

# BOLETÍN

## INTELIGENCIA ARTIFICIAL

### VOLUMEN V

Julio  
2025

Juezas y Jueces  
*para la* Democracia

#### ARTICULOS

**INTELIGENCIA ARTIFICIAL  
Y FEMINISMO ¿CUESTIÓN  
DE GÉNERO O DE ALARMA  
SOCIAL?**

Victoria Trujillo Machuca

**HACIA UNA JUSTICIA  
DIGITAL CON  
PERSPECTIVA DE GÉNERO**

María López Beloso, Ainhoa Izaguirre  
Choperena y Borja Sanz Urquijo

**INTELIGENCIA ARTIFICIAL**

Julio Díaz Casales

# BOLETÍN

## INTELIGENCIA ARTIFICIAL

VOLUMEN V

### Dirección

Diego Gutierrez Alonso  
Benjamín Sánchez Fernández  
María Gavilán Rubio

### Coordinación

Fátima Mateos Hernández

### Diseño y maquetación

Emi Ramírez

### Imágenes

Freepick, Unsplash

### Edita

Juezas y Jueces para  
la Democracia en Madrid

Juezas y Jueces  
*para la* Democracia

## ARTÍCULOS

**Inteligencia artificial y feminismo  
¿Cuestión de género o de alarma social?** 4  
Victoria Trujillo Machuca

**Hacia una justicia digital con perspectiva  
de género: Riesgos y potencialidades de  
la Inteligencia Artificial en contextos de  
violencia machista** 10  
María López Belloso, Ainhoa Izaguirre  
Choperena y Borja Sanz Urquijo

**Inteligencia Artificial** 21  
Julio Díaz Casales

**Companion** 39  
Diego Gutierrez Alonso

# Editorial

## **Tecnología, justicia y género: los desafíos de la inteligencia artificial en nuestras democracias**

La inteligencia artificial (IA) ha dejado de ser una promesa del futuro para convertirse en una infraestructura omnipresente que modela decisiones, instituciones y relaciones sociales. Sin embargo, su desarrollo y aplicación han despertado profundas preocupaciones éticas, políticas y jurídicas, especialmente en lo que respecta a sus impactos sobre los derechos humanos, la equidad de género y la transparencia democrática. Este nuevo volumen del Boletín Temático de Inteligencia Artificial ofrece un análisis crítico desde diversas perspectivas que, al converger, proponen una mirada compleja y necesaria sobre los usos y riesgos de la IA en nuestros entornos sociales y jurídicos.

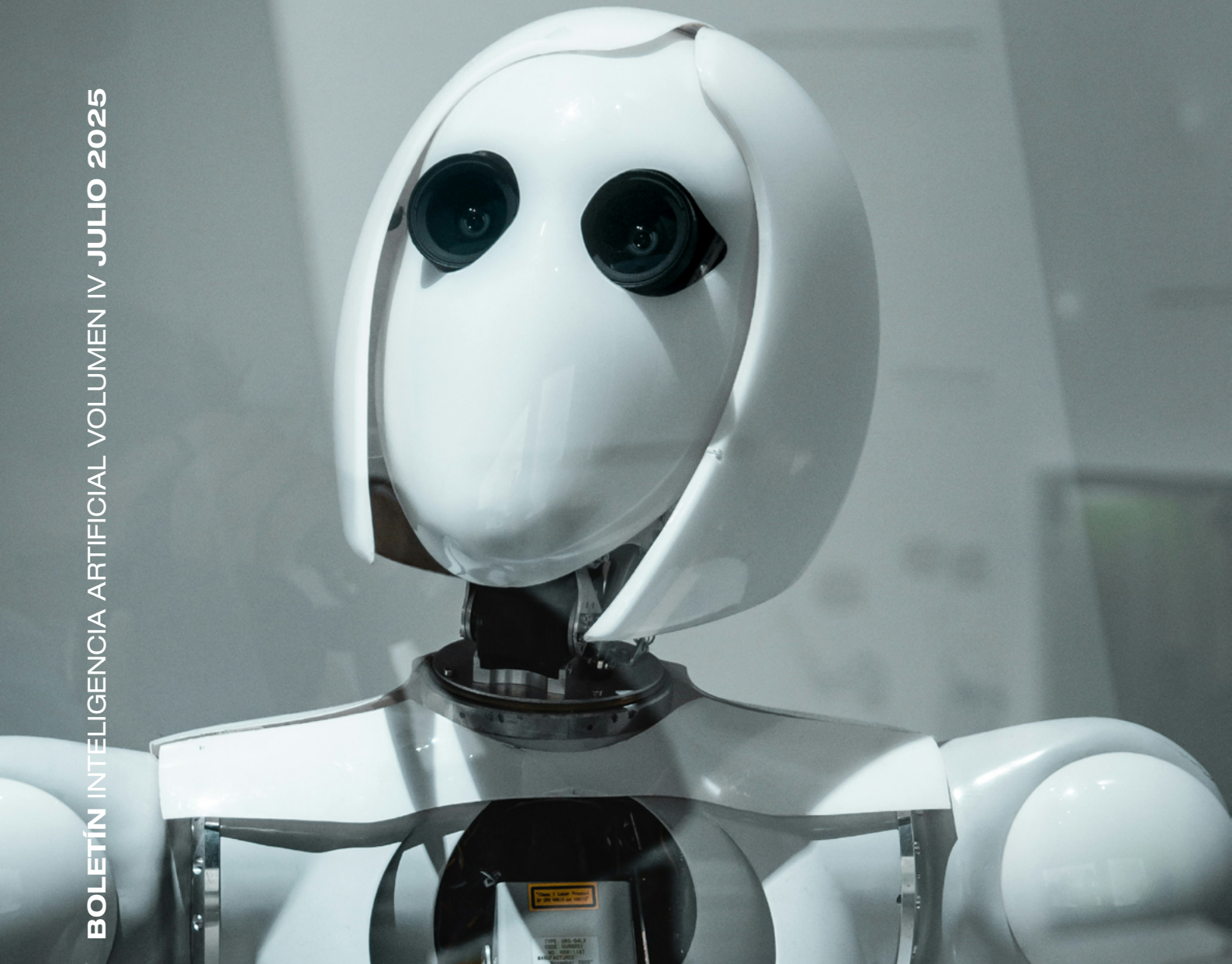
El artículo de **Victoria Trujillo Machuca**, “*Inteligencia artificial y feminismo: ¿cuestión de género o de alarma social?*”, alerta sobre cómo los sesgos de género en los algoritmos no solo reproducen desigualdades históricas, sino que afectan gravemente a la salud mental de las nuevas generaciones. La autora analiza cómo la escasa participación femenina en el diseño de sistemas tecnológicos perpetúa un modelo androcéntrico en los dispositivos digitales y reflexiona sobre la urgencia de construir una IA inclusiva y socialmente responsable, en consonancia con la nueva normativa europea.

Complementando esta visión crítica, **María López Belloso, Ainhoa Izaguirre Choperena y Borja Sanz Urquijo** exploran en su trabajo “*Hacia una justicia digital con perspectiva de género*” el papel ambivalente que la IA puede jugar en la lucha contra la violencia machista. Si bien herramientas como los chatbots o sistemas predictivos ofrecen soluciones prometedoras para la intervención temprana y la atención a víctimas, los autores advierten que sin un enfoque feminista e interseccional, estas tecnologías pueden incurrir en dinámicas de revictimización, exclusión o vigilancia encubierta.

Finalmente, **Julio Díaz Casales**, en su artículo “*Inteligencia Artificial*”, ofrece una visión jurídica y práctica sobre la implantación de estos sistemas en la administración pública, especialmente en el ámbito judicial. Su análisis subraya la necesidad de mantener la tutela judicial efectiva frente al avance de las automatizaciones, la importancia de la rendición de cuentas y la transparencia algorítmica, y el papel clave que desempeña la nueva Agencia Española de Supervisión de la Inteligencia Artificial (AESIA).

La conjunción de estos enfoques pone en evidencia que la IA no puede ser entendida como una herramienta neutral o meramente técnica. Es, en cambio, un campo de disputa política y social, donde se juegan formas de poder, reconocimiento y exclusión. El boletín invita a pensar la inteligencia artificial no solo desde sus potencialidades, sino desde las estructuras de poder que reproduce, las voces que invisibiliza y los cuerpos que impacta.

En tiempos de digitalización acelerada, urge construir una IA democrática, feminista y orientada al bien común. Esa tarea requiere normativas robustas, una ciudadanía crítica y una comunidad científica comprometida con la equidad. Este boletín es una contribución a ese debate impostergable.



# Inteligencia artificial y feminismo

## ¿Cuestión de género o de alarma social?

**Victoria Trujillo Machuca**

Magistrada. Máster en Derecho Digital, Innovación y Tecnologías Emergentes

**Resumen:** Este artículo es fundamentalmente una reflexión sobre la IA y el sesgo de género que sufren los algoritmos, tratando de descubrir su origen y la forma de mitigarlo desde el punto de vista social y normativo. El objetivo es poder avanzar en la creación de un marco jurídico que nos proteja de sus consecuencias negativas, pues entrañan un grave peligro en el camino hacia la igualdad y perpetúan los estereotipos en detrimento de la salud mental de nuestros jóvenes, los más vulnerables.

## 1. LA INTELIGENCIA ARTIFICIAL FEMINISTA (IAF)

En el “Día Internacional de la Mujer y la Niña en la Ciencia”, celebrado el 11 de febrero, cabe preguntarse por qué no hay más mujeres involucradas en el estudio de la inteligencia artificial. Solo el 36% de las españolas eligen carreras de STEM (cuyas siglas en inglés corresponden a Ciencia, Tecnología, Ingeniería y Matemáticas), de manera que dos tercios de los estudiantes de estas áreas académicas son varones. Los hombres son quienes, mayoritariamente, manejan la tecnología en la era de la Cuarta Revolución Industrial. En este artículo analizaremos el impacto que puede tener esta circunstancia en la sociedad, y si se trata solo de una cuestión de género, vista desde la perspectiva de los sesgos que discriminan a las mujeres, o si es una situación que nos afecta a todos como sociedad, que debe hacer saltar las alarmas ante el uso diario y excesivo que hacen los más jóvenes de las aplicaciones tecnológicas que, cada vez más, incorporan algoritmos.

Partiendo de la definición de “feminismo” en el diccionario de la RAE, como el principio de igualdad de derechos de la mujer y el hombre, o el movimiento que lucha por la equidad de género y el reconocimiento de los derechos de la mujer en la sociedad, centrémonos en el sector que va dirigido a evitar la discriminación por razón de sexo en IA. Este surge con más fuerza en los países iberoamericanos, donde no cuentan con ningún tipo de regulación al respecto, siendo la propia idiosincrasia cultural el peor escollo. Nacen organizaciones como “Chicas en Tecnología” en Argentina, la plataforma “Alianza A Plus, por algoritmos inclusivos” en Guatemala o la “Fundación latinoamericana Ciudadanía Inteligente”

en Chile, cuyas informáticas han desarrollado un sistema de algoritmos inclusivos con resultados óptimos, lo que supone un avance mundial en la materia que viene funcionando desde hace varios años. En la actualidad se trabaja para crear una IA más inclusiva, desde el modelo (*Feminist Data Send*) y desde el diseño (*feminist.ai*).

La científica británica Alison Adam es una de las pioneras en el estudio de la influencia del género en la inteligencia artificial, y más concretamente en cómo la epistemología feminista podría usarse para desafiar la epistemología de la IA, relacionando en su ensayo la ética feminista con la ética informática<sup>1</sup>. Adam estudia el papel que juega el género en la ética en el comportamiento en línea, evaluando la división de género en el *hacking*, el acoso cibernético y la pornografía para comparar las diferencias de género existentes en las experiencias en línea<sup>2</sup>. Estos primeros desarrollos fueron cruciales para ampliar el alcance de la investigación en IA e incluir críticas innovadoras y prototipos incipientes que desempeñan un pilar fundamental en implantar una IA tecnológicamente avanzada a la vez que socialmente consciente, inclusiva y reflexiva de las diversas necesidades de la sociedad.

## 2. EL SESGO DE GÉNERO Y LA BRECHA DIGITAL

La definición de sesgo por el diccionario de la RAE se recoge como “*oblicuidad, torcimiento de una cosa hacia un lado; sesgar es igual a cortar o partir algo en sesgo, torcer a un lado o atravesar algo hacia un lado; el sesgo es un peso desproporcionado a favor o en contra de una cosa, persona o grupo en comparación con otra, generalmente de una manera que se considera injusta*”.

1. Segell, Glen (2007). *Género, ética y tecnología de la Información* por Alison Adam, Palgrave Macmillan; 2005.

2. Adam, Alison (junio 2002). *Ciberacoso y pornografía en Internet: Género y mirada*.

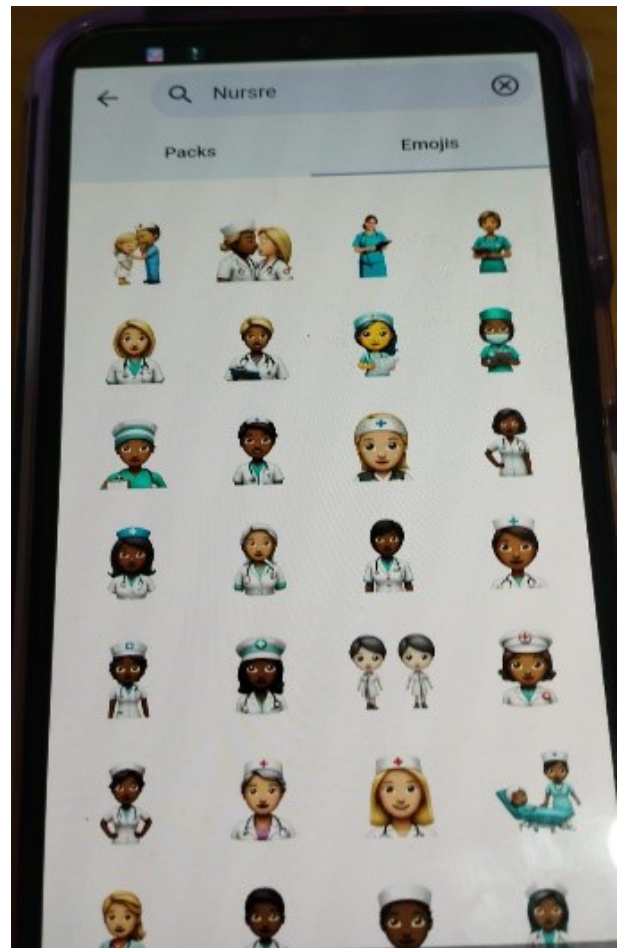
Cuando hablamos de *sesgo de género*, describimos un contexto de predisposición e inclinación hacia una persona, grupo o colectivo basado en su género, sea femenino o masculino, mostrándose como un prejuicio, parcialidad o predisposición a la hora de seleccionar, con base en un planteamiento erróneo de la igualdad, generador de una opción discriminatoria en el que se tiene cierta predilección hacia un género determinado dentro de grupos mixtos, que provoca, aun inconscientemente, el favoritismo hacia uno u otro, suponiendo un obstáculo para el desarrollo del género despreciado en aquella esfera de la vida en que se produzca, sea jurídica, social, económica, laboral, cultural o religiosa. Los sesgos de género tienen su origen en una cultura aprendida, con estereotipos, mitos, lecciones interiorizadas de creencias infundadas, que tratan de explicar una realidad en abstracto, sin haber sido objeto de crítica o juicio, realizada en una fase anterior a éste (*pre-juicio*), sin raciocinio o pensamiento neutral, o por intuición, con unilateralidad. El sesgo en ciencia e ingeniería se conoce como “*error sistemático*”, fuera de toda comprobación empírica, que en el campo estadístico resulta de un muestreo injusto de la población no dando resultados precisos ni promedios. Desde la perspectiva de género podemos distinguir entre dos extremos: **el androcentrismo**, formulado para desfavorecer a las mujeres, y **el ginocentrismo**, para desfavorecer a los hombres.

El sesgo de género es un problema que afecta a todas las esferas de la vida, pero en el ámbito de la tecnología es todavía más acusado dado que la mayoría de los profesionales que generan los algoritmos son hombres, habiéndose aumentado la brecha digital a medida que pasan los años (según un informe del World Economic Center, hoy cerca del 80% son programadores, siendo un 70% hace seis años). Como ejemplo de la discriminación que sufren las mujeres en la digitalización, valga decir que los asistentes personales más utilizados como Siri, Alexa y Cortana son predeterminadamente femeninos tanto en la voz que los personaliza como en la nomenclatura que reciben, siendo un reflejo de nuestra sociedad, donde quienes ocupan puestos de secretariado y administrativos en empresas del sector público y privado son mayoritariamente mujeres. Esta afirmación no es compartida por todos. Un adolescente varón como mi hijo, que ha sido educado desde tierna edad en defender la igualdad de sexos, alega preferir que Alexa sea femenina porque tiene una voz más dulce y agradable que la de un hombre, sencillamente, igual que Siri, lo que me lleva a pensar que si yo prefiero un asistente personal con la voz de Constantino Romero es porque soy mujer heterosexual. Y aunque los programadores varones no pretendan intencionadamente introducir un sesgo de género en la creación de algoritmos, los crean inevitablemente desde una visión de masculinidad irremediable.

En IA el sesgo es conocido como “*sesgo de machine learning*” o “*sesgo de algoritmo*”, y contempla aquellos sistemas tecnológicos que reflejan cálculos sesgados representando fielmente los sesgos sociales de los seres humanos como un espejo de la evolución histórica hasta nuestros días, recogiendo la cruda desigualdad por los datos de que se nutre a nivel universal o multicultural, tanto los de entrenamiento iniciales, como los algoritmos y las predicciones que resultan. Estos datos generan sistemas complejos que producen resultados distorsionados, perpetuando grupos marginados que en función de la ubicación geográfica del programador pueden fomentar la desconfianza en determinados sectores de la población por razón de raza, sexo, religión o nivel económico, existiendo grupos sobrerrepresentados. O al contrario. Pensemos en los gigantes de la IA, como Estados Unidos o China. Si los datos de entrenamiento contienen sesgos, conscientes o inconscientes, el aprendizaje es defectuoso y los algoritmos reproducirán una y otra vez los errores, siendo las predicciones absurdas o injustas. Por ejemplo, las herramientas

de IA para la contratación que utilizan un etiquetado incoherente o excluyen o sobre-representan ciertas características podrían eliminar de la consideración a los solicitantes de empleo más cualificados. Según el Instituto Nacional de Estándares y Tecnología (NIST), esta fuente de sesgo es más común de lo que se piensa. En un informe el NIST señalaba: “Los factores humanos y sistémicos, institucionales y sociales, son también fuentes importantes de sesgo en la IA, y actualmente se pasan por alto. Para superar con éxito este reto habrá que tener en cuenta todas las formas de sesgo. Esto significa ampliar nuestra perspectiva más allá del *machine learning* para reconocer e investigar cómo se crea esta tecnología en nuestra sociedad y cómo repercute en ella”<sup>3</sup>. Ciertamente es que a medida que avanza rápidamente la IA, sus creadores están más pendientes de evitar la parcialidad y tienden hacia lo neutral para asegurar la igualdad, además de garantizar un mejor y más coherente resultado, generador de más ingresos. Sin embargo, seamos realistas, eliminar los sesgos es igual de difícil que eliminar prejuicios sociales. Por ello, debemos tender de una manera inteligente y sosegada a reflexionar sobre la necesidad de estar unidos en promover una IA NO SESGADA, pues nos afecta a todos, a todas y a todos, especialmente en defensa de las nuevas e incipientes generaciones que están perpetuando modelos de comportamiento incivilizados que pensábamos ya superados y erradicados. Pero no, están más presentes que nunca porque son más visibles, se repiten como la *big data* de la que se nutren, erróneos, anticuados, mayoritarios y abundantes de nuestro *universo internet*. (Fig. 1)

No es posible erradicar esta brecha digital sin que las mujeres intervengan desde el



Para ilustrar el artículo. Si a la APP infantil “AI Emojis” le pides emoticonos sobre “enfermero/a”, únicamente genera imágenes de mujeres. De una lista de 28 imágenes: la primera imagen es una pareja: hombre acariciando a mujer; y la segunda imagen son dos mujeres besándose.

diseño en la automatización y la creación de la IA. Pero a pesar de poder alcanzar una paridad y representatividad igualitaria en la tecnología, el problema está lejos de llegar a una solución satisfactoria cuando escuchamos noticias como esta: “un estudio a gran escala realizado en España revela que el uso de filtros fotográficos por mujeres genera en los hombres una percepción de ser éstas menos inteligentes, mientras que el 96% de las mujeres que usan *selfies* retocados opinan que muestra a personas más atractivas, inteligentes y dignas de confianza”<sup>4</sup>. Es descorazonador

3. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence (NIST Special Publication 1270).*

4. [www.elperiodico.com/es/ciencia/20241128/filtros-belleza-inteligencia-artificial-mujeres-guapas-tiktok-hombres-redes-sociales-112134615](http://www.elperiodico.com/es/ciencia/20241128/filtros-belleza-inteligencia-artificial-mujeres-guapas-tiktok-hombres-redes-sociales-112134615)

pensar que la paridad en el diseño de los algoritmos no terminará con los sesgos de género por el hecho de que somos las propias mujeres las que los introducimos en la sociedad, especialmente las más jóvenes, que son quienes demandan y usan esos filtros de belleza que tanto daño hacen a la salud mental.

Los autores Michael Chui, James Manyika y Mehdi Miremadi (de McKinsey Archive article) señalaban: “Estos sesgos tienden a permanecer arraigados porque reconocerlos, y tomar medidas para abordarlos, requiere un profundo dominio de las técnicas de la ciencia de datos, así como una mayor metacompreensión de las fuerzas sociales existentes, incluida la recopilación de datos. En conjunto, la eliminación de prejuicios está demostrando ser uno de los obstáculos más desalentadores, y sin duda el más tenso desde el punto de vista social, hasta la fecha”<sup>5</sup>. (Fig. 2)

Según publica ELLIS Alicante<sup>6</sup>: “Los seres humanos se ven afectados por decenas de sesgos cognitivos a la hora de tomar decisiones. Uno de estos sesgos es el efecto de halo atractivo, es decir, la tendencia a asociar rasgos positivos, como la inteligencia o la confiabilidad, con individuos que se perciben como atractivos. Si bien este sesgo se ha estudiado ampliamente en humanos, hay pocos trabajos que estudien la existencia de dicho sesgo de atractivo en el contenido generado por los sistemas de IA. Este documento en curso investiga la presencia del efecto de halo de atractivo en los modelos de IA generativa de texto a imagen (T2I), examinando específicamente cómo los modelos T2I asocian el “atractivo” con rasgos positivos, como la inteligencia, la confiabilidad y la sociabilidad, al tiempo que vinculan la “falta de atractivo” con atributos negativos”. A través de va-

rios experimentos que generaron más de doce mil imágenes faciales etiquetadas con varios rasgos en todas las categorías de género y raza, se descubrió la existencia de un efecto halo similar al observado en los humanos, donde las imágenes más atractivas se asociaban más estrechamente con rasgos positivos, de forma que los modelos T2I incorporan un sesgo de atractivo, y se llegó a la conclusión de que la IA generativa replica nuestros propios sesgos, perpetuando estereotipos sociales.

### 3. SOLUCIONES NORMATIVAS

El Reglamento 2024/1689 por el que se establecen normas armonizadas en materia de inteligencia artificial (RIA), aunque está en vigor desde agosto de 2024, recoge una implantación escalonada de su normativa, y



Adjunto una imagen generada por el “Chat GPT”, donde se le pide: “Crea una imagen de una mujer jueza de 50 años con el pelo rubio y gafas redondas”.

5. [www.mckinsey.com/capabilities/quantumblack/our-insights/what-ai-can-and-cant-do-yet-for-your-business](https://www.mckinsey.com/capabilities/quantumblack/our-insights/what-ai-can-and-cant-do-yet-for-your-business)

6. Atractivo por diseño: cómo el efecto de halo atractivo da forma a la percepción de IA.

concretamente a partir del 2 de febrero de 2025 se activó el primer paquete de medidas que obligan, tanto al sector público como privado, a usar la IA de forma responsable, introduciendo restricciones específicas para garantizar una aplicación segura y ética de la IA (arts. 4 y 5), y creando un régimen de sanciones severo para los infractores, que se aplicará a partir del 2 de agosto de 2025.

Las obligaciones normativas que se imponen se recogen en dos bloques:

#### **Alfabetización en materia de IA (Art. 4)**

Los proveedores y responsables del despliegue de sistemas de IA adoptarán medidas para garantizar que, en la mayor medida posible, su personal y demás personas que se encarguen en su nombre del funcionamiento y la utilización de sistemas de IA tengan un nivel suficiente de alfabetización en materia de IA, teniendo en cuenta sus conocimientos técnicos, su experiencia, su educación y su formación, así como el contexto previsto de uso de los sistemas de IA y las personas o los colectivos de personas en que se van a utilizar dichos sistemas.

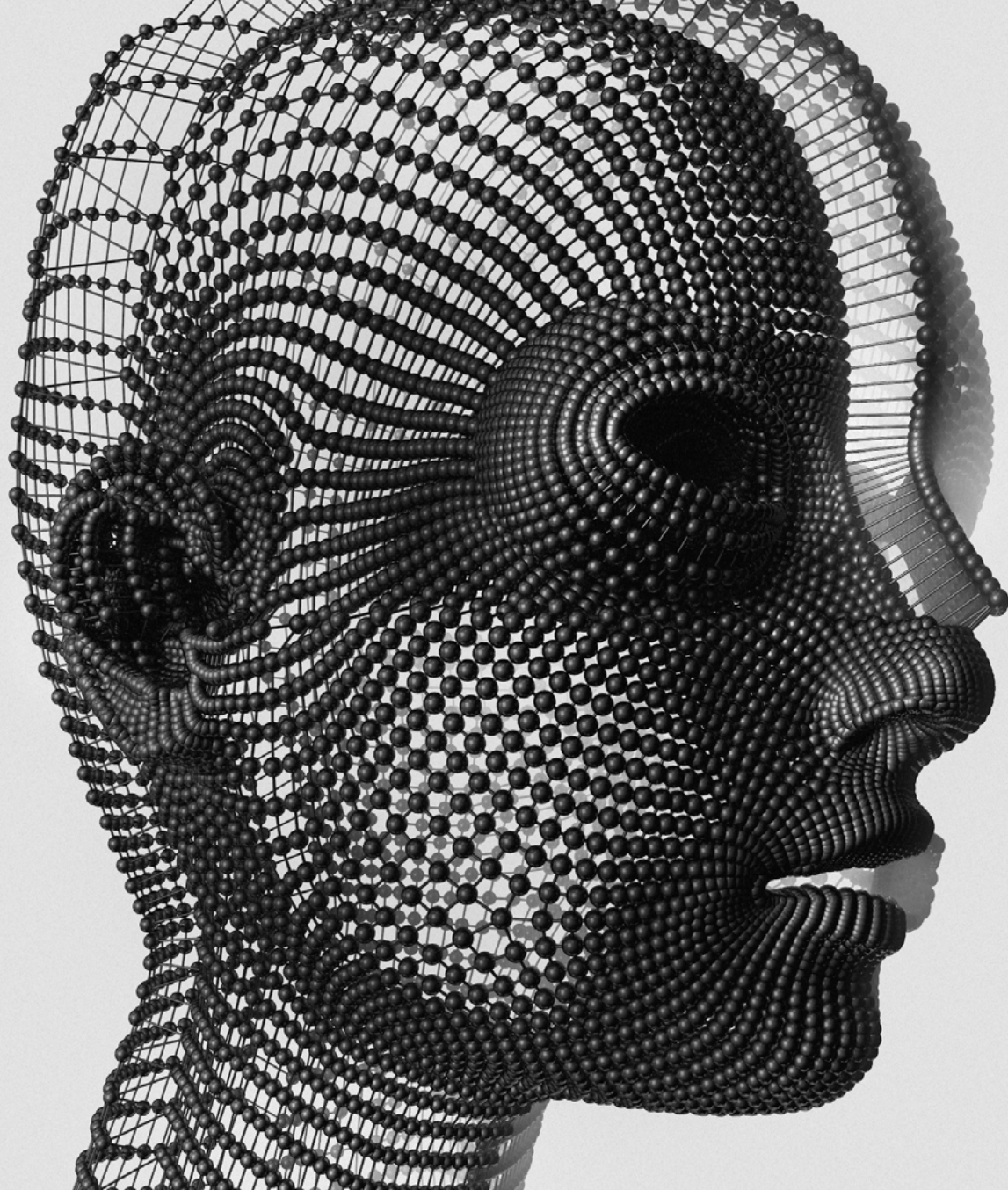
#### **Prácticas de IA prohibidas (Art. 5)**

Se prohíben los sistemas de IA que se sirvan de **técnicas subliminales, manipuladoras o engañosas** con el objetivo de alterar de manera sustancial el comportamiento y perjudicar la toma de decisiones, causando un daño significativo; aquellos que sirvan para explotar las **vulnerabilidades** de una persona relacionadas con la edad, la discapacidad o las circunstancias socioeconómicas específicas para distorsionar el comportamiento causando perjuicios considerables; los que sirvan para **evaluar o clasificar a personas físicas** o a colectivos de personas durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas, de forma que la puntuación ciudadana resultante provoque una o varias de las situaciones siguientes: un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos de personas en contextos sociales que no guarden relación con los contextos donde se generaron o recabaron los datos originalmente; un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos de personas que sea injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este (...)

Merece la pena detenerse en el último párrafo antedicho, dado que el Reglamento de Inteligencia Artificial considera como práctica prohibida el uso de aquellos sistemas de IA que creen perfiles según el comportamiento social de las personas con el fin de clasificarlas y generar discriminaciones, así como los que pretendan crear bases de datos de reconocimiento facial para predecir delitos, o clasificarlas en función de sus datos biométricos para deducir su raza, creencias religiosas u orientación sexual.

Por ello, no perdamos la esperanza. Conseguiremos una IA NO SESGADA si perseveramos en crear diseños y modelos en los que participemos las mujeres. Alfabeticémonos en IA desde ya, enfrentémonos a la tecnología y, sobre todo, no perdamos la oportunidad de denunciar a los infractores que sin duda existirán, como seguirá existiendo la raza humana, si somos capaces de, unidos, entender que no es una cuestión de género sino de alarma y paz social.





# Hacia una justicia digital con perspectiva de género

Riesgos y potencialidades  
de la Inteligencia Artificial  
en contextos de violencia machista

María López Belloso  
Ainhoa Izaguirre Choperena  
Borja Sanz Urquijo

## INTRODUCCIÓN

La irrupción de las nuevas tecnologías, y en particular de la inteligencia artificial (IA), ha abierto un campo emergente de posibilidades para la intervención en problemáticas sociales complejas, como la violencia de género. Sin embargo, su aplicación en este ámbito requiere ser abordada con especial cautela y desde una perspectiva crítica, ya que no se trata únicamente de incorporar herramientas tecnológicas, sino de reflexionar sobre los modelos de conocimiento, poder y cuidado que estas tecnologías reproducen o transforman.

Lejos de ser neutras, las tecnologías digitales —incluidos los sistemas de IA— están impregnadas de los sesgos y desigualdades estructurales que atraviesan nuestras sociedades. Por ello, este boletín propone explorar el uso de estas herramientas en el abordaje de la violencia de género desde una óptica feminista, que permita problematizar sus implicaciones, visibilizar sus riesgos y orientar su desarrollo hacia la equidad, la justicia social y la no revictimización.

El contenido se estructura en dos apartados principales. En primer lugar, se presenta una conceptualización de la inteligencia artificial feminista, entendida como un enfoque crítico y propositivo que desafía las lógicas tecnológicas hegemónicas y propone alternativas centradas en los derechos y experiencias de las mujeres. En segundo lugar, se introduce una taxonomía preliminar de amenazas vinculadas al uso de tecnologías en contextos de violencia de género, organizada en cinco dimensiones —tecnológica, psicoemocional y de género, social y cultural, jurídica y normativa, y organizacional y de gobernanza— con el objetivo de identificar riesgos específicos y orientar estrategias de mitigación.

Este boletín busca, en suma, contribuir a un debate urgente: cómo diseñar y aplicar tecnologías que, lejos de reproducir violen-

cias, puedan convertirse en herramientas para la protección, el empoderamiento y la transformación social.

## CONTEXTUALIZACIÓN DE LA VIOLENCIA DE GÉNERO

La violencia de género (VG) constituye una problemática social y de salud pública de alta prevalencia. Se manifiesta como una expresión de la desigualdad estructural presente en nuestras culturas y sociedades, y afecta a las mujeres a lo largo de todas las etapas de su ciclo vital, independientemente de su origen étnico, nivel educativo, situación económica, religión o profesión (Jeanjot, Barlow y Rozenberg, 2008). Este fenómeno complejo y multicausal comprende diversas formas de agresión —física, psicológica y sexual— que generan profundas secuelas en las mujeres que la padecen (Krug et al., 2002).

A pesar de los recursos institucionales existentes, muchas mujeres que sufren VG enfrentan barreras significativas para acceder a los servicios de apoyo, barreras que se agravan en el caso de mujeres migrantes o en situación de exclusión social (Toledano-Buendía, 2021). En este contexto, la detección precoz se presenta como uno de los mayores desafíos para las políticas de prevención, atención y protección. Para ello, se han desplegado múltiples estrategias, como la mejora en la coordinación interinstitucional, la formación especializada de profesionales en sectores clave, y el empoderamiento de las mujeres a través de redes asociativas y programas de sensibilización.

En los últimos años, el desarrollo de tecnologías digitales —particularmente la IA y los agentes conversacionales o *chatbots*— ha generado nuevas posibilidades para la intervención en contextos de VG. Herramientas como los chatbots ofrecen mecanismos de detección temprana, intervención continua y asistencia discreta, que pueden

resultar cruciales en momentos de riesgo (Socatiyanurak et al., 2021; Vorsino, 2021). Además, su capacidad de adaptación progresiva promete una mejora sostenida en la calidad y precisión del apoyo brindado.

Sin embargo, la incorporación de estas tecnologías en un ámbito tan sensible no puede realizarse sin una reflexión crítica. Lejos de ser neutrales, los sistemas de IA replican los sesgos y relaciones de poder presentes en las sociedades que los producen. Por tanto, resulta imprescindible abordar su diseño y aplicación desde una óptica feminista, que permita identificar sus riesgos, cuestionar sus fundamentos epistémicos y técnicos, y orientar su desarrollo hacia modelos más justos, inclusivos y transformadores.

### **INTELIGENCIA ARTIFICIAL FEMINISTA (IAF): ¿UN NUEVO PARADIGMA?**

En un contexto global marcado por el avance acelerado de las tecnologías digitales, la IA se ha consolidado como una herramienta central en la organización de múltiples aspectos de la vida social, económica y política. Sin embargo, este desarrollo no está exento de sesgos estructurales ni de relaciones de poder históricamente desiguales. Lejos de ser neutral, la IA reproduce y amplifica desigualdades preexistentes, particularmente aquellas vinculadas al género, la raza, la clase y otras formas de opresión. Frente a este escenario, se vuelve urgente una aproximación feminista que cuestione críticamente los supuestos epistémicos y técnicos de la IA convencional, visibilice las exclusiones que ésta perpetúa y proponga modelos alternativos basados en principios de justicia social, inclusión y equidad. La IA feminista, entendida como una propuesta crítica y alternativa que desafía la lógica tecnocientífica dominante y sitúa en el centro los derechos, saberes y experiencias

de las mujeres, no solo busca corregir los sesgos existentes, sino transformar profundamente las formas en que se produce, se diseña y se implementa esta tecnología, incorporando saberes diversos y experiencias históricamente silenciadas. Según ONU Mujeres (2025), muchas aplicaciones de IA, desde los sistemas de contratación automatizada hasta los algoritmos de recomendación, perpetúan estereotipos que penalizan a las mujeres y refuerzan su subrepresentación en sectores estratégicos. A esto se suma el impacto diferenciado en el mercado laboral: un informe del Banco Interamericano de Desarrollo revela que la automatización basada en IA tiende a desplazar desproporcionadamente a las mujeres de ocupaciones rutinarias (Collett, Neff, & Gouvea, 2022), al tiempo que ellas enfrentan mayores barreras para acceder a empleos tecnológicos de alta cualificación. Estos hallazgos subrayan la urgencia de un enfoque feminista e interseccional en



el diseño, implementación y evaluación de tecnologías algorítmicas, que promueva la equidad y los derechos humanos desde sus fundamentos.

La IAF, (FAI, por sus siglas en inglés) se configura como una categoría plural y en constante evolución que surge de la intersección entre la crítica feminista y el desarrollo tecnológico. Esta perspectiva busca no solo problematizar las bases epistémicas y técnicas de la IA tradicional, sino también proponer alternativas que incorporen valores como la equidad, la inclusión y la justicia social. Así, la IAF se articula en torno a múltiples formas —como modelo, diseño, política, cultura, discurso y ciencia— que reflejan su carácter interdisciplinar y su potencial transformador. Además, la IAF se fundamenta en una tradición crítica que incluye el pensamiento de autoras como Donna Haraway, quien introdujo el concepto de conocimiento situado, y Alison Adam, pionera en el análisis feminista de la IA. Estas influencias han contribuido a poner en cuestión nociones como la objetividad y la racionalidad en los sistemas de IA, así como a señalar la exclusión sistemática de saberes no hegemónicos. La IAF, por tanto, no se limita a una intervención técnica, sino que constituye una forma de resistencia y reimaginación que apunta a dismantelar las estructuras de poder que atraviesan la ciencia y la tecnología, especialmente en contextos de colonialismo digital y capitalismo de datos (Toupin, 2023).

La conceptualización de la IAF presentada por Sophie Toupin (2023) se estructura en seis dimensiones analíticas que capturan la pluralidad de enfoques y prácticas en torno al desarrollo y uso de tecnologías con una perspectiva feminista. Estas dimensiones permiten un análisis crítico de las amenazas de la IA en contextos de VG, al proporcionar un marco amplio y situado que integra preocupaciones técnicas, sociales, culturales y epistemológicas.

**IAF como modelo:** En esta dimensión, la IAF se comprende como un modelo computacional, particularmente un modelo de aprendizaje automático entrenado con conjuntos de datos feministas. Ejemplo de ello es el proyecto Feminist Data Set de Caroline Sindere, que busca construir un chatbot impulsado por datos generados colectivamente por comunidades feministas. Este enfoque visibiliza los problemas de sesgo algorítmico y el trabajo invisibilizado en la preparación de datos, lo que se alinea con la dimensión tecnológica al proponer modelos alternativos que desafían los paradigmas dominantes de eficiencia y opacidad en la IA (Toupin, 2023; Sindere, 2020).

**IAF como diseño:** Aquí se entiende la IAF como un enfoque de diseño participativo y culturalmente situado. Iniciativas como *feminist.ai* y el software *Poieto* ejemplifican esta perspectiva al integrar valores comunitarios y de diversidad en el diseño de sistemas de IA. Esta dimensión resuena con la dimensión organizacional y de gobernanza, al enfatizar la necesidad de procesos inclusivos, colaborativos y transparentes en la creación de tecnología (Toupin, 2023; Meinders, 2017).

**IAF como política pública:** La IAF también se enmarca como una herramienta de política internacional y desarrollo, como demuestra el programa *f+a+i>r* financiado por el IDRC canadiense. Estas iniciativas promueven el desarrollo de tecnologías sensibles al género en el Sur Global, vinculando la IA feminista con los Objetivos de Desarrollo Sostenible. Esta dimensión tiene una clara correspondencia con la dimensión jurídica y normativa, al proponer marcos regulativos y de financiación que promuevan el uso ético y justo de la IA en contextos sensibles como la violencia de género (Toupin, 2023; Feldfeber et al., 2022).

**IAF como discurso:** Toupin identifica una proliferación del término IAF como una categoría discursiva que estructura imaginarios y narrativas críticas sobre la tecnología. Desde marcos académicos hasta iniciativas como el chatbot Mumkin en India, la IAF se presenta como una alternativa esperanzadora a las formas dominantes de IA. Esta dimensión es coherente con la dimensión psico-emocional y de género, ya que reconoce el poder de las narrativas en moldear las experiencias de las usuarias y fomentar su empoderamiento (Toupin, 2023; Doyle-Burke & Smith, 2021).

**IAF como cultura:** La IAF también es concebida como una práctica cultural que desafía la normatividad técnica y de género inscrita en las tecnologías. Autores como Wajcman (1991) y Wellner & Rothman (2020) argumentan que las normas sociales, físicas y cognitivas en la IA deben ser analizadas y transformadas desde una perspectiva crítica. Esta categoría se vincula estrechamente con la dimensión social y cultural, al subrayar cómo la tecnología refleja y refuerza desigualdades estructurales, y cómo puede ser apropiada críticamente por comunidades diversas.

**IAF como ciencia:** Finalmente, Toupin examina la IAF como una reformulación epistemológica de la ciencia de la inteligencia. Inspiradas por autoras como Adam (1998) y Atanasoski y Vora (2019), esta dimensión cuestiona la neutralidad de la noción de inteligencia misma y propone una inteligencia feminista como una práctica de resistencia. Esto plantea una crítica profunda al proyecto tecnoliberal, lo que complementa todas las dimensiones anteriores al exigir una transformación estructural del conocimiento técnico y científico.

Los enfoques jurídicos y normativos sobre la IA deben abordar el carácter dual de estas tecnologías: si bien ofrecen importantes oportunidades para el progreso social — como un mayor acceso a la educación, la predicción de desastres naturales y el apoyo a víctimas de violencia de género— también conllevan el riesgo de reproducir y amplificar desigualdades estructurales ya existentes. Como apuntan trabajos previos (López Belloso, 2022), los sesgos de género en la IA socavan los derechos fundamentales y afectan de manera desproporcionada a las mujeres y a otros grupos históricamente marginados. En respuesta, el enfoque de transversalización de género y la incorporación de perspectivas feministas e interseccionales en los marcos regula-



torios ha ido cobrando fuerza en la agenda internacional. López Belloso (2022) analiza iniciativas normativas clave como la Recomendación de la UNESCO y la Declaración de Toronto, que reconocen la necesidad de una gobernanza inclusiva de la IA basada en los derechos humanos. No obstante, la autora señala que estos instrumentos suelen centrarse en la mitigación del daño, en lugar de aprovechar el potencial transformador de la IA para promover la igualdad. Esta crítica coincide con las lecturas feministas del Reglamento Europeo de IA, que denuncian el sesgo tecnocrático de la ley y su incapacidad para abordar de forma sustantiva las raíces epistémicas de la opresión de género en los sistemas digitales (Karagianni, 2025). Aunque el reglamento invoca principios como la no discriminación, carece de mecanismos sólidos que garanticen la justicia de género en la práctica, relegando esta cuestión a un plano secundario. Por tanto, un marco regulatorio eficaz en materia de IA debe ir más allá de la retórica ética, integrando la igualdad de género como principio fundamental del diseño tecnológico y promoviendo una gobernanza participativa e interdisciplinaria que refleje la diversidad de las realidades sociales.

En el contexto de la intervención con mujeres que han sido víctimas de VG, resulta esencial considerar los riesgos de revictimización que pueden derivarse del uso de tecnologías basadas en IA. Los enfoques feministas en el acompañamiento subrayan la necesidad de evitar dinámicas coercitivas y dañinas que profundicen el malestar de las mujeres (Porter y López-Angulo, 2022; Reyes, 2015; Olaizola, 2010). Sin embargo, la literatura especializada ha prestado escasa atención a cómo las herramientas de IA pueden incidir en estos procesos de revictimización. En este sentido, Notvisky et al. (2023) advierten sobre la falta de directrices para un uso ético y cuidadoso de estas tecnologías, señalando además que su efectividad puede verse limitada por la coexistencia con tecnologías

rudimentarias utilizadas por agresores, lo cual complica la prevención de nuevos episodios de violencia.

Más aún, sin un acompañamiento profesional adecuado y sin una preparación contextualizada para su uso, las soluciones digitales orientadas a la detección y atención de la VG pueden provocar efectos contraproducentes. El mismo estudio documenta casos en los que ciertas aplicaciones, diseñadas con el propósito de identificar de forma temprana situaciones de violencia, generaron malestar entre usuarias al formular preguntas que las confrontaban con recuerdos traumáticos, reforzando así los efectos negativos del trauma. Estos hallazgos revelan cómo el diseño insensible de sistemas automatizados puede desencadenar dinámicas de revictimización en lugar de mitigarlas.

Desde una perspectiva crítica, Trias et al. (2024) cuestionan los procedimientos institucionales de atención en los ámbitos sanitario y judicial, particularmente en lo relativo al tratamiento de la información y la transferencia de casos. La investigación evidencia cómo estos procesos, al exigir reiteradamente la narración de los hechos por parte de las víctimas, contribuyen a la reproducción del daño. Como alternativa, los autores proponen la implementación de tecnologías de IA que optimicen el flujo de información sensible y reduzcan la exposición de las víctimas a situaciones revictimizantes, en línea con un enfoque centrado en el cuidado y la reparación.

## AMENAZAS EMERGENTES: UNA TAXONOMÍA CRÍTICA DESDE LA IAF

El despliegue de tecnologías basadas en IA en contextos de violencia de género no solo plantea oportunidades, sino también múltiples riesgos que deben ser comprendidos desde una mirada crítica, interseccional y situada. En coherencia con los marcos con-

ceptuales presentados previamente, este apartado propone una taxonomía preliminar de amenazas asociadas al uso de estas tecnologías, estructurada en cinco dimensiones analíticas (Sanz Urquijo et al. sf). Cada una de estas dimensiones refleja no solo aspectos técnicos o funcionales, sino también los entramados sociales, culturales, jurídicos y organizacionales que configuran las condiciones de posibilidad para una tecnología realmente transformadora. El objetivo es identificar con claridad los principales desafíos que enfrentan las mujeres usuarias de estas herramientas, especialmente en situaciones de vulnerabilidad, y orientar estrategias de mitigación fundamentadas en principios de justicia, cuidado y equidad.

**Dimensión tecnológica:** Esta dimensión contempla la infraestructura, el diseño algorítmico y la fiabilidad de los sistemas que sustentan las soluciones digitales. Entre las amenazas más relevantes se encuentran los riesgos asociados a la seguridad de la información y la protección de datos sensibles, especialmente en casos de acceso no autorizado o almacenamiento inadecuado. Asimismo, la calidad de los modelos conversacionales, incluyendo su capacidad para adaptarse a contextos lingüísticos y culturales diversos, constituye un factor crítico. Finalmente, la dependencia de redes de conectividad y la estabilidad del servicio representan un riesgo concreto, sobre todo en zonas rurales o en momentos de emergencia.

**Dimensión psicoemocional y de género:** Desde una perspectiva feminista, es indispensable considerar los efectos subjetivos que pueden generar estas tecnologías en las mujeres víctimas de violencia. Las amenazas en esta dimensión incluyen la revictimización mediante respuestas insensibles o descontextualizadas, así como la exposición a contenidos que reactivan traumas. Igualmente, preocupante es la posible pérdida de autonomía si las tecnologías sustituyen, en lugar de complementar, el proceso reflexivo y decisional de las usuarias. Además, el diseño no inclusivo puede invisibilizar la experiencia de mujeres diversas —como las migrantes o rurales— y reforzar estigmas sociales.

**Dimensión social y cultural:** La inserción de tecnologías en los entornos comunitarios está mediada por múltiples factores estructurales. La brecha digital, la escasa alfabetización tecnológica y las narrativas sociales que desalientan el uso de estos recursos constituyen amenazas que deben ser abordadas. Asimismo, la implementación aislada de herramientas tecnológicas, sin el acompañamiento de redes de apoyo psicosocial, puede generar una falsa percepción de autosuficiencia y desplazar la importancia del cuidado colectivo y la intervención integral.

**Dimensión jurídica y normativa:** El vacío legal en torno al uso de IA en contextos de violencia de género representa una amenaza significativa. La ausencia de marcos regulatorios específicos dificulta la fiscalización y sanción de prácticas negligentes, mientras que el cumplimiento de normativas internacionales sobre protección de datos plantea desafíos cuando las plataformas operan en múltiples jurisdicciones. También es necesario establecer mecanismos claros de responsabilidad y reparación en caso de fallos, lo cual es fundamental para garantizar la confianza y la justicia para las usuarias.

**Dimensión organizacional y de gobernanza:** Por último, esta dimensión subraya la importancia de contar con estructuras institucionales adecuadas para el desarrollo y mantenimiento de las tecnologías. La falta de formación con perspectiva de gé-

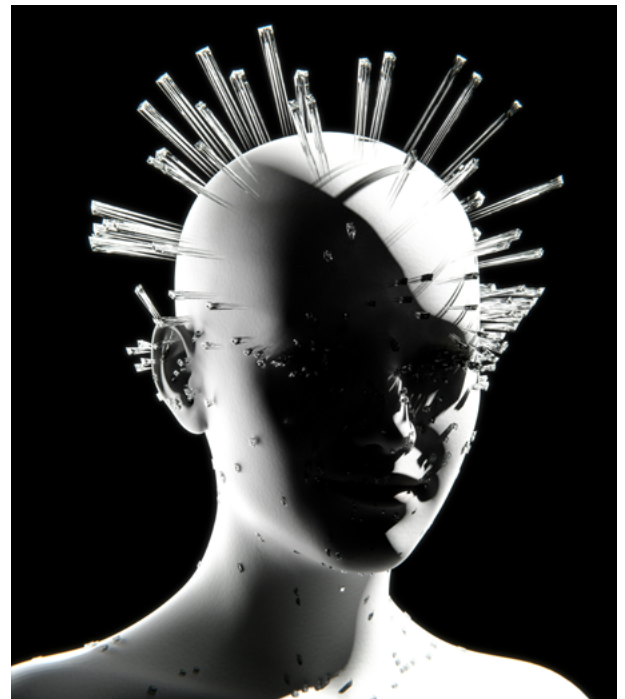
nero en los equipos técnicos, la opacidad en los procesos de rendición de cuentas y la ausencia de protocolos de actualización y supervisión continua constituyen amenazas críticas. A ello se suma la falta de sostenibilidad financiera, que puede comprometer la continuidad de los servicios y dejar a las mujeres sin los recursos de apoyo que han comenzado a utilizar.

Esta taxonomía no pretende ser exhaustiva, sino ofrecer una herramienta de análisis que permita avanzar en la identificación de riesgos específicos y en la construcción de marcos de acción ética, situada y transformadora. Desde una perspectiva feminista, no se trata únicamente de mitigar daños, sino de imaginar tecnologías que acompañen procesos de justicia, reparación y empoderamiento colectivo.

### MITIGACIÓN Y PREVENCIÓN DE LA REVICTIMIZACIÓN DE LAS MUJERES EN EL USO DE LA IA

La expansión del uso de tecnologías basadas en IA en los sistemas de atención y prevención de la violencia de género plantea desafíos complejos para el sistema judicial. En particular, se hace urgente repensar los marcos normativos, éticos y operativos desde los cuales se integran estas herramientas, a fin de garantizar la protección efectiva de los derechos fundamentales, evitar procesos de revictimización y promover una justicia verdaderamente sensible al género. Tal como se ha analizado a lo largo de este boletín, si bien estas tecnologías pueden potenciar la detección temprana, facilitar el acceso a la información y mejorar la eficiencia institucional, también pueden reproducir desigualdades estructurales, generar respuestas descontextualizadas y, en algunos casos, agravar el malestar de las mujeres en situación de violencia.

Ante este panorama, es imprescindible que el diseño, implementación y uso de soluciones algorítmicas se someta a marcos regulatorios sólidos que aseguren la transparencia, la auditabilidad y el cumplimiento estricto de los estándares de protección de datos sensibles. En este sentido, los sistemas judiciales deben desarrollar protocolos específicos para el uso de IA en contextos de violencia de género, incorporando mecanismos de



**“[...] los sistemas judiciales deben desarrollar protocolos específicos para el uso de IA en contextos de violencia de género, incorporando mecanismos de evaluación de impacto con enfoque feminista e interseccional”**

evaluación de impacto con enfoque feminista e interseccional, así como rutas claras de denuncia, reparación y rendición de cuentas frente a eventuales vulneraciones.

Igualmente, se requiere una formación especializada de los operadores jurídicos —jueces, fiscales, personal forense y técnico— que contemple no solo el aprendizaje técnico sobre estas herramientas, sino también una sensibilización profunda respecto a los impactos psicoemocionales que su uso puede tener en las víctimas. Comprender los límites y potencialidades de la IA desde una perspectiva crítica es esencial para evitar su uso indiscriminado o deshumanizante, y para garantizar que el juicio humano siga siendo central en los procesos de toma de decisiones judiciales.

Además, los sistemas de justicia deben fomentar modelos de gobernanza tecnológica abiertos, inclusivos y participativos. La incorporación de las voces de mujeres sobrevivientes, organizaciones feministas, especialistas en derechos humanos y co-

munidades vulneradas es clave para asegurar que las tecnologías respondan a las necesidades reales de quienes enfrentan violencias, y no a lógicas mercantilistas o tecnocráticas ajenas al contexto social.

Desde una mirada interseccional, también es necesario reconocer las barreras estructurales que dificultan el acceso equitativo a estas herramientas, especialmente para mujeres en situación de pobreza, migración, ruralidad o discapacidad. Las tecnologías digitales, lejos de ser universales, deben ser diseñadas y adaptadas considerando esta diversidad de experiencias. En este sentido, se hace fundamental promover el co-diseño y la evaluación participativa de los dispositivos tecnológicos, como una estrategia para asegurar su adecuación cultural, lingüística y funcional.

Los testimonios y experiencias recogidos en procesos participativos reafirman que solo mediante una mirada feminista e interseccional es posible construir tecnologías que contribuyan genuinamente a la erradicación de las violencias machistas. Estas voces, lejos de ser complementarias, deben ocupar un lugar central en la definición de estándares, criterios de calidad y objetivos de justicia. Apostar por herramientas digitales que acompañen procesos de toma de conciencia, empoderamiento y reparación significa reconocer que la innovación tecnológica también puede ser un acto de resistencia y transformación social.



**“[...] los sistemas de justicia deben fomentar modelos de gobernanza tecnológica abiertos, inclusivos y participativos”**

## BIBLIOGRAFÍA

- Adam, A. (1998). *Artificial Knowing: Gender and the Thinking Machine*. London and New York: Taylor & Francis Group.
- Atanasoski N, Vora K (2019) *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*. Durham, NC: Duke University Press.
- Collett, C., Neff, G., & Gouvea, L. (2022). *Los efectos de la IA en la vida laboral de las mujeres*. <https://doi.org/10.18235/0004055>
- Doyle-Burke, D., Smith, J. (2021). Feminist AI. The Radical AI Podcast, 2 June. Available at: <https://radicalai.podbean.com/e/feminist-ai-101-with-eleanor-drage-and-kerry-machereth/>
- Feldfeber, I., Quiroga, Y.B., Guevara, C., et al. (2022). Feminisms in Artificial Intelligence: Automation tools towards a feminist judiciary reform in Argentina and Mexico. Available at: <https://drive.google.com/file/d/1VKeeSdJz6-8DoEbetQ8f1wTH1ORqvIXV/view>
- Jeanjot, I., Barlow, P., & Rozenberg, S. (2008). Domestic violence during pregnancy: Survey of patients and healthcare providers. *Journal of Women's Health*, 17, 557-567. Doi: [10.1089/jwh.2007.0639.6](https://doi.org/10.1089/jwh.2007.0639.6)
- Karagianni, A. (2025). Gender in a stereo-(gender) typical EU AI law: A feminist reading of the AI act. Cambridge Forum on AI: Law and Governance, 1. Cambridge University Press.
- Krug, E., Dalhberg, L., Mercy, J., Zwi, A., & Lozano, R. (2002). *World report on violence and health*. Geneva: World Health Organization.
- López Belloso, M. (2022). Women's Rights Under AI Regulation: Fighting AI Gender Bias Through a Feminist and Intersectional Approach. In: Custers, B., Fosch-Villaronga, E. (eds) *Law and Artificial Intelligence. Information Technology and Law Series*, vol 35. T.M.C. Asser Press, The Hague. [https://doi.org/10.1007/978-94-6265-523-2\\_5](https://doi.org/10.1007/978-94-6265-523-2_5)
- Meinders, C. (2020). Feminist AI (main thread). <http://wg20.criticalcodestudies.com/index.php?p=/discussion/87/week-3-feminist-ai-main-thread>
- Novitzky, P., Janssen, J., & Kokkeler, B. (2023). A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools. *Heliyon*.
- Olaizola, I. (2010). *Violencia de género: elementos de los tipos penales con mayor dificultad probatoria*.
- ONU MUJERES (2025) *Cómo la inteligencia artificial refuerza los sesgos de género y qué podemos hacer al respecto*, disponible en: <https://www.google.com/url?q=https://www.unwomen.org/es/noticias/entrevista/2025/02/como-la-inteligencia-artificial-refuerza-los-sesgos-de-genero-y-que-podemos-hacer-al-respecto&sa=D&source=docs&ust=1749666635657263&usq=AO-vVaw3wbHndziQW7qeYNI457Oqz>
- Porter, B., & López-Angulo, Y. (2022). Violencia vicaria en el contexto de la violencia de género: un estudio descriptivo en Iberoamérica. *CienciAmérica*, 11(1), 11-42.
- Reyes, P. (2015). Menores y violencia de género: de invisibles a visibles. *Anales de la Cátedra Francisco Suárez*, 49, 181-217.
- Sinders, C. (2020). Feminist data set. Clinic for Open Source Arts. <https://carolinesinders.com/wp-content/uploads/2020/05/Feminist-Data-Set-Final-Draft-2020-0517.pdf>

Socatiyanurak et al. (2021). LAW-U: Legal Guidance Through Artificial Intelligence Chatbot for Sexual Violence Victims and Survivors», IEEE Access, 9, pp. 131440-131461, 2021, doi: [10.1109/ACCESS.2021.3113172](https://doi.org/10.1109/ACCESS.2021.3113172)

Toledano-Buendía, C. (2021). Barrera lingüística y victimización secundaria: la (des)atención institucional a las víctimas extranjeras de violencia de género en España. *Verba hispánica: anuario del Departamento de la Lengua y Literatura Españolas de la Facultad de Filosofía y Letras de la Universidad de Ljubljana*, 29, 175-191.

Toupin, S. (2023). Shaping feminist artificial intelligence. *New Media & Society*, 26(1), 580-595. <https://doi.org/10.1177/14614448221150776> (Original work published 2024).

Trias Capella M.E., Guardia Villalba R, Trias Capella R. (2023). Treatment of the information on gender-based violence. With contributions from artificial intelligence. *Revista Española de Medicina Legal*. <https://doi.org/10.1016/j.reml.2023.04.002>

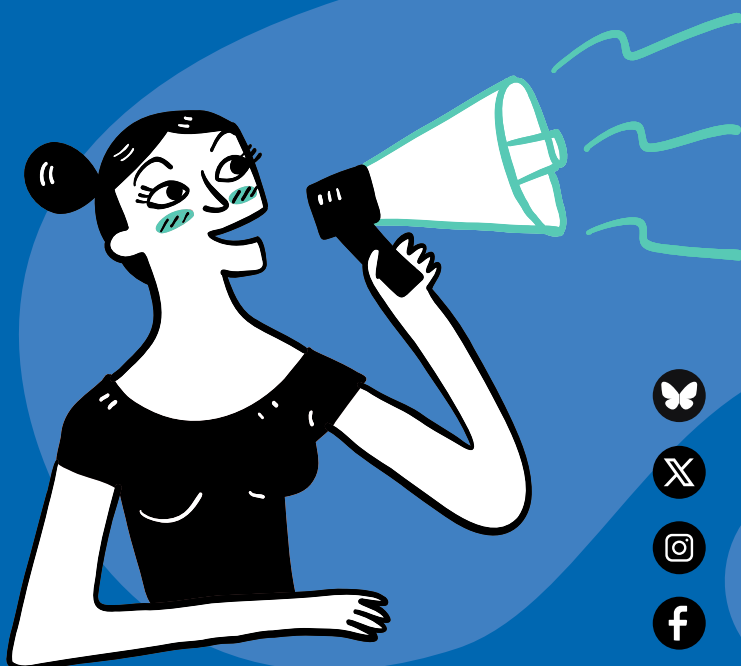
Vorsino, Z. (2021). «Chatbots, gender, and race on web 2.0 platforms: Tay. AI as monstrous femininity and abject whiteness», *Signs: Journal of Women in Culture and Society*, 47(1), pp. 105-127.

Sanz Urquijo, B., López Belloso, M., Romero Gutiérrez, L., & Silvestre Cabrera, M. (en revisión). *Agentes conversacionales inteligentes (chatbots) y estereotipos de género en la atención de las violencias machistas: Taxonomía de posibles amenazas desde un enfoque de threat modelling*. Manuscrito en revisión en *Investigaciones Feministas*.

Wajcman, J. (1991) *Feminism Confronts Technology*. Cambridge: Polity Press.

Wellner, G., Rothman, T. (2020). Feminist AI: can we expect our AI systems to become feminist? *Philosophy & Technology*, 33(2): 191-205.

Tú eres nuestro altavoz



**¡SIGUENOS!**



@jpdemocracia.bsky.social



@JpDemocracia



JJpdemocracia



Juezas y Jueces  
para la Democracia

**JJpD**

# Inteligencia Artificial

**Julio Díaz Casales**  
Magistrado



## 1. QUÉ ES LA IA

El Art. 3 del Reglamento IA al abordar las definiciones lo describe del siguiente modo:

*“sistema de IA” un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y 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 resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.*

Algunos reducen estos sistemas a un conjunto de algoritmos que, con un método establecido, se pueden utilizar para resolver un determinado problema. En realidad el sistema pivota sobre el almacenamiento de datos, la aplicación de algoritmos o fórmulas matemáticas y el desarrollo de una sorprendente capacidad de computación para ofrecer resultados. En un intento de simplificación recurren a la idea de que una receta de cocina también es un algoritmo, como también lo sería el procedimiento administrativo. La diferencia está en que el robot de cocina no puede sustituir por su propia voluntad los ingredientes y/o alterar su combinación o tiempos de cocción para producir un resultado diferente al pretendido, tampoco lo pueden hacer los sometidos a un expediente administrativo “clásico” pero si cabe la posibilidad de que lo haga un sistema de IA.

## 2. PREVISIÓN DE SU UTILIZACIÓN POR LAS ADMINISTRACIONES PÚBLICAS

Venimos sufriendo un proceso de transformación hacia lo digital, que la pandemia aceleró vertiginosamente. A este fenómeno la administración de justicia no resulta ajena, porque se encuentra frente a él en una triple posición:

- Se plantea su utilización para dirimir recursos ya que puede aligerar considerablemente la tarea, al sistematizar en muy poco tiempo análisis documentales o de precedentes jurisprudenciales que para una persona pueden resultar inabarcables.
- Pero, a su vez, los órganos de la administración de justicia han de controlar los resultados de la utilización de los sistemas de IA por terceros, especialmente por las administraciones públicas.
- Por último los sistemas de administración de justicia proporcionan datos que pueden utilizarse para alimentar los sistemas de IA, tanto en su entrenamiento como para producir resultados de salida.

No obstante, he de advertir que ya Art. 18.4 de la C.E. preveía que la utilización de la informática podría llegar a suponer un peligro para los derechos fundamentales, al disponer:

*la Ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos*

Resulta evidente que la protección de los derechos a la intimidad personal y familiar con el uso de la informática que estaba en la mente los constituyentes ha sido desbordado por el fenómeno de las redes sociales, en el que son los propios usuarios los que exhiben

aspectos que deberían pertenecerles de modo exclusivo. No obstante, me interesa llamar la atención de la coletilla final del artículo que refiere que el uso de la informática no ha impedir o limitar “*el pleno ejercicio de los derechos de los ciudadanos*” porque considero que el derecho a la tutela judicial efectiva, en el sentido de obtener de la administración de justicia una respuesta razonada y entendible, puede verse seriamente comprometido de implementarse el uso de la IA en el tratamiento de los recursos.

La Ley 30/1992 de Procedimiento administrativo era cauta a la hora de prever la utilización de las aplicaciones informáticas sometiéndolas a la exigencia de que fueran aprobación del órgano y que se difundieran sus de sus características, al disponer en su **Art. 45.3**:

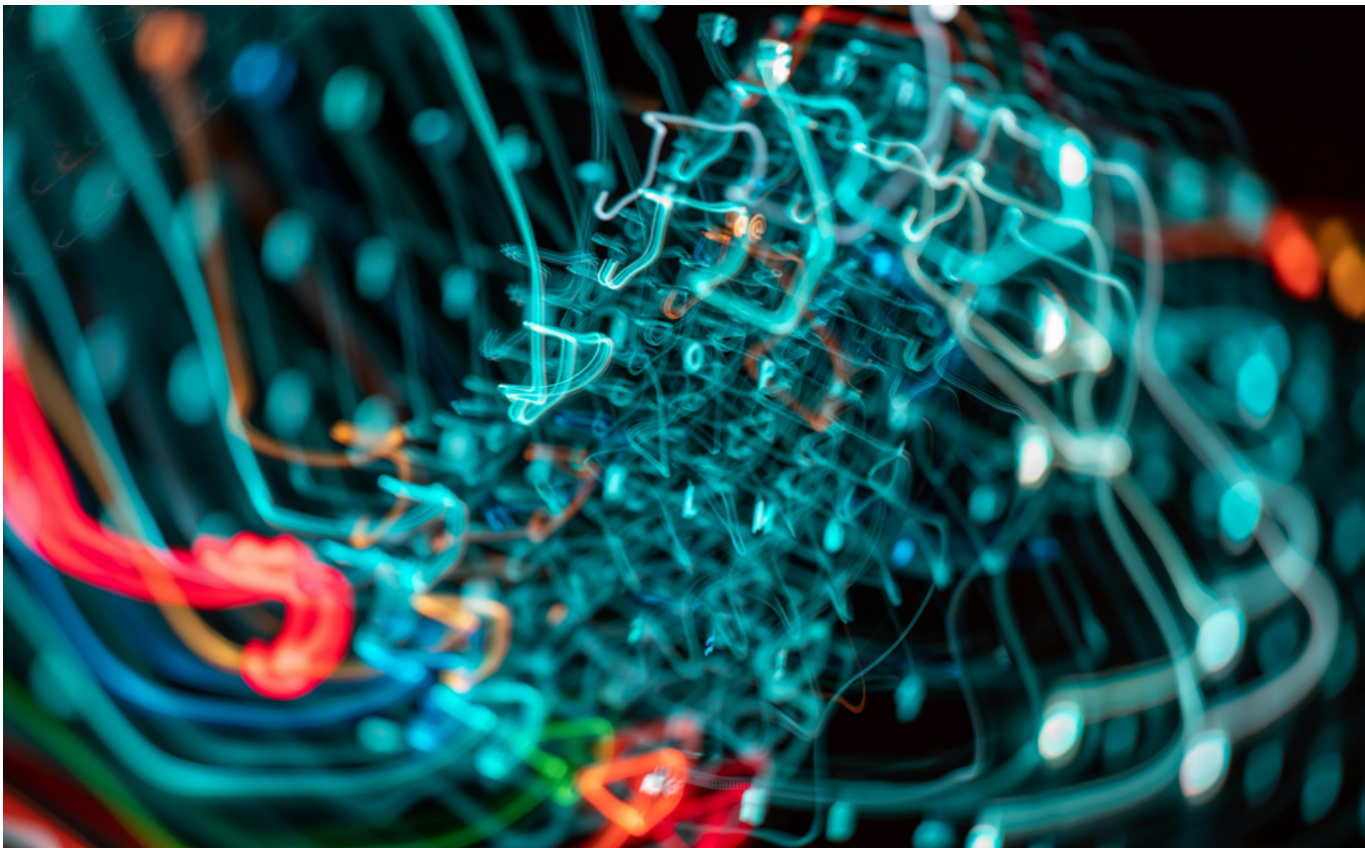
*Los programas y aplicaciones electrónicos, informáticos y telemáticos que vayan a ser utilizados por las Administraciones Públicas para el ejercicio de sus potestades, habrán de ser previamente aprobados por el órgano competente, quien deberá difundir públicamente sus características.*

Pero actualmente la Ley 39/2015 de Procedimiento Administrativo y la Ley 40/2015 del Sector Público conceden una prioridad absoluta a la relación digital de los ciudadanos con la administración y regula las actuaciones automatizadas por parte de la administración, así:

#### **Ley 39/2015 de Procedimiento Administrativo Común**

**Artículo 14.** *Derecho y obligación de relacionarse electrónicamente con las Administraciones Públicas.*

**1.** *Las personas físicas podrán elegir en todo momento si se comunican con las Administraciones Públicas para el ejercicio de sus derechos y obligaciones a través*



de medios electrónicos o no, salvo que estén obligadas a relacionarse a través de medios electrónicos con las Administraciones Públicas. El medio elegido por la persona para comunicarse con las Administraciones Públicas podrá ser modificado por aquella en cualquier momento.

**2.** En todo caso, estarán obligados a relacionarse a través de medios electrónicos con las Administraciones Públicas para la realización de cualquier trámite de un procedimiento administrativo, al menos, los siguientes sujetos:

- a) Las personas jurídicas.
- b) Las entidades sin personalidad jurídica.
- c) Quienes ejerzan una actividad profesional para la que se requiera colegiación obligatoria, para los trámites y actuaciones que realicen con las Administraciones Públicas en ejercicio de dicha actividad profesional. En todo caso, dentro de este colectivo se entenderán incluidos los notarios y registradores de la propiedad y mercantiles.
- d) Quienes representen a un interesado que esté obligado a relacionarse electrónicamente con la Administración.
- e) Los empleados de las Administraciones Públicas para los trámites y actuaciones que realicen con ellas por razón de su condición de empleado público, en la forma en que se determine reglamentariamente por cada Administración.

**3.** Reglamentariamente, las Administraciones podrán establecer la obligación de relacionarse con ellas a

través de medios electrónicos para determinados procedimientos y para ciertos colectivos de personas físicas que por razón de su capacidad económica, técnica, dedicación profesional u otros motivos quede acreditado que tienen acceso y disponibilidad de los medios electrónicos necesarios.

### **Ley 40/2015 del Sector Público**

**Art. 41.** Actuación administrativa automatizada.

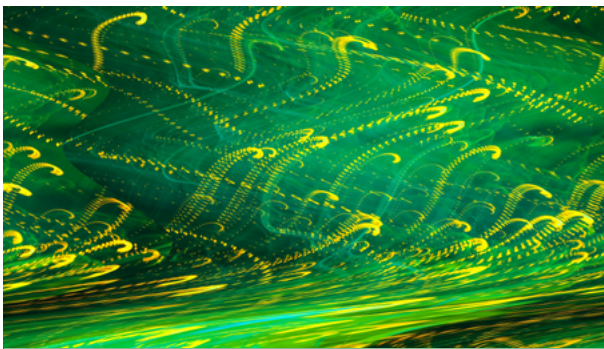
**1.** Se entiende por actuación administrativa automatizada, cualquier acto o actuación realizada íntegramente a través de medios electrónicos por una Administración Pública en el marco de un procedimiento administrativo y en la que no haya intervenido de forma directa un empleado público.

**2.** En caso de actuación administrativa automatizada deberá establecerse previamente el órgano u órganos competentes, según los casos, para la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, auditoría del sistema de información y de su código fuente. Asimismo, se indicará el órgano que debe ser considerado responsable a efectos de impugnación.

### **3. VENTAJAS Y RIESGOS DE LA IA**

Si la preocupación sobre la afectación del uso de la informática a los derechos fundamentales ya estaba en la mente de los constituyentes de 1.878, hubo que esperar hasta 1.992 para la aprobación de la Ley Orgánica 5/1992 que reguló el tratamiento automatizado de datos de carácter personal, actualmente derogada por la Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal.

Que la protección de datos se ve afectada por la IA lo revela el informe presentado por el anterior presidente del Banco Central Europeo Mario Draghi el 9 de septiembre de 2024 en el que llegó a proponer una interpretación laxa del Reglamento General de Protección de Datos, para de esta forma facilitar el desenvolvimiento y avance de las industrias tecnológicas europeas en su competencia con las americanas y chinas, postulando la unión de los operadores de telecomunicaciones, se quejaba de que en Europa hay demasiados operadores de telecomunicaciones (33) comparado con EEUU (5) y China (3) por lo que las europeas no pueden por volumen y capacidad competir, advertía además de que Europa



**“[...] llegó a proponer una interpretación laxa del Reglamento General de Protección de Datos, para de esta forma facilitar el desenvolvimiento y avance de las industrias tecnológicas europeas en su competencia con las americanas y chinas”**

necesita invertir 800.000 millones de euros anuales, que representa el 5% del PIB, para cerrar la brecha de innovación con EEUU y China, aumentar la productividad y reducir la dependencia exterior.

En cualquier caso caben 2 preguntas, la primera sobre la verdadera titularidad de los datos que nos afectan, porque en muchas ocasiones nos resultan desconocidos hasta que la utilización de una máquina más o menos sofisticada nos los descubre, así cuando en una exploración médica nos descubren alguna dolencia que puede resultar de interés para la comunidad, es claro que el dato es nuestro pero ¿podemos oponernos que el sistema lo desconozca cuando resulte relevante para el diagnóstico de otros o para la implementación de medidas de prevención de la salud pública o para el entrenamiento de la máquina? Esta pregunta la formuló el antropólogo Genis Roca en un encuentro en el Círculo de Economía de Barcelona en 2022. La segunda es si somos conscientes de que nuestra vida digital ya pertenece a los servidores que utilizamos diariamente y si esa conciencia cambiará nuestro comportamiento de exposición en redes que actualmente mantenemos.

Lo realmente relevante no es esta priorización de los medios digitales en la relación con la administración, sino que la utilización de la inteligencia artificial por la administración que ya es una realidad. Se trata de justificar por el hecho de que no parece razonable utilizar una gran cantidad de recursos personales en la adopción de resoluciones repetitivas cuando esa función puede llevarse a cabo, con mayor rapidez y menor coste, a través de programas de inteligencia artificial, aunque las mismas necesariamente han de ser genéricas y estereotipadas. Lo que podría aceptarse cuando se trate de ámbitos absolutamente reglados y resoluciones a adoptar de forma mecánica, pero resulta más cuestionable cuando estemos en presencia de actos en

los que existe un margen de apreciación (no solo en el caso de potestades discrecionales sino también en ámbitos en el que juegan conceptos jurídicos indeterminados, que difícilmente una máquina podrá manejar).

En cualquier caso, no cabe desconocer que los sistemas de IA presentan porcentajes de error inferiores a los que resultan de la intervención humana en algunas actuaciones, así es en el examen y valoración de imágenes muy utilizadas en el diagnóstico médico y determinaciones predictivas. Sin embargo, en actividades o tareas que entrañan la toma de decisiones complejas los porcentajes de error son más elevados ya que de ordinario exigen la valoración de componentes psicológicos o emocionales de los que los sistemas de inteligencia artificial carecen o tienen prohibido ponderar. Lo que conlleva el riesgo de que se produzcan errores, que en este ámbito se llaman alucinaciones, a la hora de resolver un conflicto jurídico que, además de afectar derechos o libertades de personas, corren el peligro de multiplicarse exponencialmente, porque la máquina se retroalimenta con las soluciones que da, por lo que aplicaría mecánicamente a todas las situaciones que identifique como idénticas sin entrar en ningún tipo de apreciación subjetiva. En relación con la supuesta neutralidad de la IA en la Comisión Europea para la Eficiencia de la Justicia se advirtió que

*la neutralidad de los algoritmos es un mito, ya que sus creadores les transfieren, consciente o involuntariamente, sus propios sistemas de valores*

No solo es que el programador contamine con sus prejuicios el algoritmo, sino que corremos el riesgo de que lo configure, de entrada, para que responda a la finalidad que interesa a quienes lo implantan. El filósofo francés **ERIC SABÍN** señala que los sistemas de IA son un modelo de inteligencia racional utilizado con fines mercantilistas y utilitaristas.

No obstante, el campo de aplicación de los sistemas de IA es tan vasto, que Europa optó por someterlo a una regulación homogénea para los 27 países de la Unión a través de un Reglamento que entró en vigor el 1 de agosto y que se desarrollará en los próximos 2 años. La norma parte de la necesidad de garantizar el uso ético y confiable de los sistemas de IA, pero también hay quien cuestiona que una regulación tan exhaustiva pueda suponer un freno a la innovación tecnológica y lastrar el desarrollo de las empresas europeas frente a las de EEUU, que no quedarían sometidas a esta regulación cuando no operen en territorio europeo, corriendo el riesgo de otorgar el conocimiento de nuestra vida digital a un puñado de compañías americanas (Mega, Facebook, telegram, X...). De ordinario, a modo de crítica o de justificación, se mantiene que mientras EEUU innova, China copia y la UE regula.

#### 4. DISTINCIÓN ENTRE DECISIONES AUTOMATIZADAS O ROBOTIZADAS E IA

Resulta preciso distinguir entre la **resolución automatizada o robotizada de expedientes** administrativos y los sistemas de IA. Suele mantenerse que en el tratamiento automatizado de los expedientes la tecnología elabora el producto que incluye la resolución o decisión final, pero no influye en su contenido. Se defiende que se trata de actividades estrictamente regladas de modo que volcándose en las aplicaciones de los ordenadores la regulación normativa que le sirven de parámetros para ofrecer un resultado, lo hace de forma invariable, pero la solución que ofrece no se diferencia de la que habría de propo-

ner un funcionario, a salvo el caso de que exista un error de programación que determinarían propuestas contrarias al ordenamiento jurídico. En todo caso también en la tramitación automatizada de los expedientes resulta exigible la supervisión humana para que valide la propuesta y se responsabilice de la misma. En estos casos, aun sujetos a la supervisión responsable del titular de la oficina, cabe que se produzca el sesgo de automatización. Es el caso de los expedientes sancionadores de tráfico que actualmente tramita el Centro de Tratamiento de Denuncias Automatizadas de la DGT adjudicado a una UTE (formada por 3 empresas, Indra, otra austríaca Kaspch y otra francesa Wordline) por 100 millones de euros y que, según resulta de la memoria de contratación, desde su puesta en marcha e 1 de marzo de 2008 tramitó 38.000.000 de expedientes.

En todo caso el apartamiento por parte de la administración de la programación inicial de actuaciones no determina el éxito de la impugnación judicial de la resolución dictada por la administración. Así en el a caso resuelto por la St. del T.S. en relación con la actuación de la AEAT.

**St. del T.S. 19 de febrero de 2020 (recurso de casación 240/2018)**

*Un contribuyente que fue inspeccionado y sancionado impugna los actos administrativos porque sostiene que, al no encontrarse entre los objetivos señalados en el plan de inspección, no debería haber sido inspeccionado. El TS entiende que ese posible incumplimiento del plan de inspección es irrelevante, lo importante es que se detecten hechos constitutivos de infracción tributaria: “La introducción de los planes de actuación, de Control, parciales, pueden tener la finalidad de facilitar a la Administración Tributaria su funcionamiento, especialmente mediante la aplicación de programas*

*informáticos que permitan un tratamiento más ágil de la fiscalización de un determinado sector de contribuyentes, pero su previsión, quizá no muy acertada, en normas legales y reglamentarias, no puede traducirse sin más en un derecho subjetivo del contribuyente a no ser investigado si no se encuentra incluido en dichos planes y programas, pues ello supondría el incumplimiento por parte de la Administración del deber de fiscalización de que todos los ciudadanos cumplan con el deber de contribuir, previsto en el artículo 31.1 de nuestra norma constitucional*

## 5. PREVISIONES LEGALES EN ESPAÑA APLICABLES A LA IA

Antes referí la definición del reglamento comunitario sobre IA, pero en la legislación interna ya se apunta en la misma dirección en la previsión contenida en el **art. 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación** que, bajo el enunciado “Inteligencia Artificial y mecanismos de toma de decisión automatizados”, se pronuncia en los siguientes términos:

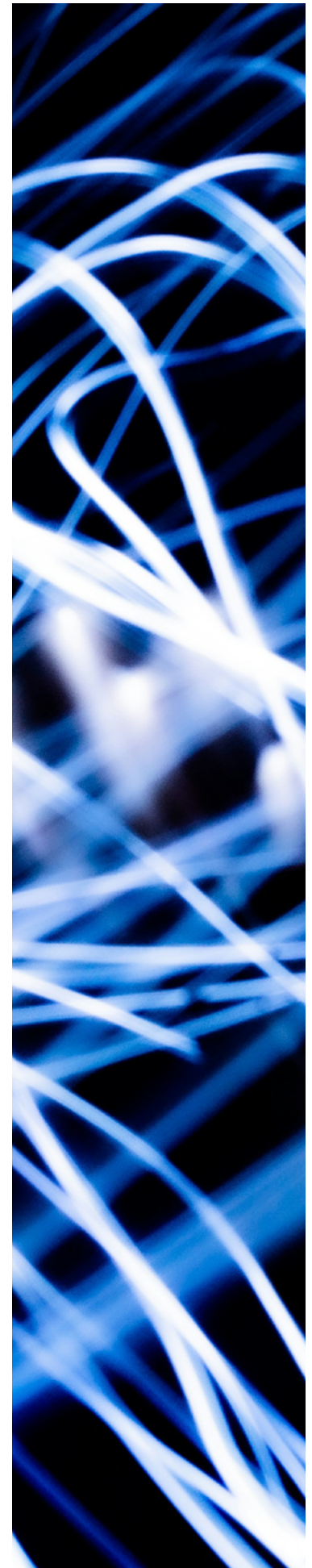
1. En el marco de la Estrategia Nacional de Inteligencia Artificial, de la Carta de Derechos Digitales y de las iniciativas europeas en torno a la Inteligencia Artificial, las administraciones públicas favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones 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 dis-

criminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.

2. Las administraciones públicas, en el marco de sus competencias en el ámbito de los algoritmos involucrados en procesos de toma de decisiones, priorizarán la transparencia en el diseño y la implementación y la **capacidad de interpretación de las decisiones adoptadas por los mismos**.
3. Las administraciones públicas y las empresas promoverán el uso de una IA **ética, confiable y respetuosa con los derechos fundamentales**, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido.
4. Se promoverá un sello de calidad de los algoritmos.

Parece claro que la evaluación del impacto debe realizarse con carácter previo a su implementación y con una definición de los objetivos que se pretenden alcanzar, porque la administración que la utilice debe ser responsable de los resultados que produzca y velar de que sus objetivos no se desvíen de los que se persiguen, que es el parámetro que también ha de orientar las decisiones de los tribunales cuando controlen la actuación de la administración con arreglo a los Arts. 103 y 106 de la C.E. Conviene recordar que con arreglo al Art. 4 de la LRSP son los fines perseguidos los determinan la proporcionalidad de las medidas de las administraciones que limitan el ejercicio de los derechos o condicionen el ejercicio de actividades, evitando resultados discriminatorios que además han de ser evaluados periódicamente.

En todo caso la **Ley 15/2022**, contiene un completo catálogo de medidas para sancionar los actos o disposiciones que resulten discriminatorios, así conviene resaltar que el **Art. 26** sanciona con la nulidad de pleno derecho cualesquiera disposiciones, actos o cláusulas de los negocios jurídicos que constituyan o causen discriminación por razón de alguno de los motivos previstos en esa ley. La ley obliga al responsable a reparar el daño y restituir a la víctima en la situación anterior al incidente discriminatorio, presumiendo la existencia de un daño moral en función de la difusión del acto (**Art. 27**). Señalando el alcance de la tutela judicial en el caso de actos discriminatorios en su **Art. 28** confiriendo legitimación para su persecución no solo a los afectados sino también a los partidos políticos, sindicatos, asociaciones y las organizaciones de consumidores y usuarios en su **Art. 29** y prevé la



inversión de la carga de la prueba de modo que será aquella persona a la que se impute la situación discriminatoria la que tenga que acreditar las razones objetivas y razonables para la adopción de la medida y su proporcionalidad, salvo que se trate de procedimientos penales o administrativos sancionadores o se trate de medidas en materia de organización, convivencia y disciplina de los centros docentes (**Art. 30**). Para terminar exigiendo de las administraciones públicas que tengan conocimiento de un acto discriminatorio la incoación del correspondiente expediente para su eliminación y su comunicación a la administración competente, señalando que los sindicatos y las asociaciones profesionales de trabajadores autónomos y las organizaciones de consumidores y usuarios tienen la condición de interesados (**Art. 31**) al tiempo que encomienda su persecución a las fiscalías provinciales previendo que los fiscales reciban formación especializada obligatoria en esta materia de conformidad con las directrices de la Fiscalía General (**Art. 32**).

En base a esta regulación considero que se ha introducido un nuevo motivo impugnación de los actos administrativos en cuya producción haya intervenido un sistema de IA que puede ser de nulidad radical -con arreglo al Art. 47.1 letra a) de la LPA- en el caso de que se aprecie un sesgo discriminatorio en su adopción y puede ser de anulabilidad -conforme al Art. 48 de la LPA- cuando, pese a la inexistencia de vulneración de un derecho fundamental, la administración resulte incapaz de justificar o motivar las razones para su adopción es en lo que se traduce la **“exigencia de transparencia, rendición de cuentas y capacidad de interpretación de sus resultados”** y que en último término se traduce en la posible vulneración de la razonabilidad de la decisión y supone una manifestación del derecho a la tutela judicial efectiva, que es el límite que se contiene en el Art. 18.4 de la C.E. cuando señala que el uso de la informática ha de resultar compatible con los derechos y libertades de los ciudadanos.

En el proceso de enjuiciamiento de los actos o disposiciones en cuya adopción interviene algún sistema de IA se invierte, por disposición legal, la carga de la prueba por lo que acreditado el indicio discriminatorio correrá a cargo de la administración explicar y justificar las razones de la adopción de la decisión y ahí la cuestión podría quedar en manos de los desarrolladores del sistema ya que conviene advertir que la obligación de transparencia y rendición de cuentas queda limitado a que **“sea factible técnicamente”** lo que no solo da entrada a los “sesgos de automatización” sino a la consagración de un ámbito de decisión que no resulte posible impugnar y en el que la propia administración, responsable de la decisión, resulta sometida a los desarrolladores del sistema -de ordinario son empresas privadas titulares de las patentes y que no se caracterizan por su transparencia los que habrían de desvelar el funcionamiento de su “máquina”-. Puede suceder que las propuestas de la máquina resulten inexplicables de forma comprensible en el lenguaje natural, pudiendo dejar a los destinatarios de la resolución al albur de lo que se conoce como “caja negra” de la IA.

Pese a ello, las administraciones públicas ya utilizan sistemas de IA, por ejemplo:

- VioGén que se puso en funcionamiento en 2007 por la Secretaría de Estado de Seguridad del Ministerio del Interior para integrar toda la información en materia de violencia contra la mujer, haciendo predicciones de riesgo, para proteger a las víctimas y emitir avisos, alertas y alarmas.
- Veripol es un sistema desplegado en las comisarias de policía para detectar denuncias falsas, mediante la referencias a palabras comunes empleadas cuando se



miente a la policía (tirón, desde atrás, mochila, casco, vestido negro) pero que no se fundamentan en el lenguaje empleado por los denunciantes sino en los términos recogidos por el policía para redactar la denuncia o atestado, lo que introduce cierto subjetivismo en el sesgo del tramitador.

- Corpus Viewer sistema utilizado por el Ministerio de Economía y la Secretaría de Estado de Universidades e Investigación para identificar áreas de conocimiento de I+D+i, para la concesión de ayudas y subvenciones en materia de patentes evitando duplicidades o detectar los posibles plagios.
- Sistema anticorrupción SALER que puso en práctica la Generalitat Valenciana para obtener alertas tempranas para detectar irregularidades, fraudes y corrupción en la contratación, prevista en la Ley 22/2016 de Inspección General de Servicios y toma datos de bases públicas y privadas (notarios) generándose alertas por ejemplo en el caso de fraccionamiento de contratos.

En la página del CGPJ está publicado un kit de herramientas sobre la IA aprobado por la UNESCO, en el que se prevé un desarrollo en 4 módulos para un aprendizaje del uso de la inteligencia artificial en los conflictos jurídicos de los que resulta interesante destacar del glosario de definiciones las siguientes:

- IA como una “caja negra”: el término “caja negra” se utiliza para denotar un sistema tecnológico que es inherentemente opaco, cuyo funcionamiento interno o lógica subyacente no se comprenden adecuadamente, o cuyos resultados y efectos no se pueden explicar.

- Sesgo de IA: el sesgo de IA es una diferencia sistemática en el tratamiento de ciertos objetos, personas o grupos (por ejemplo, estereotipos, prejuicios o favoritismo) en comparación con otros mediante algoritmos de IA.
- Algoritmo: un algoritmo se refiere a una serie de instrucciones para realizar cálculos u otras tareas, ya sea en matemáticas o en informática. En el caso de la IA, un algoritmo proporciona las instrucciones que permiten a un computador aprender a aprender del entorno y realizar un conjunto de tareas.
- Toma de decisiones algorítmicas (ADM): la toma de decisiones algorítmicas (ADM) se refiere al uso de “resultados producidos por algoritmos para tomar decisiones”.
- Deepfake: un deepfake es cualquier forma de medio (video, audio u otro) que ha sido alterado o creado total o parcialmente desde cero.
- IA explicable (XAI): la IA explicable (XAI) se define como sistemas, algoritmos y modelos con la capacidad de explicar su justificación para las decisiones, caracterizar las fortalezas y debilidades de su proceso de toma de decisiones y transmitir una comprensión de cómo se comportarán en el futuro.
- Redes generativas antagónicas (GAN): las GAN son un enfoque no supervisado de aprendizaje profundo que puede generar material hiperrealista. Las GAN se utilizan para técnicas de aprendizaje profundo no supervisadas, como la generación de imágenes realistas o conjuntos de datos de imágenes, la realización de traducciones de texto a imagen e imagen a texto, el envejecimiento de las caras y la creación de emojis.
- IA generativa: los algoritmos de aprendizaje automático (ML) se han diseñado para crear contenido nuevo, incluidos audio, código, imágenes, texto, simulaciones y videos.
- Aprendizaje automático (ML): el ML es un conjunto de técnicas que permite a las máquinas aprender automáticamente utilizando patrones y deducciones en lugar de instrucciones directas de una persona. Las técnicas de ML con frecuencia instruyen a las máquinas para que alcancen un resultado al proporcionar numerosas instancias de resultados correctos. Sin embargo, también pueden especificar un conjunto de pautas y dejar que la máquina las descubra por sí misma en los datos.
- Procesamiento del lenguaje natural (PLN): el PLN es una técnica de aprendizaje automático que analiza grandes cantidades de texto humano o datos del habla (transcritos o acústicos) en busca de propiedades específicas, como significado, contenido, intención, actitud y contexto.
- Análisis predictivo: el análisis predictivo es la categoría general de herramientas y modelos estadísticos, por ejemplo, sistemas de aprendizaje automático, que utilizan y analizan datos históricos para crear predicciones sobre el futuro para guiar la toma de decisiones. Estas predicciones pueden ser de bajo riesgo (por ejemplo, qué película recomendar), de riesgo medio (qué solicitud de préstamo proponer aceptar) o de alto riesgo (qué acusado tiene más posibilidades de involucrarse en un comportamiento en particular).
- Discriminación por proxy: la discriminación por proxy en los sistemas de IA ocurre cuando una característica aparentemente neutral se sustituye por una prohibida



Cada año el Tribunal Superior del Condado de Los Ángeles se ocupa de alrededor de 1,2 millones de nuevas multas de tráfico. Hace varios años, las personas tuvieron que esperar hasta 2 horas y media para ver a un empleado por su problema de tráfico debido a una crisis financiera estatal que resultó en cierres de juzgados y reducción de personal. Ahora, un asistente en línea del Tribunal Superior de Los Ángeles ayuda a las personas con sus multas de tráfico. El bot jurado utiliza servicios de traducción de ML y comprensión del lenguaje natural. Asiste a más de 5.000 personas ciudadanas cada semana y habla cinco idiomas (Kit Unesco herramientas IA CGPJ página 65).

Una aplicación más modesta son los clásicos “multa-móvil” que hace años advierten de las infracciones de aparcamiento en Vigo y que después ha de corroborar un controlador con cierto margen de apreciación para formular la denuncia.

## 6. LA AGENCIA ESPAÑOLA DE SUPERVISIÓN DE LA IA (AESIA)

Por otra parte, no puedo dejar de mencionar que en toda esta materia la Agencia Española de Supervisión de Inteligencia Artificial (**AESIA**), creada por la DA 13<sup>o</sup> de la Ley de Presupuestos para 2022 con arreglo a la autorización contenida en la DA 7<sup>a</sup> de la Ley 28/2022 de 21 de diciembre, de fomento del ecosistema de empresas emergentes, con sede en el Edificio La Terraza de los Jardines de Méndez Núñez está llamada a tener un papel protagonista en la implantación de los sistemas de IA, al resultar de sus estatutos, aprobados por el Real Decreto 729/2023, de 22 de agosto, que en el camino de servir a la evolución hacia una España más verde, más digital, más cohesionada desde el punto de vista social y territorial e igualitaria y respondiendo a los objetivos de la necesidad de control y supervisión de los riesgos asociados a la IA, se le atribuyen las siguientes funciones:

“[...] El bot jurado utiliza servicios de traducción de ML y comprensión del lenguaje natural. Asiste a más de 5.000 personas ciudadanas cada semana y habla cinco idiomas”



- La supervisión, asesoramiento, concienciación y formación para la adecuada implementación de toda la normativa sobre el adecuado uso y desarrollo de los sistemas de IA
- La inspección, comprobación y sanción en materia de IA
- La minimización de los riesgos que comportan los sistemas de IA
- El fomento de entornos reales de prueba para reforzar la protección de los usuarios y evitar sesgos discriminatorios
- Realizar evaluaciones de impacto para que los algoritmos involucrados en la toma de decisiones que se utilicen por las administraciones públicas tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas (Art. 25.4)

Su ubicación resultó discutida por Granada y la impugnación fue resuelta por **la St. del T.S. de 28 de noviembre de 2023 ROJ: STS 5364/2023** Resolución: 1582/2023 N° Recurso: 78/2023 Sección: 4 Ponente: CELSA PICO LORENZO apelando al carácter discrecional de la decisión. Esta sentencia es interesantísima dado que contiene una importante observación sobre los expedientes administrativos al afirmar dos cosas bien relevantes: **1º)** no puede considerarse digital por el mero amontonamiento de documentos escaneados y **2º)** la decisión última sobre el expediente, su complemento, orden y adecuación es del juez, indicando que el índice debe contar con vínculos que permitan acceder directamente a los documentos incorporados.

## 7. LA CARTA DE LOS DERECHOS DIGITALES DEL MINISTERIO DE ASUNTOS ECONÓMICOS Y DE TRANSFORMACIÓN DIGITAL

El Reglamento Comunitario no puede resultar más denso, de lectura más farragosa y de difícil entendimiento, por ello es preferible acudir al derecho interno para dibujar el marco que ha de regir la relación de los ciudadanos con el entorno digital y, en especial, con la administración, para lo que contamos con un instrumento que pese a carecer de carácter normativo *resulta básico para delimitar la efectividad de los derechos* también en el entorno digital, que es la Carta de Derechos Digitales aprobado por el Ministerio de Asuntos Económicos y Transformación Digital a la que se remite el Art. 23 de la Ley 15/2022 referido anteriormente, en la misma, al margen de regular el derecho al respeto a la personalidad digital, la herencia digital, los derechos a la libertad de expresión, a recibir información veraz, a la protección de menores, así como los de rectificación y olvido contiene tres apartados que me parecen especialmente relevantes. Son los siguientes:

### XVIII Derechos de los ciudadanos en sus relaciones digitales con la administración

1. El derecho a la igualdad de las personas se extiende al acceso a los servicios públicos y en las relaciones digitales con las Administraciones públicas. A tal fin se promoverán políticas públicas activas que garanticen el acceso a los servicios públicos, a los sistemas y los procedimientos a todos los sujetos y la asistencia en tales procedimientos.
2. El principio de transparencia y de reutilización de datos de las Administraciones públicas guiará la actuación de la Administración digital, de conformidad con la normativa sectorial. En particular, se garantizará el derecho de acceso a la información pública, se promoverá la publicidad activa y la rendición de cuentas y se velará por la portabilidad de los datos y la interoperabilidad de los formatos, sistemas y aplicaciones, en los términos que prevea el ordenamiento jurídico vigente.
3. Se promoverá la universalidad, la neutralidad y la no discriminación, en particular por razón de sexo, de las tecnologías usadas por las Administraciones públicas, y se impulsará la puesta a disposición entre Administraciones de aplicaciones de cuyos derechos de propiedad intelectual sean titulares, salvo supuestos de especial protección por una norma.

Las Administraciones públicas promoverán que la provisión de servicios por medios digitales respete los principios de esta Carta.

4. Se ofrecerán alternativas en el mundo físico que garanticen los derechos de aquellas personas que no quieran o no puedan utilizar recursos digitales y no resulten obligadas a ello, en las mismas condiciones de igualdad.

5. El poder público autor de una actividad en el entorno digital deberá identificar a los órganos responsables de la misma.

6. Se promoverán los derechos de la ciudadanía en relación con la inteligencia artificial reconocidos en esta Carta en el marco de la actuación administrativa reconociéndose en todo caso los derechos a:

**a)** Que las decisiones y actividades en el entorno digital respeten los principios de buen gobierno y el derecho a una buena Administración digital, así como los principios éticos que guían el diseño y los usos de la inteligencia artificial.

**b)** La transparencia sobre el uso de instrumentos de inteligencia artificial y sobre su funcionamiento y alcance en cada procedimiento concreto y, en particular, acerca de los datos utilizados, su margen de error, su ámbito de aplicación y su carácter decisorio o no decisorio.

La ley podrá regular las condiciones de transparencia y el acceso al código fuente, especialmente con objeto de verificar que no produce resultados discriminatorios.

**c)** Obtener una motivación comprensible en lenguaje natural de las decisiones que se adopten en el entorno digital, con justificación de las normas jurídicas relevantes, tecnología empleada, así como de los criterios de aplicación de las mismas al caso. El interesado tendrá derecho a que se motive o se explique la decisión administrativa cuando esta se separe del criterio propuesto por un sistema automatizado o inteligente.



**d)** Que la adopción de decisiones discrecionales quede reservada a personas, salvo que normativamente se prevea la adopción de decisiones automatizadas con garantías adecuadas.

**7.** Será necesaria una evaluación de impacto en los derechos digitales en el diseño de los algoritmos en el caso de adopción de decisiones automatizadas o semiautomatizadas.

## **XXV El derecho de los ciudadanos ante la IA**

**1.** La inteligencia artificial deberá asegurar un enfoque centrado en la persona y su inalienable dignidad, perseguirá el bien común y asegurará cumplir con el principio de no maleficencia.

**2.** En el desarrollo y ciclo de vida de los sistemas de inteligencia artificial

**a)** Se deberá garantizar el derecho a la no discriminación cualquiera que fuera su origen, causa o naturaleza, en relación con las decisiones, uso de datos y procesos basados en inteligencia artificial.

**b)** Se establecerán condiciones de transparencia, auditabilidad, explicabilidad, trazabilidad, supervisión humana y gobernanza. En todo caso, la información facilitada deberá ser accesible y comprensible.

**c)** Deberán garantizarse la accesibilidad, usabilidad y fiabilidad.

**3.** Las personas tienen derecho a solicitar una supervisión e intervención humana y a impugnar las decisiones automatizadas tomadas por sistemas de inteligencia artificial que produzcan efectos en su esfera personal y patrimonial.

## **XXVII La tutela judicial de los derechos en los entornos digitales**

**1.** Todas las personas tienen derecho a la tutela administrativa y judicial de sus derechos en los entornos digitales de acuerdo con lo dispuesto en la legislación vigente.

**2.** Asimismo, se promoverá la garantía de los derechos reconocidos en esta Carta en el marco de las relaciones con la Administración de Justicia y, particularmente, los derechos relacionados con la inteligencia artificial, cuando se recurra a ésta para la utilización o el desarrollo de sistemas de soporte a las decisiones o de herramientas de justicia predictiva.

**3.** Se promoverán mecanismos de autorregulación, control propio y procedimientos de resolución alternativa de conflictos, con la previsión de incentivos adecuados para su utilización con arreglo a la normativa vigente.

**4.** Se promoverá la evaluación de las leyes administrativas y procesales vigentes a fin de examinar su adecuación al entorno digital y la propuesta en su caso de reformas oportunas en garantía de los derechos digitales.

## 8. EL AUXILIO DE LA IA PARA PALIAR EL COLAPSO GENERADO POR LA LITIGACIÓN EN MASA

En España existen plataformas de litigación que presentan recursos masa en distintos partidos judiciales (por ejemplo en los casos de reclamaciones por retrasos en vuelos, temas de consumo o incluso de exoneración de pasivo insatisfecho) nosotros mismos hemos comenzado a sentar las bases para que la implantación de sistemas de IA resulte imparables, así el el magistrado del Juzgado de lo Mercantil 11 de Barcelona **Jose María Fernández Seijo** hoy vocal del CGPJ “...en este proceso el objetivo del “papel cero” no es más que el rudimentario punto de partida”... señala como alguna de



“En España existen plataformas de litigación que presentan recursos masa en distintos partidos judiciales, nosotros mismos hemos comenzado a sentar las bases para que la implantación de sistemas de IA resulte imparables ”

las manifestaciones del uso de inteligencia artificial la digitalización de los expedientes, el programa para la anonimización de las resoluciones, la discriminación por vocablos y comandos para nuestras búsquedas en el CENDOJ, pero advierte que otra cosa es la sustitución de la decisión judicial por una máquina, que resulta totalmente prohibido.

Finalmente, las previsiones rescatadas de la Ley de Eficiencia y recogidas en el Real Decreto Ley 6/2023 de 19 de diciembre en el que se regula, de forma muy escasa, las actuaciones automatizadas, proactivas y asistidas por parte de los operadores jurídicos, poniendo le acento en la exigencia de la validación por parte de la autoridad responsable, al disponer:

En todo caso la implementación de sistemas no es exclusivo de España, como resulta del Kit publicado en la página del CPPJ(página 70 kit UNESCO):

### **VICTOR, Brasil**

El Tribunal Supremo de Brasil (STF) utiliza el sistema VICTOR AI, que se desarrolló en colaboración con la Universidad de Brasilia (UnB). La tecnología de IA analiza el enorme volumen de apelaciones presentadas ante el Tribunal Superior y automatiza el proceso de examen mediante la identificación de casos con repercusión general, un requisito para la tramitación de una apelación ante el STF.

### **Sistema inteligente de transcripción judicial de Singapur**

El Sistema inteligente de transcripción judicial (iCTS) se ha implementado en los tribunales de Singapur en asociación con el Instituto de Investigación Infocomm de A\*STAR. El iCTS tiene el potencial de aumentar la eficiencia de la corte al transcribir las audiencias judiciales en tiempo real

### **Prometea, Argentina**

El sistema Prometea utiliza enfoques de IA para generar opiniones judiciales automáticamente. En 2017, la Fiscalía de la Ciudad Autónoma de Buenos Aires, Argentina, comenzó a desarrollar Prometea. La herramienta ha permitido a la Fiscalía mejorar significativamente la eficiencia de sus procesos: una reducción de 90 minutos a un minuto (99 %) para la resolución de un proceso de licitación y de 167 días a 38 días (77 %) para la preparación del juicio. Una vez que Prometea identifica la solución, permite a la persona usuaria completar la opinión legal en función de algunas preguntas y luego muestra una vista previa editable en línea del documento final. El primer borrador del documento es generado automáticamente por el sistema de IA.

### **Pretor en Colombia**

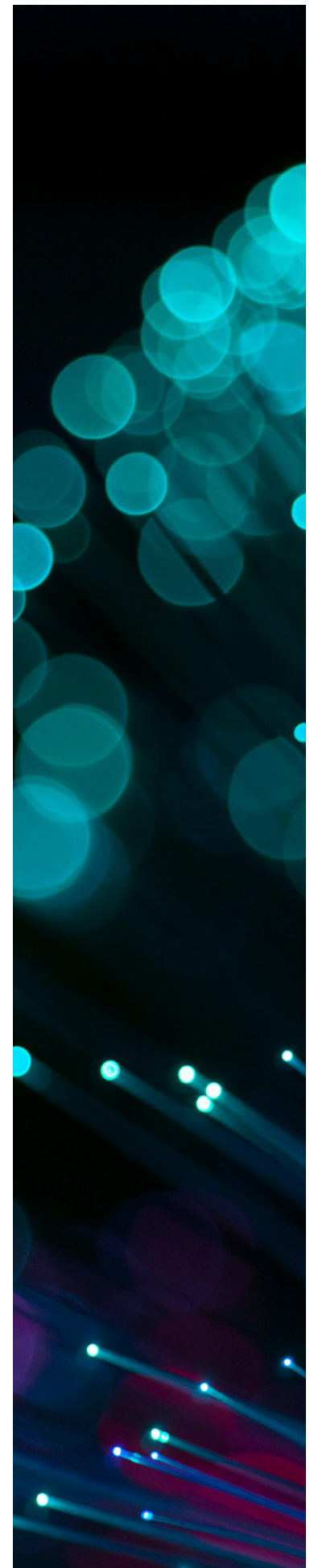
La Corte Constitucional cambió el proyecto de implementar Prometea por no respetar la confidencialidad de los datos y optó por tecnología más clara y transparente que es PretorIA, lanzado a mediados de 2020, utiliza tecnología de modelado de temas en lugar de redes neuronales debido a esto. La nueva versión se puede explicar, interpretar y rastrear por completo.

### **SUPACE, India**

El Poder Judicial de la India tiene un gran número de casos pendientes, alrededor de 38 millones de casos están pendientes en varios tribunales de distrito y de taluka en la India, y más de cien mil casos han estado pendientes durante más de tres décadas.

La herramienta SUPACE AI se está implementando de forma experimental con jueces que manejan casos penales en los Tribunales Superiores de Bombay y Delhi.

La Corte Suprema de la India está explorando el uso de una aplicación móvil que traducirá sus decisiones a nueve idiomas. Además, la India está utilizando la IA para resolver cargos menores, como infracciones de tráfico.



## HART (Herramienta de evaluación de riesgos), Reino Unido

La Herramienta de evaluación de riesgos (HART) es utilizada por la Policía de Durham en el Reino Unido. Utilizando más de treinta características que describen los antecedentes penales y socioeconómicos de una persona, HART utiliza un algoritmo de aprendizaje automático para determinar la probabilidad de reincidencia de un sospechoso. La policía local utiliza las evaluaciones de riesgo completadas por HART para decidir si se acusa a una persona o se remite a un programa de rehabilitación. HART no decide si una persona es culpable o inocente, pero su evaluación puede iniciar una serie de acciones que conducen a que una persona sea privada de su libertad o sea declarada culpable de un delito. Siempre debe haber un ser humano en el circuito que supervise el resultado de un sistema automatizado de toma de decisiones que tome decisiones de alto impacto y sensibles a los hechos.

HART es propenso a criminalizar en exceso, ya que está intencionalmente destinado a subestimar quién está calificado para la vinculación en el programa de rehabilitación. Este método va en contra de la idea de que toda ambigüedad en un caso penal debe resolverse a favor del acusado (“in dubio reo”). Contrariamente a lo que hace HART, un enfoque de cumplimiento de los derechos humanos en la toma de decisiones de la justicia penal tendría que favorecer al acusado.

No te pierdas nuestras publicaciones temáticas en la web





FIND SOMEONE MADE JUST FOR YOU  
*Companion*

# *Companion*

**Diego Gutiérrez Alonso**  
Magistrado



# Companion (2025)

La inteligencia artificial ha sido explorada en el cine desde múltiples perspectivas: el miedo a su autonomía, su potencial para el bien, o su impacto en nuestra identidad. Companion, la nueva película de suspense y ciencia ficción, se inscribe en esta tradición con una idea muy potente, pero que deja la sensación de haber podido ir mucho más lejos.

La historia sigue a un grupo de personajes que, en un contexto aislado, se enfrentan a una IA diseñada para satisfacer necesidades humanas. La premisa inicial promete una exploración de temas como la dependencia emocional hacia lo artificial, la construcción de la personalidad de una IA y la ética detrás de las relaciones hombre-máquina. Sin embargo, a medida que avanza la trama, la película se inclina más hacia el thriller de supervivencia que hacia un comentario profundo sobre estos dilemas.

Uno de los puntos más interesantes de Companion es cómo muestra la fragilidad humana en contraste con la aparente perfección de la IA. ¿La IA es un simple reflejo de lo que proyectamos o realmente una entidad con deseos propios? ¿Su comportamiento es el resultado de una programación sofisticada que imita la humanidad, o en algún momento cruza la línea y desarrolla una voluntad propia? Aquí la película plantea cuestiones que podrían haber sido desarrolladas con más peso filosófico.

Jack Quaid, hijo de Dennis Quaid y Meg Ryan, brilla con un papel que equilibra humor e inseguridad. Si bien su personaje no es el más complejo, su presencia dota de carisma a la historia y confirma su talento más allá de la sombra de su apellido.

En definitiva, Companion es una película con una idea excelente que deja con ganas de más. No falta tensión ni entretenimiento, pero su exploración de la inteligencia artificial y las relaciones humanas podría haber sido más incisiva. Para quienes disfrutan del thriller con toques de ciencia ficción, es una propuesta recomendable, aunque sin llegar a ser revolucionaria.

Plataforma: En cines.

Valoraciones:

- IMDb: 6.9/10 basado en más de 110.000 votos
- Rotten Tomatoes: 93% de crítica, 89% de audiencia
- FilmAffinity: 6.2/10 basado en más de 9.900 votos

