



APM 3.9.

Publicación digital. - Asociación Profesional de la Magistratura

JOAQUÍN DELGADO MARTÍN
COMUNIDAD DE MADRID

DECÁLOGO PARA LA HUMANIZACIÓN DE LA APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL (IA) EN LA JUSTICIA

RESUMEN: La aplicación de la IA puede aportar grandes beneficios para la calidad y la eficacia de la justicia, mejorando sus tiempos de respuesta; pero también introduce **grandes riesgos para los derechos** de las personas y las garantías del proceso. Frente a ello, debemos poner a la persona en el centro del sistema de justicia (**humanización**). Este trabajo reflexiona sobre los elementos necesarios para garantizar una aplicación “humana” de los sistemas IA, desarrollando un **decálogo** en torno a los siguientes principios: respeto a la dignidad humana; control por el usuario del sistema de justicia; respeto de los derechos fundamentales; garantía de acceso a justicia; respeto a la garantía jurisdiccional; no discriminación y ausencia de sesgos; transparencia; confianza; respeto a las garantías de proceso; y responsabilidad y rendición de cuentas.

PALABRAS CLAVE: Inteligencia artificial. Sistema de justicia. Riesgos para los derechos y garantías procesales. Humanización. Decálogo de principios. Dignidad humana; control por el usuario. Respeto de los derechos fundamentales. Garantía de acceso a justicia. Garantía jurisdiccional. No discriminación y ausencia de sesgos. Transparencia. Confianza y fiabilidad. Garantías de proceso. Responsabilidad.

1.- SOBRE LA HUMANIZACIÓN	2
1.1.- ¿QUÉ ES HUMANIZAR?	2
1.2.- ¿POR QUÉ ES TAN NECESARIO HUMANIZAR LA APLICACIÓN DE LA IA EN LA JUSTICIA?	3
2.- DECÁLOGO PARA UNA HUMANIZACIÓN	4
1º.- RESPETO A LA DIGNIDAD HUMANA: INTERVENCIÓN Y SUPERVISIÓN HUMANAS	5
2º.- PRINCIPIO DE CONTROL POR EL USUARIO DEL SISTEMA DE JUSTICIA	5

3º.- RESPETO DE LOS DERECHOS FUNDAMENTALES	6
4º.- RESPETO DE LAS GARANTÍAS DEL PROCESO DEBIDO	8
5º.- GARANTÍA DEL ACCESO A LA JUSTICIA.....	9
5.1.- <i>Personas vulnerables</i>	9
5.2.- <i>Brecha digital</i>	10
6º.- RESPETO DE LA GARANTÍA JURISDICCIONAL	11
6.1.- <i>Juez humano</i>	12
6.2.- <i>Independencia judicial</i>	12
6.3.- <i>Competencia (riesgo para el ejercicio adecuado de la función judicial)</i>	12
6.4.- <i>Decisiones judiciales automatizadas</i>	13
7º.- NO DISCRIMINACIÓN Y PROHIBICIÓN DE SESGOS	14
7.1.- <i>Principio de no discriminación</i>	14
7.2.- <i>Ausencia de sesgos</i>	14
8º.- TRANSPARENCIA.....	15
8.1.- <i>Elementos de la transparencia de los sistemas IA</i>	16
7.1.3.- <i>Identificabilidad</i>	16
8.2.- <i>Sistemas de caja negra</i>	18
9º.- CONFIANZA (FIABILIDAD)	18
9.1.- <i>Fiabilidad</i>	18
9.2.- <i>Calidad de los datos</i>	19
10º.- RESPONSABILIDAD Y RENDICIÓN DE CUENTAS	20
10.1.- <i>Rendición de cuentas</i>	20
10.2.- <i>Responsabilidad</i>	20
4.- REFLEXIONES FINALES.....	20
5.- ¿Y EN EL FUTURO?	21

1.- SOBRE LA HUMANIZACIÓN

1.1.- ¿Qué es humanizar?

Cuando hablamos de humanización nos referimos a poner a la persona en el centro del sistema de justicia. De esta manera, la inteligencia artificial (IA) no constituye un objetivo en sí mismo, sino que es un medio que debe servir a las personas, con la finalidad última de aumentar su bienestar¹.

¹ Sonia PAZ señala que necesitamos políticas y marcos normativos que aseguren que estas tecnologías beneficien a la humanidad: una IA centrada en el ser humano. Debemos proteger lo que nos hace humanos. Porque lo que nos hace humanos no es la tecnología, sino la creatividad, la intuición, las emociones. Este es uno de los principales requerimientos para el futuro: mantener al ser humano por encima de la tecnología; en “Hacia una IA humana”, Business Review (Núm. 316), TIC, Noviembre 2021, <https://www.harvard-deusto.com/hacia-una-ia-humana>

Una aplicación humana de la IA en la justicia conduce a considerar a la persona como eje del diseño e implantación de cada solución IA (enfoque antropocéntrico²), lo que se concreta en dos elementos básicos:

- La solución IA ha de ir destinada a cubrir las necesidades de la persona usuaria del sistema de justicia: una **respuesta de la justicia con eficacia y calidad** frente al derecho violado o incumplido
- Y la solución IA ha de respetar plenamente los derechos de las personas (**enfoque de derechos**). Según la Recomendación del Consejo de Ministros de la OCDE sobre la inteligencia artificial, de 22 de mayo de 2019, las aplicaciones IA deben diseñarse y desarrollarse de conformidad con los valores centrados en el ser humano y la equidad, lo que concreta en dos elementos: a) Los actores de IA deben respetar el estado de derecho, los derechos humanos y los valores a lo largo del ciclo de vida de los sistemas de IA. Estos derechos y valores incluyen libertad, dignidad y autonomía, protección de la privacidad y los datos, no discriminación e igualdad, diversidad, equidad, justicia social, así como los derechos de los trabajadores reconocido internacionalmente; y b) Para hacer esto, los actores de IA deben instituir salvaguardas y mecanismos, tales como la atribución de la capacidad final de toma de decisiones a los humanos, que se adaptan al contexto y al estado de la cuestión.

1.2.- ¿Por qué es tan necesario humanizar la aplicación de la IA en la justicia?

La necesidad de humanización salta a un primerísimo plano en la aplicación de la IA a la justicia dado el elevado riesgo que recae sobre los derechos de las personas. Recordemos que una solución IA podrá constituir un riesgo elevado para esos derechos en función de tres elementos: en primer lugar, por las **propias características del sector** en el que va a ser implementada, como puede ocurrir con la sanidad y determinados ámbitos del sector público (el asilo, la migración, los controles fronterizos, el poder judicial, la seguridad social y los servicios de empleo); en segundo lugar, por la propia **modalidad de utilización** en el sector afectado, es decir, para qué se usa la aplicación IA (no es lo mismo la gestión de citas/agenda del Juzgado que la gestión de la información de un proceso penal sometido al secreto de actuaciones en el fase de instrucción); y, en tercer lugar, por la **naturaleza de los datos** sobre los que va a operar la solución IA (datos personales y datos no personales, y dentro de estos últimos son más sensibles los datos personales de categoría especial).

El sistema de justicia, como sector, presenta una serie de características que determinan con carácter general un **riesgo elevado** de que la aplicación de las soluciones IA atenten contra los derechos de las personas, muchos de ellos protegidos como derechos fundamentales. El tratamiento de la información en el proceso presenta unas notas singulares ligadas al derecho a la tutela judicial efectiva, al derecho de defensa y a un proceso con todas las garantías (artículo 24 CE), pero también a la libertad de información (artículo 20.3 CE) y a la publicidad de actuaciones (artículo 120.3 CE); o incluso puede llegar a afectar a principios básicos del Estado de Derecho como la independencia judicial. Por otra parte, en el proceso se utilizan una gran cantidad de datos personales, muchos de ellos pertenecientes a alguna categoría especial: una gran cantidad de datos son sensibles por su propia naturaleza (por ejemplo, sufrir una determinada enfermedad) y/o por la situación de vulnerabilidad de la persona titular, quien

² Comunicación de la Comisión COM(2019) 168 «Generar confianza en la inteligencia artificial centrada en el ser humano»

puede sufrir un mayor daño derivado de la difusión de los datos personales (menores de edad, víctimas de delitos, personas con discapacidad, inmigrantes...). Y, por último, la aplicación de IA para el tratamiento de información del proceso puede dar lugar a problemas de discriminación, usurpación de identidad o fraude, pérdidas financieras, daño para la reputación, pérdida de confidencialidad de datos sujetos al secreto profesional, reversión no autorizada de la seudonimización o cualquier otro perjuicio económico o social significativo³.

La Propuesta de Reglamento UE sobre Ley de Inteligencia Artificial⁴ señala en su considerando 40 que «deben considerarse de alto riesgo ciertos sistemas de IA destinados a la administración de justicia y los procesos democráticos, dado que pueden tener efectos potencialmente importantes para la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial. En particular, a fin de evitar el riesgo de posibles sesgos, errores y opacidades, procede considerar de alto riesgo aquellos sistemas de IA cuyo objetivo es ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho y a aplicar la ley a unos hechos concretos».

Por último, los riesgos son **potencialmente más graves en el proceso penal**⁵, dada la presencia de derechos fundamentales tan relevantes como la presunción de inocencia, los derechos fundamentales a la libertad y la seguridad de la persona, así como el derecho a un juicio justo y a un recurso efectivo. La aplicación de la IA al proceso penal determina mayores capacidades para la investigación de hechos punibles y para la generación de evidencias digitales; pero genera el riesgo de volver a sistemas de adopción de medidas cautelares basadas en el perfil del acusado o demandado, así como un peligro de sesgo en sistemas automatizados de asistencia decisonal en concesión de beneficios penitenciarios⁶.

2.- DECÁLOGO PARA UNA HUMANIZACIÓN

1. Principio de respeto a la dignidad humana: garantía de intervención y supervisión humanas
2. Principio de control por el usuario del sistema de justicia
3. Principio de respeto de los derechos fundamentales
4. Principio de garantía del acceso a justicia
5. Principio de respeto a la garantía jurisdiccional: derecho a la tutela judicial efectiva por juez independiente

³ Véase el Considerando 75 del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos

⁴ Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial), 21.4.2021, COM(2021) 206.

⁵ Resolución del Parlamento Europeo de 6 de octubre de 2021

⁶ Ricard MARTÍNEZ MARTÍNEZ, «Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo», 2019, *Revista Catalana de Dret Públic*, (58). 64-81. <https://doi.org/10.2436/rcdp.i58.2019.3317>

6. Principio de no discriminación y ausencia de sesgos
7. Principio de transparencia
8. Principio de confianza: fiabilidad y calidad de datos
9. Principio de respeto a las garantías de proceso:
10. Principio de responsabilidad: rendición de cuentas y responsabilidad por las actuaciones

1º.- Respeto a la dignidad humana: intervención y supervisión humanas

En el universo IA hay que poner en un primer plano la necesidad de respeto a la dignidad humana, esto es, la estima que merece la persona en cuanto ser humano, y que se configura como fundamento axiológico de los derechos fundamentales. Se trata de garantizar tres elementos: participación humana (*human-in-the-loop*), supervisión humana (*human-on-the-loop*) y control humano (*human-in-command*). Y que ha de abarcar todo el ciclo de vida de la solución IA (fases que no son necesariamente secuenciales)⁷: planificación y diseño; recopilación y procesamiento de datos; construcción y uso del modelo; verificación y validación; despliegue; y funcionamiento y seguimiento.

La Comisión Europea⁸ considera que la **supervisión** humana ayuda a garantizar que un sistema de IA no socave la autonomía humana en la toma de decisiones ni cause otros efectos adversos. La supervisión debe lograrse a través de mecanismos de gobernanza. En este marco, cabe hablar de un principio general que ha de inspirar estas actuaciones: cuanto menor sea la supervisión que puede ejercer un ser humano sobre un sistema de IA, más extensas tendrán que ser las pruebas y más estricta la gobernanza⁹.

En todo caso, hay que garantizar que las autoridades públicas del sistema de justicia tengan la capacidad de ejercer sus competencias de supervisión conforme a sus mandatos.

2º.- Principio de control por el usuario del sistema de justicia

Este principio consiste en que el usuario de la justicia, cuando ejercita una acción o desarrolla su derecho de defensa frente a una acción, sea un sujeto plenamente informado y que controle sus elecciones.

De esta manera, el justiciable debería estar claramente informado sobre cualquier posibilidad de tratamiento de un caso por IA, tanto antes como durante un proceso judicial; y ha de tener la posibilidad de oponerse. Asimismo, debería ser informado en un lenguaje claro y

⁷ “OCDE Framework for the classification of AI Systems”, febrero 2022;

<https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1649672455&id=id&accname=guest&checksum=A4FC753EE627FB38C2FB41AEDD7E8B39>

⁸ COM(2019) 168.

⁹ Comunicación de la Comisión “Generar confianza en la inteligencia artificial centrada en el ser humano”, COM(2019) 168

comprensible sobre la naturaleza vinculante o no de las soluciones propuestas por las herramientas IA; sobre las diversas opciones posibles; así como sobre su derecho a la asistencia jurídica y a un tribunal “humano”¹⁰.

Por otro lado, el profesional de la justicia ha de poder revisar en todo momento las decisiones judiciales y datos que se utilizaron para producir un resultado. Y se ha de posibilitar que el juez se aparte del resultado del sistema IA, a la vista de las especificidades del caso concreto.

Por último, la posibilidad de decisión automatizada ha de ser prevista de forma expresa por una norma legal, que regule de forma suficiente sus presupuestos y condiciones. Y uno de los posibles supuestos puede fundamentarse en un consentimiento explícito y plenamente informado de la persona afectada.

3º.- Respeto de los derechos fundamentales

La aplicación de soluciones IA en el sistema de justicia ha de respetar necesariamente todos los derechos fundamentales en todas las fases del ciclo de vida de la IA, desde la propia fase de diseño (*human rights by design*)¹¹.

Dado que se encuentran sometidos a grandes riesgos, debemos hacer un especial hincapié en que deben garantizarse la privacidad y la protección de datos. En un primer acercamiento, hay que tener en cuenta que hay soluciones IA que no tratan datos personales en alguna o todas de las etapas de su ciclo de vida, como ocurre con los sistemas de control de calidad de productos industriales, o aquellos sistemas de toma de decisiones sobre la compra y venta de productos financieros. Sin embargo, serán numerosos los supuestos en que los sistemas IA afecten a los datos personales, y dicho grado de afectación puede resultar muy elevado. De esta manera, los riesgos de las diferentes soluciones IA sobre los datos personales han de ser evaluados y gestionados a lo largo de todo el ciclo de vida de un sistema de IA¹².

Un análisis de estos riesgos ha de tener un **enfoque integral**, de tal manera que afecte tanto al concreto sistema IA como al tratamiento en el que la solución IA se integra; es decir, no ha de referirse de forma aislada a la solución IA en cuestión, sino que ésta constituye un componente que se incluye en una o varias fases de un tratamiento de datos que también tiene otros componentes como la recogida de datos, sistemas de archivos, módulos de seguridad, interfases de usuario, y otros. En conclusión, el componente IA y el resto de los elementos que conforman el tratamiento se han de estudiar como un todo.

Si la evaluación del riesgo de un componente IA (en el contexto de un tratamiento de datos)

¹⁰ Carta Ética Europea para la utilización de la Inteligencia Artificial en los Sistemas Judiciales, CPEJ-Consejo de Europa; <https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b>

¹¹ La Recomendación (2020)1 del Comité de Ministros del Consejo de Europa, sobre las consecuencias de los sistemas algorítmicos para los derechos del hombre, recuerda que “Los Estados deberían identificar y/o desarrollar marcos y normas institucionales y reglamentarios apropiados que establezcan criterios y salvaguardias generales o sectoriales para garantizar la compatibilidad del diseño, desarrollo y despliegue continuo de sistemas algorítmicos con los derechos humanos”.

¹² En este apartado se siguen las ideas contenidas en el documento elaborado por la Agencia Española de Protección de Datos sobre «Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial» (febrero 2020).

conduce a concluir que es elevado, la responsabilidad proactiva del responsable de dicho tratamiento le obligará a aplicar medidas tendentes a suprimirlo o minimizarlo.

Siguiendo a la Agencia Española de Protección Datos¹³, estas medidas han de ir destinadas a tres finalidades: en primer lugar, la incorporación de **medidas de privacidad por defecto y desde el diseño** en el tratamiento y que sigan los principios del tratamiento; en segundo lugar, la identificación de **requisitos de seguridad** que minimicen el riesgo para la privacidad; y, en tercer lugar, la adopción de medidas específicas dirigidas a implementar un **sistema de gobernanza de los datos** personales que permitan demostrar el cumplimiento de principios, derechos y garantías para gestionar el riesgo de los tratamientos realizados. La incorporación de medidas de privacidad debería seguir los siguientes principios:

1. **Minimizar** la cantidad de datos que son tratados, tanto en volumen de información recopilada como en el tamaño de la población de estudio, así como a lo largo de las diferentes fases del tratamiento.
2. **Agregar** los datos personales en la medida de lo posible para reducir al máximo el nivel de detalle que es posible obtener.
3. **Ocultar** los datos personales y sus interrelaciones para limitar su exposición y que no sean visibles por partes no interesadas.
4. **Separar** los contextos de tratamiento para dificultar la correlación de fuentes de información independientes, así como la posibilidad de inferir información.
5. Mejorar la **Información** a los interesados, en tiempo y forma, de las características y bases jurídicas de su tratamiento para fomentar la transparencia y permitir a los interesados tomar decisiones informadas sobre el tratamiento de sus datos.
6. Proporcionar medios a interesados para que puedan **controlar cómo** sus datos son recogidos, tratados, usados y comunicados a terceras partes mediante la implementación de mecanismos adaptados al nivel de riesgo que les permita realizar el ejercicio de sus derechos en materia de protección de datos.
7. **Cumplir** con una política de privacidad compatible con las obligaciones y requisitos legales impuestos por la normativa.
8. **Demostrar**, en aplicación del principio de responsabilidad proactiva, el cumplimiento de la política de protección de datos que esté aplicando, así como del resto de requisitos y obligaciones legales impuestos por el Reglamento, tanto a los interesados como a las autoridades de control. Esto implica auditar dinámicamente el resultado/ las conclusiones de los tratamientos, evaluando las divergencias o desviaciones sobre los inicialmente previstos o evaluados como previsibles, incluidos los algoritmos ejecutados, para adoptar, en su caso, medidas correctivas, incluida, la supresión de la información y documentar detalladamente el análisis realizado y las medidas adoptadas.

Aludiendo al marco de gobernanza, cabe destacar el importante papel que está llamado a desempeñar el Consejo General del Poder Judicial, en cuanto autoridad de los datos del proceso (ficheros jurisdiccionales), en relación con los principios, derechos y garantías para gestionar el riesgo de los tratamientos inherentes a la aplicación de soluciones de inteligencia artificial al

¹³ AEPD «Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial» (febrero 2020).

sistema judicial.

4º.- Respeto de las garantías del proceso debido

La implantación de tecnologías de inteligencia artificial en el proceso judicial tiene una clara línea roja que no puede ser sobrepasada: ha de respetar todas las garantías que se agrupan dentro del concepto de proceso debido que se viene recogiendo en los distintos instrumentos internacionales. Cabe destacar el primer inciso del artículo 14.1 del Pacto Internacional de Derechos Civiles y Políticos¹⁴, según el cual «*todas las personas son iguales ante los tribunales y cortes de justicia. Toda persona tendrá derecho a ser oída públicamente y con las debidas garantías por un tribunal competente, independiente e imparcial, establecido por la ley, en la substanciación de cualquier acusación de carácter penal formulada contra ella o para la determinación de sus derechos u obligaciones de carácter civil*».

La complejidad y la opacidad de los sistemas IA pueden afectar al derecho a un juicio justo, incluida la igualdad de armas. Téngase en cuenta que la sociedad ha avanzado hacia un sistema de garantías en la resolución de conflictos, que asegure la independencia e imparcialidad del decisor y el adecuado respeto a los derechos fundamentales. Cabe afirmar con rotundidad **que la aplicación de la inteligencia artificial no puede suponer ningún paso atrás** en esta evolución.

Los sistemas IA potencialmente pueden afectar transversalmente a los diferentes aspectos del proceso, cuyo concreto análisis excede los márgenes de este trabajo. En todo caso, una parte sujeta a una decisión algorítmica ha de poder examinar y cuestionar adecuadamente su aplicación¹⁵; lo que lleva consigo al menos los siguientes elementos¹⁶:

- Es necesario que las partes conozcan que se ha utilizado una solución IA para la toma de decisión (identificación)
- La defensa ha de poder acceder a los algoritmos y datos que se utilizaron para producir un resultado (transparencia)
- Y ha de poder impugnar de forma efectiva la decisión algorítmica (impugnabilidad), lo que puede determinar una serie de consecuencias sobre las posibilidades de alegación y prueba (especialmente en materia de carga de la prueba).

Según las Directrices sobre decisiones individuales automatizadas y elaboración de perfiles a los efectos del Reglamento 2016/679, elaboradas por el Grupo de Trabajo sobre protección de datos del artículo 29 (3 de octubre de 2017 y revisadas), estas decisiones automatizadas pueden basarse en cualquier tipo de datos, citando los ofrecidos directamente por las personas afectadas (como las respuestas a un cuestionario); o los observados acerca de las personas (como los datos sobre la ubicación recogidos a través de una aplicación); o bien los derivados o inferidos como, por ejemplo, un perfil ya existente de la persona (por ejemplo, una calificación crediticia). En todo caso, **las decisiones automatizadas adoptadas en el seno de un**

¹⁴ Aprobado por la Asamblea General de la ONU el 16 de diciembre de 1966

¹⁵ David LESLIE, Cristopher BURR, Mhairi AITKEN, Josh COWLS, Mike KATELL and Morgan BRIGGS, "Artificial intelligence, human rights, democracy, and the rule of law: a primer". The Council of Europe and The Alan Turing Institute, 2021

¹⁶ Véase Antonio ABELLÁN ALBERTOS, "Juez ex machina: inteligencia artificial y justicia", columna en Conflegal.com; 26 de febrero de 2022

proceso judicial han de basarse únicamente en los datos o informaciones obrantes en el proceso, que se hayan incorporado al mismo con pleno respeto de los derechos de las partes y las garantías procesales.

5º.- Garantía del acceso a la justicia

La aplicación de la IA no puede determinar obstáculos para el acceso a la justicia, entendido como el derecho fundamental que tiene toda persona para acudir y promover la actividad del servicio público de impartición de justicia, sin discriminación alguna, con la finalidad de obtener una respuesta efectiva a sus necesidades jurídicas en un tiempo razonable.

Comprende dos espacios básicos: en primer lugar, la **resolución de conflictos por los órganos jurisdiccionales** mediante un recurso idóneo y efectivo que respete las exigencias del debido proceso; y, en segundo lugar, **otros medios de resolución de conflictos (no judiciales)**, incluyendo los medios colaborativos, la justicia restaurativa en materia penal y las distintas formas de justicia comunitarias. La IA puede aportar elementos relevantes para actuar contra las barreras de acceso en ambos espacios del sistema de justicia¹⁷, favoreciendo la inteligibilidad, accesibilidad, receptividad, responsabilidad, eficacia y transparencia de sus actuaciones¹⁸.

En este sentido, aparece el concepto de **cuarta parte** construido por la doctrina: la primera y segunda son las partes en disputa que se enfrentan para dirimir una controversia; la tercera parte hace referencia a una persona que facilita, media, arbitra o resuelve dicha controversia de alguna forma; la cuarta parte, en cambio, “toma asiento en la mesa con la primera y segunda parte (las partes en controversia) y la tercera parte (el humano neutral, como el mediator o árbitro)”. Pues bien, la cuarta parte está constituida por las herramientas algorítmicas, tecnológicas y de machine learning que apoyan (o en algunos casos resuelven) los procesos de resolución de conflictos¹⁹.

5.1.- Personas vulnerables

Las soluciones IA pueden favorecer las condiciones de acceso a la justicia de determinadas personas o grupo que, por encontrarse en situación de vulnerabilidad, tienen mayores obstáculos para la tutela de sus derechos. Recordemos que uno de los 17 Objetivos de Desarrollo Sostenible (Agenda 2030 de la ONU) se refiere específicamente a la justicia: “*promoción de sociedades pacíficas e inclusivas para el desarrollo sostenible, la provisión de acceso a la justicia para todos y la construcción de instituciones responsables y eficaces a todos los niveles*” (ODS 16). Se trata de conseguir una **justicia para todos**, de tal manera que nadie se quede atrás,

¹⁷ De conformidad con el Informe 14(2011) del Consejo Consultivo de Jueces Europeos (CCJE) del Consejo de Europa, sobre Justicia y Tecnologías de la Información (TI), aprobado en Estrasburgo los días 7 a 9 de noviembre de 2011, las tecnologías de la información y de la comunicación constituyen instrumentos para mejorar la Administración de Justicia, para facilitar el acceso de los justiciables a los tribunales y para reforzar las garantías contenidas en el artículo 6 del Convenio Europeo de Derechos Humanos, a saber, el acceso a la justicia, la imparcialidad, la independencia del juez, la equidad y la duración razonables de los procedimientos.

¹⁸ «Recomendaciones básicas sobre sistemas alternativos de organización y gestión de la justicia correspondientes a la sociedad de la información», del Proyecto «E-Justicia: la Justicia en la sociedad del conocimiento. Retos para los países iberoamericanos», desarrollado en la XIII Cumbre Judicial Iberoamericana (Santo Domingo, 21 y 22 de junio de 2006)

¹⁹ CEJA, “Resolución de conflictos en línea”, 2022, página 17; https://biblioteca.cejamericas.org/bitstream/handle/2015/5699/PUB_ODR_vf_162022.pdf?sequence=5&isAllowed=y

instaurando soluciones que permitan vencer los obstáculos de acceso a la justicia de las personas vulnerables (menores, discapacitados, personas en situación de pobreza, colectivo LGTBI, inmigrantes, víctimas de violencia de género...). Las **100 Reglas de Brasilia de acceso a la justicia de las personas vulnerables**²⁰ (Cumbre Judicial Iberoamericana²¹) constituyen su complemento axiológico (principios) y práctico (recomendaciones y soluciones adaptadas a cada grupo vulnerable) indispensable²².

5.2- Brecha digital

Se considera “brecha digital” la separación entre quienes tienen acceso efectivo a las tecnologías digitales y de la información frente a quienes tienen un acceso muy limitado o carecen de él²³. El sistema de justicia²⁴ (al igual que el conjunto de la Administración Pública²⁵) se enfrenta al reto de reducir la brecha digital (separación entre quienes tienen acceso efectivo a las tecnologías digitales y de la información frente a quienes tienen un acceso muy limitado o carecen de él), promoviendo la alfabetización digital especialmente en relación con las personas en situación de vulnerabilidad. Recordemos que esta brecha puede venir determinada por distintas variables²⁶ como el nivel de renta, el estado de salud, el nivel educativo, la edad (niños, niñas y adolescentes, pero también personas adultas mayores), localización de la residencia (“España vaciada”) etcétera²⁷.

²⁰ Texto disponible en web: <https://eurosocial.eu/biblioteca/doc/reglas-de-brasiliasobre-acceso-a-la-justicia-de-las-personas-en-condicion-de-vulnerabilidad/>

²¹ Aprobadas por la XVI Cumbre Judicial Iberoamericana (Brasilia 2008); también han sido aprobadas por la Asociación Iberoamericana de Ministerios Públicos (AIAMP) Asociación Interamericana de Defensorías Públicas (AIDEP), la Federación Iberoamericana de Ombudsmán (FIO) y la Unión Iberoamericana de Colegios y Agrupaciones de Abogados (UIBA). Han sido objeto de actualización en 2018.

²² Véase Joaquín DELGADO MARTÍN, “Guía Comentada de las Reglas de Brasilia”, Herramientas Eurosocial, 2019; <https://eurosocial.eu/biblioteca/doc/guia-comentada-de-las-reglas-de-brasilias>

²³ OECD (2021), “Bridging digital divides in G20 countries, Report for the G20 Infrastructure Working Group; OECD (2018), “Bridging the rural digital divide”; y OECD (2001) “Understanding the digital divide”; entre otros.

²⁴ Conclusión 6ª: “Debería prestarse una atención particular a las necesidades de personas que tienen dificultades para la utilización de las nuevas tecnologías”; Consejo Consultivo de Jueces Europeos (CCJE), Informe N.º 14 (2011) sobre “Justicia y tecnologías de la información (TI)” (noviembre 2011); [1 \(coe.int\)](https://coe.int)

²⁵ Juan G., CORVALÁN se refiere a la existencia de tres grandes desafíos para la Administración Pública: reducir la brecha digital, promover la alfabetización digital y garantizar la asistencia digital a las personas en condición de vulnerabilidad digital; en «Hacia una Administración Pública 4.0: Digital y basada en inteligencia artificial. Decreto de Tramitación digital completa», publicado en: LA LEY 17 de agosto de 2018. <http://aley.thomsonreuters.com/nota/837>

²⁶ Apartado XII.2 de la Carta de Derechos Digitales de España: “se promoverán políticas públicas específicas dirigidas a abordar las brechas de acceso atendiendo a posibles sesgos discriminatorios basados en las diferencias existentes por franjas de edad, nivel de autonomía, grado de capacitación digital o cualquier otra circunstancia personal o social para garantizar la plena ciudadanía digital y participación en los asuntos públicos de todos los colectivos en mayor riesgo de exclusión social, en particular el de personas mayores, así como la utilización del entorno digital en los procesos de envejecimiento activo”; https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/ficheros/Carta_Derechos_Digitales_RedEs.pdf Carta Portuguesa de Derechos Humanos en la Era Digital de 2021, el proyecto de Carta Peruana de Derechos Digitales y el Proyecto de Carta Iberoamericana de Principios y Derechos Digitales (Cumbre Iberoamericana); <https://www.somosiberoamerica.org/temas/derechos-digitales/garantizar-los-derechos-para-una-digitalizacion-centrada-en-las-personas/>

²⁷ Informe de la Cumbre Judicial Iberoamericana “El Covid19 y la Administración de Justicia Iberoamericana”, 2020, página 51

La brecha digital se refiere a dos dimensiones²⁸. Por un lado, el acceso a un software, un hardware y un acceso a internet (ancho de banda) adecuados para la realización de la actuación online, aplicable tanto a la realización de trámites escritos (notificación electrónica, presentación telemática de escritos...) como a la asistencia de actos judiciales (asistencia telemática mediante videoconferencia o sistema similar²⁹). Y, por otra parte, la tenencia de habilidades suficientes para el uso de los instrumentos tecnológicos.

Los obstáculos de acceso a la justicia derivados de la brecha digital pueden tener consecuencias en tres elementos básicos: el acceso a la información; la realización de trámites y la práctica de actos de comunicación; y la participación telemática en actuaciones del sistema de justicia³⁰. En este sentido, la brecha puede afectar tanto a la propia posibilidad de realización del acto (acceso a la justicia) como a las probabilidades de éxito de la pretensión ejercitada (acción) ante los tribunales: en primer lugar, por la falta de realización de un trámite o su práctica defectuosa; y, en segundo lugar, porque puede tener consecuencias negativas sobre la valoración judicial de la asistencia telemática a actos judiciales, ya que puede afectar a la capacidad de la parte de trasladar veracidad y persuasión al juez, así como a la propia capacidad de transmitir emociones y sentimientos.

Esta problemática ha de ser enfrentada mediante políticas públicas destinadas a mejorar las infraestructuras de acceso a internet; y a promover la alfabetización digital, entendida como el conjunto de destrezas, conocimientos y actitudes que necesita una persona para poder desenvolverse funcionalmente dentro de la sociedad de la información y tiene por objetivo el desarrollo de habilidades y conocimiento que les permitan utilizar la tecnología de manera efectiva, especialmente en relación con las personas en situación de vulnerabilidad. En todo caso, las distintas soluciones de IA que puedan implantarse deberán tener en cuenta esta realidad para no afectar al derecho a la tutela judicial efectiva, contemplando vías que posibiliten el ejercicio de los derechos.

6º.- Respeto de la garantía jurisdiccional

Se ha de respetar la denominada garantía jurisdiccional, es decir, la intervención de los jueces y tribunales como mecanismo para garantizar el cumplimiento de las leyes, especialmente el disfrute de un derecho, el cumplimiento de una obligación o el sometimiento de los poderes públicos a la ley³¹.

²⁸ ALBERT FOX CAHN, ESQ. & MELISSA GIDDINGS se refieren a que existen dos grandes brechas digitales: una brecha de acceso y una brecha de habilidades; en «Virtual Justice: Online Courts During Covid-19», de 23 de julio de 2020, página 9

²⁹ Sobre los efectos de la brecha digital en la asistencia telemática a actos judiciales, es muy interesante el trabajo de Antonio ABELLÁN, “Brecha digital y derecho a la asistencia jurídica gratuita”, Confilegal.com, 19-11-20

<https://confilegal.com/20201119-opinion-brecha-digital-y-derecho-a-la-asistencia-juridica-gratuita/>

³⁰ La brecha digital es y debe ser punto de partida, medio y fin de las estrategias de sociedad de la información como la del camino hacia una administración electrónica de la Justicia. Y destaca que afecta a tres elementos básicos: el acceso de la información, la participación en procesos digitales y la monitorización de la Administración; en Estudio comparado sobre E-Justicia: la Justicia en la Sociedad del Conocimiento, UOC, encargado por el Grupo de Trabajo E-Justicia de la XIII Cumbre Judicial Iberoamericana (junio 20026); página 11 y ss

³¹ Diccionario Panhispánico del Español Jurídico

6.1.- Juez humano

Desde algunos ámbitos se defiende que el artículo 24 CE, en conexión con el principio de exclusividad jurisdiccional (art. 117.3 CE), garantiza a los ciudadanos el derecho a una resolución fundada en Derecho dictada por un Juez o Tribunal, esto es, el derecho a que su caso sea resuelto por un Juez- persona³².

6.2.- Independencia judicial

Esta independencia tiene un doble significado: personal, porque afecta al propio estatuto del Juez (conjunto de derechos y obligaciones); y funcional, por cuanto el Juez debe estar exento o libre de influencias e intervenciones extrañas en el cumplimiento de su tarea, tanto si provienen del gobierno, del parlamento, del electorado o de la opinión pública³³, o incluso frente a los propios órganos de gobierno de poder judicial³⁴ y frente a los otros jueces³⁵.

La **dimensión funcional de la independencia judicial cobra un especial sentido en la construcción de soluciones de inteligencia artificial**, dado que el diseño y aplicación de los algoritmos ha de realizarse de tal forma que se eviten sesgos (desviaciones inadecuadas en el proceso de inferencia) e influencias indirectas ilegítimas. Por otra parte, en los supuestos en los que la solución IA ofrece al juez una recomendación de resolución, una adecuada observancia de la independencia judicial exige que sea posible un **apartamiento del resultado** de la solución IA a la vista de las especificidades del caso concreto. Y, aunque este elemento sea garantizado por el sistema, pueden concurrir riesgos ligados a una posible asunción acrítica de la recomendación (objeto del siguiente apartado). Sin embargo ¿realmente es posible que el juez de pueda apartar de forma efectiva de la solución ofertada por un sistema tecnológico basado en datos objetivos?.

6.3.- Competencia (riesgo para el ejercicio adecuado de la función judicial)

Si bien la IA puede reducir la arbitrariedad, las presiones sobre los jueces³⁶ y la acción discriminatoria, las decisiones judiciales apoyadas en la IA pueden afectar negativamente a la independencia normativa y decisoria del Poder Judicial. Los actores judiciales deben tener un nivel suficiente de comprensión sobre la IA que utilizan para garantizar responsabilidad por las

³² Informe CGPJ al Proyecto de Ley de Eficiencia Digital

³³ Karl LOEWENSTEIN, «Teoría de la Constitución», obra citada, página 295.

³⁴ Según el artículo 4 del Estatuto del Juez Iberoamericano, «en el ejercicio de la jurisdicción, los jueces no se encuentran sometidos a autoridades judiciales superiores, sin perjuicio de la facultad de éstas de revisar las decisiones jurisdiccionales a través de los recursos legalmente establecidos, y de la fuerza que cada ordenamiento nacional atribuya a la jurisprudencia y a los precedentes emanados de las Cortes Supremas y Tribunales Supremos»; dicho Estatuto fue aprobado por la VI Cumbre Judicial Iberoamericana, que tuvo lugar en Santa Cruz de Tenerife (España) los días 23 y 25 de mayo de 2001.

³⁵ El artículo 7 del Código Iberoamericano de Ética Judicial dispone que «al juez no sólo se le exige éticamente que sea independiente sino también que no interfiera en la independencia de otros colegas»; este Código fue aprobado por la XIII Cumbre Judicial Iberoamericana, que tuvo lugar en República Dominicana los días 21 y 22 de mayo de 2006.

³⁶ Juan Carlos HIGUERAS destaca sus beneficios para evitar las presiones que pueden recibir los jueces, fomentando la separación efectiva de poderes; en “Cuando la máquina supere al hombre”, Business Review (Núm. 316), TIC, noviembre 2021 página 24, <https://www.harvard-deusto.com/cuando-la-maquina-supere-al-hombre>

decisiones tomadas con su asistencia

Otra cuestión preocupante surge en la interpretación de los resultados de la IA, que consiste en que el juez acepte, sin espíritu crítico, los resultados de una IA como ciertos e inamovibles, asumiendo un «principio de autoridad» derivado de las expectativas creadas por dichos sistemas.

6.4.- Decisiones judiciales automatizadas

Determinadas soluciones avanzadas de inteligencia artificial³⁷ pueden llegar a **sustituir la decisión del juez humano** (toma de decisión automatizada). Cabe acudir a la normativa UE de protección de datos, en concreto al artículo 22 Reglamento UE 2016/679 (RGPD), así como al artículo 11 Directiva UE 2016/680 en relación con prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales. Ambos preceptos, de contenido similar, contienen la **prohibición de las decisiones basadas únicamente en un tratamiento automatizado** (incluida la elaboración de perfiles) que produzcan efectos jurídicos negativos para el interesado o le afecten significativamente, **salvo que concurren cuatro requisitos**:

- Que estén autorizadas por el ordenamiento jurídico.
- Que el responsable del tratamiento establezca **medidas adecuadas para salvaguardar los derechos y libertades del interesado**; y al menos el derecho a **obtener la intervención humana** por parte del responsable del tratamiento.
- Que **no se basen en las categorías especiales de datos personales**, salvo que se hayan tomado las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado; y
- En las actuaciones penales, la Directiva establece además que, cuando se trate de elaboración de perfiles, que **no dé lugar a una discriminación** de las personas físicas basándose en las categorías especiales de datos (art. 1.3 Directiva)

En todo caso, los resultados de las decisiones individuales automatizadas en el proceso han de poder ser supervisados y controlados por las personas humanas. Según la Recomendación del Consejo de Ministros de la OCDE sobre la inteligencia artificial, de 22 de mayo de 2019, las aplicaciones IA deben diseñarse y desarrollarse de conformidad con los valores centrados en el ser humano y la equidad, lo que concreta en dos elementos: a) Los actores de IA deben respetar el estado de derecho, los derechos humanos y los valores a lo largo del ciclo de vida de los sistemas de IA. Estos derechos y valores incluyen libertad, dignidad y autonomía, protección de la privacidad y los datos, no discriminación e igualdad, diversidad, equidad, justicia social, así como los derechos de los trabajadores reconocido internacionalmente; y b) Para hacer esto, los actores de IA deben instituir salvaguardas y mecanismos, tales como la atribución de la capacidad final de toma de decisiones a los humanos, que se adaptan al contexto y al estado de la cuestión.

³⁷ Javier ERCILLA GARCÍA habla de la existencia de dos modelos de Juez IA, un Juez IA que aprenda de la jurisprudencia emitida por el Tribunal Supremo (AlphaJugde), o un Juez IA que sólo aprenda las leyes que son de aplicación (no habría un límite más allá de la jurisdicción) y actuara en consecuencia (AlphaJudge Zero); en «La inteligencia artificial en la Justicia. Asistentes IA judiciales y Jueces IA», *Revista Aranzadi de Derecho y Nuevas Tecnologías*, n.º 53, mayo-agosto 2020.

Asimismo, hay que exigir que **la norma legal regule de forma suficiente sus presupuestos y condiciones**. En relación con las especificaciones técnicas de estas actuaciones judiciales automatizadas, y en espera de una futura Ley de Eficiencia Digital, el artículo 42 de la Ley 18/2011, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la Administración de Justicia, establece que «*en caso de actuación automatizada, deberá establecerse previamente por el Comité técnico estatal de la Administración judicial electrónica la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso la auditoría del sistema de información y de su código fuente*». De esta manera, otorga protagonismo a las actuaciones del **Comité técnico estatal de la Administración judicial electrónica (CTEAJE)**. Y su apartado 2º contempla que «*los sistemas incluirán los indicadores de gestión que se establezcan por la Comisión Nacional de Estadística Judicial y el Comité técnico estatal de la Administración judicial electrónica, cada uno en el ámbito de sus competencias*».

7º.- No discriminación y prohibición de sesgos

7.1.- Principio de no discriminación

Es necesario prevenir específicamente la creación o agravación de la discriminación entre individuos o grupos de individuos, debiéndose mantener una especial atención cuando se fundamente en de datos personales de categoría especial. Esta necesidad de prevención afecta a dos elementos básicos:

- En primer lugar, los **conjuntos de datos utilizados** por los sistemas de IA. Los datos usados en las aplicaciones de IA (tanto para el entrenamiento como para el funcionamiento) no deben contener sesgos, y estas aplicaciones deben ser periódicamente auditadas para detectarlos, en su caso. Asimismo, se debe estar razonablemente seguro de que se nutre a la aplicación de la cantidad, variedad y calidad de datos más exigentes³⁸. Téngase en cuenta que estos datos pueden verse afectados por la inclusión de sesgos históricos involuntarios, ya sea por ser incompletos, ya sea por modelos de gobernanza deficientes. La persistencia en estos sesgos podría dar lugar a una discriminación indirecta.
- En segundo lugar, la **forma en la que se desarrolla el concreto sistema de IA** (por ejemplo, cómo está escrito el código de programación de un algoritmo) también puede estar sesgada. Se trata de cuestiones que han de ser abordadas desde el propio inicio del diseño del sistema y durante todo su desarrollo.

7.2.- Ausencia de sesgos

Una de las cuestiones más problemáticas en las soluciones IA es la posible existencia de **sesgos**, que cabe definir como desviaciones inadecuadas en el proceso de inferencia, que son particularmente graves cuando, por ejemplo, derivan en discriminaciones de un grupo en favor de otro³⁹. Como recuerda Ricard MARTÍNEZ, el sesgo puede deberse a muy distintos factores:

³⁸ Ernest SOLÉ y Susana DOMINGO, “Los siete retos de la inteligencia artificial en el entorno empresarial”, Business Review (Núm. 316) · TIC · Noviembre 2021; <https://www.harvard-deusto.com/los-siete-retos-de-la-inteligencia-artificial-en-el-entorno-empresarial>

³⁹ Documento elaborado por la Agencia Española de Protección de Datos sobre «Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial» (febrero 2020).

en unas ocasiones la calidad de los datos que se toman como marco de análisis, en otras las preconcepciones del programador; y en algunos casos son las propias elecciones de sus usuarios las que incorporan el sesgo al insertar datos cuando el sistema debe retroalimentarse con nueva información⁴⁰.

Se trata de un problema que no afecta únicamente a las soluciones de IA, sino que puede afectar cualquier proceso de toma de decisión, ya sea humano o automatizado, pero que en el contexto IA puede devenir especialmente grave. En la doctrina se habla del principio de **no discriminación algorítmica** en el diseño y/o implementación de los algoritmos inteligentes. Este principio consiste⁴¹ en impedir que los sistemas de IA procesen la información o los datos bajo sesgos o distinciones frente a los seres humanos, por motivos de raza, color, sexo, idioma, religión, opinión política o de otra índole, origen nacional o social, posición económica, nacimiento o cualquier otra condición social (artículo 2 del Pacto de los Derechos Económicos Sociales y Culturales); es decir, evitar que las operaciones inteligentes utilicen en sus circuitos de procesamiento ciertas distinciones basadas en categorías prohibidas (categorías algorítmicas sospechosas) para efectuar rankings, scoring o categorizaciones violatorias del debido proceso y de los derechos humanos.

Javier ERCILLA⁴² distingue entre sesgos intencionales o conscientes, que buscan un objetivo concreto; y sesgos fortuitos que son aquellos que se introducen de manera involuntaria en el propio algoritmo, o que se va originando a consecuencia del aprendizaje máquina no supervisado con que el mismo ha sido configurado para ir mejorando en sus respuestas.

Deben ser adoptadas garantías que impidan que **sesgos intencionales** accedan de forma ilegítima a la aplicación de IA al proceso judicial, mediante la utilización de instrumentos de auditoría pública y la introducción de elementos de transparencia (que se examinan más adelante). Y los **sesgos fortuitos** han de ser evitados en la fase de toma de datos, que deben proceder de fuentes de datos adecuadas y dotadas de fiabilidad (calidad del dato); así como en la fase de diseño y programación del algoritmo, y en el ulterior entrenamiento del sistema.

8º.- Transparencia

En un primer acercamiento, el Diccionario de la Real Academia de la Lengua Española define una de las acepciones de «transparente» como «*claro, evidente, que se comprende sin duda o ambigüedad*». La transparencia resulta un elemento nuclear de la actuación del sector público, también del sistema de justicia.

⁴⁰ Ricard MARTÍNEZ MARTÍNEZ, «Inteligencia artificial, derecho y derechos fundamentales», Capítulo 11 de la obra «Sociedad Digital y Derecho», editada por BOE, el Ministerio de Industria, Comercio y Turismo y Red.es, 2018, página 269.

⁴¹ Concepto de Juan Gustavo CORVALÁN, «L'algorithme et les droits de l'homme», en Conseil d'État, Étude annuelle 2017. Puissance publique et plateformes numériques: accompagner l'ubérisation»; páginas 179 y ss. Disponible en: <https://www.dgdr.cnrs.fr/daj/archiv-actus/2018/Docs/174000714.pdf> .

⁴² Javier ERCILLA GARCÍA, «La inteligencia artificial en la Justicia. Asistentes IA judiciales y Jueces IA», *Revista Aranzadi de Derecho y Nuevas Tecnologías*, n.º 53, mayo-agosto 2020, página 105.

8.1.- Elementos de la transparencia de los sistemas IA

La solución IA ha de ser transparente en sus decisiones, lo que resulta especialmente relevante en la justicia. La transparencia, si bien cuenta con el mismo fundamento que el resto de las actuaciones de poderes públicos, en su aplicación a la justicia presenta una mayor intensidad derivada de las especificidades propias de las actuaciones judiciales. De esta manera, deberían ser exigibles los siguientes elementos en el diseño y aplicación de soluciones IA en la justicia⁴³:

8.1.1.- Explicabilidad

Que se pueda inferir o deducir una “explicación comprensible” sobre los criterios que fundamentan la obtención del resultado. Ello está directamente ligado al adecuado respeto al derecho a la tutela judicial efectiva y al deber de motivación de las resoluciones judiciales; lo que permite el control de la actividad jurisdiccional por parte de la sociedad (opinión pública), contribuye al convencimiento de las partes sobre la corrección de la decisión (eliminando arbitrariedad) y posibilita un adecuado ejercicio del derecho a los recursos (revisión por tribunal en segunda instancia y en casación).

La utilización de sistemas IA ha de resultar compatible con el derecho a la motivación de las resoluciones jurisdiccionales (integrado en el contenido material del derecho a la tutela judicial efectiva que garantiza el art. 24.1 de la CE): el mismo difícilmente puede satisfacerse con una motivación automatizada en la que el algoritmo sustituye al razonamiento judicial⁴⁴; lo que resulta aún más grave cuando se trata de resoluciones que tienen consecuencias sobre los derechos fundamentales donde la exigencia de justificación del resultado (motivación de la decisión) se encuentra reforzada.

Y también puede ser contraria al derecho de defensa. Como recuerda MARCHENA GÓMEZ⁴⁵, la falta de transparencia algorítmica se presenta así como una quiebra indiscutible del derecho de defensa, ya que la falta de conocimiento sobre los parámetros que conforman el algoritmo, así como sobre sus resultados intermedios, impedirán que alguien pueda defenderse de una conclusión que no puede conocer en sus fundamentos.

8.1.2.- Trazabilidad

Que se puedan explicar, paso a paso, las operaciones técnicas que la solución IA realiza desde el inicio hasta el fin de un proceso determinado. De esta manera, es importante registrar y documentar tanto las decisiones tomadas por los sistemas como la totalidad del proceso (incluida una descripción de la recogida y el etiquetado de datos, y una descripción del algoritmo utilizado) que dio lugar a las decisiones.

8.1.3.- Identificabilidad

Los sistemas de IA deben ser identificables como tales, garantizando que los usuarios sepan que están interactuando con un sistema de IA y qué personas son responsables del

⁴³ Juan Gustavo CORVALÁN, «L’algorithme et les droits de l’homme», en Conseil d’État, Étude annuelle 2017. Puissance publique et plateformes numériques: accompagner l’ubérisation»; páginas 179 y ss. Disponible en: <https://www.dgdr.cnrs.fr/daj/archiv-actus/2018/Docs/174000714.pdf> .

⁴⁴ Manuel MARCHENA GÓMEZ, ‘Inteligencia artificial y jurisdicción penal’, ponencia con motivo de su ingreso, como académico de número, en la Real Academia de Doctores el 26 de octubre de 2022.

⁴⁵ Manuel MARCHENA GÓMEZ, ‘Inteligencia artificial y ...’

mismo. Y es importante **comunicar** adecuadamente las capacidades y limitaciones del sistema de IA a las distintas partes interesadas afectadas de una manera adecuada al caso de que se trate⁴⁶.

8.1.4.- Acceso algorítmico

Cuando se utiliza una aplicación IA en un sector de la **actividad** de las entidades públicas, se ha de plantear el derecho de acceso a la información algorítmica, es decir, el acceso al sistema de procesamiento de información usado por la solución IA⁴⁷. Nos encontramos con un problema relevante: la propiedad intelectual (desarrollo del sistema IA por empresas con ánimo de lucro) frente a la total transparencia técnica (conocimiento de algoritmos completos y del código fuente).

Resulta necesario encontrar un equilibrio entre la propiedad intelectual de determinados métodos de tratamiento y los requisitos de transparencia (acceso al procedimiento diseño), neutralidad (ausencia de sesgo), la lealtad y la integridad intelectual (dando prioridad a los intereses de la justicia), debido a la legalidad o el impacto que estos métodos pueden tener en las personas⁴⁸. Como recuerda expresamente la Carta Ética Europea para la utilización de la Inteligencia Artificial en los Sistemas Judiciales, la transparencia técnica total (por ejemplo, open source del código y de la documentación) constituiría una primera posibilidad, a veces limitada por la protección de los secretos industriales. Y añade que el sistema también podría explicarse en un lenguaje claro y vulgarizado para describir la forma en que produce sus resultados, comunicando, por ejemplo, la naturaleza de los servicios ofrecidos, las herramientas desarrolladas, los rendimientos y los riesgos de error. Autoridades o expertos independientes podrían encargarse de la certificación y de auditar el procesamiento, o de suministrar asesoramiento.

¿Realmente es necesaria la total transparencia técnica? Es interesante la propuesta contenida en el estudio del MSI-NET del Consejo de Europa titulado “Algoritmos y derechos humanos”⁴⁹: *“La solución consistente en exigir la divulgación pública de algoritmos completos o de su código fuente es utópica en este contexto, en que las empresas privadas consideran sus algoritmos como softwares propietarios estratégicos y, por lo tanto, protegiéndolos en consecuencia. Sin embargo, parece posible exigir la publicación de información parcial, pero no obstante importante, como variables utilizadas, los objetivos perseguidos por la optimización de los algoritmos, los conjuntos de datos de aprendizaje, los valores medios y desviaciones estándar de los resultados obtenidos, o la cantidad y el tipo de datos procesados por el algoritmo”*.

⁴⁶ Comunicación de la Comisión Europea “Generar confianza en la inteligencia artificial centrada en el ser humano”, COM/2019/168

⁴⁷ Juan Gustavo CORVALÁN, «L’algorithme et les droits de l’homme», en Conseil d’État, Étude annuelle 2017. Puissance publique et plateformes numériques: accompagner l’ubérisation»; páginas 179 y ss. Disponible en: <https://www.dgdr.cnrs.fr/daj/archiv-actus/2018/Docs/174000714.pdf>.

⁴⁸ Carta Ética Europea para la utilización de la Inteligencia Artificial en los Sistemas Judiciales, CPEJ-Consejo de Europa; <https://rm.coe.int/charte-ethique-fr-pour-publication-4-decembre-2018/16808f699b>

⁴⁹ COMMITTEE OF EXPERTS ON INTERNET INTERMEDIARIES (MSI-NET) del Consejo de Europa, “Algorithmes et droits humains. Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires”, Préparée par le comité d’experts sur les intermédiaires d’internet (MSI-NET) del Consejo de Europa, 2017), página 41; <https://rm.coe.int/algorithms-and-human-rights-study-on-the-human-rights-dimension-of-aut/1680796d10>

8.2.- Sistemas de caja negra

En la justicia no cabe admitir los **sistemas de caja negra**, porque no pueden ofrecer una explicación suficiente de cómo la aplicación de la solución IA llega a un determinado resultado.

Se puede argumentar que la decisión automatizada puede ser sometida a la decisión de un juez humano en vía de recurso o revisión a instancia de la persona afectada. Sin embargo, ¿qué posibilidades efectivas de revisión tiene dicha persona frente a decisiones adoptadas por sistemas de aprendizaje automático y/o de caja negra?. En todo caso, esta posibilidad de revisión humana efectiva debería quedar totalmente garantizada; pero abre interrogantes relevantes en materia de forma de impugnación, de medios probatorios, de reglas de distribución de la carga de la prueba⁵⁰... Como recuerda el «Libro Blanco sobre la inteligencia artificial, un enfoque europeo orientado a la excelencia y la confianza» de la Comisión Europea⁵¹, este tipo de decisiones mediante IA *“aportan obstáculos al acceso efectivo a la justicia frente a situaciones en las que una decisión resultante de aplicación de tecnología IA afecte negativamente a la persona afectada”*.

9º.- Confianza (fiabilidad)

Cualquier sistema sustentado en aplicaciones de IA debe ofrecer el máximo nivel de confianza posible, no solamente para despejar estas dudas, sino para que resulte realmente útil y justo⁵², lo que resulta especialmente relevante en la aplicación de la IA en la decisión judicial.

9.1.- Fiabilidad

Los algoritmos aplicados en el sistema de justicia han de ser suficientemente seguros, fiables y sólidos para resolver errores o incoherencias durante todas las fases del ciclo vital del sistema de IA y hacer frente adecuadamente a los resultados erróneos⁵³:

- **Exactitud.** Sus decisiones deben ser acertadas o, como mínimo, reflejar su nivel de acierto, y sus resultados
- **Ausencia de sesgos.** Se ha examinado en otro lugar.
- **Seguridad.** Los sistemas de IA deben ser lo bastante seguros para ser resilientes, tanto frente a los ataques abiertos como a tentativas más sutiles de manipular datos o los propios algoritmos, y deben garantizar un plan de contingencia en caso de problemas. Asimismo, deben integrar mecanismos de seguridad desde el diseño para garantizar que sean verificablemente seguros en cada fase.

⁵⁰ Vid. Vanessa GARCÍA HERRERA, “Panorama legislativo de la inteligencia artificial en la Unión Europea”, en “La robótica y la inteligencia artificial en la nueva era de la revolución industrial 4.0 (Los desafíos jurídicos, éticos y tecnológicos de los robots inteligentes)”, editorial Dykinson, 2021

⁵¹ COM(2020) 65 de 19 de febrero de 2020.

⁵² Ernest SOLÉ y Susana DOMINGO, “Los siete retos de la inteligencia artificial en el entorno empresarial”, Business Review (Núm. 316) · TIC · Noviembre 2021; <https://www.harvard-deusto.com/los-siete-retos-de-la-inteligencia-artificial-en-el-entorno-empresarial>

⁵³ Comisión Europea COM(2019) 168.

- **Calidad.** Deben utilizarse fuentes certificadas y datos intangibles con modelos diseñados de forma multidisciplinaria; lo que se desarrolla en el punto siguiente

9.2- Calidad de los datos

La Comisión Europea⁵⁴ considera que la calidad de los conjuntos de datos utilizados es primordial para el funcionamiento de los sistemas de IA. Cuando se recopilan datos, pueden reflejar sesgos sociales, o contener inexactitudes o errores. Esto debe resolverse antes de entrenar un sistema de IA con un conjunto de datos. Además, debe garantizarse la **integridad** de los datos. Los procesos y conjuntos de datos utilizados deben ponerse a prueba y documentarse en cada fase, como la planificación, el entrenamiento, el ensayo y el despliegue. Esto debe aplicarse también a los sistemas de IA que no han sido desarrollados internamente, sino que se han adquirido fuera. Por último, el **acceso** a los datos debe estar adecuadamente regulado y controlado.

Los datos sobre los que trabaje la IA suponen un presupuesto básico para su éxito, de tal manera que ha de contar con un **volumen de datos suficiente** y con un **nivel óptimo de calidad**. Como recuerda BARTHE⁵⁵, el valor creado por la inteligencia artificial proviene de los datos necesarios para aprender mucho más que del algoritmo, que se desarrolla en código abierto, de tal manera que «los datos son más importantes que el *software*».

De esta manera, puede afirmarse que la **calidad de los datos** es una condición previa fundamental para todos los proyectos de IA. La OCDE⁵⁶ recuerda que se ha observado que la importancia de los datos de calidad (o la falta de ella) se plantea a menudo como el aspecto más importante que contribuye al éxito o fracaso de una iniciativa AI; y que algunos incluso han ido tan lejos como para decir que la mayoría de los gobiernos simplemente no están listos para la IA, y que primero deben centrarse en conseguir sus datos con el fin.

El sistema de justicia, al igual que ocurre con otros sectores públicos, no se encuentra configurado para manejar y aprovechar un gran volumen y variedad de datos; y tiene una comprensión de sus activos de datos (los datos que poseen y la infraestructura que los contiene) poco adecuada; incluso no es fácil responder a preguntas básicas tales como cuántas bases de datos existen dentro de la organización judicial, qué información concreta contiene cada base de datos, o cómo los datos son recopilados. Este problema se agrava si tenemos en cuenta que el marco de gobernanza de los datos en el sistema judicial español no resulta óptimo, contando con un amplio margen de mejora. Téngase en cuenta que las organizaciones que no poseen las capacidades para comprender y administrar sus datos no pueden aprovechar la IA⁵⁷.

⁵⁴ COM(2019) 168.

⁵⁵ Emmanuel BARTHE, «Intelligence artificielle en droit : derrière la «hype», la réalité». Disponible en web: <http://www.precisement.org/blog/Intelligence-artificielle-en-droit-derriere-la-hype-la-realite.html>

⁵⁶ «Estado de la técnica en el uso de tecnologías emergentes en el sector público», *Documentos de Trabajo de la OCDE sobre Gobernanza Pública*, n.º 34.

⁵⁷ Julián TORRES SANTELI, Sabine GERDON, World Economic Forum, «5 desafíos para la adopción gubernamental de IA,» 16 de agosto de 2019; <https://www.weforum.org/agenda/2019/08/artificial-intelligence-government-public-sector/>

10º.- Responsabilidad y rendición de cuentas

Deben instaurarse mecanismos que garanticen la responsabilidad y la rendición de cuentas de los sistemas de IA implantados en la justicia, así como de sus resultados⁵⁸.

10.1.- Rendición de cuentas

La posibilidad de **auditar** los sistemas de IA es fundamental, puesto que la evaluación de los sistemas de IA por parte de auditores internos y externos, y la disponibilidad de los informes de evaluación, contribuye en gran medida a la fiabilidad de la tecnología. La posibilidad de realizar auditorías externas debe garantizarse especialmente en aplicaciones que afecten a los derechos fundamentales⁵⁹, como ocurre en gran parte de los casos del sistema de justicia.

Según la Comisión Europea, los potenciales impactos negativos de los sistemas de IA deben señalarse, evaluarse, documentarse y reducirse al mínimo. El uso de las evaluaciones de impacto facilita este proceso. Estas evaluaciones deben ser proporcionales a la magnitud de los riesgos que plantean los sistemas de IA. Los compromisos entre los requisitos —que a menudo son inevitables— deben abordarse de una manera racional y metodológica, y ser tenidos en cuenta.

10.2.- Responsabilidad

Cuando se produzcan efectos adversos injustos, deben estar previstos mecanismos accesibles que garanticen una reparación adecuada⁶⁰: ¿Cuál será el régimen de responsabilidad derivado de errores de la aplicación de soluciones de IA que generan daños materiales o sobre las personas?

4.- REFLEXIONES FINALES

1. Debemos tener un **debate serio** en la sociedad⁶¹, donde reflexionemos sobre las ventajas aportadas y los diferentes peligros derivados de la implantación de soluciones de IA en el sistema de justicia, cuyos resultados puedan guiar las políticas públicas judiciales en una norma con rango de ley.
2. Es necesario un **marco normativo sectorial** específico en la justicia, con los criterios propios del sector, que regule los presupuestos y condiciones de aplicación de la IA en el sistema de justicia; que establezca de forma adecuada un marco de **evaluación de riesgos** para ayudar a reducirlos y mitigarlos; que contemple un marco común para informar sobre los incidentes de IA; y que establezca las actividades relativas a **gestión**

⁵⁸ Comunicación de la Comisión Europea “Generar confianza en la inteligencia artificial centrada en el ser humano”, COM/2019/168

⁵⁹ Comunicación de la Comisión “Generar confianza...”

⁶⁰ Comunicación de la Comisión “Generar confianza...”

⁶¹ La Recomendación (2020)1 del Comité de Ministros del Consejo de Europa sobre las consecuencias de los sistemas algorítmicos para los derechos del hombre afirma lo siguiente: “*Los Estados deben participar y apoyar debates continuos, inclusivos, interdisciplinarios, informados y públicos para identificar áreas de servicios públicos que impiden el disfrute de los derechos humanos que no pueden ser determinadas, decididas u optimizadas por sistemas algorítmicos*”.

de riesgos, es decir, la mitigación, el cumplimiento y la aplicación a lo largo del ciclo de vida del sistema de IA⁶²

3. No se puede cerrar la puerta a toda aplicación de la IA a la justicia, porque no podemos prescindir de los grandes beneficios para la **calidad y la eficacia de la justicia**, mejorando sus tiempos de respuesta.
4. Pero tampoco podemos asumirla de forma indiscriminada y acrítica⁶³, porque puede presentar elementos de **deshumanización de la justicia**, afectando negativamente a derechos de las personas y garantías del proceso.
5. Debemos analizar el **balance riesgos/beneficios de la concreta solución tecnológica**, porque las aplicaciones de robotización e inteligencia artificial afectan de forma diferente a estos derechos y garantías
6. Cuando se formulen medidas de implantación tecnológica, es necesario introducir **enfoque basado en el riesgo**⁶⁴: valoración integral de riesgos/beneficios de cada solución IA, que afecte tanto al diseño como a la implantación, diferenciando distintos niveles de riesgo⁶⁵.

5.- ¿Y EN EL FUTURO?

Esta necesidad de humanización deberá mantenerse con gran rigor en un horizonte temporal en que las soluciones IA serán cada vez más desarrolladas, especialmente con motivo de la computación cuántica, de tal manera que los riesgos sobre los derechos de las personas serán cada vez mayores. Como señala HIGUERAS⁶⁶, el gran dilema que debería ocupar y preocupar a los seres humanos tiene que ver con la explosión de una inteligencia artificial que, tarde o temprano, llegará a superar las capacidades físicas, mentales y emocionales de las personas y que podría suponer que los ordenadores se convirtiesen en entidades superiores, que nos controlasen o que, simplemente, nos hicieran desaparecer, al dejar de ser imprescindibles.

⁶² "OCDE Framework for the classification of AI Systems", febrero 2022; <https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1649672455&id=id&accname=guest&checksum=A4FC753EE627FB38C2FB41AEDD7E8B39>

⁶³ El Parlamento Europeo resalta "*las consecuencias negativas potencialmente graves, particularmente en el ámbito de las actividades policiales y judiciales, que pueden derivarse de una confianza excesiva en la naturaleza aparentemente objetiva y científica de las herramientas de IA*"; apartado 15 de la Resolución de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales [2020/2016(INI)]

⁶⁴ Propuesta de Reglamento del Parlamento Europeo y del Consejo de 21 de abril de 2021, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial)

⁶⁵ Carta Ética Europea sobre utilización de la inteligencia artificial en los sistemas judiciales y su entorno, aprobada por el plenario de CEPEJ en diciembre 2018

⁶⁶ Juan Carlos HIGUERAS, "Cuando la máquina supere al hombre", Business Review (Núm. 316), TIC, noviembre 2021, <https://www.harvard-deusto.com/cuando-la-maquina-supere-al-hombre>