chamorro, M.J.: L*a justicia digital en España. Retos y desafío*s, Tirant Lo Blanch, Valencia, 2023.

Cerdán Martínez, V. y Padilla Castillo, G.: "Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso", *Historia y comunicación social*, 2019, núm. 24 (2).

Constanza Gamboa, N.: "La inteligencia artificial como herramienta al servicio de la erradicación de la Violencia de Género", Observatorio violencia, septiembre 21, 2020.

Danesi, C.: "Sesgos algorítmicos de género con identidad iberoamericana: las técnicas de reconocimiento facial en la mira", *Revista Derecho de Familia*, 2021, núm.100.

Delfino, R. A., "Deepfakes on trial: a call to expand the trial judge's. Gatekeeping role to protect legal proceedings from technological fakery", *Hastings Law Journal,* 2023, vol. 74, núm. 2, pp. 293- 348.

De Luis García, E., "Justicia, inteligencia artificial y derecho de defensa", *IDP: revista de internet, derecho y política*, 2023, núm. 39.

Ei, D., y Moser, G.: "Human arbitrators (the undisputed champion) v (the robots challenger)", *Hong Kong L.J*, 2020, vol. 50.

Fernández, A.: "Inteligencia artificial en los servicios financieros", *Boletín económico - Banco de España*, 2019, núm. 2.

Gil, P.: "La perspectiva de la mujer víctima del sistema judicial ajeno al género", en AA.VV.: *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018.

González Pulido, I.: "El uso de la inteligencia artificial generativa en la investigación de la ciberdelincuencia de género: ante el auge de los deepfakes", *Ius et Scientia*, 2023, vol. 9, núm. 2.

González-Álvarez, J.L., Santos-Hermoso, J. y Camacho-Collados, M.: "Policía predictiva en España. Aplicación y retos futuros", *Behavior & Law Journal*, 2020, núm. 6(1).

Hao, K., "This Is How AI Bias Really Happens - and Why It's So Hard to Fix", *MIT Technology Review*, 4 febrero 2019.

Kweilin, L.T.: "Deepfakes and domestic violence: perpetrating intimate partner abuse using video technology", *Victims & Offenders*, 2022, vol. 17, núm. 5.

Llorente Sánchez-Arjona, M.: "La inteligencia artificial como nueva estrategia de prevención en los delitos de violencia sexual", en AA.VV.: *Uso de la información y de los datos personales en los procesos: los cambios en la era digital* (dir. por I. Colomer Hernández), Aranzadi, Cizur Menor, 2022.

Lloria García, P.: *Violencia sobre la mujer en el siglo XXI. Violencia de control y nuevas tecnologías: habitualidad, sexting y stalking*, Iustel, Madrid, 2020.

LO, M.: "A Domestic Violence Dystopia: Abuse via the Internet of Things and Remedies Under Current Law", *California Law Review*, 2021, vol. 109.

Lorente Acosta, M.: "Justicia, género y estereotipos", en AA.VV.: *Análisis de la Justicia desde la perspectiva de género*, Tirant Lo Blanch, Valencia, 2018.

Macchiavelli, N.: "La violencia de género y el uso de algoritmos como herramienta efectiva para la protección de los derechos fundamentales", 2022, *AFD*, (XXXVIII).

Macchiavelli, N.: "Perspectiva de género en las nuevas tecnologías. El problema de los sesgos", *Diario Suplemento Derecho y Tecnología*, 2021, núm. 84.

Magro Servet, v.: "La inteligencia artificial para mejorar la lucha contra la violencia de género", en AA.VV.: *Inteligencia artificial legal y Administración de Justicia* (dir. por S. Calaza López y M. Llorente Sánchez-Arjona), Cizur Menor: Aranzadi, 2022.

Marchena Gómez, M.: "Inteligencia artificial y jurisdicción penal", Discurso con motivo de su ingreso como Académico de Número de la Real Academia de Doctores de España el 26 de octubre de 2022, separata de la Real Academia de Doctores de España, Madrid.

Marcos Francisco, D.: "Sistema arbitral de consumo: algunas propuestas 'inteligentes' de lege ferenda", *InDret*, 2024, núm. 1, pp. 114-150.

Marks, P., "Bangladesh: Sex harassment app helps women map abuse", *NewScientist,* mayo 2014.

Martin, P., "España usará la IA y el 'big data' para proteger mejor a las víctimas del machismo", *Diario el Periódico*, 23 de septiembre de 2023.

Martínez García, E.; Borges Blázquez, R. y Simó Soler, E.: "Inteligencia artificial y perspectiva de género en la justicia penal", *Diario* La *Ley*, Sección Ciberderecho, 20 de enero de 2021, núm 47.

Mayson, S. G.: "Bias In, Bias Out", *Yale Law Journal*, 2018, núm. 128, pp. 2218-2300.

Miguel Freita, P.: "Deepfakes, conteúdo gerado por inteligência artificial e verdade processual", en AA.VV.: *El proceso penal ante una nueva realidad tecnológica europea* (dir. por C. Arangüena Fanego, M. De Hoyos Sancho y E. Pillado González), Thomson Reuters Aranzadi, 2023.

Miró Llinares, F.: "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", *Revista de Derecho Penal y Criminología*, 2018, núm. 20.

Navas Navarro, S.: "La perspectiva de género en la inteligencia artificial", *Diario La Ley*, sección Ciberderecho, 8 de marzo de 2021, núm. 48.

O'neil, C.: *Armas de destrucción matemática*, Capitán Swing, Madrid, 2017.

Ortiz De Zárate, L. y Guevara gómez, A.: Inteligencia artificial e igualdad de género. Un análisis comparado entre la UE, Suecia y España, Fundación alternativas, 2021, núm. 101.

Pfefferkorn, R.: «Deepfakes" in the Courtroom», *BU Pub. Int. LJ*, 2020, vol. 29.

Rivas Vallejo, P.: "Sesgos de género en el uso de inteligencia artificial para la gestión de las relaciones laborales: análisis desde el derecho antidiscriminatorio", *e-Revista Internacional de la Protección Social(e-RIPS)*, 2022, vol. VII, núm. 1.

Romeo Casabona, C. M.: "Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad", *Revista de Derecho, Empresa y Sociedad* (REDS), julio-diciembre 2018, núm. 13.

Sanchis Crespo, C., "Inteligencia artificial y decisiones judiciales: crónica de una transformación anunciada", *Scire: Representación y organización del conocimiento*, 2023, vol. 29, núm. 2, pp. 65-84.

Senden Xenidis, R. y Senden, L.: "EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination", en AA.VV.: *General Principles of EU law and the EU Digital Order* (ed. por U. Bernitz *et al.*), Kluwer Law International, Países Bajos, 2020.

Simó Soler, E.: "Retos jurídicos derivados de la Inteligencia Artificial Generativa Deepfakes y violencia contra las mujeres como supuesto de hecho", *InDret*, febrero 2023, núm. 2, pp. 493- 515.

Simó Soler, E., *Estereotipos de género en procesos por violencia sexual,* Tirant lo Blanch, Valencia, 2023.

Simón Castellano, P., *Justicia Cautelar e inteligencia artificial. La alternativa a los atávicos heurísticos judiciales*, Bosch, Barcelona, 2021.

Soriano Arnanz, A.: "Discriminación algorítmica: garantías y protección jurídica", en AA.VV.: *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas* (dir. por L. Cotino Hueso), Aranzadi, 2022, pp. 139-169.

Soriano Arnanz, A.: "Creating non-discriminatory Artificial Intelligence systems: balancing the tensions between code granularity and the general nature of legal rules", *Revista de Internet, Derecho y Política*, 2023, núm. 38.

Soto Santana, M.: "Justice for Women: Deep fakes and Revenge Porn", 3rd Global Conference on woman´s studies, 25-27 septiembre 2022.

T Trend Micro Research, United Nations Interregional Crime and Justice Research Institute (UNICRI) Europol's European Cybercrime Centre (EC3), "Malicious Uses and Abuses of Artificial Intelligence", 2020.