

**ÉTICA E INTELIGENCIA ARTIFICIAL:
UNA MIRADA DESDE EL PROCESO JURISDICCIONAL¹**

***ÉTICA E INTELIGÊNCIA ARTIFICIAL:
UM OLHAR A PARTIR DO PROCESSO JURISDICCIONAL***

Lorenzo M. Bujosa Vadell

Catedrático e Doutor em Direito Processual pela Universidade de Salamanca, Espanha. E-mail: lbujosa@usal.es.

RESUMO: No presente artigo, o autor aborda sobre o uso da inteligência artificial pelo processo jurisdiccional e trata das questões éticas envolvidas no tema, bem como a magnitude dos riscos que deverão ser enfrentados para aproveitar as virtudes dos avanços tecnológicos no processo, sem cair em violações dramáticas éticas de garantias consolidadas no ordenamento jurídico.

PALAVRAS-CHAVE: Inteligência artificial; ética; revolução tecnológica; segurança; privacidade; dignidade humana.

RESUMEN: En este artículo, el autor discute el uso de la inteligencia artificial por parte del proceso jurisdiccional y aborda las cuestiones éticas implicadas en el tema, así como la magnitud de los riesgos que se deben enfrentar para aprovechar las virtudes de los avances tecnológicos en el proceso, sin caer en dramáticas vulneraciones éticas de garantías consolidadas en el ordenamiento jurídico.

PALABRAS CLAVE: Inteligencia artificial; ética; revolución tecnológica; seguridad; privacidade; dignidade humana.

ABSTRACT: In this paper the author discusses the use of artificial intelligence by the jurisdictional process and deals with the ethical issues involved in the topic, as well as the

¹ Artigo recebido em 17/11/2021, sob dispensa de revisão.

magnitude of the risks that must be faced to take advantage of the virtues of technological advances in the process, without falling into dramatic ethics violations of guarantees consolidated in the legal system.

KEYWORDS: Artificial intelligence; ethics; technological revolution; security; privacy; human dignity.

1. INTRODUCCIÓN

Lejos de colocarnos en el ámbito de la ciencia ficción, cuando hablamos de inteligencia artificial nos referimos a los dilemas y a las dudas que se nos plantean en el presente en multitud de aspectos de la realidad que nos envuelve. Todos hemos visto en televisión cómo se realizan complejas operaciones quirúrgicas sin apenas presencia humana visible, cómo se realizan de manera automatizada cálculos sobre órbitas satelitales que se demostrarán milimétricamente correctos o, de manera más simple y cercana a nuestra actividad cotidiana como juristas, cómo en pocos segundos podemos tener a nuestra disposición la jurisprudencia reciente sobre determinada materia específica sobre la que tenemos que interponer una demanda.

La enorme inversión en determinados campos científicos produce continuamente avances que contribuyen notablemente a modificar en pocos años el contexto científico y tecnológico hasta llevarnos a cambios de paradigma² o, con mayor rotundidad, a nuevas revoluciones industriales³. Pero, en realidad, no estamos hablando de cambios repentinos sin fundamento previo. Hace muchos decenios que el desafortunado Alan M. TURING, a partir de los postulados de la lógica matemática, dictó una conferencia en la que planteaba algo tan actual como la cuestión de si puede pensar una máquina⁴. Las novedades están más bien en la generalización de la aplicación de las tecnologías digitales como consecuencia de su

² KUHN, T.S., *The Structure of Scientific Revolutions*, 50th Anniversary edition, Chicago-London, 2012, sistematizaba la idea de progreso a través de revoluciones científicas.

³ Son conocidas las consideraciones de SCHWAB, K., *La cuarta revolución industrial*, Barcelona, 2016, por las que, desde el punto de vista sociológico, habla de las consecuencias de la digitalización: “Las innovaciones tecnológicas más importantes están a punto de generar un cambio trascendental en todo el mundo, algo inevitable”.

⁴ TURING, A.M., “Computing Machinery and Intelligence”, *Mind*, vol. LIX, núm. 236, 1950, pp. 433-460; traducida al español como *¿Puede pensar una máquina?*, Oviedo, 2012.

continua transformación y perfeccionamiento. Su inmersión en nuestra vida diaria, por un lado, nos produce un inevitable efecto de embelesamiento y, por otro, una cierta reacción de temor y desconfianza.

Si nos limitamos a centrar nuestra observación al campo del Derecho, intuitivamente podemos enumerar interesantes ventajas en la aplicación de las tecnologías digitales, que aligeran el trabajo, aceleran las tareas, aumentan la exactitud de los resultados -si es que en Derecho se nos permite hablar en estos términos-, ... Pero, también *a priori* se nos plantean dudas ante la posible “deshumanización” del trabajo de los juristas, que nos aleje de las valoraciones de equidad en el caso concreto⁵ o que multiplique los efectos de sesgos muy discutibles⁶.

Desde la perspectiva procesal nos interesan todas las vías que puedan sustraer las decisiones del juzgador de la excesiva subjetivización y, por consiguiente, de la irracionalidad y la arbitrariedad. Pero, por el camino de la objetivación a la que nos conducen las aplicaciones digitales, podemos llevarnos por delante consideraciones específicas, no tenidas en cuenta por los programadores y que son ni más ni menos que los nexos de enlace entre la actividad jurisdiccional y el valor superior de la justicia.

En todo ello, por supuesto, es imprescindible atender a las cuestiones éticas implicadas, que han sido puestas en primera línea por diferentes organismos que reflexionan sobre la aplicación de la inteligencia artificial en el proceso; lo cual confirma la magnitud de los riesgos a los que tenemos que hacer frente, tratando de transitar por el justo medio:

⁵ Recordemos como ARISTÓTELES, *Ética a Nicómaco*, Madrid, 2008, afirmaba: “... cuando la ley presenta un caso universal y sobrevienen circunstancias que quedan fuera de la fórmula universal, entonces está bien, en la medida en que el legislador omite y yerra al simplificar, el que se corrija esta omisión, pues el mismo legislador habría hecho esta corrección si hubiera estado presente y habría legislado así si lo hubiera conocido. (...). Y tal es la naturaleza de lo equitativo: una corrección de la ley en la medida en que su universalidad la deja incompleta” (5:10).

⁶ Es conocido el caso de COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), un algoritmo utilizado en la justicia penal para predecir la probabilidad de reincidencia a partir de unos cien factores. *Vid.* la sentencia del Tribunal Supremo del Estado de Wisconsin: *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016), en la que se deniega la pretensión de Loomis respecto a la vulneración del derecho al proceso con todas las garantías (*due process fo law*), fundada, entre otras razones, en que la decisión judicial condenatoria se basaba en un instrumento de apreciación del riesgo en que se tenía en cuenta el género y la raza y que no era susceptible de ser discutido por tratarse de un secreto mercantil. A su vez el Tribunal Supremo de los Estados Unidos denegó el *writ of certiorari* el 26 de junio de 2017. *Vid.* también “Would You Trust An Artificially-Intelligent Expert?”, *The National Law Review* July 11, 2020, Volume X, Number 193: <https://www.natlawreview.com/article/would-you-trust-artificially-intelligent-expert> Desde un punto de vista más crítico *vid.* TASHEA, J., “Courts Are Using AI to Sentence Criminals. That Must Stop Now”, en <https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/>.

aprovechando en el proceso las virtudes de los avances tecnológicos y evitando caer en dramáticas vulneraciones de garantías consolidadas en nuestros ordenamientos.

2. LA GRAN INCÓGNITA DE LA INTELIGENCIA ARTIFICIAL

A pesar de la frecuencia con que en los medios generales oímos hablar de “inteligencia artificial”, para los profanos se trata aún de una expresión difícil de definir. Se trata de un concepto complejo por la amplitud y la heterogeneidad de los elementos que contiene y, además, por ser inequívocamente inestable, por su constante movilidad y evolución. Se trata, tal vez, del mejor ejemplo actual de *work in progress* del que ya es obligado valorar sus variadas consecuencias jurídicas⁷, pero del que todavía no sabemos bien adónde nos va a llevar, con lo cual nos queda el campo abierto a múltiples elucubraciones que, estas sí, pueden entrar en muchos casos aún en el ámbito de lo irreal o de la ciencia ficción.

En efecto, se habla de la existencia de oleadas o generaciones de inteligencia artificial, según la evolución de los instrumentos a los que se refiere y, sobre todo, de las posibilidades de actuación respecto a la gestión de una multitud de datos, sustituyendo con ventaja a la inteligencia del ser humano⁸. Se habla incluso de “superinteligencia”⁹, en lo que es una continua comparación con el objetivo de superación de las capacidades del cerebro biológico humano¹⁰. Nos encontramos aquí con la gran paradoja de que se pretende superar algo que todavía en buena parte constituye un misterio para nosotros, a pesar de los grandes avances de la neurociencia¹¹. Parece que nos situamos en la senda histórico-cultural del

⁷ SUSSKIND, R., *Online Courts and the Future of the Justice*, Oxford, 2019, p. 263, lo expresa de manera llamativa: “Scarcely a week passes without news of ‘an AI’ or a ‘robot lawyer’ that is outperforming or poised to replace traditional human lawyers in some legal task or other”.

⁸ Cfr. SUSSKIND, R., *Online Courts...*, *op. cit.*, pp. 264-275.

⁹ BOSTROM, N., *Superinteligencia. Caminos, peligros, estrategias*, 2016, p. 22, la define tentativamente como “cualquier intelecto que exceda en gran medida el desempeño cognitivo de los humanos en prácticamente todas las áreas de interés”.

¹⁰ BOSTROM, N., *Superinteligencia...*, *op. cit.*, p. 60: “En la actualidad, el poder computacional del cerebro biológico todavía se compara favorablemente con el de las computadoras digitales, aunque las supercomputadoras punteras están alcanzando niveles de rendimiento que están dentro del rango de estimaciones plausibles de potencia de procesamiento del cerebro. Pero el hardware está mejorando rápidamente, y los límites últimos de rendimiento para el hardware son muy superiores a los de los sustratos biológicos de computación”.

¹¹ *Vid.* COBB, M., *The Idea of the Brain. The past and the future of Neuroscience*, New York, 2020.

romanticismo inglés¹², cuando aparecen estudios que aluden específicamente a la “creación de una mente”¹³. Y no olvidemos que tanto el concepto de mente, como el de inteligencia, son también conceptos elusivos, complejos y con discutidas delimitaciones¹⁴.

No es nuestra intención terciar en polémicas especializadas para las que distamos de estar preparados, pero sí va a ser necesario acoger una definición de inteligencia artificial, que aún con sus defectos e inseguridades, nos permita seguir adelante hacia lo que más nos interesa, que son las consecuencias jurídico-procesales de esta nueva realidad, y sobre todo, como veremos enseguida, los límites éticos de esas consecuencias.

Así pues, acogemos la definición amplia y simple, por lo menos en apariencia, que se ha sostenido en la Unión Europea, por la cual se afirma que “El término “inteligencia artificial” (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción –con cierto grado de autonomía– con el fin de alcanzar objetivos específicos”¹⁵. Destaca en esa perspectiva una preocupación principal que nos atañe directamente: la de generar confianza en la inteligencia artificial centrada en el ser humano.

En tiempos en que la subjetividad jurídica se ha ampliado notablemente, y así se ha reconocido capacidad para ser parte en un proceso no sólo a los patrimonios separados, al río Atrato en Colombia, a la mismísima naturaleza o a la Madre Tierra en otras partes de América Latina, se nos plantea si un robot, dotado de inteligencia artificial¹⁶, podrá convertirse algún día en sujeto procesal también, e incluso si lo tendremos que contar algún

¹² Es inevitable aquí el recuerdo de FRANKSTEIN, de Mary SHELLEY, publicado el 1 de marzo de 1818, exactamente un año y medio después del famoso año en que no hubo verano.

¹³ Valgan los ejemplos de KURZWEIL, R., *Cómo crear una mente. El secreto del pensamiento humano*, Berlin, 2013, o en sentido crítico, FODOR, J., *La mente no funciona así. Alcance y límites de la psicología computacional*, Madrid, 2003. Es obligado también citar aquí a PENROSE, R., *La nueva mente del emperador*, Madrid, 2019, publicado originalmente en 1989.

¹⁴ OLIVEIRA, A., *The Digital Mind. How Science is Redefining Humanity*, London, 2017, p. 88: “... intelligence is a more elusive concept than had once been thought. Though it is commonly accepted that intelligence is required in order for a human to address any of the problems mentioned (...), it isn’t at all clear that the techniques computers used to solve those problems endowed them with general human-like intelligence. In fact, those problems were tackled with specialized approaches that were, in general, very different from the approaches used by humans”. Se refiere a la resolución de complicados teoremas matemáticos, diseñar correctamente movimientos de ajedrez, planificar actuaciones, etc.

¹⁵ *Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia artificial para Europa*. Bruselas, 25 de abril de 2018 COM (2018) 237, final.

¹⁶ Vid. LACRUZ MANTECÓN, M.L. *Robots y personas. Una aproximación jurídica a la subjetividad cibernética*, Madrid, 2020. De una manera interdisciplinaria, BARRIO ANDRÉS, M., (Dir.), *Derecho de los Robots*, Madrid, 2018 y ROGEL VIDE, C., (Coord.), *Los robots y el Derecho*, Madrid, 2018.

día entre el personal jurisdisciente, capaz de dictar resoluciones judiciales que satisfagan pretensiones de manera motivada¹⁷. Desde luego no cabe confundir la robótica con la inteligencia artificial, pues todos tenemos conocidos mecanismos que funcionan de manera automatizada -por ejemplo, los robots de cocina-, pero a algunos de ellos se les pueden aplicar algoritmos propios de la inteligencia artificial, con lo que ya también entrarían indiscutiblemente en nuestro concepto¹⁸.

El Parlamento Europeo se ha pronunciado en diversas ocasiones acerca de esa conjunción entre inteligencia artificial y robótica, partiendo de la idea de que “la inteligencia artificial (IA) y una robótica transparentes y que integren consideraciones éticas tienen el potencial necesario para enriquecer nuestras vidas y consolidar nuestras capacidades, tanto en el plano individual como para el bien común”, pero que al mismo tiempo son susceptibles de un uso malintencionado que vulnere o, por lo menos, ponga en riesgo los derechos fundamentales y, para ello pide una evaluación periódica de la legislación con el fin de garantizar que sea adecuada para su propósito en relación con la inteligencia artificial y que respete al mismo tiempo los valores fundamentales de la Unión¹⁹.

No está de más recordar -y sorprendernos una vez más, por su agudeza- que Isaac ASIMOV, en una fecha tan temprana como 1942, ya se preocupó por la estrecha relación entre la inteligencia artificial y las directrices éticas infranqueables, al formular sus Leyes de la Robótica:

1. Un robot no hará daño a un ser humano o, por inacción, permitirá que un ser humano sufra daño.
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entrasen en conflicto con la primera ley.

¹⁷ Vid. BARONA VILAR, S., *Algoritmización del Derecho y de la Justicia. De la inteligencia Artificial a la Smart Justice*, Valencia, 2021, pp. 610-664.

¹⁸ Como dice SALAZAR, I., *La revolución de los robots. Cómo la inteligencia artificial y la robótica afectan a nuestro futuro*, Gijón, 2019, p. 51, “la revolución de los robots, de una era tecnológica sin precedentes, de un mundo completamente interconectado, está sucediendo en un tiempo record”.

¹⁹ Cfr. *Resolución del Parlamento Europeo*, de 12 de febrero de 2019, *sobre una política industrial global europea en materia de inteligencia artificial y robótica* (2018/2088(INI)), y, con anterioridad, la *Resolución del Parlamento Europeo*, de 16 de febrero de 2017, *con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica* (2015/2103(INL)).

3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley²⁰.

La ética, por tanto, ya tenía un protagonismo primordial en estos primeros ejercicios de elucubración sobre las posibilidades de las innovaciones tecnológicas.

Como veremos más adelante, la centralidad del ser humano seguirá siendo el criterio fundamental frente a los avances en la aplicación de la inteligencia artificial en el proceso. Por tanto, las grandes y pasmosas ventajas que ya observamos en las distintas perspectivas procesales no deben obnubilar nuestra mente crítica y por tanto no pueden alterar la esencia de lo que consideramos fundamental en la función jurisdiccional y en el ejercicio de los derechos de los justiciables.

3. EL PROCESO ANTE LOS DILEMAS DE LA INTELIGENCIA ARTIFICIAL

Por supuesto no es nueva la aplicación de las tecnologías de la información y la comunicación al proceso con finalidades muy variadas, todas ellas dirigidas a facilitar la actividad procesal. Desde la reforma de 1994 la Ley Orgánica del Poder Judicial aludió, por vez primera en nuestro ordenamiento jurídico, a la posibilidad de utilizar medios técnicos, electrónicos e informáticos para el desarrollo de la actividad y el ejercicio de las funciones de juzgados y tribunales. Pero la verdad es que se tardó mucho en proceder a una digitalización de la actividad de la administración de la justicia: hasta 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, no se dio un paso amplio y significativo a este respecto.

Aunque en todo ello pueden surgir amenazas para la integridad de los derechos fundamentales, puede afirmarse que en la previsión de ninguna de las aplicaciones tecnológicas previstas por la legislación en diversos ámbitos (videoconferencias, expediente digital, medidas de investigación tecnológica...) hay un sistema de inteligencia artificial “fuerte”, en el sentido de que haya “máquinas con conciencia propia”²¹, ni mucho menos

²⁰ ASIMOV, I., “Círculo vicioso”, publicado con el título *Runaround*, en la revista *Astounding Science Fiction*, en marzo de 1942.

²¹ SUSSKIND, R., *Online Courts...*, *op. cit.*, p. 265: “The term ‘intelligence’, for many observers suggests perhaps that the latest systems are in some sense actually ‘conscious’. In the philosophical jargon of AI, a system that is conscious would be an exhibit of what is known as ‘strong AI’”.

con capacidad de decisión autónoma. Es usual hablar de distintos grados de inteligencia artificial: fuerte, débil, específica, general, como es propio de un concepto amplio y plural. Esta vaguedad nos somete a retos continuos, que debemos tratar de superar.

Así, por el momento parece descartable la total sustitución del ser humano por la máquina pensante, en una aplicación de la inteligencia artificial fuerte y general, con capacidad de aplicar conocimientos generales a las tareas que se le planteen. De este modo nos situaríamos ante el fascinante debate sobre el aprendizaje de las máquinas y de adoptar decisiones conforme a ese aprendizaje -aunque, obviamente, ello dependerá de lo que entendamos por aprendizaje-²². Pero parece indiscutible que está ya entre nosotros la inteligencia artificial débil, basada no en el reconocimiento de patrones independientes que, a su vez son integrados, sino en el aprendizaje automático, con el procesamiento de gran cantidad de datos y con verdaderos razonamientos deductivos a partir de aprendizajes inductivos²³.

Como consecuencia de todo ello SUSSKIND nos habla de tecnologías disruptivas para el abogado del mañana, en el sentido de que desafían y cambian fundamentalmente el funcionamiento de un sector determinado²⁴. En nuestro ámbito podemos decir que hay una evolución en el acceso a la justicia²⁵ que cambia de medio, por lo menos parcialmente: del presencial y físico al virtual y electrónico, supone muchas más facilidades para obtener información, e incluso podemos aventurarnos a afirmar que ciertas decisiones basadas en criterios objetivos son fácilmente susceptibles de automatizarse²⁶. Posiblemente un ejemplo claro puedan ser las decisiones relativas al procedimiento monitorio.

²² LÓPEZ MORATALLA, N., *Inteligencia artificial ¿Conciencia artificial?*, Madrid, 2017, p. 52: “Las máquinas construidas con la llamada tecnología cognitiva, simulando el cerebro humano, podrían tener un Aprendizaje Automático o un Aprendizaje Profundo. De manera simplificada se puede decir que, en el primer caso, simulan el funcionamiento de la neurona, trabajando en una capa o varias capas. En el segundo, simularían las redes neuronales complejas”.

²³ Cfr. LÓPEZ MORATALLA, N., *Inteligencia artificial...*, *op. cit.*, p. 74.

²⁴ SUSSKIND, R., *El abogado del mañana. Una introducción a tu futuro*, 2.ª ed., Madrid, 2020, p. 77.

²⁵ Vid. SOLAR CAYÓN, J.I., *La inteligencia artificial jurídica. El impacto de la innovación tecnológica en la práctica del Derecho y el mercado de servicios jurídicos*, Cizur Menor, 2019, pp. 306-313.

²⁶ Justamente el artículo 42 de la ya mencionada Ley 18/2011 se refiere a las actuaciones judiciales automatizadas: “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. 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”. NIEVA FENOLL, J., *Inteligencia artificial y proceso judicial*, Madrid, 2018, p.

Salvo estos casos excepcionales de automatización, parece que es más realista limitarnos a los casos en que la inteligencia artificial se utiliza como instrumento auxiliar para el juzgador²⁷, aunque no es razonable desdeñar los instrumentos que permiten una gran predictibilidad en las decisiones judiciales²⁸, con los problemas que ello lleva consigo, como son los sesgos implícitos tanto en la configuración del algoritmo como en los criterios de selección de los datos que se introducen para la decisión. A los sesgos inadvertidos que caracterizan el razonamiento humano y que nos han sido determinados por algunos prestigiosos psicólogos como relevantes también para las decisiones judiciales²⁹, se suma aquí un factor de multiplicación por la influencia de los ingenieros informáticos, matemáticos, etc., que al diseñar el algoritmo y organización la gestión masiva de los datos introducen también, consciente o inconscientemente, patrones que en absoluto son neutrales³⁰.

Ya SARTOR y BRANTING en 1998 planteaban un análisis de las aplicaciones judiciales de la inteligencia artificial³¹ y apuntaban una tendencia estable a la expansión hacia áreas más complejas que implican importantes desafíos, avanzando hacia una mayor influencia en la decisión judicial. Se partía del examen de la confección automática de documentos judiciales complementarios, para pasar a observar la creación de un sistema de apoyo y suministro de información para una adecuada adopción de resoluciones, la aplicación de sistemas de inteligencia artificial que puedan ayudar a regular la

33-41, valora lo que denomina las “decisiones automatizables”, atendiendo a “los aspectos más obvios, que de hecho en algunos lugares ya son objeto de automatización”.

²⁷ SUSSKIND, R., *Online Courts...*, *op. cit.*, p. 265: “instead, my attention here is on ‘weak AI -functionally, these systems seem to be doing some of the work of lawyers but without the cognitive states enjoyed by humans, such as our self-awareness and emotional satisfaction”.

²⁸ Para una perspectiva histórica de estos avances, *vid.*, PALIWALA, A. (Ed.), *A history of legal informatics*, Zaragoza, 2010. Una perspectiva actual en SAN MIGUEL CASO, C., “Las técnicas de predicción judicial y su repercusión en el proceso”, CONDE FUENTES, J., y SERRANO HOYO, G., (Dir.), *La justicia digital en España y la Unión Europea*, Barcelona, 2019.

²⁹ TVERSKY, A.; KAHNEMAN, D., “El juicio bajo incertidumbre: heurísticas y sesgos”, publicado originalmente en inglés en *Science*, vol. 185, 1974, y en español, en KAHNEMAN, D., *Pensar rápido, pensar despacio*, 3.ª ed., Barcelona, 2012, pp. 545- 567.

³⁰ *Vid.* ARMENTA DÉU, T., *Derivas de la justicia. Tutela de los derechos y solución de controversias en tiempos de cambios*, Madrid, 2021, pp. 228-238: “La elección de los ‘algoritmos de recomendación y de clasificación’ son cruciales en la medida en que modifican masivamente la influencia de unos individuos sobre otros”.

³¹ *Vid.* SARTOR, G., ad BRANTING, K., (Eds.), *Judicial Applications of Artificial Intelligence*, Dordrecht, 2010.

discrecionalidad judicial, la adecuada selección y aplicación de la doctrina de los precedentes y, en definitiva, el apoyo informático para llegar a la “mejor” decisión judicial³².

Por otra parte, la existencia de programas informáticos que introduzcan los más avanzados sistemas de inteligencia artificial no nos debe llevar al error de pensar que van a ser directamente aplicados en la actividad jurisdiccional cotidiana. Es más, existe una queja fundada acerca de la falta de utilización de los sistemas ya disponibles en la gestión judicial diaria³³, lo cual no necesariamente es una consecuencia de la brecha digital, sino más bien de la tendencia conservadora del trabajo en los juzgados y tribunales que también debe ser considerada, si queremos adoptar una perspectiva realista.

4. LA RELEVANCIA INEVITABLE DE LAS CUESTIONES ÉTICAS

La profunda alteración que provoca la aplicación, en alguna medida, de la inteligencia artificial a algún aspecto del proceso corre el riesgo de producirnos un efecto de espejismo o embelesamiento, como consecuencia del cual es posible que minusvaloremos las consecuencias negativas y las inseguridades en las que podemos incurrir. Por ello se insiste, como ya avanzamos, en situar al ser humano en el centro y en articular una serie de reglas éticas que deben respetarse para evitar que toda esta problemática se nos vaya de las manos.

En muchas de las aplicaciones informáticas la clave de su eficacia se encuentra en el manejo de una cantidad exorbitante de datos, bastantes de ellos cercanos a la privacidad³⁴,

³² Especial interés tiene el último capítulo, en el que TARUFFO, M., “Judicial decisions and Artificial Intelligence”, SARTOR, G., ad BRANTING, K., (Eds.), *Judicial Applications...*, *op. cit.*, pp. 207-220. El jurista italiano afirma (p. 212) que “If one considers the evident features of complexity, variability, flexibility and discretion that are typical of judicial decisions, any approach aimed at interpreting the judicial reasoning according to logical rules and models may appear as doomed to failure”. Y sigue (p. 214): “The so-called ‘easy cases’ often are not easy enough to be standardized in terms of computerized models. A *fortiori* such a standardization is extremely difficult or impossible to achieve in the ‘average’ judicial case, let alone in *hard cases*. It seems, therefore, that the type of computerization we are considering may be useful but only in a relative narrow are of judicial practice”.

³³ Cfr. SUSSKIND, R., *El abogado del mañana...*, *op. cit.*, p. 147, donde habla, justamente por esa causa, de “avances decepcionantes”.

³⁴ Vid. MAYER-SCHÖNBERGER, V., y CUKIER, K., *Big Data. La revolución de los datos masivos*, Madrid, 2013, y HOFFMANN-RIEM, W., *Big Data. Desafíos también para el Derecho*, Madrid, 2018.

lo cual ya nos sitúa ante unas cuestiones específicas desde el punto de vista constitucional³⁵. Así, recurriendo a la conocida metodología de las generaciones de derechos humanos, VASAK habló hace tiempo de la tercera generación³⁶, en la que PEREZ LUÑO situó las consecuencias de la revolución tecnológica, siendo consciente de que se trata de un fenómeno dialéctico, con constantes avances, retrocesos y contradicciones, lo cual implica, sin embargo, la certeza del reconocimiento sobre el carácter abierto del catálogo de libertades y de que “una sociedad libre y democrática debe mostrarse siempre sensible y abierta a la aparición de nuevas necesidades, que fundamenten nuevos derechos”³⁷.

En esta línea se entronca la Carta de Derechos Digitales presentada este mismo año por el Gobierno español, con el fin “proteger los derechos de los ciudadanos y las ciudadanas en la nueva era de Internet y la Inteligencia Artificial”, en la que estos derechos presentan “vulnerabilidades nuevas y de extrema gravedad”. Se destaca el valor informador de esta declaración de derechos respecto al tratamiento normativo de estas materias en un futuro próximo, pues la Carta “pretende servir de guía para futuros proyectos legislativos y desarrollar políticas públicas más justas, que nos protejan a todos”³⁸.

Sin embargo, hay que tener en cuenta un paso previo, por el que se pretende ofrecer una serie de reflexiones morales, e incluso unos principios éticos, que sirvan de guía al legislador. Estamos, pues, en el ámbito de la ética aplicada³⁹, en que se trata de averiguar cómo pueden los principios éticos ayudar a orientar los distintos tipos de actividad. En efecto, la aplicación es redundante, porque toda ética, por definición, debería poder aplicarse, pero sí que supone entender la ética de una manera distinta, como nos dice Victoria CAMPS, “más atenta a las interpelaciones de nuestro mundo”⁴⁰. Dicho de otro modo, se trata del “giro

³⁵ GARCÍA COSTA, F. M., “Perfiles constitucionales de la justicia electrónica”, en GÓMEZ MANRESA, M.F., y FERNÁNDEZ SALMERÓN, M., *Modernización digital e innovación en la Administración de Justicia*, Cizur Menor, 2019, pp. 23-35.

³⁶ VASAK, K., *Pour les droits de l’homme de la troisième génération*, Strasbourg, 1979.

³⁷ PÉREZ LUÑO, A.E., *Los derechos humanos en la sociedad tecnológica*, Madrid, 2012, p. 38.

³⁸ *Vid.* https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf.

³⁹ Como dicen CORTINA, A., y MARTÍNEZ, E., *Ética*, 4.ª ed., Madrid, 2008, p. 151: “Entre las tareas de la Ética (...) no sólo figura la aclaración de lo que es la moralidad y la fundamentación de la misma, sino la aplicación de sus descubrimientos a los distintos ámbitos de la vida social”.

⁴⁰ CAMPS, V., *Breve historia de la ética*, Barcelona, 2013, pp. 392-406: “Ha contribuido a la evolución de la ética hacia terrenos más prácticos; por una parte, la complejidad y el desconcierto de un mundo que ya no tiene los puntos de apoyo que, en otro tiempo, proporcionaron la religión, ideologías fuertes o formas de organización social más homogéneas que las actuales. Por otra, el desarrollo extraordinario de la ciencia y la

aplicado” de la filosofía y sobre todo de la ética: “profesionales, ciudadanos y gobiernos exigen un mayor nivel ético en las distintas esferas sociales y sobre todo la institucionalización de ese nivel en comités, comisiones, *guidelines* y documentos bien perfilados, de suerte que es la realidad social misma la que exige a la filosofía moral comprometerse con la vida corriente”⁴¹.

La evidente heterogeneidad de las sociedades, sin entrar siquiera en multiculturalidades, no impide que debamos buscar en la racionalidad de la ética una intersubjetividad que haga posible la convivencia entre todas las diversas perspectivas acerca del bien y de lo bueno⁴², y en nuestro caso, de manera necesaria y urgente para delimitar los principios éticos aplicables a los avances tecnológicos más punteros, y más en concreto, ante las disrupciones que provoca la aplicación de instrumentos de inteligencia artificial. Por consiguiente, es fundamental tener en cuenta que “la tarea de la ética aplicada a una actividad concreta estará centrada en la deliberación sobre los medios que debemos utilizar para llegar a los fines de la actividad, es decir, sobre qué valores y virtudes es preciso cultivar para decir que se está practicando esta actividad de manera correcta a nivel moral, o sea, con vistas a alcanzar su finalidad”⁴³. Es importante, pues, plantear una ética de mínimos en este ámbito complejo y cambiante de la inteligencia artificial, también respecto a su aplicación en el ámbito de la administración de la justicia⁴⁴.

De este modo, deberemos volver a la Unión Europea, que ya va teniendo una larga tradición en lo que desde el Programa de Estocolmo se viene conociendo, en términos muy amplios, como “e-Justicia” o justicia electrónica⁴⁵. En los últimos años se ha destacado una

técnica han exacerbado la perplejidad y también el temor hacia una deriva que parece humanamente incontrolable”.

⁴¹ CORTINA, A., *Justicia cordial*, Madrid, 2010, pp. 41-53, capítulo que lleva el significativo título de “La Ética es de este mundo”.

⁴² RUIZ TRUJILLO, P., *Ética de las nanotecnologías*, Barcelona, 2020, p. 144.

⁴³ RUIZ TRUJILLO, P., *Ética de las nanotecnologías, op. cit.*, p. 157.

⁴⁴ CORTINA, A., *Ética mínima. Introducción a la filosofía práctica*. 15.ª ed., Madrid, 2010, p. 335: “... algunos éticos nos hemos refugiado humildemente en una ética de mínimos, y nos limitamos a decir a nuestros oyentes y lectores: al decidir las normas que en su sociedad van a regular la convivencia, tenga en cuenta los intereses de todos los afectados en pie de igualdad, y no se conforme con los pactos fácticos, que están previamente manipulados, y en los que no gozan todos del mismo nivel material y cultural ni de la misma información”.

⁴⁵ *Programa de Estocolmo – Una Europa abierta y segura que sirva y proteja al ciudadano* [Diario Oficial nº C 115 de 4 de mayo de 2010]. Vid. SENÉS MOTILLA, C., (Coord.), *Presente y futuro de la e-Justicia en España y la Unión Europea*, Cizur Menor, 2010, y DE LA OLIVA SANTOS, A.; GASCÓN INCHAUSTI, F.; AGUILERA MORALES; M., (Coords.), *La e-Justicia en la Unión Europea. Desarrollos en el Ámbito Europeo y en los Ordenamientos Nacionales*, Cizur Menor, 2012.

preocupación por la dimensión ética de la inteligencia artificial, como se señala en la Comunicación de la Comisión Europea, de 8 de abril de 2019, que lleva precisamente como encabezamiento el de “Generar confianza en la inteligencia artificial centrada en el ser humano”⁴⁶. En este texto, justamente, se destaca que “... la Inteligencia Artificial no es un fin en sí mismo, sino un medio que debe servir a las personas con el objetivo último de aumentar su bienestar. Para ello, la fiabilidad de la inteligencia artificial debe estar garantizada ...”.

4.1. Los principios ASILOMAR sobre inteligencia artificial

Desde la perspectiva ética, fuera del ámbito específico de la Unión Europea puede ser interesante citar los conocidos Principios ASILOMAR sobre inteligencia artificial, aprobados en el seno del *Future of Life Institute*, en la reunión de más de cien expertos de distintas disciplinas, que tuvo lugar en Pacific Grove (California), en enero de 2017⁴⁷. De una forma genérica se enumeran una larga serie de exigencias que deben cumplir los sistemas de inteligencia artificial. Desde el punto de vista del proceso nos interesan sobre todo los siguientes:

“Seguridad: Los sistemas de inteligencia artificial deberían ser seguros a lo largo de su vida operativa y verificables donde sea posible.

Transparencia en los fallos del sistema: Si un sistema de inteligencia artificial causa daño debería ser posible determinar la causa.

Transparencia en las decisiones: Cualquier intervención de un sistema autónomo en una decisión debería ir acompañada de una explicación satisfactoria y controlable por parte de una autoridad humana competente.

(...)

Respeto a los valores: Los sistemas de inteligencia artificial altamente autónomos deberían ser diseñados para que sus metas y comportamientos puedan alinearse con los valores humanos a lo largo de sus operaciones.

⁴⁶ Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Bruselas, 8.4.2019 COM(2019) 168, final.

⁴⁷ *Vid.* <https://futureoflife.org/ai-principles/>.

Valores humanos: Los sistemas de inteligencia artificial deberían ser diseñados y operados para que sean compatibles con los ideales de dignidad humana, derechos, libertades y diversidad cultural

Privacidad personal: La gente debería tener el derecho de acceder, gestionar y controlar los datos que se generan por la aplicación de sistemas de inteligencia artificial (...).”

Aunque todos estos criterios y requisitos se formulan de manera general, naturalmente son aplicables a las diversas modalidades que pueda adoptar la inteligencia artificial en la administración de la justicia. Es fundamental que tales instrumentos sean también fiables y que para ello se arbitren las necesarias cautelas respecto a la seguridad, transparencia, privacidad, no discriminación y dignidad humana. Si hay dudas respecto a alguno de estos *ítems* será mejor desechar las ventajas que pueden derivarse de su utilización, pues existen demasiados riesgos para poder afirmar con fundamento que el respeto al ser humano sigue estando en el centro de todo ello. Parece evidente, no obstante, que para eso es preciso un desarrollo normativo adecuado, que está todavía por determinar⁴⁸.

4.2. Las directrices éticas para una inteligencia artificial fiable

La Comisión Europea en la ya citada Comunicación de 25 de abril de 2018 *Inteligencia Artificial para Europa*, inició un planteamiento coordinado dirigido a aprovechar al máximo las oportunidades que ofrece la inteligencia artificial y abordar los retos que conlleva desde una perspectiva genérica⁴⁹. En este documento ya se resalta la necesidad de que las nuevas tecnologías estén basadas en valores y para ello debe desarrollarse en un marco adecuado “que promueva la innovación y respete los valores y derechos fundamentales de la Unión, así como principios éticos tales como la obligación de

⁴⁸ LATORRE, J.I., *Ética para máquinas*, Barcelona, 2019, pp. 206-207: “La insensatez campa a sus anchas. Las inteligencias artificiales avanzan sin códigos éticos consensuados. Ningún político habla de qué programar, de qué criterio sobre el bien y el mal rige las acciones de un dron, un robot o el gestor de tráfico de una ciudad. Este error no es nuevo. Los humanos no hemos sabido establecer normas éticas consensuadas para los avances anteriores”. En este sentido, COECKELBERGH, M., *AI Ethics*, Cambridge (Massachusetts), 2020, pp. 144-165, se refiere a “qué es necesario hacer y otras cuestiones que los políticos deben responder”.

⁴⁹ Ya la reunión del Consejo Europeo en octubre de 2017 invitó a la Comisión a proponer un planteamiento europeo respecto de la inteligencia artificial, en el que se tuvieran en cuenta, entre otros elementos, las normas éticas. *Vid.* <https://www.consilium.europa.eu/media/21604/19-euco-final-conclusions-es.pdf>.

rendir cuentas y la transparencia”, de modo que la inteligencia artificial pueda beneficiar a las personas y a la sociedad en su conjunto. En esta iniciativa europea ya ocupaba un lugar prominente la exigencia de garantizar el establecimiento de un marco ético y jurídico apropiado y el mandato de establecer unas directrices éticas específicas, de modo que se urgía a adoptar propuestas concretas con el fin inmediato de que “tanto los ciudadanos como las empresas necesitan confiar en la tecnología con la que interactúan, disponer de un entorno jurídico predecible y contar con la garantía efectiva de que van a protegerse los derechos y libertades fundamentales”⁵⁰.

En esta amplia estrategia europea⁵¹ se inscribe el documento titulado “Directrices éticas para una inteligencia artificial fiable”⁵², elaborado por el Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial, constituido en junio de 2018 como centro del diálogo entre el gobierno de la Unión Europea y las personas involucradas en el futuro de la inteligencia artificial. El 18 de diciembre de 2018 se sometió a consulta pública un borrador y el 8 de abril de 2019 se ofreció ya la formulación acabada de estas directrices, de gran interés porque se da un paso importante hacia la concreción de lo que hasta ese momento eran solamente declaraciones generales. Un concepto clave es el de “fiabilidad”, y de este modo se destaca que “la inteligencia artificial fiable” es la aspiración fundacional de este grupo de expertos: “puesto que los seres humanos y las comunidades solamente podrán confiar en el desarrollo tecnológico y en sus aplicaciones si contamos con un marco claro y detallado para garantizar su fiabilidad”⁵³.

La idea de fiabilidad es contemplada como presupuesto (“requisito previo”) para todo el desarrollo posterior de exigencias éticas que se pretenden formular. Conviene tener en

⁵⁰ En este sentido se disponía la elaboración de un proyecto de directrices éticas en relación con la inteligencia artificial en el que se debía examinar el impacto en los derechos fundamentales, en particular, a la intimidad, la dignidad, la protección de los consumidores y la lucha contra la discriminación”.

⁵¹ Es preciso añadir la referencia a la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. *Plan coordinado sobre la inteligencia artificial*. Bruselas, 7 de diciembre de 2018. COM(2018) 795 final: “Para ganar la confianza, que es necesaria para que las sociedades acepten y utilicen la IA, la tecnología debe ser predecible, responsable, verificable, respetar los derechos fundamentales y observar las reglas éticas. De lo contrario, el uso de IA puede conducir a resultados indeseables, como crear una cámara de eco donde las personas solo reciben información que corresponde a sus opiniones, o reforzar la discriminación, como en el caso de un algoritmo que se volvió racista en un plazo de 24 horas debido a la exposición a material racista”.

⁵² *Vid.* <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.

⁵³ Se proclama: “A través de una IA fiable, los ciudadanos europeos podremos tratar de cosechar los beneficios de esa tecnología de un modo acorde con nuestros valores fundacionales: respeto de los derechos humanos, democracia y estado de Derecho”.

cuenta que la vocación de estas reflexiones es muy amplia, no reducida solo al ámbito de la Unión europea, sino dirigida a la creación de “un marco global para una inteligencia artificial fiable” y propiciar un consenso internacional, así como “promover y defender el enfoque fundamental basado en derechos”. En esta dirección, la expresión acerca de la posición fundamental de la confianza va acompañada de una valiosa declaración respecto al refuerzo imprescindible que proviene de los parámetros éticos aplicables⁵⁴. De este modo, las directrices que se van a pormenorizar se apoyan en tres componentes que deben satisfacerse a lo largo de todo el ciclo de vida del sistema:

- a) la inteligencia artificial debe ser lícita, es decir, cumplir todas las leyes y reglamentos aplicables;
- b) ha de ser ética, de modo que se garantice el respeto de los principios y valores éticos; y
- c) debe ser robusta, tanto desde el punto de vista técnico como social, puesto que los sistemas de inteligencia artificial, incluso si las intenciones son buenas, pueden provocar daños accidentales.

Estos tres elementos esenciales no son de aplicación independiente, sino conjunta y armónica, a pesar de las tensiones que puedan resultar de esa consideración simultánea y, por supuesto, variará en su determinación respecto a los campos específicos de la vida social en que deba aplicarse⁵⁵. Como señalaremos más adelante, una de las urgentes necesidades que desde la perspectiva adoptada en este estudio debe satisfacerse es la de reflejar las consideraciones generales en normas de Derecho positivo que eviten una brecha excesiva entre una realidad en continua evolución y unos conjuntos normativos tendentes a la estabilidad. Contamos con numerosas normas generales e indirectas que atienden de algún modo a las exigencias de aseguramiento de la confianza en la inteligencia artificial, pero es perentorio dar un paso más hacia normas de contenido mucho más directo y coercitivo. En ellas se deben reflejar, por supuesto, los criterios éticos que con minuciosidad se han ido

⁵⁴ Así pues, se proclama la ética como pilar fundamental para garantizar y expandir la inteligencia artificial fiable.

⁵⁵ De ahí que se afirme que “debería explorarse la necesidad de adoptar también un enfoque sectorial que complementa el marco horizontal más general propuesto en este documento”.

determinando en estos recientes años. Con todo, no debe perderse de vista el reforzamiento de la seguridad en la aplicación de los instrumentos de inteligencia artificial, desde luego intrínsecamente desde el punto de vista técnico, pero también de manera adaptada a las necesidades del ámbito en que se apliquen, por ejemplo, en el contexto jurisdiccional que nos ocupa prioritariamente.

El desarrollo pormenorizado de las consideraciones en torno a la fiabilidad de la inteligencia artificial lleva a traer a colación los derechos fundamentales como derechos morales recogidos en las más elevadas proclamaciones normativas, por supuesto jurídicamente vinculantes, pero también incompletos para hacer frente a los efervescentes cambios de la realidad. Sin embargo, eso no impide la colocación de esos derechos como base fundamental de la inteligencia artificial fiable, junto con la consideración central del respeto a la dignidad humana, a lo que nos referiremos más adelante.

Como es lógico, estas materias tan delicadas no deben separarse de las manos de los especialistas, y por ello, el documento que estamos considerando, toma a su vez como punto de partida, la aportación del llamado “Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías”, que actúa como asesor de la Comisión Europea⁵⁶. Entre sus trabajos, nos interesa especialmente la “Declaración sobre inteligencia artificial, robótica y sistemas ‘autónomos’”, emitida el 9 de marzo 2018⁵⁷, pues parte de unos principios éticos y presupuestos democráticos elementales y, por tanto, de un evidente alcance constitucional:

- Dignidad humana
- Autonomía
- Responsabilidad
- Justicia, equidad y solidaridad

⁵⁶ En el DOUE L 46, de 10 de febrero de 2021, se publicó la Decisión (UE) 2021/156 de la Comisión de 9 de febrero de 2021 por la que se renueva el mandato del Grupo Europeo de Ética de la Ciencia y de las Nuevas Tecnologías. Su misión es formulada en el artículo 2: “La tarea del GEE consistirá en proporcionar a la Comisión asesoramiento independiente sobre cuestiones en las que las dimensiones éticas, sociales y de derechos fundamentales estén interrelacionadas con el desarrollo de la ciencia y las nuevas tecnologías, ya sea a petición de la Comisión o por iniciativa propia, expresado a través de su presidente y con el acuerdo del servicio de la Comisión responsable. En particular, el GEE: a) identificará, definirá y analizará las cuestiones éticas planteadas por los avances de la ciencia y las tecnologías; b) proporcionará orientación esencial para el desarrollo, la aplicación y el seguimiento de las políticas o la legislación de la Unión en forma de análisis y recomendaciones, presentados en dictámenes y declaraciones, cuyo objetivo será promover una elaboración ética de las políticas de la Unión, de conformidad con la Carta de los Derechos Fundamentales de la Unión Europea”.

⁵⁷ *Vid.* http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.

- Democracia
- Estado de Derecho
- Seguridad e integridad corporal y mental
- Protección de datos y privacidad
- Sostenibilidad

A partir de todo ello, la declaración del documento de Directrices Éticas determina cuatro imperativos “arraigados en los derechos fundamentales, que deben cumplirse para garantizar que los sistemas de IA se desarrollen, desplieguen y utilicen de manera fiable”:

- I) respeto de la autonomía humana;
- II) prevención del daño;
- III) equidad;
- IV) explicabilidad.

El desarrollo de estos cuatro principios constituye una parte central que conviene examinar con mayor detalle. En efecto, se resalta que las personas que deban interactuar con sistemas de inteligencia artificial deben poder mantener una autonomía plena y efectiva sobre sí mismas y deben diseñarse de modo que “aumenten, complementen y potencien las aptitudes cognitivas, sociales y culturales de las personas”⁵⁸, dejando amplias oportunidades para la elección humana, y manteniendo por tanto la función de supervisión y control fuera del instrumento de inteligencia artificial, en manos del ser humano.

Como ya había intuido ASIMOV, la preocupación por evitar daños producidos por la inteligencia artificial sigue vigente: “Los sistemas de IA no deberían provocar daños (o agravar los existentes) ni perjudicar de cualquier otro modo a los seres humanos”. El reforzamiento de la seguridad y la garantía frente a usos malintencionadas adquiere relevancia también en el ámbito de la administración de la justicia.

La equidad debe ser, por otra parte, un principio de aplicación transversal, desde la configuración de los sistemas de inteligencia artificial, hasta la aplicación de los mismos. Se destacan dos dimensiones complementarias: la sustantiva y la procedimental. Conforme a la

⁵⁸ “Los sistemas de IA no deberían subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de manera injustificada”.

primera, se exige garantizar una distribución justa e igualitaria de los beneficios y costes, y asegurar que las personas y grupos no sufran sesgos injustos, discriminación ni estigmatización, lo cual, como hemos recordado, ya ha tenido sus lamentables manifestaciones en el ámbito de la jurisdicción, como reflejo de las inequidades sociales. Con lo cual se demuestra que los sistemas de inteligencia artificial no son espacios de inmunidad respecto a los problemas de la realidad social, sino que además pueden tener un efecto multiplicador respecto a planteamientos injustos originados fuera de estos instrumentos tecnológicos. Frente a ello deben arbitrarse garantías suficientes para asegurar la igualdad de armas en el proceso y aplicando el principio de igualdad de oportunidades, así como el de proporcionalidad entre medios y fines ante circunstancias concretas.

En cuanto a la perspectiva procedimental, es de gran interés, pero a la vez de gran dificultad práctica, el asegurar la facultad de oponerse a las decisiones adoptadas por los sistemas de inteligencia artificial y por las personas que los manejan, lo cual exige una transparencia que no siempre se puede obtener y suficientes conocimientos previos para fundamentar adecuadamente la argumentación. De todas formas, en caso de comprobarse una aplicación inadecuada, sesgada, injusta, debe poderse obtener una respuesta adecuada y efectiva. Entendemos que no sólo como compensación -dando a entender implícitamente quizás un mero resarcimiento económico-, sino una restauración de la posición en que se encontraba el interesado antes de sufrir la decisión inequitativa, con la aplicación de las correspondientes técnicas de nulidad.

Con el discutible término de “explicabilidad”, el documento que comentamos se refiere a la transparencia, a la comunicación de las posibilidades y fines de los sistemas de inteligencia artificial en concreto. Es, desde el punto de vista procesal, una novedosa dimensión del principio de contradicción, pues los justiciables deberían poder estar informados del funcionamiento de los instrumentos que inciden en su situación jurídica, ya sea directa ya indirectamente. En efecto, muy difícilmente podrán discutir su aplicación, si no se les transmite en lenguaje llano el funcionamiento y las implicaciones de los sistemas de inteligencia artificial. Sin embargo, esta exigencia parece rozar la utopía, pues la ciencia de los algoritmos ha demostrado ser un arcano cuyo dominio se limita a los ingenieros informáticos o matemáticos, con lo que éstos pasan a ocupar una posición decisiva muy discutible en el procedimiento de adopción de decisiones que, constitucionalmente, exigen

las garantías de la jurisdicción (art. 117 CE)⁵⁹. Todo ello no demuestra sino la delicada posición de vulnerabilidad en que se encuentra el ciudadano medio ante el avance de las nuevas tecnologías, especialmente grave cuando estamos tratando de la posición del justiciable ante el ejercicio de la función jurisdiccional trufada de decisiones adoptadas con la asistencia de la inteligencia artificial⁶⁰.

Con el objetivo de avanzar hacia una inteligencia artificial fiable, se establecen siete requisitos esenciales que se formulan como exigencias no exhaustivas para la aplicación de la inteligencia artificial:

1. Acción y supervisión humanas, incluidos los derechos fundamentales, la acción y la supervisión humana.
2. Solidez técnica y seguridad, incluida la capacidad de resistencia a los ataques y la seguridad, un plan de repliegue y la seguridad general, precisión, fiabilidad y reproducibilidad.
3. Gestión de la privacidad y de los datos, incluido el respeto de la privacidad, la calidad y la integridad de los datos, así como el acceso a estos.
4. Transparencia Incluidas la trazabilidad, la explicabilidad y la comunicación.
5. Diversidad, no discriminación y equidad, incluida la ausencia de sesgos injustos, la accesibilidad y el diseño universal, así como la participación de las partes interesadas.
6. Bienestar social y ambiental, incluida la sostenibilidad y el respeto del medio ambiente, el impacto social, la sociedad y la democracia.
7. Rendición de cuentas, incluidas la auditabilidad, la minimización de efectos negativos y la notificación de estos, la búsqueda de equilibrios y las compensaciones.

Estos requisitos deben ser tenidos en cuenta tanto por quienes diseñen y desarrollen mecanismos de inteligencia artificial, por quienes sean responsables del establecimiento de

⁵⁹ En este sentido, se afirma: “Esos casos, que se denominan algoritmos de «caja negra», requieren especial atención. En tales circunstancias, puede ser necesario adoptar otras medidas relacionadas con la explicabilidad (por ejemplo, la trazabilidad, la auditabilidad y la comunicación transparente sobre las prestaciones del sistema), siempre y cuando el sistema en su conjunto respete los derechos fundamentales. El grado de necesidad de explicabilidad depende en gran medida del contexto y la gravedad de las consecuencias derivadas de un resultado erróneo o inadecuado”.

⁶⁰ El propio documento alude a cuestiones de gran relevancia jurídico procesal: “Considérese el ejemplo de la utilización de sistemas de IA para la «actuación policial predictiva», que puede ayudar a reducir la delincuencia, pero de formas que incluyan actividades de vigilancia que vulneren la libertad y la privacidad individuales. Además, los beneficios globales de los sistemas de IA deberían ser sustancialmente superiores a los riesgos individuales previsible”.

los nuevos instrumentos y por los usuarios finales y la sociedad en su conjunto, que necesitan ser informados sobre ellos y sobre la posibilidad de exigir que se cumplan las garantías que deben acompañarlos. Para su cumplimiento cabe utilizar tanto medios técnicos⁶¹ como de otro tipo⁶² y se recomienda una evaluación constante de los métodos empleados y de los cambios introducidos en su aplicación. Además, se establece una lista no exhaustiva para la evaluación de la fiabilidad de tales medios con el fin de poner en práctica la inteligencia artificial fiable, previendo una fase piloto de aplicación, con dos procesos paralelos: un proceso cualitativo que debe garantizar la representatividad, y un proceso cuantitativo, en el que todas las partes interesadas podrán inscribirse para experimentar la lista de evaluación y dar su opinión a través de una consulta abierta.

4.3. La generación de confianza en la inteligencia artificial centrada en el ser humano

En el trasfondo de las preocupaciones éticas que se abordan respecto a los avances de la inteligencia artificial es fácil vislumbrar la base kantiana del ser humano racional como fin en sí mismo, que el filósofo de Königsberg formuló en su *Fundamentación para una metafísica de las costumbres* en 1785: “el hombre y en general todo ser racional existe como un fin en sí mismo, no simplemente como un medio para ser utilizado discrecionalmente por esta o aquella voluntad, sino que tanto en las acciones orientadas hacia sí mismo como en las dirigidas hacia otros seres racionales el hombre ha de ser considerado siempre al mismo tiempo como un fin”⁶³. Y frente a las críticas y discusiones frente al antropocentrismo, que son frecuentes por ejemplo desde planteamientos ecológicos, aquí precisamos de una reivindicación de la centralidad del hombre, que no puede ser usurpada por la máquina, por muy “inteligente” que esta sea. Es pertinente la afirmación de Adela CORTINA cuando

⁶¹ Así, se determina que “los requisitos de una inteligencia fiable deben «traducirse» en procedimientos (o en la imposición de restricciones sobre estos) que deben integrarse en la arquitectura de los sistemas de IA. Esto puede lograrse a través de un conjunto de normas de tipo «lista blanca» (comportamientos o estados) que el sistema debería seguir en todo momento, restricciones («lista negra») sobre determinados comportamientos o estados que el sistema jamás debería transgredir y combinaciones de ambas, o garantías demostrables más complejas sobre el comportamiento del sistema. El control del cumplimiento de dichas restricciones por parte del sistema durante su funcionamiento puede llevarse a cabo a través de un proceso separado”.

⁶² En este sentido, se describen diversos métodos dirigidos a garantizar y mantener esa confiabilidad en la inteligencia artificial, entre ellos: la introducción de normas nuevas, la adopción de códigos de conducta, la certificación por organizaciones *ad hoc*, la rendición de cuentas a través de marcos de gobernanza, etc.

⁶³ KANT, I., *Fundamentación para una metafísica de las costumbres* (Ed. R. R. Aramayo), Madrid, 2002, p. 137.

defiende que “a este tipo de seres -las personas- es imposible entonces intercambiarlos por equivalentes que realizarían la misma función, porque no tienen equivalente”⁶⁴.

La estrategia europea, trazada como acabamos de ver en 2018 e inicios de 2019, se acoge en la ya mencionada Comunicación de la Comisión de 8 de abril de 2019⁶⁵, que parte de la constatación del potencial de la inteligencia artificial para mejorar nuestro mundo, beneficiando a la sociedad y a la economía en su conjunto, pero también observa el planteamiento de nuevos retos y novedosas cuestiones jurídicas y éticas. De manera significativa se resalta la opción por la inteligencia artificial centrada en el ser humano desde el propio encabezamiento de este documento⁶⁶. Se informa del plan coordinado que la Comisión puso en marcha en cooperación con los Estados miembros, promoviendo amplias inversiones públicas y privadas, todo ello en una agenda estratégica común de investigación, innovación y despliegue.

De este modo, toca partir de lo que debería ser evidente: la inteligencia artificial no es un fin en sí mismo, como ya apuntábamos, y la confianza es un presupuesto para la garantía de un enfoque de la inteligencia artificial centrado en el ser humano. Estamos de nuevo ante una perspectiva general, y por tanto muy anterior a la específica adaptación de la inteligencia artificial a las necesidades del proceso, pero la conclusión no es ni mucho menos de menor entidad para las aplicaciones que incidan en el ejercicio de la función jurisdiccional y, en definitiva, en la satisfacción de la tutela judicial efectiva: “La dimensión ética de la inteligencia artificial no es un lujo ni algo accesorio: ha de ser parte integrante del desarrollo de la inteligencia artificial”.

Ello se inscribe en el marco normativo que se va construyendo con la finalidad de reforzar la confianza en el mundo digital, desde luego en el marco del respeto de la dignidad humana, la libertad, la democracia, la igualdad, el Estado de Derecho y el respeto de los

⁶⁴ CORTINA, A., *Las fronteras de la persona*, Madrid, 2009, pp. 24-25.

⁶⁵ Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. *Generar confianza en la inteligencia artificial centrada en el ser humano*. Bruselas, 8 de abril de 2019. COM(2019) 168, final.

⁶⁶ Según se señala en su introducción, se “adopta un planteamiento triple para potenciar la capacidad tecnológica e industrial de la Unión Europea e impulsar la adopción de la IA en todos los ámbitos de la economía, prepararse para las transformaciones socioeconómicas y garantizar el establecimiento de un marco ético y jurídico apropiado”.

derechos humanos, como valores en los que indudablemente se fundamenta la Unión Europea. Pero todo ello además con una expresa vocación universal⁶⁷.

La Comisión toma como patrón de actuación las directrices elaboradas por el grupo de expertos de alto nivel sobre inteligencia artificial, destacando la necesidad de tener en cuenta el contexto específico en el que se apliquen a efectos de su aplicación concreta y proporcionada, “adoptando un enfoque basado en el impacto”. Desde luego, en las diversas variantes en las que puede incidir la inteligencia artificial en el proceso hay diversos grados de intensidad, y por tanto diversos niveles de riesgo, incluso cuando nos limitamos a considerar la aplicación asistencial y no sustitutiva de las respectivas decisiones procesales: no supone la misma incidencia un instrumento que ayude a prever la conducta de los menores a los que se les ha aplicado una medida como consecuencia de la declaración de su responsabilidad en una infracción criminal con el fin de establecer relajaciones en cuanto al internamiento decretado⁶⁸, que la aplicación de mecanismos biométricos de identificación en un aeropuerto⁶⁹. Bien es cierto que en todos estos supuestos debe respetarse la centralidad de la dignidad humana y la protección de los derechos fundamentales como exigencias ineludibles⁷⁰.

Desde esta perspectiva, la Comisión apoya los requisitos esenciales formulados por el grupo de expertos y destaca los elementos fundamentales de cada uno de ellos:

1. Intervención y supervisión humanas, en el sentido de que debe reforzarse la intervención humana y la protección de los derechos fundamentales, con la finalidad de que la autonomía humana no sea disminuida, limitada o desorientada y no se causen otros efectos adversos. Por tanto, son fundamentales medidas de control (“incluida la

⁶⁷ “La UE se asienta sobre un sólido marco normativo, que constituirá la referencia mundial para la IA centrada en el ser humano”.

⁶⁸ Vid. PILLADO GONZÁLEZ, E., “Algoritmos predictivos del comportamiento y proceso penal de menores”, en *Justicia algorítmica y neuroderecho. Una mirada interdisciplinar*, (Ed. S. Barona Vilar), Valencia, 2021, pp. 421-442, y CANO PLACER, C., “La predicción y la gestión del riesgo como base de una intervención especializada en el comportamiento criminal”, en *Menores y justicia juvenil* (Ed. L. Bujosa Vadell y F. Martín Diz), Cizur Menor, 2021, pp. 99-119.

⁶⁹ Vid. FARALDO CABANA, P., “¿Tiene la biometría sesgos de género? Sobre la falsa neutralidad de los mecanismos biométricos de identificación”, en *Justicia algorítmica y neuroderecho...*, *op. cit.*, pp. 229-242, y en el mismo volumen colectivo, ETXEBERRÍA GURIDI, J.F., “Inteligencia artificial aplicada a la videovigilancia: Tecnologías de reconocimiento facial”, pp. 443-468.

⁷⁰ Vid. MARTÍN DIZ, F., “Inteligencia artificial y proceso: garantías frente a eficiencia en el entorno de los derechos procesales fundamentales”, en *Justicia: ¿garantías versus eficiencia?* (Dir. F. Jiménez Conde y R. Bellido Penadés), Valencia 2019, pp. 815-827.

adaptabilidad, la exactitud y la explicabilidad de los sistemas de inteligencia artificial”) y la aplicación de mecanismos de participación, de supervisión y control.

2. Solidez y seguridad técnicas, pues la fiabilidad implica que los algoritmos puedan resolver errores o incoherencias durante todas las fases del ciclo vital del sistema de inteligencia artificial y hacer frente a resultados erróneos. La seguridad aplicada a los sistemas debe ser verificable, y debe tener en cuenta la minimización y, cuando sea posible⁷¹, la reversibilidad de las consecuencias no deseadas o errores en el funcionamiento del sistema.
3. Privacidad y gestión de datos, y ello en todas las fases del ciclo vital del sistema de inteligencia artificial. El ser humano debe mantener el control sobre sus propios datos y tener la seguridad de que sus datos personales no se utilizarán para su perjuicio o discriminación. Pero hay un elemento más a considerar que se refiere a la garantía de la calidad de los datos utilizados, evitando sesgos e inexactitudes, lo cual implica asegurar la integridad de los datos y la regulación y control del acceso a los mismos.
4. Transparencia, para proteger la trazabilidad de los sistemas de inteligencia artificial, registrando y documentando las decisiones tomadas por los sistemas y en el conjunto de su funcionamiento, explicando también cómo se toman las decisiones a partir del algoritmo, destacando su influencia en la adopción de las decisiones, además de comunicar las posibilidades y limitaciones del sistema. Es importante que el justiciable sepa que interviene un sistema de inteligencia artificial en la decisión que finalmente se adopte.
5. Diversidad, no discriminación y equidad, como concreción de las exigencias generales de contenido constitucional. Debe tenerse cuidado con los sesgos, voluntarios o involuntarios, por incompletos o por atender a modelos deficientes. La discriminación puede proceder tanto de la configuración del algoritmo, como de la forma sesgada en

⁷¹ Posibilidad física, más que jurídica, pues las normas deben articular estas posibilidades de manera adecuada y efectiva.

que los datos sean incorporados. Frente a ello se propone arbitrar equipos diversificados y mecanismos de participación.

6. Bienestar social y medioambiental, pues en la inteligencia artificial fiable es preciso tener en cuenta el impacto ambiental y el respeto a todos los seres sensibles, también desde la perspectiva de impacto social.
7. Rendición de cuentas, pues es necesaria la existencia de mecanismos para exigir responsabilidad y rendición de cuentas respecto al funcionamiento y los resultados de los diversos sistemas de inteligencia artificial, evaluando y reduciendo impactos negativos y estableciendo vías de reparación adecuada.

Más recientemente, de manera aún más institucional, la Comisión Europea utilizó la técnica del Libro Blanco para abrir un debate más amplio sobre la inteligencia artificial⁷². De nuevo el planteamiento es genérico, por tanto no referido específicamente a la aplicación de la inteligencia artificial a la justicia, pero aquí ya hay consideraciones mucho más cercanas al objeto de nuestra preocupación, pues, por un lado, se habla de la existencia de riesgos para los derechos fundamentales⁷³, especialmente la protección de los datos personales y de la privacidad y, de nuevo, la no discriminación⁷⁴. Se afirma que tales riesgos

⁷² *Libro Blanco sobre la inteligencia artificial* - Un enfoque europeo orientado a la excelencia y la confianza. Bruselas, 19 de febrero de 2020 COM (2020) 65, final.

⁷³ Vid. DE HOYOS SANCHO, M., “Premisas y finalidades del Libro Blanco sobre Inteligencia Artificial de la Comisión Europea: Perspectiva procesal del nuevo marco regulador”, en *Justicia algorítmica y neuroderecho...*, *op. cit.*, pp. 129-156: “A la luz del contenido del *Libro Blanco* sobre la materia aprobado por la Comisión Europea y del resto de argumentos puestos de relieve en los epígrafes precedentes, consideramos que tales instrumentos podrían ser empleados por el aparato judicial, en particular por lo que respecta al orden penal, si se pueden garantizar al menos los siguientes requisitos y presupuestos”, y enumera la previsión normativa del ámbito de aplicación y la forma de utilización de estos sistemas de inteligencia artificial; el control previo de la legalidad /admisibilidad del concreto algoritmo y del tratamiento de los datos que procesa el sistema de inteligencia artificial, con la debida supervisión de su aplicación y funcionamiento; la necesaria intervención humana en la adopción de la decisión final y la posibilidad de “acción judicial efectiva” contra la decisión judicial basada en sistemas de inteligencia artificial; transparencia y trazabilidad, así como seguridad y verificabilidad de los datos empleados, sin sesgos ni discriminaciones y necesidad de una regulación armonizada en el espacio de libertad, seguridad y justicia de la Unión Europea.

⁷⁴ “El uso de la inteligencia artificial puede afectar a los valores sobre los que se fundamenta la UE y provocar la conculcación de derechos fundamentales, como la libertad de expresión, la libertad de reunión, la dignidad humana, la ausencia de discriminación por razón de sexo, raza u origen étnico, religión o credo, discapacidad, edad u orientación sexual, y, en su aplicación en determinados ámbitos, la protección de los datos personales y de la vida privada, el derecho a una tutela judicial efectiva y a un juicio justo, o la protección de los consumidores”.

pueden ser resultado de defectos en el diseño general de los sistemas de inteligencia artificial o del uso de datos que puedan ser sesgados sin una corrección previa (por ejemplo, se entrena un sistema utilizando única o principalmente datos relativos a hombres, y ello se traduce en resultados peores con relación a las mujeres). Pero también hay otra dimensión de riesgos para la seguridad, en relación con el funcionamiento eficaz del régimen de responsabilidad civil⁷⁵.

Es importante subrayar la constatación de supervisión humana, pues ayuda a garantizar que un sistema de inteligencia artificial no socave la autonomía humana o provoque otros efectos adversos. Se trata, en definitiva, de compensar de algún modo algunas características particulares de algunas de las tecnologías de inteligencia artificial, como la opacidad, la complejidad, la imprevisibilidad y un comportamiento parcialmente autónomo, que pueden complicar la aplicación efectiva de la normativa sobre derechos fundamentales.

4.4. El vigente enfoque europeo sobre inteligencia artificial

El 21 de abril de 2021 la Comisión Europea, siguiendo la senda ya trazada por los textos mencionados, propuso nuevos documentos para reforzar el enfoque de la inteligencia artificial basada en la excelencia y en la confiabilidad. En concreto son tres los textos que fueron publicados:

- La Comunicación sobre la promoción de un enfoque europeo acerca de la inteligencia artificial⁷⁶.
- El plan coordinado actualizado para los Estados miembros⁷⁷, y
- La propuesta de reglamento sobre inteligencia artificial⁷⁸.

⁷⁵ “Las tecnologías de IA pueden presentar nuevos riesgos de seguridad para los usuarios cuando estén integradas en productos y servicios”.

⁷⁶ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *Fostering a European approach to Artificial Intelligence*, Brussels, 21.4.2021 COM(2021) 205, final.

⁷⁷ Annexes to the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *Fostering a European approach to Artificial Intelligence*, Brussels, 21.4.2021 COM(2021) 205, final. Es oportuno destacar el capítulo III dirigido a asegurar que la inteligencia artificial se aplica en beneficio de las personas y del bien común de la sociedad.

⁷⁸ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Brussels, 21.4.2021 COM/2021/206, final.

El planteamiento sobre la eventualidad de desafíos y riesgos se mantiene y también la constatación de las múltiples ventajas en la aplicación de instrumentos de inteligencia artificial. Los derechos fundamentales y la seguridad siguen necesitando una protección específica frente a los avances tecnológicos. No es el propósito de este apartado el de analizar todas las novedades que se incluyen en los tres documentos mencionados. Desde la perspectiva de las exigencias éticas puede afirmarse que se observa un ambicioso desarrollo que da pasos significativos para una mucho mayor concreción normativa. Partiendo del importante trabajo previo y de los distintos pronunciamientos internacionales, a algunos de los cuales hemos hecho detenida referencia en estas páginas, se pretende profundizar en las exigencias de confiabilidad en los mecanismos de inteligencia artificial, a sabiendas de que la normativa vigente es claramente insuficiente, aunque se disponga ya de un marco sólido a partir del que debe procederse a su cualitativo perfeccionamiento, basado en criterios de proporcionalidad y de atención a los riesgos específicos.

De ahí la novedad de la clasificación en categorías de riesgos, con la determinación de que entre los más elevados se sitúa la aplicación de estos mecanismos a la adopción de decisiones judiciales (*high-risk*). En estos casos las exigencias se plantean naturalmente de un modo más grave y necesitan una respuesta técnica, ética y jurídica fortalecida⁷⁹. Destaca la prohibición de determinados usos por considerar que contravienen los valores en los que se asienta la Unión Europea o vulneran los derechos fundamentales. En concreto, se establecen restricciones importantes respecto a las identificaciones biométricas y se proponen otros usos de los sistemas de inteligencia artificial sujetos especialmente a exigencias mínimas de transparencia.

4.5. La inteligencia artificial en la Carta de Derechos Digitales

Directrices de contenido ético para el tratamiento de los instrumentos de inteligencia artificial se contienen en la ya aludida Carta de Derechos Digitales. En efecto, en la sección quinta relativa a “derechos digitales en entornos específicos” el artículo vigésimo quinto

⁷⁹ “... *high-risk AI systems need to respect a set of specifically designed requirements, which include the use of high-quality datasets, the establishment of appropriate documentation to enhance traceability, the sharing of adequate information with the user, the design and implementation of appropriate human oversight measures, and the achievement of the highest standards in terms of robustness, safety, cybersecurity and accuracy*”.

sintetiza algunas de las más importantes exigencias. La formulación general está en la línea de lo expuesto en las páginas anteriores, pues se concibe la inteligencia artificial desde un enfoque centrado en la persona y en su inalienable dignidad, manteniendo el principio kantiano e incorporando la perspectiva teleológica de responder al bien común, y no solo a los beneficios individuales, además de respetar el principio de no maleficencia, es decir, el principio bioético que exige no hacer nada que innecesariamente pueda perjudicar al ser humano⁸⁰.

Pero estos principios generales, se ven acompañados por otros más específicos, a los que también nos hemos referido: la garantía de la no discriminación en relación con las decisiones, el uso de datos y procedimientos basados en inteligencia artificial; el establecimiento de condiciones de transparencia, auditabilidad, explicabilidad, trazabilidad, supervisión humana y gobernanza, y desde luego, la accesibilidad y comprensibilidad de la información que se facilite; así como la accesibilidad, usabilidad y fiabilidad respecto a los sistemas de inteligencia artificial. Todo un programa que exige una mayor concreción que pueda garantizar una verdadera fuerza normativa⁸¹.

En el apartado tercero, se destaca la proclamación de dos derechos importantes, que antes o después debieran adquirir el carácter de fundamentales: el derecho a solicitar una supervisión e intervención humana en la aplicación de instrumentos de inteligencia artificial y el derecho a impugnar las decisiones automatizadas tomadas por estos y que puedan producir efectos en su esfera personal y patrimonial.

4.6. La Carta Ética Europea para el uso de inteligencia artificial en sistemas judiciales

Los diversos documentos a los que nos hemos referido se caracterizan por una formulación general de las necesidades éticas que ya se aplican a los sistemas informáticos o están a punto de aplicarse en los diversos ámbitos de la realidad social. En algunos puntos hemos señalado algunos comentarios desde la perspectiva jurídico-procesal. Pero en el seno del Consejo de Europa, ya en 2017 se aprobó un texto dirigido a un ámbito mucho más

⁸⁰ Cfr. KRAUS A., y PÉREZ TAMAYO, R., *Diccionario incompleto de bioética*, México, 2011.

⁸¹ Nos parece especialmente pertinente la cita de CAMPS, V., *Breve historia...*, *op. cit.*, p. 397, cuando afirma que “La consideración de las consecuencias complementa a una ética de principios en la medida en que ayuda a contextualizarlos y a no ignorar una realidad social que es cambiante y exhibe más rasgos irracionales que racionales”.

específico: la Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales y su entorno⁸², por tanto ya con una delimitación en cuanto a su incidencia justamente a lo que nos interesa desde nuestra perspectiva procesalista.

El punto de partida nos es ya conocido. Se proclaman cinco principios sobre el uso de la inteligencia artificial en el ámbito judicial:

1. El principio de respeto a los derechos fundamentales en el diseño e implementación de las herramientas y servicios de inteligencia artificial.
2. El principio de no discriminación entre individuos y grupos de individuos⁸³.
3. El principio de calidad y seguridad respecto a la adopción de decisiones judiciales y utilización de datos⁸⁴.
4. El principio de transparencia, imparcialidad y equidad, asegurando la accesibilidad⁸⁵ y comprensibilidad de los métodos y autorizando auditorías externas.
5. Principio “bajo control del usuario”.

Como se observa, se recogen las principales ideas generales para un planteamiento ético de la aplicación de sistemas de inteligencia artificial, pero ya dirigidas en concreto al ámbito de la administración de justicia. Así, ya hay mención explícita de la exigencia de respetar el derecho de acceso a la jurisdicción y el derecho a un proceso con todas las garantías, también las reglas del Estado de Derecho y de la independencia judicial.

De todas formas, aunque se dirijan ya a nuestro campo específico se trata de formulaciones amplias de carácter programático, que deben ser tenidas en cuenta en los desarrollos normativos a los que es necesario que procedan los diversos Estados miembros. El documento se completa con cuatro apéndices valiosos para estudios más concretos que el nuestro: en el primero se contiene un análisis de los sistemas de inteligencia artificial aplicados en el ámbito judicial, en particular respecto a la adopción de decisiones y gestión

⁸² Documento aprobado por el adoptado por el CEPEJ durante su 31ª reunión plenaria en la ciudad de Estrasburgo, durante los días 3 y 4 de diciembre de 2018. <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>

⁸³ “This could include alleged racial or ethnic origin, socio-economic background, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health-related data or data concerning sexual life or sexual orientation”.

⁸⁴ “Forming mixed project teams in short design cycles to produce functional models is one of the organisational methods making it possible to capitalise on this multidisciplinary approach”.

⁸⁵ “A balance must be struck between the intellectual property of certain processing methods and the need for transparency (access to the design process), impartiality (absence of bias), fairness and intellectual integrity (prioritising the interests of justice) when tools are used that may have legal consequences or may significantly affect people’s lives”.

de datos personales; en el segundo, sobre los usos de la inteligencia artificial en los sistemas judiciales; en el tercero se incluye un glosario, y en el cuarto, unas amplias listas para la comprobación de aplicación de los principios.

5. CONSIDERACIONES FINALES

Todas las anteriores consideraciones, los documentos preparatorios, las reflexiones doctrinales, los continuos avances tecnológicos no nos pueden más que conducir a unas conclusiones abiertas. Estamos en un momento en que nos planteamos dudas, dilemas, necesidades, pero también exigencias. No se trata, por tanto, de una situación totalmente nebulosa e indefinida. Aunque está claro que se trata de un objeto de estudio en continua evolución, ya tenemos unos referentes éticos insoslayables, que deberían concretarse más y servir de límites para los desarrollos legislativos.

La visión realista de la aplicación de la inteligencia artificial al proceso se centra en la actualidad en la consideración de métodos de apoyo al juzgador, por ejemplo, permitiéndole una rápida selección de jurisprudencia sobre el caso. Sin embargo, en algunas decisiones nos acercamos ya a la aplicación de instrumentos que, por lo menos parcialmente, pueden suponer una cierta sustitución del juzgador en alguna de sus decisiones, como en los requerimientos de pago automatizados; en la apreciación de prueba electrónica (*blockchain*); en valoración de probabilidades de reincidencias...

Todo ello no nos debe ocultar la persistencia de problemas serios que están pendientes de resolución: la cuestión de la construcción neutral del algoritmo; el problema de los sesgos en los datos que se van introduciendo en el sistema informático; el aseguramiento de los datos (*big data*) y la privacidad; o la cuestión de la autonomía en el aprendizaje y su control. Todo lo cual nos sitúan ante una etapa apasionante para el estudioso del Derecho, que deberá permanecer atento a todos los avances y resistir los espejismos de la eficiencia para mantener el predominio de las consideraciones éticas que en las páginas anteriores hemos esbozado.

REFERÊNCIAS:

ARISTÓTELES, *Ética a Nicómaco*, Madrid: Alianza Editorial, 2008.

- ARMENTA DÉU, T., *Derivas de la justicia. Tutela de los derechos y solución de controversias en tiempos de cambios*, Marcial Pons, Ediciones Jurídicas y Sociales: Madrid, 2021.
- ASIMOV, I., *Círculo vicioso*, publicado con el título *Runaround*, Revista Astounding Science Fiction: EUA, marzo de 1942.
- BARONA VILAR, S., *Algoritmización del Derecho y de la Justicia. De la inteligencia Artificial a la Smart Justice*, Tirant lo Blanch: Valencia, 2021.
- BARRIO ANDRÉS, M., (Dir.), *Derecho de los Robots*, Madrid: Wolters Kluwer, 2018.
- BOSTROM, N., *Superinteligencia. Caminos, peligros, estrategias*, Teell Editorial, 2016.
- CAMPS, V., *Breve historia de la ética*, RBA Libros: Barcelona, 2013.
- CANO PLACER, C., La predicción y la gestión del riesgo como base de una intervención especializada en el comportamiento criminal, en *Menores y justicia juvenil*, Ed. L. Bujosa Vadell y F. Martín Diz, Thomson Aranzadi: Cizur Menor, 2021.
- COBB, M., *The Idea of the Brain. The past and the future of Neuroscience*, New York: Basic Books, 2020.
- COECKELBERGH, MARK, *AI Ethics*, MIT Press: Cambridge (Massachusetts), 2020.
- COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. *Fostering a European approach to Artificial Intelligence*, Brussels, 21.4.2021 COM(2021).
- COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. *Inteligencia artificial para Europa*. Bruselas, 25 de abril de 2018 COM(2018) 237.
- COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES. Plan coordinado sobre la inteligencia artificial. Bruselas, 7 de diciembre de 2018. COM(2018) 795.
- COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS

- REGIONES. *Generar confianza en la inteligencia artificial centrada en el ser humano*. Bruselas, 8 de abril de 2019. COM(2019) 168 final.
- CONDE FUENTES, J., y SERRANO HOYO, G., (Dirs.), *La justicia digital en España y la Unión Europea*, Atelier: Barcelona, 2019.
- CONSEJO EUROPEO, *EUCO 14/17*, Bruselas, 2017, <https://www.consilium.europa.eu/media/21604/19-euco-final-conclusions-es.pdf>.
- CONSEJO EUROPEO, *Programa de Estocolmo – Una Europa abierta y segura que sirva y proteja al ciudadano*, (2010/C 115/01), Diario Oficial nº C 115, de 4 de mayo de 2010, <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:115:0001:0038:ES:PDF>.
- CORTINA, A., *Ética mínima. Introducción a la filosofía práctica*. 15.ª ed., Madrid: Editorial Tecnos, 2010.
- CORTINA, A., *Justicia cordial*, Madrid: Editorial Trott, 2010.
- CORTINA, A., *Las fronteras de la persona. El valor de los animales, la dignidad de los humanos*, Madrid: Taurus, 2009.
- CORTINA, A., y MARTÍNEZ, E., *Ética*, 4.ª ed., Madrid: Akal, 2008.
- DE HOYOS SANCHO, M., *Premisas y finalidades del Libro Blanco sobre Inteligencia Artificial de la Comisión Europea: Perspectiva procesal del nuevo marco regulador*, en VILAR, S. B., *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*, Valencia: Tirant Lo Blanch, 2021.
- DOUE L 46, de 10 de febrero de 2021, *Decisión (UE) 2021/156* de la Comisión de 9 de febrero de 2021.
- ETXEBERRÍA GURIDI, J.F., *Inteligencia artificial aplicada a la videovigilancia: Tecnologías de reconocimiento facial*, en VILAR, S. B., *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*, Valencia: Tirant Lo Blanch, 2021.
- EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ), *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, Council of Europe, February 2019, <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

- FARALDO CABANA, P., *¿Tiene la biometría sesgos de género? Sobre la falsa neutralidad de los mecanismos biométricos de identificación*, en VILAR, S. B., *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*, Valencia: Tirant Lo Blanch, 2021.
- FODOR, J., *La mente no funciona así. Alcance y límites de la psicología computacional*, Madrid: Siglo XXI de España Editores, S.A, 2003.
- FUTURE OF LIFE INSTITUT, ASILOMAR AI principles, 2017, <https://futureoflife.org/ai-principles/>.
- GARCÍA COSTA, F. M., *Perfiles constitucionales de la justicia electrónica*, en GÓMEZ MANRESA, M.F., y FERNÁNDEZ SALMERÓN, M., *Modernización digital e innovación en la Administración de Justicia*, Cizur Menor: Aranzadi, 2019.
- GOBIERNO DE ESPAÑA, *Carta de Derechos Digitales*, Madrid, 2021, https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf.
- GRUPO EUROPEO SOBRE ÉTICA DE LA CIENCIA Y LAS NUEVAS TECNOLOGÍAS, *Declaración sobre inteligencia artificial, robótica y sistemas 'autónomos'*, Bruselas, 2018, http://www.bioeticayderecho.ub.edu/archivos/pdf/EGE_inteligencia-artificial.pdf.
- GRUPO INDEPENDIENTE DE EXPERTOS DE ALTO NIVEL SOBRE INTELIGENCIA ARTIFICIAL, *Directrices éticas para una inteligencia artificial fiable*, Bruselas, 2019, <https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.
- HOFFMANN-RIEM, W., *Big Data. Desafíos también para el Derecho*, Madrid: Cuadernos Civitas, 2018.
- JEFATURA DEL ESTADO, *Ley 18/2011*, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la Administración de Justicia.
- KAHNEMAN, D., *Pensar rápido, pensar despacio*, 3.^a ed., Barcelona, 2012.
- KANT, I., *Fundamentación para una metafísica de las costumbres*, Ed. R. R. Aramayo, Madrid, 2002.
- KRAUS A., y PÉREZ TAMAYO, R., *Diccionario incompleto de bioética*, México: Taurus, 2011.

- KUHN, T. S., *The Structure of Scientific Revolutions*, 50th Anniversary edition, Chicago-London, 2012.
- KURZWEIL, R., *Cómo crear una mente. El secreto del pensamiento humano*, Berlin: Lola Books, 2013.
- LACRUZ MANTECÓN, M.L. *Robots y personas. Una aproximación jurídica a la subjetividad cibernética*, Madrid: Reus, 2020.
- LATORRE, J.I., *Ética para máquinas*, Barcelona: Ariel, 2019.
- Libro Blanco sobre la inteligencia artificial - Un enfoque europeo orientado a la excelencia y la confianza. Bruselas, 19 de febrero de 2020 COM(2020).
- LÓPEZ MORATALLA, N., *Inteligencia artificial ¿Conciencia artificial? Una perspectiva desde las ciencias de la vida*, Madrid: Digital Reasons, 2017.
- MARTÍN DIZ, F., *Inteligencia artificial y proceso: garantías frente a eficiencia en el entorno de los derechos procesales fundamentales*, en *Justicia: ¿garantías versus eficiencia?* (Dir. F. Jiménez Conde y R. Bellido Penadés), Valencia: Tirant lo Blanch, 2019.
- MAYER-SCHÖNBERGER, V., y CUKIER, K., *Big Data. La revolución de los datos masivos*, Madrid, 2013.
- NIEVA FENOLL, J., *Inteligencia artificial y proceso judicial*, Madrid: Marcial Pons, Ediciones Jurídicas y Sociales, 2018.
- OLIVEIRA, A., *The Digital Mind. How Science is Redefining Humanity*, London: MIT Press, 2018.
- PALIWALA, A. (Ed.), *A history of legal informatics*, Zaragoza: Prensas de la Universidad de Zaragoza, 2010.
- PENROSE, R., *La mente nueva del emperador: En torno a la cibernética, la mente y las leyes de la física*, Madrid, 2019.
- PÉREZ LUÑO, A.E., *Los derechos humanos en la sociedad tecnológica*, Madrid: Editorial Universitas, S.A, 2012.
- PILLADO GONZÁLEZ, E., *Algoritmos predictivos del comportamiento y proceso penal de menores*, en VILAR, S. B., *Justicia algorítmica y neuroderecho: una mirada multidisciplinar*, Valencia: Tirant Lo Blanch, 2021.

RESOLUCIÓN DEL PARLAMENTO EUROPEO, de 12 de febrero de 2019, sobre una política industrial global europea en materia de inteligencia artificial y robótica (2018/2088(INI)).

RESOLUCIÓN DEL PARLAMENTO EUROPEO, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)).

ROGEL VIDE, C., (Coord.), *Los robots y el Derecho*, Madrid: Editorial Reus, 2018.

RUIZ TRUJILLO, P., *Ética de las nanotecnologías*, Barcelona: Herder, 2020.

SALAZAR, I., *La revolución de los robots. Cómo la inteligencia artificial y la robótica afectan a nuestro futuro*, Gijón, 2019.

SAN MIGUEL CASO, C. *Las técnicas de predicción judicial y su repercusión en el proceso*. En CONDE FUENTES, J. y SERRANO HOYO, G. (Dirs.), *La justicia digital en España y la Unión Europea: Situación actual y perspectivas de futuro*, Atelier, 2019.

SARTOR, G., BRANTING, K., (Eds.), *Judicial Applications of Artificial Intelligence*, Dordrecht, 2010.

SCHWAB, K., *La cuarta revolución industrial*, Barcelona: Debate, 2016.

SENÉS MOTILLA, C., (Coord.), *Presente y futuro de la e-Justicia en España y la Unión Europea*, Cizur Menor, 2010, y DE LA OLIVA SANTOS, A.; GASCÓN INCHAUSTI, F.; AGUILERA MORALES; M., (Coords.), *La e-Justicia en la Unión Europea. Desarrollos en el Ámbito Europeo y en los Ordenamientos Nacionales*, Cizur Menor, 2012.

SOLAR CAYÓN, J.I., *La inteligencia artificial jurídica. El impacto de la innovación tecnológica en la práctica del Derecho y el mercado de servicios jurídicos*, Cizur Menor, 2019.

SUSSKIND, R., *El abogado del mañana. Una introducción a tu futuro*, 2.^a ed., Madrid: La Ley, 2020.

SUSSKIND, R., *Online Courts and the Future of the Justice*, USA: Oxford University Press, 2019.

TARUFFO, M., *Judicial decisions and Artificial Intelligence*, en SARTOR, G., BRANTING, K., (Eds.), *Judicial Applications of Artificial Intelligence*, Dordrecht, 2010.

- TASHEA, J., *Courts Are Using AI to Sentence Criminals. That Must Stop Now*, 2017, en <https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/>.
- THE NATIONAL LAW REVIEW, *Would You Trust An Artificially-Intelligent Expert?*, July 11, 2020, Volume X, Number 193: <https://www.natlawreview.com/article/would-you-trust-artificially-intelligent-expert>.
- TRIBUNAL SUPREMO DEL ESTADO DE WISCONSIN: *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016).
- TURING, A.M., *Computing Machinery and Intelligence*, *Mind*, vol. LIX, núm. 236, 1950; traducida al español como *¿Puede pensar una máquina?*, Oviedo, 2012.
- TVERSKY, A.; KAHNEMAN, D., *El juicio bajo incertidumbre: heurísticas y sesgos*, publicado originalmente en inglés en *Science*, vol. 185, 1974.
- VASAK, K., *Pour les droits de l'homme de la troisième génération*, Strasbourg, 1979.